SELLING PRINT SHORT: THE NEED TO RE-ASSESS READING AND READERSHIP

Brian Shields, Worldwide Research Director, International Herald Tribune

"Do we need a new definition of what we mean by readership?" First, the question rather implies that we have an agreed old one - and we don't. Whilst we do not have to consult a dictionary to concur on what we are talking about when discussing readership, in practice we go about measuring the supposed same phenomena differently. And in measuring differently we effectively change, or create, a variety of detailed meanings of readership. For the way we measure readership cautions and qualifies its specific meaning. The reality is probably more complicated still for we are not interested in readership values *per se* but measure them only as a surrogate for gauging the advertising communication value of publications - both in respect of print titles relative to other print titles and as a collective medium against, say, television. If there was no supposed high positive correlation between readership and communication value then we wouldn't even bother to measure the former. This, sometimes neglected, issue will be returned to later in the paper.

Further, reading does not, for our commercial purposes, exist as a singular fixed-point experience: it also has to be understood and measured by frequency over time. Thus the need to specify average issue readers, regular readers, infrequent readers, casualness *et al.* Again this gets back to readership being a surrogate measure for communication. Advertisers want to know schedule reach, average frequency or opportunities to see (a convenient gloss over the issue of communication) and gross ratings points. Given that all of these are a function, not of average or single issue readership alone, but of frequency of reading, then the measures used to establish these, and the subsequent modelling of the data, determine a schedule's gross and net readership.

All of these only matter if different measures of readership used to calculate different aspects of it – single/average issue and frequency – produce different measures of "reading". If, by and large, they don't, then the debate is merely semantic driven. If they do, then by operating different measures we are, in effect and at best, measuring different things and are operating with different definitions. Alternatively if we do have the same definitions and objectives but get different results then some measures must be judged to be better than others are.

We know from Erhard Meier's terrific work over the years on comparative readership methodologies that there are many measures in operation. Quoting from his 2001 paper in Venice we see that:

Table 1
A: Number of surveys by sequence of Readership Questions *

S, F, R	33
S, R, F	15
F, R	10
R, F	2
S, R	2
R only	1
F only	1

* S = Screen, F = Frequency, R = Recency

B: Number of surveys by Type of Frequency Question

Frequency scale common for all publications

Numeric: number out of 12	2
Numeric: number out of 6	5
Numeric: number our of 4	10
Verbal, no numeric explanation	3
Verbal, with numeric explanation	15

C: Number of scale positions (35 surveys with common frequency positions for all publications)

3 positions	2
4 positions	15
5 positions	7
6 positions	7
7 positions	2
12 positions	2

The variables given above are only some of the ones that drive measurement and hence, I would argue, inform "definition". Others will be dealt with later. In all, however, we need to see what affects, if any, are produced by variously applied measures, for in effect if there are no affects there are no problems – only in the presence of differences does the question of which is 'right' or 'better' become pertinent. For the sake of brevity, let us take a few variables.

1) Recency versus Through-The-Book. Though now largely dead as a practical issue, its history might suggest, to some, that in a battle between commercial expediency and best practice the latter will always lose. Also the two approaches clearly demonstrate that in attempting to measure the same thing - 'reading' - very different answers are produced. No consideration of a new definition of reading makes sense without acknowledging this for, casting aside the late Timothy Joyce's notion that a perfectly conducted Through-The–Book survey would produce congruent results to a Recency measure, the two approaches did produce radically different results. Of course, only one, at most, could be right if it was accepted that both had an agreed and common definition of 'reading'.
2) The ordering affect of Recency and Frequency.
3) The scales used to establish Frequency.
4) The constraints, if any, put on a respondent to make a 'reading' claim as relating to the wording of the 'readership' question.

What I'm looking for in the brief excursion through current and previous applied readership measures is not a sense of which might be 'right' but whether significant differences accrue to measures intended to assess the same phenomena. If they do then we have to accept that some are better than others are. Or, alternatively, that all could be good, but they are not measuring the same thing. And given that 'readership' is but a path to assess intra-medium and inter-media communication values, does it matter that various outcomes are produced or that different definitions are in reality in operation?

1) Measuring readership: recency versus through-the-book

At the Florence Symposium, whilst arguing that recency measures inflate weekly and monthly magazine scores relative to daily newspapers, Neil Shepherd-Smith showed the following table. It is re-produced here without comment on the relative merits of the two methods, but rather to suggest that the differences produced between the two methods are both real and significant.

Table 2
A move from Through-The-Book to Recent Reading

Canadian National Readership Survey (readers per copy)

	TTB Scores	Recency Scores
Gardening	5.0	20.0
House and Garden	4.5	13.9
Toronto Life	3.7	11.2
TIME	5.4	9.7

2) The ordering effect and Recency and Frequency

To test the effect of any variable, the ideal is to keep constant all other variables. An opportunity nearly presented itself in the last six months to look at the ordering effect of the frequency and recency questions and deserves examination.

Ipsos-UK has conducted the European and Asian Business Readership Surveys (known respectively as EBRS and ABRS) since 1975 (EBRS). 2003 saw them propose to embark upon an USBRS study. Clearly many of the details of a US survey would be different from its European and Asian counterparts, and they need not detain us here. However, the putative sponsors were concerned that the self-completion method with average issue readership established by reading in last publishing interval proceeding a four-point verbal frequency scale was something that would require justifying in their market. Therefore a pilot study was commissioned whereby the sample was to be split with half of the potential respondents receiving a questionnaire where the recency question was first, and half where the frequency questions proceeded the recency ones. For reasons of cost, and maybe attitude, what would have proved to be an admirable test was abandoned, but the US sponsors were not going to agree to a Recency first, Frequency second model. Therefore the pilot went ahead – testing response rates, contact methodology et al – but with a reversal of the Recency/Frequency model that Ipsos-RSL employs on EBRS and ABRS. Assuming that more than desire to be contrary drives the piloted order, it has to be supposed that the sponsors believe that different results may accrue to asking the same questions. Given that the same questions should, in isolation, give the same answers it has to be supposed that the ordering effects either the answers given so as to produce a different claim, or that the response or number of those answering the second question is affected by the very asking and answering of the one prior to it. We know that many who are asked the recency question leave blank the Yes or No option; to this we must add those that answer Yes to the recency question and then fail to complete any of the options in the frequency scale (leaving usually the option to re-distribute their frequency claims in line with those that have answered both recency and frequency questions – in short we make their answers up for them). Given that both EBRS and ABRS use this recency/frequency option amongst roughly similarly defined occupational groups, albeit on very different continents, we can express a range of likely non-response to frequency after a positive recency claim. The new USBRS pilot will allow us to examine the level of non claims of frequency (no positive claim and no "never" claim) with the EBRS/ABRS method, and will further show the level of incomplete recency claims – Yes/No – after a positive frequency claim. These analyses will be included in the verbal paper.

Given the arena for EBRS and ABRS is multi-country, the issue of ascertaining frequency is perhaps further complicated by using a verbal four-point verbal scale with only the first (the clearly mislabelled "regularly") having a numerical guide – at least three out of four issues. Obviously, "regularly", "often", "sometimes" and "never" can be translated into the eleven or so languages used in the various EBRS questionnaires, but do the words mean the same to respondents in different countries? [Numbers do have the benefit of universality that Carl Sagan believed might be the only initial common ground for communication between those inhabiting different parts of the Cosmos. Further, one is reminded of a scene from the film Manhattan where the characters played by Woody Allen and Mia Farrow are separately asked by their psychiatrist how often they have sex: Allen answers "Hardly ever…two or three times a week" ; Farrow answers, "All the time…..two or three times a week" . A similar numerical judgement of frequency produces polar opposite verbal assessment of the same phenomenon.

3) The scales used to establish Frequency.

At the Venice symposium Gary Morgan and Michele Levine from the Roy Morgan Research Centre introduced a new way of asking about frequency that promised to eliminate faulty measures of the past and produce more accurate reader per copy figures. The paper was controversial, its delivery more so, and the announcement of patents and copyrights of the "new" questions and approach were both precedent setting and potentially damaging for developing best practice in readership research. However, the contentious nature of the new method should not blind us to the potential significance of the work – those caring to read or re-read the full symposium paper are unlikely to feel that they have wasted their time.

Essentially it was said that "old" frequency scales inherently gave too low a measure of "casualness" and hence reach projections. The re-interview method could establish "real" casualness figures and that the new Roy Morgan approach produced virtually the same results as the re-interview method. And by implication the re-interview method was always likely to be better in respect of casualness measures than any other, but was just largely impractical for reasons of cost. For brevity, I'll try to state to basis of the case in two tables.

Table 3 Comparisons of Casualness Estimates

	Re-interview (Establishment Survey and diary)	Old Single Interview*	Difference
BRW	66.3	48.3	18.0
National Geographic	57.6	46.9	10.7
Reader's Digest	47.4	39.1	8.3
Sunday Life	48.0	24.5	23.5
Vogue Australia	79.3	63.7	15.6

Table 4 Comparisons of Casualness Estimates

	Re-interview (Establishment Survey and diary)	New Single Interview**	Difference
BRW	66.3	64.0	2.3
National Geographic	54.7	47.1	7.6
Reader's Digest	44.1	41.5	2.6
Sunday Life	44.1	37.4	6.7
Vogue Australia	73.7	68.8	4.9

*** Four point scale – interviewing April 198 – December 1998**
**** Two point scale (0, 1 or 2+) – interviewing October 1999 – March 2000**

First to be noted is that replacing four publication intervals with two (and using different mathematical treatments) very varying results accrue to attempts to measure the same phenomena within a survey series. Secondly, Gary Morgan presented data to show that using the new technique can eliminate previous title discrepancy across surveys and region.

Table 5
Readership Currency Readers Per Copy

Magazine	Australia RMR (18+)	New Zealand Nielsen (20+)	USA MRI (18+)
People/Who	4.3	8.5	9.8
Reader's Digest	2.4	3.9	3.4
Cosmopolitan	3.0	n/a	6.1
TIME	3.4	5.7	5.1
Newsweek/Bulletin	4.0	n/a	6.1

Table 6
Roy Morgan Research Readers Per Copy (18+) estimates across countries

Magazine	Australia	New Zealand	USA
People/Who	4.3	4.7	4.4
Reader's Digest	2.4	2.5	2.9
Cosmopolitan	3.0	3.5	3.4
TIME	3.4	3.7	4.3
Newsweek/Bulletin	4.0	not published	5.1

The main reasons cited for the differences are "replicated reading using the 'recency' model and questionnaire/title confusion". (In this, Morgan echoes the arguments of those supporters of Through-The-Book, though the solutions are different). The claim made for the "new method" was "not an alternative but a necessity" (Australians and modesty are rarely common bedfellows). However, this was only half the story; the real revelation was *"The Roy Morgan Centre Pty has applied to patent this new method, and the question is copyright".*

Assume for the moment that the Roy Morgan method is better. Assume we are all seeking to improve measurement (notwithstanding that most magazine titles end up with lower average issue readers per copy, though higher cumulative figures, and that there would be commercial implication for this). Logic, and Gary Morgan, suggests a movement to the two-point scale. Does this mean that all companies wishing to adopt it have to pay a fee for the right? Could the right be refused? Or can only Roy Morgan Research operate the "better" method?

If we are seriously looking at the need for "*a new definition of what we mean by readership"*, we had best hope that no one party has patented it or will do so. If 'best practice' becomes the domain of copyright and patent then there is no possibility of universal best practice.

4) The constraints, if any, put on a respondent to make a 'reading' claim as relating to the wording of the 'readership' question.

If any irrefutable example were still required that we are directly or indirectly operating with several definitions of readership, then it can be supplied here. It relates to what the respondent is asked, or not asked, about a readership claim. On some surveys it is along the lines of "read or looked at a copy of x, it doesn't matter for how long, where it was or whose it was". Others are less generous "read or looked at for at least two minutes". Most surveys - certainly on self-completion questionnaires - give the respondents no guidance; they are simply asked if they "read" a copy – they are left to decide what reading is.

So should a bold headline appear in national newspaper proclaiming "Readership Symposium agrees on new definition of readership" (presumably an April 1st issue), and I am sat on a train opposite the owner of the copy and glance at the headline, what are the current realities of my claiming to be a reader of that copy? For the first set of wording given above, I am a reader. For the second, I am not. For the third – should I be interviewed the next day with a recency model employed - I get to decide if I am a reader or not, and if I decide yes, I am as much an average issue reader as the purchaser of the paper, though to an advertiser in that issue I am worthless.

This is important. Some "readership" definitions (the broadest given above – anywhere, no matter how long, whose copy etc. - being a good example), have detached themselves from the purpose of measuring readership. Seeing a headline on someone else's newspaper whilst sat on a train has a communication value of nil to advertisers within the publication. The tightest definition I am aware of is "read or looked at for at least 2 minutes". Is this a good enough starting point to express communication value? Would "read or look at 25% or more of the pages" (or something more elegantly worded) produce a more realistic measure when set against the point of collecting readership data? Wherever we go with this issue, clearly by using different (or no) wording to qualify what is meant by reading means we are operating with different definitions of reading.

Which leads us back to the reason we measure readership – to establish communication values for advertisers. Even if a new definition was agreed upon and more importantly a set of common methodologies and practices to illuminate it equally concurred, we would only have a "better agreed" measurements of a surrogate.

Early in the history of readership research the term "opportunities to see" was adopted. It was a neat way around saying that in "reading" a publication an advertisement would/could be exposed to a reader. The assumption being that the relationship across titles between seeing advertisements and reading publications is constant. Clearly, no one actually believed this so the working assumption became that the relationship between reading and advertising exposure is secondary to the original measure. Over time, the condescending nod to the communication value of print titles has been the adoption of a variety of "qualitative measures". These include amongst others: time spent reading, most important read, number of pick-ups, proportion read, proportion read per pick-up, place of reading and source of copy. It will be remembered that many readership surveys have no "qualitative" reading questions.

Print is different in kind, not degree, from television audience measurement. If TV research measures viewing as accurately as it can (on an agreed definition of viewing) then it gives pretty much real exposures to advertising messages – it is a viewer who is watching the television when the advertisement runs. The best print will give is that a reader had an opportunity to see an advertisement in so far as they "read" the publication in which the ad appeared. TV research as it currently stands can give a far better numeration of the communication value (as given by "seeing" an advertisement) of the medium whilst measuring exposure to the medium itself than print research can. Most print readership research lays no claims to measuring readership of advertising, only reading of titles within the medium.

To this extent TV advertising can be seen to be *episodic* (the viewing of advertising can be collected in the measure of viewing of channels at specific times); print advertising is *contextual* (the readership or sight of all advertising cannot be measured by the process of measuring readership of titles). This leads to varying assessments of the value of print.

1) In equating OTS to 1, print can overstate its communication value. If the OTS figure of 1 is not discounted, then print maybe under-invested in because communication goals are believed to have been achieved when they have not.
2) Agencies and advertisers know the communication value is overstated and require more OTSs than they would real ad exposures. But they don't know how many more – nor do they know by title the relationship between ad exposure and OTS.
3) Print has, therefore, to provide generalised "proof" of communication value. Or specific issue "proof" along the lines of the type of studies conducted by Starch and by Harvey Research (et al). These however tend to be collected in different circumstances, at different times and with different methodologies than the readership research itself.
4) Print advertising, by being *contextual* as opposed to *episodic* may generate additional benefits over and above "reaching" a target audience. Environment can enhance, or detract, from an ad differently in print than on television.

Recently The Canadian Newspaper Association has published a document entitled "Factors Influencing Newspaper Advertising Effectiveness". In essence it is a synthesis of "qualitative" or "effectiveness" readership research dating from 1961 to most recent times from Canada, USA, Australia, Denmark, Germany, Netherlands, Norway, Sweden, UK and international studies. More than 30 studies in total are used in the analysis. The basic methods used in the collection of data for all of the studies were reading and noting and eye-camera measures. The document deserves thorough study though there is not space here to do so.

My first point of interest though is that this document – and more importantly the numerous research studies that it is predicated on – exists. This research has been conducted to show that which readership (how many) research doesn't show.

In giving a synopsis of the key findings, we see the areas of advertising communication that are not covered off by reader-measure research.

1) Good ads always score best irrespective of size, colour, shape or position
2) Ad noting increases increases as size increases
3) Readership of larger size ads has remained steady over time
4) Colour increases ad noting
5) Colour increases ad reading by about 20% in Canada and the US
6) The impact of colour has dropped historically (from 1966–1979) but has not changed in the past 20 years
7) Colour increases ad noting among both men and women
8) Most studies indicate there is no difference in ad noting between left and right-hand pages
9) There are no differences between the way ads on left and right-hand pages are noted by men and women
10) Noting of ads is constant in weekday issues
11) Vertical ads were noted somewhat more than horizontal ones
12) Newspapers are effective for both image and sales

These conclusions are drawn from newspaper research only, though not dissimilar studies have been conducted for magazines, too. The interesting point for me is that "effectiveness" relates almost exclusively to advertisements having been seen or read.

The readership research focuses on titles having been seen or read. The two are clearly parallel but separate and distinct. Neither research strand measures the phenomena that are the focus of the other; 'readership' measurement is the starting point upon which 'effectiveness' measures are layered.

In the main, publications pay for the lion's share of Readership Research. The end users at the buying-end may have some formal (JIC) or informal input as to how the research is conducted, but their greatest choice is if and how to use the resultant studies. There is the possibility for media owners to gain a competitive edge dependent on how surveys are conducted – the Canadian experience on the move to recency from Through-The-Book shows this, and the comprehensive adoption of the Roy Morgan method would probably produce movement in the opposite direction for a.i.r.. That is, magazines generated higher reader-per-copy scores when TTB was replaced by recency and the Roy Morgan examples tend to give lower RPC scores for magazines for single issue scores, though not ones where casualness and cumulative reach are taken into account. Agencies tend not to gain a competitive edge over each other when readership methodology or models change; their interest is therefore more disinterested and, it could be argued, determined by getting genuinely better rather than bigger numbers. But given their commercial nature, agencies are not going to spend vast sums of money to get "better" data, even if they could. What they can do is commission research that gives them a better understanding of readership and how it connects with communication, and which can give them a competitive advantage.

In my area, that of international (multi-country) media research, that is exactly what a number of agencies have done, and I believe will increasingly seek to do. Prominent in this arena have been Mindshare, Doremus and Universal McCann and the latter have recently published a study – funded also by The Economist, Financial Times, International Herald Tribune, TIME, BBC World and CNN International – that looks at media consumption patterns amongst C-suite executives in the USA, Europe and Asia. The interview method took the form of face-to-face depth interviews. The thinking behind it was this: the C-suite executive is for many of Universal McCann clients and for owners of international media, a prime target market. Whilst their "reading" habits are measured on several purposive international surveys, there are too many communication and usage questions left unanswered by standard readership measures. Rather than just "what do they read?" the question becomes "how do they read and why? And how are their media consumption patterns affected by (international) travel?" The fundamental question driving this study (and others like it) is: what can be learnt from this type of research that informs the debate on readership and advertising effectiveness and of media usage and effectiveness?

The survey is called "International C-Level Media Study" and has separate reports for the USA, Europe and Asia. The research was conducted by NFO WorldGroup. The summary of findings document for Europe alone, excluding appendices, runs to 64 pages, so clearly consideration in any meaningfully manner here is impossible – indeed it could be said to deserve a paper dedicated to it exclusively. But, however perfunctory its reference here, I am inclined to believe it will be a form of "readership" research that will burgeon. In being predicated on in-depth interviews, they are "qualitative in nature, and therefore no statistical inference should be drawn". Thus they will obviate much of the discussion of this symposium, including a new or old definition of reading.

The research asks these senior level corporate respondents to re-trace their most recent workday and the media (print, TV, radio and the Internet) consumption within it. Each media exposure is probed for reasons for reading and values attached to the media. Categories – in this case national versus international media being a prime consideration – are probed in greater depth. As too are "frequently untypical" periods like 'last international trip' and even non-working weekends. The results is more of a fully composed picture (as opposed to a 'join the dots' representation that a quantitative readership survey may be said to present) of media consumption. I give some of the major findings below, but would ask the reader to note that these are somewhat stark summaries of an already condensed executive summary.

1. International media consumption frequently is part of multi-tasking. Because they are so busy and time pressured, travel time is seen as an opportunity to multi-task, that is getting somewhere while simultaneously consuming media. Multi-tasking spills over into lunchtime when they may also consume media (especially the Internet) while eating.

2. Interest in International Media is very high just prior to an international trip. International travel frequently leads to opportunities to consume new media. Part of the attraction is curiosity. They want to know more about what is going on in the place they will be visiting.

3. Decisions concerning specific International Media to be consumed are mostly made prior to the actual flight. Media are often taken from the office or are selected at the airport.

4. There is a steady rise and in use and importance of the internet medium

5. Workplace media consumption varies by type – media consumption during the workday appears to be influenced by a pre-existing sense of appropriateness to the office setting.

6. Time is the executives most precious commodity

7. Media convergence in beginning to happen
8. Perception of bias increases scepticism
9. International print medium often holds advantages over national in terms of perspective, objectivity and style

Now, some of these may seem to be rather anodyne, but remember they are they are summaries of summaries and underpinning all the insights are aggregates of consumption patterns of specific titles, channels and websites. And in these, the usual suspects re-appear time and time again, as does categories of media e.g. business magazines, international newspapers, quality national business or general interest newspapers. There is not a pre-set media list in the interviews, those mentioned are ‘findings’, but the latter number in the tens, not the hundreds.

Given the results, albeit for a small market niche and sponsored by a small market media segment (internationals), is this a phenomenon we are likely to see growing? If yes, are agencies going to be more reliant on their own research – whether entirely proprietary or conducted with certain “invited” media – rather than the currently dominant readership surveys? In short, in going to the answers they want, rather than the answers they are given, will agencies produce their own definitions and measures of “readership” that will be supplemented by traditional counts? The attractions are obvious: they determine the questions and issues and oversee the research that is designed to give answers; they can gain a competitive edge over other agencies on the back of market and media expertise and by their sole (agency) access to the data; in doing so they still have access to the conventional readership measures. This research is important then not just for what it might show given its specifics, but where it might lead.

It is unlikely to be the case that standard readership surveys will be entirely eschewed for the exclusive use of the types of surveys similar to the Universal McCann study. But particularly where important niche markets (be they very senior executives or income elites who are geographically dispersed) are the focus, then these studies grow in importance, not least because the vagaries of the conventional studies render “quantitative” measures less reliable. For example, on Europe 2003 (a 16 country study of individuals with income in excess of approximately US$55,000) most English language print titles can see their readership decline by around 50% without it being statistically significant. Add to this the difficulties of measuring “reading” (however it is defined) that I have dealt with above, and the use of quantitative research as a benchmark only - a starting place – seems more and more sensible. For all its lack of “statistical inference” qualitative research can directly link readership to communication value in a way that is more illuminating than most readership surveys. When the “statistical inferences” on leading studies come with a +/- 50% variance, it could be argued that statistical inference is something not worth having.

Is a new definition worth having? Probably the answer is that it would be neither here nor there; for it seems clear that different methodologies, questions, and techniques will be employed to pursue measurement of any similarly (or dissimilarly) defined entity – reading. For those of us who are concerned with accurate measurement – and perhaps more than our commercial interests should dictate – there is the frustration that inappropriate and faulty models remain largely unchallenged outside these symposia discussions. Define as you will, but measure as you have/want to is unlikely to change. The long-term collective loser is, of course, print.

Whither Definitions, Methodologies and Surveys?

I will conclude with some predictions

1) We will have to start by defining what we mean by “definition”
2) We can agree on a definition of “reading” – within a range
3) We will continue to adopt a vast variety of methods, techniques and questions to establish reading scores but we will continue to learn even if we chose not to put the learning into practice (see 4)
4) Most changes will be commercially driven, not best practice driven
5) Those who patent questions will have their largely ignored
6) We will ascribe, model and fuse more despite questionnaires getting longer
7) Response rates will continue to fall
8) Agencies will increasingly rely on their own readership research
9) Few agency planners/buyers will ever attend these Symposia
10) We will all re-convene again in two years to discuss largely the same issues (and so we should)

Sources

The Ideal Readership Survey, Neil Shepherd-Smith Worldwide Readership Symposium 1999

What's Been Happening and What's New – Twenty Years of Looking at Readership Research, Erhard Meier Worldwide Readership Symposium 2001

A New Method to Measure Media Casualness for Magazines and Newspapers, Gary Morgan and Michele Levine Worldwide Readership Symposium 2001

Factors Influencing Newspaper Advertising Effectiveness, Canadian Newspaper Association 2003

Cosmos, Carl Sagan, 1983

Europe 2003, Ipsos-RSL

C-Level Media Study, NCO WorldGroup and Universal McCann New York, 2003