

## 2.1 Magazine readership and its measurement in 1983

---

Although readership is easy to define, it cannot easily be measured. In fact, it seems fair to say that whenever any of us measures readership, we really do not know exactly what it is we do measure. Here I discuss some of the reasons for, as well as some ways out of, this difficulty. The four basic questions I address are:

- 1 What is readership?
- 2 What is it we measure?
- 3 Why do we have problems?
- 4 What are some solutions?

### WHAT IS READERSHIP?

Simply put, readership is *exposure to the inside of a magazine*. No matter which method of measuring readership we use, the empirical event we are attempting to capture is exposure to the inside of a magazine.\*

The three principal models used to measure readership in the form of average issue audience involve:

Model	What is measured
Recognition:	Specific issue readership
Recency:	Specific time frame readership
Frequency:	Probabilistic readership (without reference to a specific issue or a specific time)

\* This definition does not apply to newspapers such as The Wall Street Journal where exposure to the front page also involves readership.

Each model derives audience estimates from a different data base and different assumptions. But the goal of each is to provide an estimate of the number of different people who are exposed to the inside of an average issue of a magazine.

Let us examine this definition of readership in more detail. Readers are people who have spent some amount of time with some amount of the inside of a magazine in their normal visual field. How are the terms in this definition usually used? *Some amount of time* usually means any amount of time, however brief. *Some amount of the inside of a magazine* usually means any

amount, even if it is only one pair of facing pages. *Normal Visual Field* means that a person's eyes were pointed in a direction that *allowed* viewing or reading of an open magazine.

Notice that our definition does not include qualifications regarding: the amount of material seen; whether the reader was reading or just glancing at pages in an issue; whether the reader was attentive to the material seen or pre-occupied with some other matter; whether the purpose of reading was to gain information or to escape; whether the reader is a very frequent or a one-time reader of the magazine; where the reading took place.

I mention these factors as they are some of the major ones which determine to what extent a given technique succeeds or fails in capturing readers and avoids capturing non-readers.

### WHAT IS IT WE MEASURE

Reading includes an extremely wide range of human behaviour even if we examine only the factors mentioned above.

The readers we attempt to capture in our average issue audience estimates include those who differ in terms of a number of dimensions. Here are six and for the sake of brevity, consider each of these variables as being only two values or states. On this basis we have:

- |                        |  |
|------------------------|--|
| 1 Amount of Exposure   | — which can be substantial vs. limited   |
| 2 Type of Exposure     | — which consists of reading vs. glancing |
| 3 Mental Set           | — being attentive vs. pre-occupied       |
| 4 Purpose of Reading   | — to seek information vs. escape         |
| 5 Frequency of Reading | — defined as frequent vs. infrequent     |
| 6 Place of Reading     | — being at home vs. at a public place    |

## 2.1 Magazine readership and its measurement in 1983

Exposure			Substantial				Limited			
Exposure Type			Reading		Glancing		Reading		Glancing	
Mental Set			At- tentive	Pre- oc- cupied	At- tentive	Pre- oc- cupied	At- tentive	Pre- oc- cupied	At- tentive	Pre- oc- cupied
Information	Frequent	Home	(1)							
		Public Place								
	Infrequent	Home				(2)				
		Public Place								
Escape	Frequent	Home								
		Public Place					(3)			
	Infrequent	Home								
		Public Place								(4)

Even with this over-simplified scheme, we have 64 combinations of readers and reading conditions. The enormous variation in reading we attempt to measure can be made clear by describing only four of these combinations.

Reader **(1)** is a frequent reader who was seeking information and, attentively read all of this issue while at home.

Reader **(2)** is an infrequent reader who glanced at some of this issue at home looking for information while other things were on his mind.

Reader **(3)** is a frequent reader who attentively read a little of this issue in a public place while killing time.

Reader **(4)** is an infrequent reader who for escape glanced at a few pages in a public place while his mind was elsewhere.

For the sake of simplicity, we have side-stepped one major complicatoin issue, namely, some people have poor memories.

We, each with our own system of measurement, hope to include all of these different kinds of readers in our average issue audience estimates — without

## 2.1 Magazine readership and its measurement in 1983

incorrectly including non-readers as well. I am afraid there is substantial evidence that we do not succeed too well.

Before we review the evidence, it should be pointed out that errors of measurement extend well beyond readership research into all areas of opinion surveys and the social sciences in general. **(1)** In fact, even the physical sciences are plagued by measurement error.

For example, between the 1960's and 1970's there were reports of a significant geological uplift in Southern California followed by a sinking of the land. These reports caused great concern as they presumably reflected events which presaged major earthquake activity. It now appears that the only problem was improperly calibrated instruments.

In a ten year period ending 1974, some 500 technical journal articles discussed a new substance called 'polywater' which was declared to be denser than regular water as well as having lower freezing and higher boiling points. Polywater turned out to be impure regular water.

The first point of these examples is that our measurement problems in audience research are not unique. The second point is that with diligence and imagination we, like people in other disciplines, should be able to come up with solutions.

But first, back to our problems. The Recency method does not lend itself to full direct validation because it is not issue specific and we have no methods for observing people 24 hours a day for the period involved.

The Through-The-Book method lends itself to direct validation because it is issue specific. But the measures we have are only half-way measures. We only know how to measure the level of under-claiming. We do not know how to measure the level of over-claiming. Furthermore, the validation experiments conducted in the past, such as the ARF studies reported in New Orleans, yield high levels of under-claiming for public place reading. **(2)**

The work by *Newsweek* and the work by *Time*, reported in sessions 3a and 3b, does show some promise for future validity testing. In the meantime, however, the validity problem remains.

The most compelling evidence that none of us is sure of what we are and what we are not measuring with current audience estimation techniques, however, is the fact that: changes in almost any aspect of our measurement systems have, on occasion produced dramatic and surprising changes in audience estimates.

### VARIABLES THAT HAVE PRODUCED MAJOR CHANGES IN AUDIENCE ESTIMATES

TECHNIQUE  
CONTRACTOR  
FILTERS  
ROTATION  
QUESTION WORDING  
EXHIBIT CARDS  
INTERVIEWER MATERIALS

#### Technique variation

In the United States, we have two syndicated audience measurement systems: MRI, which uses a version of the recency method and SMRB, which uses a version of the Through-The-Book method. For weeklies, the average difference in audience estimate is 22%, while for monthlies it is 56%. **(3)**

#### Contractor variation

In the United Kingdom, numerous changes in contractors from 1968 to 1977 were associated with large changes in audience size. Although some of the variations were no more than one would expect across time, the consensus seems to be that these events constitute evidence of a large sized contractor effect. **(4)**

#### Filters

In the United States, one of the ARF validity studies reported a 12% loss in the capture of observed readership by the Through-The-Book method when the traditional type of filter question was used. Moreover, this occurred when the interview was conducted within two weeks of the observed reading. **(5)**

In Italy, as reported by Liliana Denon, it was found, contrary to general expectations, that the use of two filter questions rather than one produced higher audience estimates. Specifically, from 1975 to 1976, audience estimates for weeklies and monthlies declined by 15% and 21%, respectively. The only reported change was the elimination of one of two filter questions. **(6)**

#### Rotation

Edward Whitley reported in New Orleans, based on JICNARS data for January — June, 1979, that across four rotations, general weekly magazine audience estimates show an extreme spread of 8%. For monthlies, the extreme spread is 42%. For example, for every 100 readers generated by general monthlies when they were in the fourth or last position, there were 142 readers for these same magazines when they were in the first position. **(7)**

# 2.1 Magazine readership and its measurement in 1983

## Question wording

In 1974, as reported by Wally Langschmidt, the South African readership question was modified so that the word "read" was changed to "read or paged through". This change was associated with marked increases in audience estimates: 21% for weeklies, 35% for fortnightlylies and 77% for monthlies. **(8)**

## Exhibit cards

In a West German study by Friedrich Tennstädt and Jochen Hansen, it was reported that answer categories provided to respondents on exhibit cards can have a very significant impact on audiences estimated by the recency method. Specifically, they reported a very high correlation between the proportion of answer categories that qualify one as a reader (as opposed to a non-reader) and average issue audience size. For example, in these experiments, the average issue audience of weekly news magazines increased by 33% when the number of qualifying answer categories on the exhibit cards went from 1 out 5 to 2 out of 7. **(9)**

## Interviewer materials

In the same West German paper, we find that changes in answer categories provided only to *interviewers* can have a major effect on audience estimates. In one experiment two variations in answer categories were tested in connection with a question on recency of reading for a weekly magazine.

	VARIATION A	VARIATION B
Qualifying Categories	1	1
Non-qualifying Categories	1	4

Under alternative B, where the pre-coded answer options — known only by the interviewers — were 4 to 1, favouring the non-qualify category, readership was 30% lower than under the other alternative.

## WHY WE HAVE PROBLEMS

Clearly, we have established, then, that changes in almost any aspect of our measurement system can produce major changes in audience levels.

Therefore, something is amiss and we should understand what it is, if it helps us on toward solving our measurement problems.

A major part of our problem, I believe, is our attempt to include in average issue audiences all people who fit our simple definition of readership, namely, 'all individuals exposed to the inside of a magazine.' However, changing our definition of readership prior to knowing what parts of our current definition are

troublesome would be less than judicious.

When we consider some of the other major problems we face in establishing average issue audiences, the problem of memory has shown that recall or recognition of even an important event, such as being a victim of crime, is subject to substantial memory distortion or loss. **(10)** Nevertheless, we require readers to remember reading events across time spans that vary from 1 day to 3 months.

A partial review of research on memory provides a number of clues about why we have problems when we ask respondents to remember events such as reading, which usually have less than earth-shaking significance.

At the most general level, we can say that memory of readership events is less than perfect because:

1. Memory of even important events is known to suffer distortion and loss..and important events are remembered better than unimportant ones.
2. The ability to retrieve past events from memory decreases as the length of time since the past event increases.
3. If people pay attention to an event, they remember it better than if they do not. **(11)**
4. Memories of past events are sometimes modified by subsequent events. **(12)**

A large body of reasonably recent experimental work has contributed to some new theories of memory. These theories generally separate the subject memory into three operations.

ENCODING	STORAGE	RETRIEVAL
Registering an event	Maintaining what was registered	Later reporting on the event

In the past, discussion of faulty memories usually concentrated only on 'storage' problems. But research has shown that memory distortions or failures can occur in connection with any of the three operations; encoding, storage or retrieval.

Some of the most interesting work which, I believe, has relevance to measuring readership deals with a phenomenon called the 'context effect'. This effect can be expressed as follows: memory is facilitated if the context in which retrieval is attempted is like the context in which the event occurred.

## Mood and memory

Extensive experiments have established that to some degree, memory is mood-dependent. As shown in the following table, events experienced when an individual is in a particular mood, such as being happy or sad, are better remembered if the individual is in the same mood when recall or recognition is attempted.

## 2.1 Magazine readership and its measurement in 1983

Recall Mood	% RECALL OF LEARNED ITEMS	
	Learning Mood <i>Happy</i>	<i>Sad</i>
Happy	78%	46%
Sad	45%	80%

This connection between mood and memory is extensive. It has been shown to be significantly related with such diverse activities as recalling word lists, remembering childhood experiences and recalling recent emotional events. (13)

Other work has shown that memory is also facilitated if extrinsic factors such as physical surroundings are the same for both encoding and retrieval. (14)

### Using landmark events

A landmark event is an occurrence of some consequence to a respondent which clearly marks the beginning of a reference period for him. Numerous experiments have shown that the use of landmark events has a major positive impact on recall.

One experiment used as a landmark the well-publicised event of a major volcanic eruption of Mt. St. Helens, which occurred six months prior to the experiment. The question involved the timing of the US rescue mission to free the American hostages in Iran, which had in fact occurred more than six months prior to the experiment. Individuals were asked whether or not the US rescue mission had occurred within the last six months, and they were also asked whether it has occurred since the eruption of Mt. St. Helens. Twenty-three percent of respondents incorrectly said the US failure to free the American hostages occurred in the past six months but these same respondents correctly said this had not occurred since the eruption of Mt. St. Helens.

The article containing this experiment was named after another experiment showing the value of landmarking. The title of the article was, "Since the eruption of Mt. St. Helens, has anyone beaten you up?" (15) It showed that landmarking even improved people's recall of whether or not they had been physically attacked in the past six months.

### Respondent performance and interviewing variables

Our final reference to relevant research on memory is a fascinating experiment reported in an article entitled, "Striving for Response Accuracy" by Cannell and others. (16) Five levels of respondent instruction were tested to measure their impact on respondent accuracy in reporting personal medical information.

#### GROUP

**A** Control (no special instructions)

**B** Special instructions to respondents

**C** Special instructions and positive feedback when recall was achieved

**D** Special instructions and respondent commitment to answer accurately and completely

**E** Special instructions and respondent feedback and respondent commitment

In general, accuracy improved step-by-step as results are compared from Group A to B, from B to C, etc. For example, respondents associated with Group E supplied 20% more information about their medical history than did those associated with Group A.

Based on just this limited review of recent research on memory, we can easily put memory at or near the top of our list of mischief makers. However, as we will see shortly, this research suggests some possible directions to take to solve our problems.

### Model biases

Another probable source of difficulty in audience research is model bias. The Through-The-Book method does not recognize first time issue reading after about five weeks in the life of a weekly and after about three months in the life of a monthly. Studies of the inventories of magazines in public places in the US show that a non-trivial component of this inventory consists of issues well past the age when they would be used in a typical Through-The-Book audience study.

The recency method, of course, assumes that publication by publication, both weeklies and monthlies which generate few or many average reading days per reader — all exhibit replicate and parallel reading at levels which cancel out each other. Einstein once said that, "God is not malicious," but I doubt that He is this kind.

### SOME SOLUTIONS

So far, we have defined readership, reviewed our measurement problems and discussed some of the reasons these problems exist. The final ideas I present address themselves to some of the directions we might

## 2.1 Magazine readership and its measurement in 1983

take towards obtaining some solutions.

The goal as I see it is to define and measure readership with a method which, with reasonable consistency, yields audience estimates that fall in the same ballpark as 'TRUTH'.

We may ultimately decide that truth is beyond our grasp: that reading and readers are so diverse and the event is so ephemeral that we cannot come suitably close to estimating actual average issue audience levels.

We may have to settle on selecting a method which yields acceptable (rather than correct) results to the users of such data in each of our countries. We would then proceed to improve our various approaches to eliminate or resolve known inconsistencies and biases. In short, we may have to give up the quest for truth and only pursue reliability.

Another approach would be to replace average issue audience with some other measure such as reading days or magazine page exposure. Presumably, we can measure those dimensions more accurately than average issue audience. But it is premature to give up on the concept of audience reach and frequency and settle for estimating only gross rating points.

In fact, making any of these compromises at this time is extremely premature.

So the balance of my comments deal with some avenues we can pursue in quest of a reasonably valid measure of average issue audience. Here are eight specific suggestions:

**1** We have yet to directly measure underclaiming for the recency method. This should be done to establish

minimum underclaiming levels. The design would consist of observing reading behaviour and subsequently measuring readership after a time period equal to the publication interval.

Variations in the design would deal with using issues of varying ages at the time of the observed reading.

**2** The First Time Read Yesterday method shows some promise as a candidate for becoming a validity method. This method should be further examined through the experimentation discussed by Steve Douglas in session 3a.

**3** If yesterday recall shows promise, then using 'Today Recall' should be better.

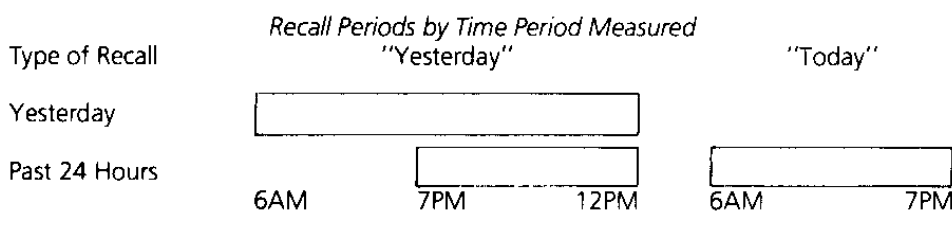
As shown in **Figure 1** and **Table 1** below, if we interview respondents at 7PM in the evening and query them about the past 24 hours, the average elapsed time between the interview and the behaviour being measured is 12 hours. With yesterday recall, the average elapsed time is 29 hours.

In May of this year, we at Audits & Survey completed experiments designed to find a method to maximize recall of leisure activities. We found that today recall was significantly more accurate than yesterday recall.

This certainly suggests that the accuracy of the first-time-read-yesterday method can also be improved if the average length of the recall period is reduced from 29 hours to 12 hours.

**4** We also found in measuring leisure activities that recall was maximized when we had respondents reconstruct the past 24 hours around events which they thought to be significant.

**FIGURE 1**



**TABLE 1**

Type of Recall	Memory Period in Hours	
	Average	Range
Yesterday	29	19 to 39
Past 24 Hours	12	0 to 24

## 2.1 Magazine readership and its measurement in 1983

In most consumer research involving yesterday or the past 24 hours, the time span is split into day-parts based on meals and the times between meals. We have found that this maximizes recall if the activity being measured is consumption of liquids such as soft drinks or beer. This is in contra-distinction to our findings involving leisure research, where recall is maximized if respondents structure their own day.

Both of these findings seem to be predictable from the experimenting on memory we mentioned earlier under the heading of 'Context Effects'. A substantial amount of beverage consumption is centred around meal times but leisure time activities are not.

This, in turn, suggests that the work conducted in academic circles on Context Effects should be tested as part of an overall programme to develop a valid method against which commercial techniques can be evaluated. An obvious experiment would involve the use of the 'yesterday' or 'today' recall technique centred around having respondents provide a personally meaningful structure to the time period involved. An important variation of the basic experiments would have respondents first describe the surroundings associated with each self-structured day-part including their own mood. Only then would they be asked to report on magazine reading.

**5** Landmarking could well aid in the accuracy of reporting under the recency method if the publication interval expressed as the last 'x' days is associated with a landmark of consequence to the respondent.

**6** The experiments reported by Cannell and others dealing with interviewing variables appear to be directly translatable to a readership experiment involving measures of both overclaiming and underclaiming. Their basic approach seems promising and appears to be equally applicable to all existing measurement techniques.

**7** In Session 4 *Time Inc.* reports on some developmental work on a validity method which is based on electronic or sonic measurement devices. This avenue of research should also be pursued.

**8** One major problem area we have not mentioned is the frequent occurrence of surprising results and confusion which follows procedural, methodological or question changes which are made in national surveys without any or without proper experimentation. This should be the easiest problem to solve. Untested changes should be outlawed.

### IN SUMMARY

The solution to our audience measurement problems could well come about by directing an imaginative and thorough frontal attack on the problem of validity. We

would then know how to define readership to include those kinds of readers and those kinds of reading which can be accurately measured.

At the present time, I feel it would be premature to redefine readership because we do not know how our current definitions contribute to measurement problems. It would also be premature to give up on validity testing and turn either to striving only for consistency or to some measure of reading that does not involve average issue audience.

Until recently — in the interest of improving the reliability of our estimates — we have spared almost all effort to obtain a valid measure of readership. I recommend that in the next few years, we spare no effort.

The interests of the industry, I believe, lie in having a valid measure of average issue audience. In fact, it seems not unlikely that — if advertisers knew as much about the subject of magazine audience research as we do — they would absolutely demand that we develop an audience measurement approach of specifiable validity, as well as high reliability.

### BIBLIOGRAPHY

- 1** Turner, CF and Martin, E, eds., (1981). *Surveys of Subjective Phenomena. Summary Report*, National Research Council, National Academy of Sciences, National Academy Press, Washington, D.C.
- 2** Lysaker, RL, (1982). The ARF certitude tests, in *Readership Research: Theory and Practice*, Henry, H. ed., Sigmatext Limited, London.
- 3** 1982 Audience Reports of MRI and SMRB
- 4** Bird, M, (1982). Does the mother of national readership surveys need rejuvenating?, *Admap*, May.
- 5** *Measures of Magazine Exposure, Certitude Test No. 1, Laboratory Waiting Room*, Advertising Research Foundation, New York, 1978.
- 6** Denon, L, (1982). The effects of changing the filter, in *Readership Research: Theory and Practice*, Henry, H., ed. Sigmatext Limited, London.
- 7** Whitley, E, (1982). Some rotation effects in the British survey, in *Readership Research: Theory and Practice*, Henry, H, ed. Sigmatext Limited, London.
- 8** Langschmidt, W, (1978). *Reliability of response in readership research*, S.A. Advertising Research Foundation, Sandton, Republic of South Africa.
- 9** Tennstädt, FWR and Hansen, J, (1982). Validating the recency and through-the-book techniques, in *Readership Research: Theory and Practice*, Henry, H, ed. Sigmatext Limited, London.
- 10** Bailar, B, (1982). Nonsampling research errors at the Bureau of the Census, in *Improving Data Quality in Sample Surveys*, Wiseman, F, ed. Marketing Sciences

## 2.1 Magazine readership and its measurement in 1983

---

Institute, Cambridge, Mass.

**11** Salamaso, P, Baroni, MR, Job, R and Peron, ER, (1983). Schematic information, attention and memory for places. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 263-268.

**12** Loftus, ER, (1981). Mentalmorphosis: Alterations in memory produced by the mental bonding of new information to old. In J Long and A Baddeley (eds.) *Attention and Performance IX*. Hillsdale, N.J.: L. Erlbaum Associates, Inc.

**13** Bower, GH, (1981). Mood and memory, *American Psychologist*, Vol. **36**, No.2, 129-148.

**14** Baddeley, AD, (1982). Domains of recollection, *Psychological Review*, Vol. **89**, No.6.

**15** Loftus, EF and Marburger, W, (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events, *Memory and Cognition*, Vol. **11 (2)**, 114-120.

**16** Cannell, CF, Oksenberg, L and Converse, JM,(1977). Striving for response accuracy: experiments in new interviewing techniques, *Journal of Marketing Research*, Vol. **14**, 306-15.