

8.1 Reading and 'Readership' – can the correlation be improved?

HISTORY

In the years before 'readership' everyone in the business lived in a neat physical world, and dealt in a currency which all understood – the number of copies of an issue of a publication which were sold to the public. The situation was like science before Heisenberg – before crystalline predictability was clouded over by uncertainty and probability.

The first relevant survey in Britain was indeed called 'Press Circulations Analysed' (1) and was much more geared to giving a demographic picture of each publication's audience than towards estimating *how many* people saw each issue.

At that stage we are on the other side of one of the two major divides – *time*. Gradually the advantages of being able to talk about the *people* who read a publication began to tempt more and more to leave the featureless, straight, 'canal of undelineated certainty' and cross to the more scenic waterways which readership surveys could represent.

The second great divide is probably a cultural one, but certainly a geographical one. Europe adopted one set of dominant approaches, North America another. It is reasonable to talk about different survey cultures. I offer to anyone interested the idea of a thesis on the legal, political, scientific, literary and social backgrounds which give rise to the mystic acceptance of 'recent reading' in one continent, and 'through-the-book' in another, and assessing the impact of such influences, as well as those of powerful personalities and power-blocks and simple accident.

WHERE WE ARE NOW

Both survey cultures are thus rooted in a collective past – the concept of the 'average issue audience' stemming from the circulation criterion. If publications had been invented *after* the broadcast media can we believe that we would have attempted to measure, directly, the numbers exposed to an average issue throughout its life?

Further, if publications had been invented at a later stage in the development of *survey research* would we measure readership in any of the ways so far generally accepted in either survey culture? This is not to imply that broadcast media audience surveys are perfect.

It would be highly likely that an industry accustomed to broadcast measures would say that Average Issue

Audience would be too remote from their needs (an indication of exposure to white space) and so unlikely, conceptually and practically, to be measurable, that it would be pointless to try.

Do we continue with Average Issue Readership because it is the only possible way, because it is the cheapest way, or is it perhaps because it is what we are used to?

Originally, because of the circulation framework, survey research was told what the desired currency was, and went out to measure what it was told. It was rather like one of those old legends where would-be heroes go off to find some mystic relic. When they ask the elders what it is they seek they are told 'Our ancient rites forbid us to tell you – you wouldn't comprehend it until you found it'. What they fail to add is 'and nor will we'. What they do, have, and will recognise, are the various results of applying certain words and/or pictures to samples of the population, later counting the numbers who respond in agreed ways. They will recognise it as 'Readership'.

Once the package of stimuli has been applied, and the results analysed for a period of years without serious challenge, it becomes the accepted norm for trading. It is important to get this into perspective. If one continues to trade only nationally, does it really matter that other survey cultures use different techniques and hence different operational definitions of readership?

WHY SURVEYS?

Let me go a stage further. We here are dominantly concerned with the measurement of 'readership' and thus have a vested interest in continuing to do it via surveys. What about the people who buy and sell media? Picture yet another culture where a team of experts reads through a sample of issues of each publication and opines on the pattern of its editorial appeal in terms of age and sex. Analysis of content for average word and sentence length would be the basis of each title's class or cultural positioning. Used in conjunction with circulation figures would the target effectiveness of schedule choice based on such criteria be very different from its current level?

A PERSONAL NOTE

We have listened to learned discussions of the effects of group order and of rotation, filters, of the numbers of

8.1

Reading and 'Readership' – can the correlation be improved?

boxes an informant can fit himself into, and title confusion; elaborate ways in which our data can be analysed; ways of categorising readers according to their relationship with the publications they read; and so on.

But there has not been much discussion of the fundamental issue. That is: why should we *expect* any of the questions we put, and answers we get, to have any very close relationship to the actual behaviour we want to know about?

In my first job in 1955 I remember asking why one would expect the number of 'readers' in the last week to represent the readership of an average issue (because of what is now known as replicated and parallel readership) and being told not to be silly. Soon after that, I moved into publishing, and have been involved in practically every method of assessing 'readership' known to man. I have validated readership by page traffic, studied rotation effects, attempted to trace through people's daily life and probed when reading was mentioned, repeated Politz's original car load studies, carried out 'first reading' studies and the first published reading frequency work in Great Britain (2), used interviewers disguised as assistants to establish hair salon readership by observation or counting broken glue spots, as well as studied variations of wording, prompt aids, etc. All that within the company. In addition, I have been on the Technical Sub-Committee of our national survey since 1964 and have been involved in all its experimental work during that period as well as the press owners' corporate experimental work. Thus if my remarks seem naïve, it is a naïveté borne of some experience.

However, through all this time the measurement of audiences has not been the major part of my work. That has been research into the editorial appeal of newspapers and magazines.

THE VALIDITY OF SURVEY DATA

In my view these two sets of experience have as much to suggest about survey research in general as they do about readership measurement. By and large people will answer any questions we put to them. The fact that some survey questions produce estimates which tally with independent non survey observation may cheer us, but it should not cause us to believe that it is for the reasons we expect. Nor does such cross checking give any basis for equal confidence about the things we cannot independently check. The point is not an academic one. It has hard commercial relevance.

Readership research suffers more than product research in that there is no clear unambiguous act which is measurable or definable in other than survey terms.

In much survey research total accuracy or 'truth' is

not a dominant requirement. We are often seeking only a working model of interaction and process. Even if it has only as much relationship to 'reality' as a toy train to the real thing, it may represent as much of the market dynamics as we can hope to adjust or control, or possibly even more. It is thus operationally 'right' even if not 'true'.

Readership surveys provide input for very simple conceptual models which will work better with some data than with none. They are much safer than market surveys because if the sauce advertiser puts a bit too much spice in his *schedule* (say *Playboy*) and a bit too little sugar (say *Reader's Digest*) he won't go off the market, and nor will the two titles. If his *product* research is wrong and he puts too little sugar and too much spice in his sauce he may well take a bad knock.

But we all know he wouldn't rely on survey research alone. Its results would be appraised for congruence with other knowledge, albeit some of it also faulty. In readership measurement there is no outside criterion, only another survey. Other surveys will be rooted in the same belief system about how to measure, and what it is that one seeks to measure, so even if they have different design systems in *detail* (like TTb vs RR) there is no rationale for endorsing one rather than the other apart from emotional or historic preference. All one can say about most of the attempts at validation that one knows of is that they corroborate claims of a particular section of the ostensible audience – often likely to be a minority (such as prime buyers), or a minor part of reading (such as observed public reading or experimental reading conditions). The often-quoted Belson study (3), 'helpfully' cross-examining people in a second interview, did nothing, in my view, to suggest what the right levels were. It was invaluable in showing that discussion of experience in a manner often peripheral to the personal concerns of the informant can get the memory to turn cartwheels, as any good lawyer knows.

MEDIA IN PEOPLE'S LIVES

This may be blindingly obvious. It took me some years of editorial research to realise that the detailed things people say about publications – whatever form of interviewing one uses – were not really sufficient to explain why they buy and read to the extent they do, *nor explain their expressions of overall enthusiasm*. Many people will always be able to express themselves in ways which give rich clues, but many don't or can't. It is not a problem of verbal skills, or of education. On the other side of the fence, editors can be very expressive about why they think something will or will not appeal. They also have models of their readers. However, as they are dominated by their task, just as researchers are, and by thinking about what

8.1

Reading and 'Readership' – can the correlation be improved?

they have *done*, or might *do*. (4) Thus, in both editing and researching, whether about audience numbers or reaction to editorial the dominant frame of reference is the publication as object.

With temporally fixed media the preferred audience measurement systems are contemporaneous with the experience. That is to say the measurement is simultaneous, or prior, to the media experience being absorbed and integrated by the viewer or listener. The medium is still relatively, 'object' when measured.

By the time we get round to measuring press audiences, the exposure we are after may be several layers down in the mind's jellygraph, and interlaced with a schematic set of related constructs and attitudes, much of which has nothing to do with publications of any sort. This is exactly the problem one has in researching response to editorial. Whatever specific questions one puts tend to conjure up publications as objects more than as experience.

The following is only a brief summary of points to be developed elsewhere. The character of much of *life* experience, particularly routine experience, is such that it is not neatly filed in the conscious memory. It melds into existing mental schema without any specific time or source coding. Perhaps more important it serves as an indicator of lapsed time, and of personal time location. Life's daily experience tells us we are not stationary in time. My contention is that because we think and talk of media as objects we drastically underestimate their rôle. We think of them as channels down which are communicated 'bits of data' which are consciously received, and hence consciously and separately stored. Our questioning approaches can only be explained on the basis of such a set of beliefs.

My belief is that the bulk of media experience (apart from the deliberate search for particular information) is not intrinsically different from real-life experience. It may sometimes be less intense than life and it may well carry a 'colour coding' which treats it as tints rather than hues or shades, but much of it fulfils similar functions. But for a lot of people a lot of the time the accumulation of possibly 'unreal' media experience will outweigh *in sheer quantity* of 'bits' the real experience they have. For most of the rest it will at least be substantial.

RECALL OF MEDIA EXPERIENCE

If this hypothesis is *correct*, it goes some way to explaining the problems we have in trying to access media experience in surveys. While the *general* source may be more recallable (ie newspapers, magazines, TV), the *specific* origin ie, *which* magazine, *which* issue, and/or *when*, ie *media events*, will be directly recallable

according to the extent to which the experience has *not* been integrated into existing schema. If we accept that much media experience has the character of life experience then we must face the fact that personally significant reading occasions may be recallable less well than relatively insignificant ones.

Thus if people answer *event* questions *literally in the terms that they are put*, in recent reading the last reading occasion identified will be one where the experience was so uniquely coded that it couldn't have come from any other source and probably has not been fully integrated into existing schema. In practice, there will be a lot of different experiences within the issue, and so if any are recalled as occurring within the last week or month the informant *should* qualify. But some people will have no recallable title – or time-coded experience even though they saw in the appropriate period, and these will be missed.

Through-the-book *should* fare better in these terms. Are we not showing the content of a variety of issues, to prompt recall of the initial experience, so that it might seem analogous to personally metered TV? Yes, but there are four problems: (a) much editorial material is similar in appearance to that in other issues/titles; (b) many topics (if not also their appearance) will be featured in a variety of titles so people who limit their reading to a narrow range of topics may not be able to identify it when they see it; (c) integrated experience not consciously linked with a particular title will have been screened out by the modified TTB filter; (d) advertisement experience and some editorial experience will be screened out in modified TTB because of stripped issues.

Of course, people do not respond literally to the wording of the questions. They scan their habits and use other cues. But insofar as much discussion centres around the literal differences in question wording and structure, even at this level there is good reason to expect readership 'loss' from both of the prevailing methods, and also artificial 'gain'.

WHAT DO WE DO IN AN INTERVIEW?

We start by telling informants we are finding out what people read. From then on what we get are entirely functions of how the informant relates his behaviour, and recall of it, to what we ask him. The prevailing model is that there is a set of recallable behaviour 'in there', and our questions elicit it. It would seem more likely, however, that what we really do is impose upon him our structure, of how we think about reading (because of commercial requirements and tradition). This applies to all survey questioning, even the best. We interpose a cataracted, astigmatic lens between his conscious thought-accessing

8.1

Reading and 'readership' – can the correlation be improved?

and the memories and habits he is trying to access. This happens in areas which are very clear-cut, like the buying of bananas. In an area where there is no absolute objective generally understood definition of the activity we are trying to recapture, such effects must be at a maximum, and we are almost entirely structuring the informant's scrutiny of his experience in our terms rather than his.

Thus when we say 'read or look at', 'When did you last', 'how many issues out of six', 'have you seen this copy?', we are imposing frameworks of our own, which may have little to do with how our informant sees his experience. Our prime interest is in classifying people as 'readers' or 'non-readers'. If we are to have any chance of doing this well, our approach must be to use as clear and even a lens as possible, to minimise the generation of artefactual experience, and the wiping out of actual.

ROTATION

Rotation variations have been a major cause of worry about the British recent reading technique. However, all the discussion of rotation that I can recall compares the results in terms of the total number of AIR (or 'any frequency') readers each rotation yields. In effect, therefore, the readers from one rotation are regarded as being the same as the readers from any other. Clearly, however, they must be different in some ways. The usual discussion implies that they are only different in the way they respond to the technical aspect of where the title comes in the set of prompt-aids they are questioned with.

I would suggest that we perhaps start to look at the 'extra' readers from high-yield rotations in terms of other data we have about our informants, to find out who the added readers are, as well as trying to find out why they are added, and then whether we want them! It is counter-productive and buries what we are interested in to do it for groups of titles.

While such analyses are worth doing, a fundamental question is begged, upon which light might be thrown only by inference. We tend to assume that 'high' or 'low' yielding rotations simply add or subtract (depending on one's standpoint) readers. Past analyses of rotation, and those suggested above, can only tell us about the *net* difference. Cumberland Lodge (5) recognition tests, and the Canadian PMB (6) work both suggest, however, that some people move in or out according to the technique used.

Rotation variations may well be subject to similar principles, further confirming that our data are in some part generated by our stimuli i.e. prompt-aids, questions, interview method, etc. Order of presentation of groups of titles may operate in at least two ways: expanding or

contracting the perceived total number of titles read; and, allocation of an experience to one or other title.

ASPECTS OF THE PROBLEM OF READERSHIP ESTIMATION

Saliency of title vs saliency of experience

Because we want to find out about particular titles we ask about particular titles. Because of correct or incorrect assumptions about people's perceptual frameworks, memory schema, etc, we ask about titles one at a time. If we were simply trying to establish for each title whether the informant had ever heard of it, or ever seen it, this might be reasonable. But we are asking him about his past behaviour in our own set of terms.

His chance of giving the *same* set of answers, whatever the order titles are presented in, depends upon:

- (a) all the titles he reads having *high* saliency.
- (b) saliency being *equal* across all the titles read, and across all titles not read (ideally across all titles).
- (c) all the titles he does not read having high, or very low, saliency.

Clearly, these conditions are rarely, if ever, met. Any book on perception is full of examples of how perceived experience varies according to expectation and preceding and contextual stimuli. Indeed, the more one thinks about the way we do surveys, one would be highly disturbed about the effect on the informant's mind if we did *not* get variations according to rotations and techniques used. Some evidence of the effect of preceding stimuli has been indirectly adduced in the 'Drive' study (7) and by Tom Corlett (8) after he found similar patterns in his rotation studies, i.e. answering frequency questions within the framework of the *preceding* publication frequency period.

But I doubt if we have sufficiently taken on board the implications. Limiting them for the moment only to titles, each rotation offers, and then eliminates from further consideration, titles in varying sequences. This alone means that the frame of reference and perception of the task is progressively changed as the interview proceeds. For example, 'Now I've been through the ones I read, I wonder what I'm going to be asked about next?' or 'when is she going to mention something I've heard of?'

In Britain we have had some discussion of the possible need to handle different titles in different ways. In this we are acknowledging that 'objectively' identical treatment may not be equally fair to all titles. It is just as reasonable to posit that rigidly identical treatment of informants may be just as 'unfair'. This is another point which is occasionally acknowledged and then shelved. While it is hard to envisage a quantifying survey based on 'what would you like to tell me about what you read and

8.1

Reading and 'readership' – can the correlation be improved?

how often you do it?', do we perhaps normally go too far in the other direction and impose an unnecessary rigidity.

Recall of reading events – a hypothetical readership survey

Let us assume, artificially, that answers to a readership survey would be based solely on *recall of specific reading events* and ignore any other factors such as knowledge of one's own habits, conceptually a quite different thing.

The salience of a *unit reading event* eg, an article, a news item, an advertisement, etc, will obviously depend upon the emotional or factual interest it has, its novelty, its distinctness.

The salience of a *publication reading event* will depend upon the quantity of such interesting items absorbed on one or more occasions of exposure to that same issue *plus* the extent to which they cohere in the mind as being found in the *same* issue of the *same* publication (whether or not its title is recalled).

The salience of a *specific title reading event* will depend therefore on:

- (a) the interest of each unit reading event.
- (b) the aggregate level of interesting reading events.
- (c) their original, and recalled, 'family' connection.
- (d) the distinctness of this connection from other 'family' connections found in other titles.

If each reader reads an equally representative cross-section of the total content of each title then we could hope that at least the claimed title would be in the right *group* of titles even if not the correct single title. But readers tend to follow their own particular set of interests and pay attention to other content. Thus the reader's experience from two titles in groups normally considered quite distinct may be quite similar for him or her. This has two important implications. Firstly, the groups we put titles in may be quite irrelevant for some readers. Secondly, the chance of locating a reading experience in the right title may be low.

To bring these points nearer home, how often do we hear 'I was told about this scandal at the Town Hall the other day – now who on earth was telling me?'. The situation is strictly analogous. However, in conversation there may be sufficient time eventually to recall the name. In an interview there usually is not.

The foregoing relates to problems we might expect to encounter if our survey offered no title prompt-aids and was based entirely on recall of reading acts, and in this section we are ignoring entirely the placing of events in time.

Actual readership surveys

Of course, we don't work without aids and other questions, but it is a relevant starting point to consider what we do do. Additionally, at a minimum, we offer:

(a) some kind of a description of what we mean by reading, etc.

(b) mastheads, or covers, or stripped issues, or in former times, complete issues.

Going further we may also enquire about:

(c) habits, ie, frequency.

(d) when reading last took place.

(Here we are considering 'd' only insofar as it affects the informant's interpretation of the kind of act we are enquiring about – the accuracy of the location is a different point.)

Thus in practice, our set of questions operates in a very different way from the hypothetical survey based on recall of reading events. In reality, we offer the informant a set of jig-saw pieces to help him construct a model of what he does. But we never give them all to him at once. Some surveys first give him bits of edge, one section at a time, others might give him all the bits of sky, then some faces, and so on.

Whether we use habit questions or not the informant is at best erecting his own probabilistic model based on what he thinks he encounters, buys, has delivered, sees around the house, etc, and only partly on actual recall of reading events.

His model is also probably constructed within an overall, perhaps quite erroneous self-limitation as to the number of different magazines he imagines that he reads. On top of that, since the number of titles is unknown when he starts, the *number and range* of the universe of magazines and newspapers he is considering expand as the interview continues. Likewise, the *nature* of the definition of that universe will change, along with the size, *according to the order of titles he is offered*.

Accuracy of temporal location and of habit self-assessment

READING IN THE ISSUE PERIOD

The foregoing relates to people's chance of accurately recalling *whether* something happened at all. The recent reading technique requires that we also establish *when* something happened, and most techniques enquire about how often it happens.

Whereas in the British NRS we at least let the total universe of titles being examined eventually emerge (albeit in different orders) so that the informant knows by the end of the first filter question all that we want to know about, we never tell him the *timescale* in which we wish to know about *when* he last read. The Cumberland Lodge (5) variants did so, but perhaps not in very normal terms, eg two days to one month ago.

Another practice which is the inevitable consequence of not letting him know the periods is the use of *when last* read. We don't know how people understand it *in the*

8.1

Reading and 'readership' – can the correlation be improved?

context of the NRS. This and many other traditional survey practices fall under the heading of researchers' fear of 'conditioning' and 'conditioning to negative'. We have, of course, no idea of how much what we currently do conditions to positive!

Accurate location in time is of differential significance for different groups of readers. People who know that they receive or buy *every* issue of a title and read it at all will almost inevitably claim some reading sufficiently within the issue period. Those who claim high frequency and not to have read in the last issue period are anomalies of the RR model (ie they are no longer current readers) or simply don't recognise the title when it is *read out* at the AIR question (long after their response at the filter stage).

The problem really arises with intermittently regular and with occasional readers. While these are also the groups where replication can operate this is not here the point at issue. From the point of view of location in time of the last reading occasion, the question we should really ask ourselves is why we should believe there is any chance of people being able to do it accurately, particularly beyond or within the last month.

The attempt we have made in Britain to sub-divide reading *within* the issue period does not appear to have worked, in that *last* reading yesterday appears to be far too small a fraction of reading in the issue period to be acceptable.

We must admit that the question may be at fault, and that '*last read*' may in itself be a concept which implies greater historic distance than is associated with '*yesterday*'. This is not necessarily incompatible with the hypothesis sometimes advanced that people bring reading forward over the issue period hurdle. Even though people don't know what the hurdle is, there may be a tendency to centralise reading occasions within a band of time perceived as appropriate. Indeed, if some title relationships are perceived as attitudinal self-attributes then very recent or non-habitual behaviour may be screened out until it has become sufficiently historic or repeated to be incorporated into the self-image.

While it is somewhat chimerical to judge the reasonableness of data sets by theoretical mathematical models, the only avenue likely to be open is the relationship of reading claims during progressively longer intervals of time, eg 'in the last seven days' to 'over seven days, in the last month', though not *asked* in that way.

FREQUENCY OF READING

Here we are on somewhat stronger ground particularly in that there are potentially two types of experience with a good chance of giving similar results.

Firstly, we have a simpler reading act to recall, or rather an act to recall which corresponds more closely

with what we ask about because we are asking not about precisely *when* a reading act took place, but simply whether any issue was read, and how many. This point probably has more weight if the question relates to a number of issues rather than a time period.

Secondly, it is quite feasible for the informant to solve the particular puzzle which this question represents by reference to the number of issues he believes pass through his hands or home. We may not like this but I think we would have sizeable data gaps if we were able to exclude answers given on this basis. I would hazard a guess that cumulative coverages would not differ much whichever method the informant used.

We still have the problem of whether we should attempt to assess the 'current' rate, or the historic rate (number read nowadays, or number out of last six issues). There is a strong case for limiting either the period or number of issues – not necessarily the four-cell split based on last but one and/or last one. It might be that a compromise such as 'every issue/about every issue/less than that/none' would be more realistic perhaps in a 'Do you generally see' setting.

Filters

The nature of the first filter is obviously critical and one would only wish to say that it should be very broad – perhaps ever seen in the last year, whatever terms may be used for later questions.

Informant load

There is no sensible way of maintaining an equal *total* interview load. However, I think that there is a strong case for an equal load per title passing the initial filter. I do not hold this view because of any belief in 'conditioning to negative' through informants wishing to 'avoid work'. The data which led to that hypothesis could stem just as much from cumulative redefinition of the task as the informant perceives it. Rather I feel it should be considered as part of the process of giving equal total attention to each title worthy of being questioned about, ie all passing the initial filter.

Prompt-aids

My conclusions from studying the RR and TTB-A sample of the Canadian PMB study (6), subject to confirmation, are that mastheads alone are not equally effective filter aids for all titles.

The alternative prompt-aids, reproduced covers and stripped issues also perform differently. Reproduced covers are used as the filter which allows an informant to get to the stripped issue readership question and the frequency question in through-the-book. It seems likely that stripped issues are in some cases a worse prompt-aid than covers, possibly partly because of the age of issue

8.1

Reading and 'readership' – can the correlation be improved?

problem.

The differences between the PMB through-the-book filter levels (using two covers) and the RR frequency filter levels using mastheads vary substantially according to title, and cut across weeklies and monthlies.

Further, it would seem undesirable that any later questions about a title are posed in the absence of the prompt-aid. This need not mean completing all questions about a title before going on to the next.

Criterion 1 – acceptability

Historically, the dominant criterion. 'Success' is more likely if the technique 'looks' reasonable and if the results are not oversold (particularly when there is survey competition). If the industry stopped talking about 'readers' and presented the results as a 'Press Negotiation index' we would delude ourselves less, and conflicting results from different sources might have a better chance of being profitably understood instead of shaking people's mental foundations.

Criterion 2 – subjective evaluation

The first way in which we might better understand the results of our survey(s) would be to think through how real people with various reading experiences might be expected to respond to the questions put, hence what our figures might represent. Some such thinking exists in survey *design*, but more is also needed at that stage.

Criterion 3 – calibration/description

Here one has in mind supplementary survey work which would better quantify or at least describe, the profile of behaviour and relationships which are included in the 'Index' level (and, hopefully, which *excluded*). Examples are to be found in the JICNARS '1968 Development Work' (9) and 'Reader Categorisation Study' (10).

Discussion at the symposium prompted the writer to draw attention to the wastefulness of many studies where gentle exploratory discussion of the responses to the basic interview, *as soon as completed*, would explain a lot more than reinterviews on another occasion with a different method. The result enables a within-person comparison. Even more extravagant is the application of different techniques to different samples for no within-person comparison is possible – only net differences emerge.

Criterion 4 – validation

Already referred to on page 447. One would add here that the elements of 'audience' which are 'in' or 'out' according to technique are likely to be those most susceptible to the different operational definitions which any validating technique must constitute.

Satisfactory grossing up of circulation cannot be a validation of non-buyers' reading, and thus gives little

ground for confidence or depression about the total audience estimate.

The 'intensive interview' as a means of validation

There was a fairly strong reaction to this paper's comments on the intensive interview approach, not surprisingly from Dr Belson and the following should be added to explain the position taken. (Essentially the method means re-interviewing the informant about a few titles, both 'read' and 'not read' (see reference three for full details). Dr Belson refers to it as a 'detective' process.) The justification for a detective type interrogation is that there are 'facts' known to the person questioned. The detective has to cross-examine to aid the memory or see whether the person is lying. Nor is he necessarily worried about whether the informant realises that he may be trying to prove something!

To apply this technique to readership is justified for, and only for, those memories which are *deliberately* falsified or withheld from the interviewer at the original interview, or if the informant grossly misunderstood the first interview. This is one unstated assumption of the intensive interview.

The second assumption is that all, or all significant, readership events are recallable with a title, time and/or issue coding.

The alternative thesis is that much reading will be forgotten, or that its issue or time coding will be lost by the time the first interview is carried out. Attached to this is the fact that the interview structures what acts are to be treated as operationally relevant 'reading' in the first interview. That is to say, the interview defines or redefines the behaviour inquired about. The result of the *first* interview is a *mix*, variables *between* titles *within* the same person, and *within* title *between* the members of the sample, of original behaviour, recalled behaviour, and operational definition.

To take the findings of the first interview and make a within-person comparison with a new form of interview represents a second and different form of operational definition, *especially* if only a few titles are singled out. Guesses or attempts to solve the puzzle at the first interview are changed or confirmed as a result of the new set of stimuli and puzzles which the second interview represents.

It is fairly easy to understand why the intensive interview appeals. Social scientists originally brought up in the context of verification (in the pre-Heisenberg model) in the physical sciences unconsciously or consciously apply the same model to the human sciences. The lability of much of their data is psychologically disturbing. They are perhaps sometimes the sort of people who would emotionally prefer specific location of motor and sensory

8.1

Reading and 'Readership' – can the correlation be improved?

function in the brain cortex to *relative* location.

The low importance or significance of:

- (a) the way we discuss reading in interviews.
- (b) some of the original experience or behaviour which we are asking about mean that we are foolish if we do not acknowledge that a part of our findings are effectively the constructs or artefacts of the techniques we apply to people.

DESIGN OF FUTURE SURVEYS – AVERAGE ISSUE STUDIES

If we insist on trying to estimate Average Issue Audience the following would appear essential if improvement is to be achieved.

Better understanding of task at all stages

It would appear essential that the informant knows the whole universe of titles being considered before any strict filtering occurs. (There is no good reason for supposing that the TTB-M first filter is more or less subject to title confusion than any RR first filter.)

Functional equivalence of prompt-aids

The only rationale for the stripped issue is that it has been used for a long time. If *issues* are wanted, then whole issues should be used, and even these may need additional aids. If *covers* are used, more than one should be shown. Some titles may need far more than others eg, if covers change a lot issue to issue. Further description may be necessary 'it comes every Saturday with the but it doesn't look like a separate magazine'. If mastheads are used even more description may be appropriate for some titles.

Assisting the informant

If we were less afraid of 'leading' the informant we could probably help them to be more 'accurate'. One way would be to ask assisting questions (not to be recorded) eg, 'If you see where do you see it? how do you get it? what sort of occasions? what do you like about it?'. Quite apart from possibly helping to reconstruct events, it would give the informant more time to think about which titles he or she sees.

Reading event definition

If one is after uniform response there is probably not much real choice other than between a very loose, vague definition (probably less intense in implication) and a tight one (probably more intense than any we now have in USA or Europe). Anything in between is more likely to be open to differential interpretation by informants.

Short recall periods

While not overestimating the chances of our questions really reconstructing events, there must be opportunity for improvement if at least part of our estimating base is limited to very recent time periods, minimising the amount of 'sourceless' and 'dateless' integrated experience, or forgetting.

FUTURE SURVEYS – OTHER APPROACHES

As indicated in other writings there is a case for considering other units than average issue, or ones where average issue might be a roughly estimated byproduct. Such measures might include Reading Days, and Page exposures per day, measured directly, not as qualifiers of Average Issue Readership.

The basic stance could perhaps be seen as starting to find out what we can measure reliably, and what it represents, and training ourselves and the market place to use it. It could offer a much better level of equivalence to the audiences estimated for other media. The resources are probably not available in smaller countries. They may be in the USA.

THE STANCE IN FUTURE DEVELOPMENT THINKING AND WORK

- (1) none of the foregoing suggests any easy new answers.
- (2) though the discussion has been in terms of techniques and effects, it has as much relevance to the goals or targets we set ourselves, as to the detail of method.
- (3) we must make our 'public' clear as to the level of improvement we can *hope* to achieve. In this sense they can better judge whether they really want it, and want to pay for it, or keep quiet.
- (4) putting this another way, we and our public must be clear whether we are:
 - (i) trying to design surveys without the defects of which we are *now* aware.
 - (ii) trying to establish *all* the defects which may exist and design surveys to eliminate them.
 - (iii) trying to quantify the importance of all the established defects and eliminate those beyond a certain level of importance.
- (5) we should try to agree on how much we are after stability, and how much after validity or, to put it in cruder terms, where we aim to be between acceptability and truth.
- (6) thus we should not oversell what we are providing.
- (7) equality of treatment across titles is a goal, but will never be fully achieved, even if publications only bore

8.1

Reading and 'Readership' – can the correlation be improved?

codes instead of names.

(8) we must question more vigorously some of the largely or totally unverified hypotheses advanced to 'explain' data such as 'fatigue effects', 'conditioning to negative', 'prestige claims', etc. If we do not we may block our chance of using techniques which might appear to suffer from them.

As an alternative to the concept of lazy unco-operative lying informants we have to recognise that they may simply be confused people put into a quiz game with no prize for getting the right answer.

(9) we should consider other measures of press exposure than direct estimation of AIR, if the reaction is 'we don't believe you will be able to improve AIR methods and we still don't like the current method'.

REFERENCES

- 1 *Press Circulations Analysed* (1928). London Research Bureau. The survey asked housewives which titles anybody in their household read. No prompt-aids were used.
- 2 *A new Measurement Study of Women's Weekly Magazines*, 1964, Odhams Press, London.
- 3 Belson, William A (1962). *Studies in readership* Business Publications for Institute of Practitioners in Advertising.
- 4 Allt, Brian (1974). *Research for Editors* ESOMAR Seminar, Verona, 'Editorial Research in the Publishing Industry'.
- 5 NRS Development. A report on the 'Cumberland Lodge' research programme, 1980, JICNARS, London. This compared the current NRS method with a 'display' questionnaire, and a card-sorting method. The last two showed the informant the time bands relevant to the recency question.
- 6 PMB pilot study, July 1978, for PMB III committee, Toronto.
- 7 NRS reading measures and the 'Drive study'. Report of working party on irregular publications. 1975, JICNARS, London.
- 8 Corlett, Tom (1974). *Unpublished report on Rotation Effects* for JICNARS, London.
- 9 Development Research for the 1968 National Readership Survey. 1975 JICNARS London.
- 10 Reader Categorisation Study: *Indirect measures of exposure to contents of publications*, 1972, JICNARS, London.