

# OPTIONS OR ANSWERS? WHAT ARE WE ALL DOING HERE?

Andrew Green, ZenithOptimedia

*“Media research is not designed to find out the truth. It is a treaty between interested parties.”*

Rodney Harris (quoted by Sigurd Bennick, Hong Kong Symposium)

*“...when you improve methodology, you put at risk a number of business models of companies that have flourished based on non-transparent market conditions.”*

Stephan Loerke, World Federation of Advertisers. July 2005

This paper could have been written about television or radio audience research. It is about the conflict between pure, researched standards and the glibby commercial considerations of the real world.

There is, for example, a clear and consistent methodological bias in the radio and television diary method, one which benefits larger, longer-established stations and penalises newer, smaller ones. Yet the diary continues to be the main form of radio audience measurement around the world, while the television diary survives in most of the US local markets.

Perhaps things are changing however. Over the summer, the US radio industry has been debating what journalists have dubbed ‘Version 2.0 of audience measurement’<sup>1</sup>. How, in other words, to update or replace the current diary measurement system.

The same debate has been going on elsewhere in the world – some markets, including Canada, Belgium, Singapore and Switzerland have already moved forward with electronic measurement.

In July Forrester published what it dubbed an economic impact study concluding that the US radio business stood to gain \$414 million of incremental revenue if it switched over to electronic measurement, while it would likely lose \$282 million if it failed to move ahead.

On the television side, also in the United States, the big commercial battle between companies who want to keep the ratings system unchanged because it benefits them and newer, smaller stations which tend to gain from more granular and accurate measurement has entered the political arena. The Fairness, Accuracy, Inclusivity and Responsiveness in Ratings Act was being discussed as I prepared this paper. As the President of Comcast Spotlight, America’s largest cable operator commented:

*“While there has been a persistent effort by certain broadcasting companies – and organizations they have funded – to portray this debate as affecting the public interest, the simple fact is that those companies are trying to hold back technological progress because it threatens their bottom line”<sup>ii</sup>.*

Only in America you might say. Or will we in the print measurement business have US Congressman at our next symposium? Will Erhard Meier be asked to give evidence at a Senate hearing? Will they start looking at what we do with the kind of scrutiny broadcasters are starting to come under after so many years of basically getting away with it?

Michael Brown states in his book *Effective Print Media Measurement*<sup>iii</sup> that the start point for readership research is ‘to provide a “currency” in which advertising space may be traded and which is mutually acceptable both to sellers... and to buyers... That is to say, any readership survey’s first aim is to put a credible valuation on the advertising opportunities provided by print media.’ (my underlining)

He goes on to state: *‘We select some particular (readership) definition in the light of the uses to which the data are to be put and, as a matter of negotiation between sellers – who may wish to maximise reported audience size by adopting a widely-embracing definition – and buyers – who may be suspicious of the valuation of some newspaper or magazine when based on a relatively loose audience criterion.’*

This paper is about the methodological choices we have made over the past 20-30 years and about the outcome of the ‘negotiation’ between sellers and buyers over whether to use a broad or a narrower definition of readership. There is more to readership than size of course. The intensity as well as the level of reading needs to be – and has been - studied. But we will say little about qualitative measures in this paper, important though they are.

My opening hypothesis is that, whenever it has been practical and possible to change (or start from scratch) any part of the readership research methodology in a survey, the chosen method will tend to be one that favours higher rather than lower readership numbers. This is mainly because although the principal funding for readership research is provided by advertisers, it is largely controlled and channeled by the publishers.

On this cynical view, research conferences and symposia are as useful to media owners seeking techniques to maximise their audiences (while warning them of any that might point in the opposite direction) as they are voyages of discovery for research professionals.

As long as techniques evolve, research companies can make money. As long as readership figures rise or don't decline too much, the publishers are happy. Agencies just want an easy-to-access set of numbers. The great thing about readership research is that the 'truth' is not a precise thing. There is therefore plenty of room for variation in techniques between markets and sectors without one approach being 'right' and the other 'wrong.'

The trouble is, this is not good enough any more. Advertisers are increasingly demanding a demonstration of Return on Investment. It is the theme of this Symposium. It is what we should be focusing on when we design and execute readership measurement methods.

Key techniques to help marketers identify and maximise ROI include:

- Market mix modelling;
- Continuous tracking; and
- Integrated communications planning processes. These describe various ways of measuring the impact of multiple channels through which consumers come into contact with brands.

One way or another, the print medium (and all other media) needs accurate and granular measures of likely audience exposure to an advertiser's message in order to feed into the market mix models. This is the medium's number one priority today.

'Accurate and granular' may not be the same thing as 'consistent,' 'credible' or whatever other words have been used to guide industry currency definitions in the past. It means not only trying to be fair between different types of publication, but also being fair vis a vis other media.

Every research technique will have pros and cons, so whichever method is finally adopted can be wrapped in a veil of scientific respectability. Most of the dialogue will be ignored by the press, by advertisers and even by media planners because our discipline is neither popular nor user-friendly. Perhaps this is a relief...but we cannot live in this cloud cuckoo land for too much longer. We will be found out!

As an industry, we have been either unable or unwilling to make judgements on exactly what 'best practice' is on the various methodological choices available to us. We make no statement on whether – all things being equal - TTB, FRY, FRIPI or RR is the 'best' way to measure readership. Instead we say that all things are *never* equal.

There are a host of decisions to be made.. First we must decide which of several basic *models* of readership measurement should be adopted.

From there we need to consider the many permutations that are possible within our chosen approach. How, for example, should we present titles to respondents? What is the right order of asking questions and how exactly should we word them? One key learning over the years is that many of these decisions have consequences for readership results.

Readership results translate into revenue and profits for publishers who, of course, fund the measurement services in the first place.

Our symposia, our White Papers and our learned discourse revolve around what the effects might be of using this or that technique, all other things being equal. There is then a natural tendency to move towards adoption of those techniques which increase readership generally without upsetting the established balance between different publication groups on a survey. In other words, it is quite hard to change the readership measure or the questionnaire.

Should we be searching for truth or as near to truth as we can get? Or are the symposia merely forums where we discuss the effect of this or that methodological change but carefully avoid taking sides? In short, are we there to provide options or answers?

If the first, will this not encourage publishers to pick and choose from amongst the various methodological choices until they find a combination that maximises readership rather than one that provides advertisers with as close an approximation to accurate readership measurement as possible?

What is the evidence? From papers given at various symposia in the past, here are some methodological choices for maximising (or minimising) readership numbers for magazines (they may be slightly different for newspapers):

### 1. Basic technique

One of the oldest ways of measuring readership was to take respondents through actual issues of a magazine – or Thru The Book (TTB). This was used in both the US and Canada until both finally dropped the technique a few years ago.

There were certain theoretical concerns with the method. For example, there were found to be instances where there might be a certain amount of over-claiming.<sup>iv</sup> There were also worries that the ‘stripped-down’ issues interviewers needed to carry around (due to the sheer number of publications being asked about) would not generate as full an answer as complete issues would, although plenty of evidence suggested that the differences were actually quite small.

Other concerns were expressed as to the right age of issue to be used in the interview; the possibility, for example, that monthlies could be under-counted due to the time taken by them to build up readership.

However, as Michael Brown states: ‘*a clear balance of Symposia contributions have presented evidence that TTB leads to understatement of readership*’.

The Recent Reading method was adopted in the UK in the early 1950s and was eventually introduced into the United States in 1972.

Much debate raged in the late 1970s and early 1980s and two US research companies fought over which was the best technique. The industry funded a large scale Comparability Study in 1979 which quantified the difference in results without commenting on which were right and which were wrong.

Comparisons of AIR levels using TTB and RR (same titles)	
Publication Type	Index (RR vs. TTB)
Weekly Magazines	127
Tri-weekly Magazines	141
Large Circulation Monthlies	180
Smaller Circulation Monthlies	196

Source: ARF Comparability Study(USA), 1979

Other comparisons were made and investigations undertaken over the years into why the effects manifested themselves as they did and which was more valid. But perhaps the death knell for TTB lay in the sheer scale of the difference in readership results.

The RR technique involves asking respondents about their reading of *any* issue of a given title and therefore relies on their *recall* of reading a publication rather than *recognition* of a specific issue.

The technique raises questions about how we deal with title confusion, order effects, telescoping, parallel and replicated reading and a host of other things where much experimentation has been done.

Two attempts were made to modify the biases in the technique to address what was considered to be a key problem with the RR model: replicated reading. FRIPI (First Read in the Publication Interval) was introduced on the South African readership survey in 1988, while FRY (First Read Yesterday) originated in the Netherlands in the early 1980s.

Again there were problems. It is difficult, for example, to ask people about their ‘first’ reading of a magazine. It is unlikely to be a concept they think about much, especially for longer ago than a few days. Also, because the number of ‘first’ reading claims – especially of monthly titles ‘yesterday’ - will be much lower than total reading, very large sample sizes will be needed in a survey.

But most importantly of all, both methods reported lower readership numbers than Recent Reading.

The Dutch SummoScanner organisation, for example, ran a large-scale test comparing the FRY and RR methods on two matched samples of 4,000 respondents. Although the only substantial drop in readership levels occurred amongst monthly magazines with a high number of infrequent readers, it is probably worth noting that monthly magazines make up more than half of all the magazine sponsors of the survey...

### Netherlands SummoScanner Test, Q4 1995. Split Run Results

Average Issue Reach (summed percentages)			
Publication Type	FRY	RR	Index
TV Magazines	88	92	104
Womens' Weeklies	96	98	103
Other Weeklies	121	123	101
Monthlies (infrequent reading profile)	117	146	125
Monthlies (frequent reading profile)	139	141	101

Source: From FRY to FRIP (1997). Costa Tchaoussoglou. Vancouver Symposium.

The Dutch, the Danish and the Norwegian Newspaper Surveys all briefly flirted with FRY, but soon reverted to standard RR questions. FRIP is still used in South Africa today.

The hypothesis in this paper is that methodological changes resulting in higher readership numbers will be more likely to be adopted than change that pushes the numbers down. Almost all the evidence concerning the basic readership model – i.e. between TTB, RR, FRIP or FRY – suggests that the RR technique will generate the highest numbers. This is particularly so for monthly titles and those of lower frequency.

Recent Reading has not won all the technical and methodological arguments by any means, although it certainly has practical advantages in its application. There is clearly some model bias in favour of monthlies over weeklies that has been raised many times.

But 65 out of 67 readership surveys in Erhard Meier's latest *Summary of Current Readership Research* (2003) employ the recent reading technique. Simmons in the US and the PMB in Canada have given up using TTB in recent years. The Canadian NADBank – which measures only newspapers (largely unaffected by all the issues discussed) – and the South African AMPS survey remain the only exceptions.

Is it perhaps coincidence that six out of every ten magazines measured across sixty-two surveys around the world in 2003 are of monthly or lower frequency?

## 2. Screening

In a perfect world of perfect respondents, you would not need to 'screen' or filter publications to include in an interview. But in the practical world of buying and selling space and of limited time in which to engage a respondent's attention, ways have to be found of ensuring best use is made of this time.

Media planners and buyers want more and more titles to be included on readership surveys so they have comparable information on as many of them as possible. Publishers, as well, are desperate to get their own titles included on official readership surveys in order to ensure they get considered by media planners and buyers – whose first and often only port of call is this very survey.

The problem is one of massive over-supply: in the UK just 179 of the 3,343 consumer magazines published are squeezed onto the National Readership Survey. In the United States, MRI asks about readership of just 251 of the 7,188 consumer titles there.

Nobody can read or look at this many titles nor can they be expected to endure an interview that asks them to consider so many. So ways need to be found that allow respondents to quickly filter out the those they never read from a survey.

There is little consensus, however, on exactly *how* the filtering or screening should be administered.

Screening has the disadvantage for publishers that it will generally lead to the loss of some readers and that this loss will vary by publication group. Experimental research reported at various symposia has attempted to quantify these declines. Michael Brown<sup>vi</sup> quotes several papers which report 'losses' in readership varying from 6% to as high as 39% of respondents initially failing to 'screen in' and then subsequently claiming that they had, in fact, read a particular title.

Hastings Withers summarised a number of these studies in his own paper in Berlin.<sup>vii</sup>

<b>Summary of Experiments to Gauge the Impact of Screening Questions</b>		
<b>Experiment</b>	<b>'Lost' Readers</b>	<b>Screening Interval</b>
Belson (1962)	9-13%	3 months
Lysaker (1979)	12%	6-7 months
PMB (1995)	8%	6-8 months
Vorster (1984)	7-11%	6 months
Allensbach (1979)	4%	12 months

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

In Finland, a change in the screening question from a common 6-month period to one that varied by publication type in the early 1970s led to a fall in readership claims for weekly magazines (where the screening period had been reduced) and not much change for monthlies (where it remained unchanged) - though it was still introduced<sup>viii</sup>. (Currently, the Finnish survey uses a common twelve-month screener).

Appel and Stein<sup>ix</sup> reported a US experiment in 1993 where newspaper reading claims were significantly impacted by using either a 6-month or a 7-day screening hurdle – with resulting Average Issue Readership declines of 40-60% when the shorter period was used.

In Hong Kong in 1994, AC Nielsen experimented with adding a screening question to the Media Index survey. The pilot showed a big drop off in readership claims and so was quickly suppressed.

For the hypothesis of this paper to hold, we would therefore expect to see *fewer* surveys employing screening or hurdle questions or, where they do, to make them of longer rather than shorter duration.

Given an in-built bias against less frequently published titles, we might also expect to see these trends more prominently where monthly magazines can exert a strong influence over the methodology of the relevant readership survey.

In fact, the hypothesis is only partly supported. The proportion of surveys not fielding a screening question has fallen. In 1988 14 of 38 readership surveys failed to employ a screening question while 24 did.\* Fifteen years later with another 29 surveys included in the compilation, the number of surveys not fielding a screening question had risen to just 15, while 52 did employ one.

For surveys that did employ a screener the greatest increase was seen in the use of the fixed twelve month period. This will tend to have a positive impact on recorded reading levels versus shorter periods and supports our hypothesis. But the majority of surveys now use screening questions that vary with publication frequency.

\* Some surveys not formally asking a screening question used either the recency or frequency question as a filter– for the purposes of this paper we use the formal definitions of whether or not there is a screening question as published in Erhard Meier's *Summary of Current Readership Research* series.

<b>Screening Questions by Type</b>				
	<b>1988</b>	<b>1995</b>	<b>2003</b>	<b>Index: 2003 vs. 1988</b>
<b>Total # of surveys</b>	<b>38</b>	<b>43</b>	<b>67</b>	<b>176</b>
Number of surveys <b>not</b> using screening/hurdle question	14	14	15	107
Number of surveys using screening/hurdle question	24	29	52	217
Fixed screener/hurdle (across all publications)	12	12	21	175
12 months	2	5	13	650
6 months	4	2	6	150
Variable screener/hurdle (by publication type)	12	17	31	258

Source: *Summary of Current Readership Research*. (1988). Barcelona (Table 6/column 2); (1995). Berlin (Table 5/column 2); (2003): Cambridge (Table 5/column 2)

A 'best practice' which most of the evidence shows is likely to depress readership numbers has now become firmly established. On the other hand, there are almost as many versions of screening as there are readership surveys. So we haven't yet decided exactly how it should be done.

So there has clearly been a commercial trade-off: on the positive side, more titles can be added and are thereby advantaged by appearing on the official readership survey. On the less positive side, there may be marginal negative effects on the readership results of titles already on the survey. In every case, it seems, the argument in favour of more titles has won out.

### 3. Questionnaire length

Which brings us neatly onto the question of interview length. All things being equal, adding titles to a readership survey would tend to lengthen the interview. A limited amount of experimental research has shown that this can well have the effect of reducing average issue readership levels for individual titles. Our hypothesis suggests, therefore, that publishers will not be rushing to add questions or titles to the surveys.

But adding titles is just what has been happening for many years, as publishers scramble to get their publications onto the official surveys. The average number of publications included on a survey has almost doubled in the last fifteen years (although, obviously, not every title is asked of every respondent).

<b>Number of Titles Included By Survey</b>				
	<b>1988</b>	<b>1995</b>	<b>2003</b>	<b>Index: 2003 vs. 1988</b>
<b>Total # of surveys</b>	<b>36*</b>	<b>43</b>	<b>65*</b>	<b>181</b>
Total Number of Titles Included	4,321	7,405	15,110	350
Average Number of Titles per Survey	120	172	232	193
Average Length of interview (minutes)	45	43	48	107

NB: the number of titles included on a survey is not the same the number presented to an individual respondent.

\*Excludes surveys where no answers given to # of titles question

Source: *Summary of Current Readership Research* (1988, 1995, 2003)

Michael Brown quotes an example where the readership of *Der Spiegel* across several different surveys of progressively greater length ranged from 11.9% to 14.9%. Extrapolating results from this and other titles from four surveys where the main difference was in the total length of the interview suggested that the addition of 40-50 publications to a survey could reduce AIR estimates by as much as 10%.

But adding titles *without* increasing interview length – as, for example, was achieved in the UK when it moved over to a grouped titles approach – does not seem to have a perceptible effect on readership levels. In this case, it is probably fair to say that, because other things were not equal – i.e. the number of titles on average went up, but the interview length did not – our hypothesis is simply unproven either way.

#### 4. Title presentation

Because memory is imperfect, readership surveys employ various forms of prompting to help respondents remember whether or not they have read or looked at particular titles. Many more surveys now use a ‘grouped titles’ approach, which allows the number of titles asked about to be increased without lengthening the interview.

<b>Presentation of Titles: Individually vs. Grouped</b>				
	<b>1988</b>	<b>1995</b>	<b>2003</b>	<b>Index: 2003 vs. 1988</b>
<b>Total # of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>	<b>176</b>
Title by Title	30	36	49	163
Grouped Titles	7	7	18	257
Not specified	2	0	0	-

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

Grouping titles allows respondents to quickly discard whole sets of similar titles which they feel certain they never read or look at. It is possible that some infrequent readers of certain titles may be excluded in this way, but this disadvantage seems to have been outweighed in publishers’ minds by the opportunity offered to add more publications onto a survey.

A second advantage offered by the grouped titles approach is that respondent confusion is reduced through the simple expedient of placing two or more easily confusable titles onto the same card.

A good example of this was the case in the UK of *Homes & Garden* and *House & Garden*. Prior to introducing the Grouped Cards technique in 1984 there was reported to be a 95% discrepancy in readership claims according to which of the two magazines came first in the interview.

Other UK data showed that womens’ monthly magazines could be shown to have a gross readership 29% higher if they appeared early rather than late in the original masthead booklet; general monthlies generated differences of 38% between early and late positions<sup>4</sup>.

Given that individual publishers could not specify the positions of their titles in the order of presentation, it made sense to everybody to support a system that was equally fair to all titles. In this case therefore, the commercial imperative was to increase the number of titles on the survey while maintaining overall levels of readership and maintaining fairness between titles – not simply to increase readership scores.

#### 5. Visual stimulus

A number of different approaches to the way publications are visually presented in face-to-face interviews have been made. Simple typescript, title logos/mastheads, full covers or photographs of full covers are all used to varying degrees around the world.

When logos or mastheads are used, yet further choices are available: they can be presented in black and white or colour and can be shown singly or grouped. Thus far, no common approach has been adopted, although usage of mastheads (three quarters of all surveys) has become even more dominant than it was twenty years ago – the use of colour mastheads in particular has grown considerably.

<b>Presentation of Titles: Visual Devices</b>				
	<b>1988</b>	<b>1995</b>	<b>2003</b>	<b>Index: 2003 vs. 1988</b>
<b>Total# of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>	<b>176</b>
Mastheads/Title logos	27	30	51	196
- Colour	7	7	17	243
- Black & White	12	19	23	192
- Unspecified	8	4	11	157
Typed	6	4	8	133
Oral	2	6	6	300
Mixed	1	1	1	100
Covers	2	2	1	50
Not specified	1	0	0	-

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

The growth in ‘oral’ presentation simply reflects the existence of several telephone surveys on the list, and represents a small minority of all surveys.

But is the dominance and growth in the use of masthead prompts – and in particular colour prompts - based on any evidence that they produce higher readership claims than other methods?

When the Norwegian magazine readership survey shifted from simple printed lists of titles to a shuffled deck of mastheads in 1979, average issue readership claims across twenty magazines rose by 40% over 1978. The smallest titles gained the greatest number of readers proportionally, suggesting that the old technique of reading out titles in alphabetical order with no visual prompts had missed plenty of the most infrequent readers. Although other survey changes occurred simultaneously, the change in the questionnaire and the introduction of visual prompts was felt to be largely responsible for the surge in readership numbers<sup>xi</sup>.

When the UK NRS moved from individual masthead prompts to grouped, typeset lists in 1984 there were few noticeable effects on readership levels, despite a very large increase in the number of titles asked about. The typed lists were later changed to mini-masthead prompts, following experimental work which showed no significant (and certainly no negative) impact on readership levels.

We can therefore say that our hypothesis is weakly supported by this evidence.

As to whether the mastheads should be presented in colour or black & white, there is almost no data showing advantage to either technique<sup>xii</sup>.

Erhard Meier’s overall conclusion was that:

*“Mastheads were in general found to be superior as recall aids, especially amongst low frequency readers...However, recall aids are only one element of a survey’s design; and sources of possible errors underlying readership results can only with great difficulty be attributed to the design of recall aids alone. There are normally other design aspects which have far greater effects on a survey’s results<sup>xiii</sup>.”*



## 6. Question type

There are two principal ways of asking the recent reading question:

- The ‘hidden’ form, whereby respondents are asked *when* they last read a copy of a given title (without prompting any particular period of time) and
- The ‘disclosed’ or direct form where they are prompted as to whether the last reading event was in the past week, month, yesterday or whatever.

There does not seem to have been much methodological research on this subject. Val Appel reported on a small experiment amongst two national newspapers in the US where asking the question in these alternative ways generated no statistically significant effect<sup>xiv</sup>.

Michael Brown has stated that he would ‘marginally prefer asking a direct (disclosed) question.’

The trend in favour of this has been quite striking in recent years with two-thirds of readership surveys now preferring disclosed questioning, up from one third ten years ago. The hidden form has lost favour, falling from just under half of all surveys in 1995 to a quarter in 2003.

Question Type			
	1988	1995	2003
<b>Total # of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>
Disclosed	15	15	44
Hidden	12	20	16
Other/Not Classified	12	8	7

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

Once again, there does not seem to be any evidence either way in favour of our hypothesis – that such consensus in favour of a particular methodological option (in this case the disclosed form of questioning) has been prompted by any expectation that it would deliver a higher set of AIR outcomes than would result from using alternative forms of questioning.

## 7. Question order

I have not been able to discover, either, any experimental work looking at the effect of whether the frequency or the recency question should be asked first in the interview.

In 1967 the UK National Readership Survey removed a three month filter question asked of all 83 titles on the survey in order to add extra reading intensity questions.

In order to retain some sort of filter, the order of questioning was changed so that the frequency question was asked first and itself used as the filter followed by the recency question. The overall readership numbers did not show any significant changes between 1966 and 1967. However, the following year (when several other changes had also been made) readership levels for monthly magazines improved, much of which was credited to the change in filter question.

Question Order			
	1988	1995	2003
<b>Total # of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>
Frequency asked before Recency	24	21	49
Recency asked before Frequency	13	22	18
Mixed order	1	0	0

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

In Singapore the order of questions was changed in 1995, putting frequency before recency. At the same time a common twelve month screener was introduced. There was no perceptible effect on readership levels there, although it may be that there was some cancelling out of positive and negative effects.

In short, there is little to indicate that one order or the other will have any significant overall impact on AIR levels, though surveys are clearly converging on the frequency before recency approach.

**8. Response Rates**

Response rates are unfortunately not covered in the regular *Summaries* of Current Readership Research. They are not, of course, a methodological choice as such. But they are likely to be impacted, often in known ways, by many of the methodological decisions made in designing a survey.

Most countries are known to be struggling to maintain these rates at reasonable levels, including – as in the US – by paying quite large sums of money to respondents. In some areas of the UK, response rates are rumoured to be reaching alarmingly low levels.

But as long as AIR levels are maintained, it is possible to argue that the alarm bells may not ring. On television, where peplemeter response rates have long been shrouded in mystery and subterfuge, it is suggested by many that low response rates are actually encouraged, because ‘easy-to-recruit’ people tend be heavier viewers, thus pushing up average viewing levels.

I am not aware whether any such effect exists in print audience measurement. A study looking at screening and reading levels in relation to response rates (for example, by comparing these levels according to the number of call-backs needed to recruit a respondent) would be an interesting and welcome one, but cannot be pursued here without data.

**9. Reading Frequency**

Average Issue Readership figures provide an estimate of the number of readers to an average issue of a magazine or newspaper; however an estimate of reading *frequency* is crucial in order to analyse the coverage of print schedules over time.

Amongst the challenges that arise in making such estimates are how to ensure very infrequent readers are properly accounted for, whether to ask people about their actual or ‘normal’ behaviour and whether to ask them using a numerical or verbal scale of alternatives – in other words, how many out of the last four issues of a title are read (0,1,2,3 or 4) versus asking people whether they (for example) ‘always,’ ‘almost always,’ ‘sometimes’ or ‘very seldom’ read a given magazine.

Most of the debate occurs around this last subject of numeric versus verbal scales. The proportion of surveys using a numeric scale peaked in 1995 at almost three-quarters of the total. Verbal scales have recently become more popular however; by the time of the 2003 *Summary*, they were used on just under 30% of surveys.

<b>Frequency Scales</b>			
	<b>1988</b>	<b>1995</b>	<b>2003</b>
<b>Total# of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>
Numeric Scale	25	31	43
Verbal Scale	8	7	19
Neither/Mixed/No answer	6	5	5

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

There does not seem to be a great deal of evidence suggesting that one technique or the other generates higher reading frequency probabilities, although UK developmental work in 1964/65 and Dutch work in the 1980s quoted by Michael Brown<sup>xx</sup> both suggested that a numerical scale could exaggerate ‘true’ frequency levels. In this limited sense, our hypothesis can be said to be mildly supported.

**10. Title Rotation**

It is well known that there are important order effects according to when in the interview a given title is asked about. Data from several countries in different periods suggested that individual monthly magazines in particular benefitted considerably from being asked about early in an interview. This was especially important in the case of confusable titles like the UK’s *Homes & Gardens* and *House & Garden*, where the relative position in the interviewing sequence became crucial.

Rotation effects, as well as title confusion have been dampened partly by growing use of a grouped titles approach (see above).

<b>Rotation of Publication Groups/Titles</b>			
	<b>1988</b>	<b>1995</b>	<b>2003</b>
<b>Total# of surveys</b>	<b>39</b>	<b>43</b>	<b>67</b>
Publication Groups			
Fixed Order	25	25	30

Rotated	11	18	34
Other/Not Stated	3	0	3
Titles Within Publication Groups			
Fixed Order	8	11	11
Rotated	28	32	54
Other/Not Stated	3	0	2

Source: *Summary of Current Readership Research (1988, 1995, 2003)*

So in this case, where a clear methodological effect can be seen on readership results, there has been almost universal adoption of what is probably the best solution: some kind of rotation of the titles asked about within publication groups on four out of five of the world's major readership surveys. Rotating publication *groups* is less common, but seems also to be less important.

In this case, fairness between publications has triumphed over any attempt to push to the front of a questionnaire.

### ***Summary and Conclusions***

I set out on this journey of discovery with a clear and perhaps cynical expectation that, given two or more methodological choices, an industry readership survey primarily funded by publishers would opt in favour of whichever technique delivered the highest readership numbers. I also assumed that, with several hundred methodological papers read at Readership Symposia over the years, some kind of closure would have been given to many of these fundamental issues.

The assumption that such a decision was clear cut in the majority of methodological decisions was a heroic one however: in most cases the kind of pure, experimental work needed to prove that one or the other technique will invariably deliver superior readership numbers has simply not been undertaken. It is also fair to say that counting readers is not quite so straightforward as counting sheep.

Take the cases of whether to employ a 'hidden' or a 'disclosed' form of recency question or whether to ask about frequency first or about recency. Here, there has been no clear finding from the limited work been done to date.

On issues such as the use of logos or typed cards on the one hand or on whether to use numeric or verbal frequency scales, research has indicated a small positive effect on readership levels by choosing one or the other of the methods. But the evidence is not overwhelming and may have been confounded by effects from other variables.

Higher Average Issue Readership is not the only goal of publishers in readership surveys. Sometimes other commercial criteria are more important.

Screening, for example, has a fairly clear negative effect on AIR. Yet it has still been adopted by most surveys because it reduces the burden on interviewer and respondent and allows more titles to be included on a survey.

Similarly, the grouped titles technique has not been shown to generate higher reader numbers, but it has enabled surveys to include more publications than if they continued asking about each title separately.

Title rotation has been almost universally adopted, presumably because the proven effects on readership claims of different positions in the order of the questionnaire can hit any title. Nobody wants the potential problems associated with appearing late on a survey.

In just one respect – that of the choice of basic technique – have there been both clear positive effects on readership claims versus other methods and an almost universal adoption of the method. But even here, there are plenty of more pragmatic explanations of the move, including the burden on interviewers of carrying actual issues around and other major criticisms of the TTB and FRY methods outlined earlier.

One conclusion is, however, inescapable, we need to adopt measurement standards that allow print to compete on the same playing fields as television, the internet and other competitive media. Let us focus on ROI and the measurement questions that throws up.

<b>Summary of Key Methodological Choices</b>				
<b>Issue</b>	<b>Main Options</b>	<b>Most Popular Option</b>	<b>% of Surveys Using</b>	<b>Does Chosen Method Deliver Higher or Lower AIR?*</b>
1. Basic Technique	RR, TTB, FRY, FRIPI	RR	97%	Higher
2. Common Screener	Yes or No	Yes	78%	Lower
3. Title Presentation	Grouped or Individually	Individually	73%	Unproven
4. Visual Stimulus	Logos or Typescript	Logos	76%	Mildly Higher
5. Question Type	Hidden or Disclosed	Disclosed	66%	Unproven
6. Question Order	Frequency or Recency first	Frequency first	73%	Unproven
7. Reading Frequency	Numeric or Verbal Scale	Numeric	64%	Mildly Higher
8. Title Rotation	Fixed or Rotated	Rotated	83%	Benefits titles asked about early in the interview

\* On the basis of studies quoted elsewhere in this paper.

- <sup>i</sup> *Mediapost*. 22<sup>nd</sup> July 2005
- <sup>ii</sup> *Mediaweek*. 25<sup>th</sup> July 2005
- <sup>iii</sup> Brown, Michael (1999). *Effective Print Media Measurement*.
- <sup>iv</sup> Schreiber, Robert and Schiller, Clark (1984). *The Effects of Candour in Respondent Instructions on Overclaiming of Magazine Readership*. 2<sup>nd</sup> Worldwide Readership Symposium, Montreal.
- <sup>v</sup> Brown, Michael (1999). *Op cit*.
- <sup>vi</sup> Brown, Michael (1999). *Op cit*.
- <sup>vii</sup> Withers, Hastings (1995). *Loss of Readers at Screening Stage*. 7<sup>th</sup> Worldwide Readership Symposium, Berlin
- <sup>viii</sup> Haukatsalo, Jean (1982). *Experiments with Filter Questions in Finland*. 1<sup>st</sup> Worldwide Readership Symposium, New Orleans
- <sup>ix</sup> Appel, Valentine and Stein, Michael G. (1995). *Newspaper Screening Intervals: Six Months vs. Seven Days*. 7<sup>th</sup> Worldwide Readership Symposium, Berlin
- <sup>x</sup> Meier, Erhard (1985). *The Extended Media List: Grouped Titles, No Mastheads*. 3<sup>rd</sup> Worldwide Readership Symposium, Salzburg.
- <sup>xi</sup> Langhoff, Per (1982). *Norway: The Development of Industry Surveys of Consumer Media*. 1<sup>st</sup> Worldwide Readership Symposium, New Orleans.
- <sup>xii</sup> Schneller, Johannes (2001). *The Effects of Different Masthead Card Designs: Colour vs. Black-and-White*. 9<sup>th</sup> Worldwide Readership Symposium, Florence.
- <sup>xiii</sup> Meier, Erhard (1985). *Op cit*.
- <sup>xiv</sup> Appel, Valentine and Robinson, Thomas D. (1997). *Equal Measures: Why National Newspapers Should Be Measured Like Magazines*. 8<sup>th</sup> Worldwide Readership Symposium, Vancouver.
- <sup>xv</sup> Brown, Michael (1999). *op cit*. P103