

SELF-ADMINISTERED QUESTIONNAIRES AND FREQUENCY OF READING: AN OPTIMAL COMBINATION TO MEASURE READERSHIP?

Ingemar Lindberg, Sifo Research & Consulting AB / Montilo Media AB

Summary

Would you like to substitute your expensive face-to-face interviews with a self-administered postal questionnaire? Would you like to drop the contentious last issue question and derive readership estimates from a single frequency question? A joke or a provocation, in a world where face-to-face and recent reading are words of honour? Not at all.

This paper describes how and why reliable AIR (Average Issue Readership) figures can be established by postal self-administered questionnaires and a single frequency of reading question, in combination with high response rates and (single source) collection of vast amounts of target market data.

Self-administered data collection has not been extensively used in the past. One important reason has of course been the unavailability of random samples. Other reasons have been low response rates and inability to manage the combination of recent reading questions with daily random samples.

It is the author's experience that a properly handled self-administered data collection procedure, which carefully considers such essential elements as introductory letters, accompanying letters, follow-up letters, instructions, incentives, and, not least, inviting and pedagogical question formulation and questionnaire lay-out, will achieve high quality information and outstanding response rates.

Frequency of reading questions seem to be mostly used only as arbitrary carriers of reading probabilities, derived from AIR figures by recent reading questions, to be used with computerised media analysis programs. As a consequence frequency of reading questions have not explicitly been developed or fully explored in the past, as a method in its own right to estimate AIR.

Frequency of reading questions seem to work best in combination with self-administered questionnaires, according to a major pilot survey. This combination of research procedures has rarely been investigated elsewhere, and maybe therefore not considered as a potential alternative.

A self-administered questionnaire, in combination with a frequency of reading question, has many advantages: The exact day of the interview is not important. The respondent can himself decide when it suits him to fill in the questionnaire: He can also do it on several occasions. There is no time pressure so he can take as much time as he likes to consider his answers. The questionnaire gives him an overview, a frame of reference, which helps him create a 'composite standard' in his answers. It can handle publications with odd publishing frequencies and irregular publishing intervals. Lots of classification (target market) data can be collected in the same questionnaire, with high response rates, and, not least, very cost-efficiently, which may allow for increased sample sizes.

Another positive element of considerable importance is, that the margin of error of the AIR estimates, using frequency of reading questions, is on average only HALF compared to what you get with a recent reading question, with its considerably lower number of observations.

A general observation: In order to improve readership research methodology, research components cannot be compared, analysed and evaluated independently, one by one. Their respective functionality is dependent upon their interaction with each other.

This interaction between all research elements is very important to consider, when deciding upon a survey procedure set-up. This is of course easier said than done, but in the end it is the *combination* chosen that counts.

The author is convinced that most of the findings and experience presented here can successfully be applied in many markets, while only partly, or perhaps not at all, in other environments. It is also to be expected that the future development in media research will go in the self-administered direction, while opening up many new optional research approaches, including frequency of reading questions.

Some general media research methodology considerations

When starting a survey project with the purpose of measuring and assessing AIR (Average Issue Readership), a methodological set-up has to be decided upon. In so doing several explicit and/or (probably mostly) implicit judgements and decisions have to be made, concerning each of the many individual research components involved.

The need to consider the interaction between research components

As early as in his 1967 thesis the late Dr Jarko Cerha¹ expressed concern that so little of a holistic view is taken in communication research, including media research: "...Methodological development requires that one considers the interaction between ALL methodological components of the research operation".

He made this statement in disappointment, finding that in spite of the many surveys, experiments etc. relating to mass communication research done over several decades world-wide, it was not possible to come up with any synthesis about the way mass communication works. He related this to the fact that basically all research projects do not link into each other, each project being based on its own framework.

From the Readership Research Symposia since 1981, and lots of other media seminars over the last 30 years, I cannot recall many papers having explicitly taken such a wider approach in discussions and/or experiments concerning media research methodology. Jean-Louis Chandon and Carole Fagot² are giving an example under the heading 'Mode of (data) Collection': "...From this preliminary review of literature, it appears that the effects of the mode of collection (face-to-face versus self-administered) are not well known and have not been studied extensively..."

I think that you agree, that research findings presented normally deal with only one research dimension at a time. Of course, that is in line with classic experimental tradition and theory, but it does not consider the interaction between the research elements.

It seems logical to argue, that we should try and work more multi-dimensionally, i.e. compare and analyse results between **totalities** of research methodology set-ups. The number of possible combinations of research set-ups is, of course, considerable, and may well be in the hundreds, even when combining (x) only the most common elements:

- form of data collection
- x type of readership question(s)
- x scale alternatives
- x question wording
- x kind of stimuli (prompts)
- x etc.

Even if we try to apply a more holistic view, we may still overlook to test potentially successful methodological combinations in a multitude of research set-up options. To avoid that is of course desirable but unfortunately much easier said than done.

Avoiding Interview shock and achieving a Composite standard effect

There are many subtle ingredients in media research methodology, not commonly discussed or referred to but highly relevant. Two of them, 'interview shock' and 'composite standard', are explained below.

Most readership surveys are made with face-to-face and, to an increasing extent, with telephone interviews. During most of these interviews (particularly telephone interviews) the respondent gives his answers on a question by question basis, without being explicitly given an overview of, or being supplied with very much of a frame of reference between, the questions he replies to.

Jarko Cerha¹ (p 135) discusses in relation to this phenomenon:

"(The)...contrast between the situation created by the interview and the respondent's usual situation influences response behaviour. It creates *interview shock*..." "Interview shock generally produces answers, most of which lack terms of reference and reflect the effect of factors forming the interview situation rather than the effect of the respondent's customary situation."

¹ Cerha, Jarko "Selective Mass Communication", 1967

² Chandon, Jean-Louis & Fagot, Carole "Personal in-home interviews versus self-administered mail panel: Does it really matter for measuring print audiences?" Worldwide Readership Research Symposium VII, 1995

"...The interview shock, as defined here, (1) *creates response deviations from the respondent's true response behaviour*, (2) *is influenced by the entire methodological set-up in the response situation*, and (3) *influences the rest of the process, the interpretation*."

Cerha holds the opinion that "...the only way of suppressing the effect of interview shock is to use instruments which discretely provide the respondent with a frame of reference when answering the question."

He continues: "According to psychometric theory, a respondent who consecutively evaluates *a series of stimuli through the same instrument*, develops a subjective norm - a "composite standard" - which enables him to assess the stimuli in meaningful relationships to one another".

..."It should be noted that not only stimuli but also the actual way of handling the instrument are included in this development of a norm (self-administration of the instrument). From this may be derived the conclusion that our sociological instruments cannot generate absolute measures but only relative ones."

Later in the text he emphasises that "...*interview shock, instruments of measurement, methods of contact and problems of interpretation should be discussed together*. In this context the use of metrical relativism may be taken to soften the shock and promote interpretation. The application of this relativism would seem to reduce the importance of exact formulation of the questions at the same time as it would increase the burden on the respondent. Self-administration is therefore required to develop the subjective standard of assessment (composite standard). Self-administration in its turn reduces or eliminates the need for an interviewer and hence his influence on the respondent. *Instruments that are suitable for self-administration can be used to advantage in combination with postal contact*. This eliminates in its turn the disadvantages of a fixed time for the interview (the respondent can answer when he has time and also in several steps) as well as the heterogeneity of the interviewer effect. Response bias is expected to be more constant. The treatment of a metrical detail will thus automatically involve adjustment of other metrical details and initiate a modification of the approach's total construction. In my opinion, this type of "total planning" is essential for a more rapid development of sociological methodology."

A methodological pilot survey

We had the above methodological considerations in mind when we were to start doing readership surveys in the mid '60s, and needed to find a combination of research methodology that would cope with the demands and conditions we had to work from:

- Random sampling easy
- Large number of interviews required
- Low population density in Sweden
- Large amounts of target market data required (hundreds of items)
- Coverage of all main print media necessary
- Many publications having odd publishing intervals
- AIR should be consistent and 'acceptable'
- High response rates essential
- Limited research budget

In order to try and adhere to the ideas in the first section above, we started off with a major pilot survey¹, designed to test the suitability of various methodological set-ups. The combination of face-to-face interviewing with recent reading questions was not included in the pilot test design, since this was the NRS procedure at the time. In our pilot test was matched (1) personal, postal and telephone interviewing against (2) two frequency of reading questions, (3) each with three scale alternatives, and (4) using open ended questions and check lists respectively. Even this limited experimental set-up was a major project in itself.

We wanted primarily to find out how type of questioning interrelated with data collection mode from a response rate point of view (which was extremely important at that time).

We knew, that if we would get a positive result for the self-completion and frequency of reading question combination, we would automatically solve lots of other problems. Among those were, for example: avoid the need for daily random sampling and interviewing, coping with odd publishing frequencies and irregular publishing intervals, limiting the effects of interview shock, creating frames of reference to promote composite standard effects, allow simultaneous collection of classification (target market) data, obtain cost-efficiency, which could be used to increase sample size, and getting lower margins of error for the AIR estimates (being the benefit of having many more observations per title with the frequency of reading scale).

¹ Cerha, Jarko: as above

In summary, the outcome of this pilot study was that a combination of postal (self-administered) questionnaires and a 7-point frequency of reading question would be the optimal solution from a response rate perspective:

	Response rates with different methods of contact:			
	Telephone	Personal	Self-admin.	Total
Reading frequency question:	%	%	%	%
5-point scale	72	86	87	82
7-point scale	53	75	95	74
16-point scale	44	70	91	74
Total	56	77	91	75

Testing reliability of frequency of reading in combination with self-completion

In subsequent studies¹, before the survey was launched commercially, an extensive control programme was drawn up in consultation with our Scientific Advisory Board. The reliability of the methodology was checked (with a test/re-test method) with very positive results. The internal individual consistency in the results obtained was remarkably high, not least for the frequency of reading questions.

Several validation attempts and plausibility analyses were also made, which further supported the belief that this methodological set-up complied with our demands.

This research procedure has now been successfully used, basically identical, for almost 30 years. A methodological adjustment and improvement of the frequency of reading scale was made in 1984, primarily to correct for suspected influence of parallel reading.

The results of this survey (ORVESTO) have been very stable over the years (see findings in my second Symposium paper). The rest of this paper is devoted to describe and motivate the research procedures in detail, with supplementary findings from other researchers.

Details of the ORVESTO operational survey procedures

The Swedish ORVESTO survey differs in most respects from other readership surveys around the world:

1. The sample

The sample for this survey is a *pure systematic random sample* directly drawn (in one step) from the continuously updated official Swedish population register. It is a systematic sample as far as age is concerned, since the register is in order of age and every *n*:th entry is selected. Because of the sample size (30,000 completed interviews/year) the relation male/female is also basically in perfect balance, as is the geographic distribution. It seems fair to say that the sample is second to none.

2. The data collection procedure: self-administered questionnaires

The information (media consumption as well as the vast amounts of TGI, lifestyle and other classification data) is collected in one step (a procedure followed basically unchanged since 1971), by means of a *self-administered questionnaire*, mailed to and from the respondent. (In later years detailed information on TV and radio is also collected via a subsequent 8 days diary to the same respondents.) The result is a true single source survey.

The quality of the data has been checked and approved on several occasions over the years in different ways, both with regard to reliability and to validity (to the extent that it is possible to check). Non-response structure has been studied, not only by sex, age and geography, but also by Acorn and Mosaic type classifications, with only insignificant deviations being noted.

The possibility to collect data in this way (with response rates of 65-70% of the gross sample) is probably unique to the Nordic countries, and Sweden in particular. It rests on two conditions: a good ability and habit of reading and a positive attitude and tradition to 'reporting' to authorities (or those who can be taken to be somewhat of authorities).

¹ Cerha, Jarko: as above

Newspaper readership figures (daily reach according to national readership surveys) confirm the first condition:

Finland	87%	Austria	75%	Belgium	56%
Switzerland	87%	Denmark	74%	France	53%
Norway	85%	Holland	72%	Italy	44%
Sweden	83%	Ireland	63%	Greece	47%
Germany	81%	United Kingdom	60%	Spain	38%

The second condition, a positive attitude to 'reporting', is also there (in Sweden, that is), even if response rates have fallen by approximately 0.5% units/year in the last 30 years (in 1965 we were sometimes getting up to 90%!).

The most important characteristic about newspaper reading is, that there are very small differences between young and old, between low and high educated etc. Those not subscribing to their own copy may well read at work etc. This is reflected in our survey, in that response rates do not differ in any systematic way between subsets of the population.

3. Research findings supporting the self-administration approach

There has always been a lot of scepticism about self-administered interviewing. However, in later years the acceptance seems to grow widely, when researchers discover the benefits of the amounts of reliable data possible to collect, in one single interview and the economy connected with that (while avoiding the problems of fusion as well as interviewer bias and the need for daily random samples).

(1) One example is Jean-Louis Chandon & Carole Fagot², who in their paper referred to two studies: "...De Leeuw and Van der Zouwen (1992) compared systematically postal, face-to-face and telephone collection modes. They concluded, after many comparisons, that...the answers to postal questionnaires are more precise and less sensitive to social prestige biases (*my italics*). Returned postal questionnaires show very few omissions or non response. Lejeune and Bied-Charreton (1992) demonstrated, in a comparative study on bank data, that the postal mode gives more precise estimates (*my italics*)."

Chandon & Fagot² report from their own research the hypothesis that "... in front of the interviewer, respondents might have a tendency to inflate their answer "Yes" to the screening question." ... "With the self-administered format there is less time pressure and no 'social desirability', therefore there should be no reason to screen in a magazine that was not actually read."

They found that the hypothesis was correct: "...magazines with the largest discrepancies (=higher figures for face-to-face) are either well known and/or prestigious."

They also found, contrary to their expectations, that the absence of title rotation and interviewer probing (in self-administered surveys) did not cause respondents to miss or under-report "less-known" titles.

(2) Dr. Scott C. McDonald³ reported, relating to problems of memory (referring to experiences from cognitive psychology): "...The evidence is that the pressure to respond quickly reduces accuracy of recall behaviour (Cannell et. Al., 1977; Neter and Waksberg, 1964; Sudman and Bradburn, 1982)."

" If people have time to calculate rather than estimate, their responses will generally be more accurate. The cognitive survey approach encourages people to "think out loud" or to "count out loud", recognising that while memory itself may be fallible, it is the pressure to respond quickly that prevents people from fully accessing their memories.

The problem of response time interacts with mode effects. *Time pressure is negligible for self-administered questionnaires, and is most acute for telephone surveys (my italics)*, where conversational delaying signals are more limited than in face-to-face (limited, indeed, to utterances like "um"). On the telephone people give shorter answers and allow fewer pauses.

Chances are, they do less calculation in responding to retrospective behaviour questions, so there is reason to fear that this mode of data collection yields the least reliable estimates of recalled behaviour unless explicit steps are taken to encourage people to slow down and reflect on their answers."

(3) An odd personal reflection: It is perhaps overlooked that in *personal readership measurement interviews, using Recent Reading, the respondent is in fact often largely doing a self-administered interview, albeit being supervised*. He has to sort cards with publication titles on, normally first in some kind of filter question procedure, and then (again by sorting) which titles he has read or not within the last publishing interval. Or, as in the recent German experiments, where the respondent actually keys in his answers on a portable PC with a touch-screen-technique.

² Chandon, Jean-Louis & Fagot, Carole: as above

³ McDonald, Scott C. "Response effects in survey measures of behaviour: Insights from research in other fields" Worldwide Readership Research Symposium VI, 1993

4. Odd magazine publishing frequencies: the unsuitability of the Recent Reading technique in Sweden

In most countries the AIR figures are the result of a Recent Reading (RR) question (read in the last publishing interval). This result is then used (as a 'true' value) to attribute probability values to the various groups in the frequency question being used. (This procedure is in itself dubious. If I remember my textbook correctly, it is not 'allowed' to validate one method by another. This is a relevant observation insofar as we know that results from RR questions are sometimes far from the 'truth', even for dailies, as proven by Belson's studies of reliability in readership long ago).

The RR questioning is normally based on a combination of: (1) personal interviews, (2) the usage of visual prompts (often mast-heads), (3) readership questions presupposing regular and 'normal' publishing frequencies: daily, weekly, fortnightly and monthly, and, under optimal circumstances (4) daily representative samples.

Apart from shortcomings with the RR technique as such, which have been exhibited and heavily debated over several decades, the method was (and is) not very suitable for Sweden for other reasons. Firstly: personal interviews are extremely costly, when population density is as low as in Sweden.

Secondly, and more importantly: the number of issues published vary a lot between titles. Actually, the fact is that *in two thirds of the cases the publishing intervals do not conform correctly either to weekly, fortnightly or monthly:*

Issues/year	No. of titles	Aver. publishing interval (in days)	
4	9	91	
5	3	73	
6	12	61	
7	2	52	
8	8	45	
9	2	40	
10	17	36	
11	12	33	
12	24	30	(Monthlies: read in the last 30 days)
13	4	28	
14	1	26	
16	1	23	
20	2	18	
21-22	5	17	
23	1	16	
24-25	2	15	
26	7	14	(Fortnightlies: read in the last 14 days)
30	2	12	
33	1	11	
40-42	5	9	
46-48	2	8	
52	13	7	(Weeklies: read in the last 7 days)

Out of 135 titles in total, only 44 (= 33%) appear with an issue frequency, that complies exactly with the issue frequency alternatives in the normal RR questions! How can you potentially handle this situation correctly with a RR technology?

1) Using the individual average publishing interval is hardly an alternative: "*Did you read or look into publication X in the last 23 days?*" It is perhaps possible to attempt to come closer to the actual intervals, rather than by using the 'normal' alternatives, but the problem still exists. In the case of publications with a low publishing frequency there is also sometimes the problem of irregular publishing over the year, which makes it even more difficult.

2) Could one disregard the problem and use the nearest response alternative? Does not seem fair:

- A publication with 10 issues/year has an average publishing interval of 36 days, which actually is 20% longer than for a regular monthly. Treated as a monthly it would then be at a 20% disadvantage AIR-wise.

- A publication with, say, 20 issues/year has an average publishing interval of 18 days. Treated as a monthly it would be at a 40% advantage, and treated as a fortnightly it would be at 29% disadvantage.

3) What about compensating by some corrective weightings? Considering the general problems with RR (replicated and parallel reading, telescoping etc), and relating those to the individual differences in the build-up of a readership over time (often continuing long after the publishing interval has expired), seems to potentially make corrective weighting procedures a hazardous exercise, particularly for low-number-of-issues-per-year-magazines.

From Canada, Hastings Withers⁴ reported problems of a similar nature, where 20 out of 70 publications had an irregularity in publishing intervals, that prevented the use of RR technique, and lead to the introduction of frequency of reading measurements.

5. Increasing problems with odd publishing frequencies and irregular publishing intervals

As indicated above only 33% of the 135 magazines measured have 'normal' publishing frequencies: weekly, fortnightly and monthly. Incidentally, in 1987 the corresponding figure was 44%, which means that the problems has been aggravated over the last ten years. A similar development is probably emerging also elsewhere, making RR measurements more difficult and unsuitable.

At the same time, the actual publishing *intervals* are getting increasingly irregular over the year. Issues tend to be more concentrated to periods of heavy advertising spending. What is even worse is, that scheduled issue dates are not always respected by publishers, thus increasing the irregularity.

Both of these circumstances seem to aggravate and reinforce the unsuitability of the recent reading approach, while our experience over the years is that a frequency of reading question is not sensitive to these difficulties.

6. Using a non-equidistant semi-nominal 7-point symmetric frequency scale

In the end the ultimate purpose of a readership measurement exercise is to get individual reading probability figures to be used in computerised media analysis. It was therefore tempting to try and see if it would be possible to collect reading frequency information, to be used both to get the probabilities and the AIR.

As reported above we found in our pilot study that people were quite capable of answering frequency questions about readership, in a consistent way, with self-administered questionnaires, with very high response rates. This basic procedure has been followed ever since.

Since 1984 we are using a 7-point frequency of reading scale, which is semi-nominal and non-equidistant. It is printed, in our questionnaire, together with mini-mastheads of the publication names, and the respondent is asked to state:

"Of the issues published I read:"

- No issues
- Almost no issues
- Approximately 1 out of 4
- Approximately 2 out of 4
- Approximately 3 out of 4
- Almost all issues
- All issues

		Av de nummer som ges ut läser jag:						
		Inga nr	Nästan inga nr	Ca 1 nr av 4	Ca 2 nr av 4	Ca 3 nr av 4	Nästan alla nr	Alla nr
AFTONBLADET	Antal nr/år (ca) vard 300 sönd 50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
EXPRESSEN	Antal nr/år (ca) vard 300 sönd 50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

For each title the number of issues per year is stated (between the logo and the boxes) for the respondent's reference.

Contrary to most other frequency scales our scale is not mathematical, i.e. asking "how many out of the last 4 (or 6 or 8 or 12 etc) issues etc did you read or look at". It is neither explicitly retrospective, nor time related. It simply asks for what proportion of the issues published "...do you read?". Our contention is that the scale is interpreted as 'usual/near-future intended' behaviour.

The reason for choosing the scale we use was that it should as well as possible (a) relate to the respondent's natural frame of reference, and (b) be adapted to the fact that most reading frequency distributions are J-shaped (or reverse J-shaped, depending upon the direction of the scale).

(a) When creating the frequency question and the frequency scale we tried to ask ourselves, what our frame of reference would be, and how we would normally express ourselves, if we were to describe our reading habits. We then had several informal discussion groups with various people in our staff. The conclusion was, that asking "...how many of the last x issues did you read or look at?..." was a way of asking that did not match the way in which the respondent would think about it. However mathematically correct, asking about 'the last x issues' was regarded as unnatural and difficult to imagine. A direct question about proportion of issues read was considered to be a more understandable alternative.

(b) About the J-shape (or reverse J-shape) of the distribution of the answers to the frequency scale:

For most publications the concentration of answers lies in the Non-reader category. The second typical category consists of those claiming to read all or almost all issues (or i.e. 5-6 out of 6). Reading 2-3 out of 4 issues is a less typical behaviour.

Our conclusion then was that the scale needed to be more detailed at the ends than in the middle. Having only the possibility to choose between 'No issues' and '1 out of 4' may lead the respondent to either overestimate his reading (by choosing 1 out of 4) or to underestimate (by choosing 'No issues'). Likewise, there is a gap between reading 3 out of 4 and reading all issues.

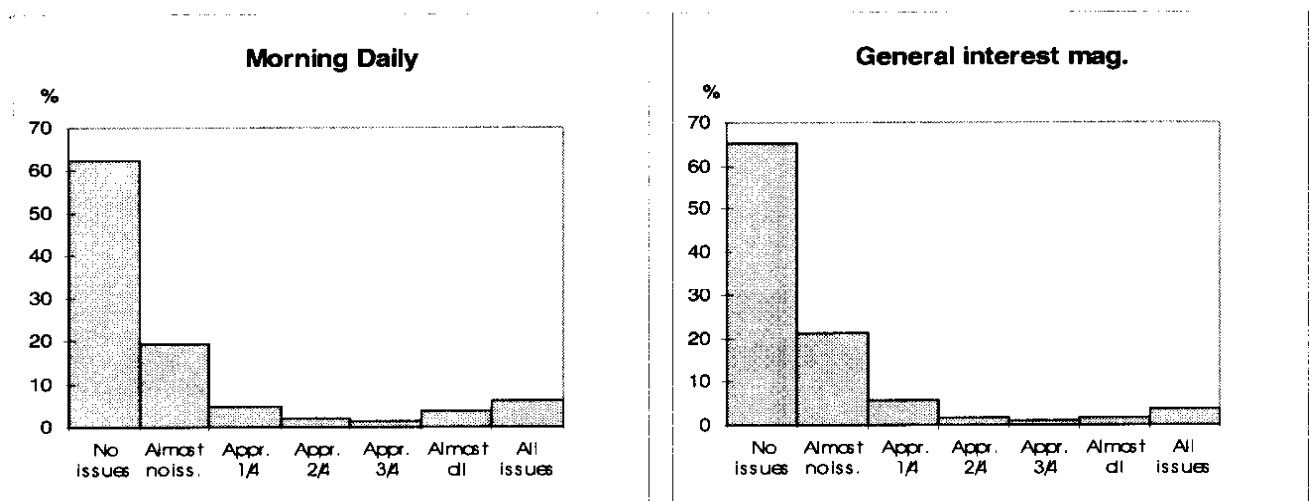
⁴ Withers, Hastings "Irregularly published magazines: Frequency as a means of estimating readership in a recent reading environment" Worldwide Readership Research Symposium VII, 1995

We therefore introduced and added the alternatives "read almost all issues" and "read almost no issues" respectively, to give the respondent a possibility to be more specific about his regular and his sporadic reading. In doing that we still kept the scale symmetrically balanced, which would help the respondent's interpretation of the scale (Composite Standard).

We also decided to put the "no issues" alternative first on the scale and "all issues" to the far right. Reason? The most commonly used alternative should be closest to the masthead, and it should take a serious movement of your hand to reach to the far right end with the alternative "all issues" (remember: self-completion questionnaire).

Some examples of frequency distributions:

	Morning daily %	Evening daily %	General int- erest mag. %	Business mag. %	Car mag. %
No issues	62.5	39.6	65.2	86.2	76.2
Almost no iss.	19.2	34.5	21.2	8.1	15.2
Approx. 1/4	4.9	13.1	5.6	3.1	4.8
Approx. 2/4	2.0	4.7	1.6	0.9	1.6
Approx. 3/4	1.5	2.5	0.9	0.5	0.7
Almost all	3.6	3.8	1.8	0.7	0.7
All issues	6.2	1.9	3.6	0.6	0.9
	100	100	100	100	100



From the graphs we see how the 'almost no' and 'almost all' issues alternatives fill the gap between 'no issues' and '1 out of 4' and between '3 out of 4' and 'all issues'.

7. Establishment of AIR (Average Issue Readership)

The AIR figures are calculated directly from the results achieved from the frequency scale, using apparent probabilities. The probabilities are, however, slightly different for the two verbal alternatives at the ends of the scale, by publishing frequency:

	Reading probabilities by publishing frequency		
	'Monthlies' - 29 iss/yr %	'Weeklies' 30-100 iss/yr %	'Dailies' 150+ iss/yr %
No issues	0.0	0.0	0.0
Almost no issues	12.5	10.0	7.5
Approx. 1/4	25.0	25.0	25.0
Approx. 2/4	50.0	50.0	50.0
Approx. 3/4	75.0	75.0	75.0
Almost all issues	87.5	90.0	92.5
All issues	100.0	100.0	100.0

The 'almost no issues' position is weighted marginally higher for monthlies and lower for dailies, than for weeklies, and vice versa for the 'almost all issues' alternative. The weights were established as a result of several considerations. Logically we felt that i.e. 'almost all issues' would be closer to 'all issues' for dailies than for monthlies and that it should motivate a slightly higher weight. This contention was supported by the fact that the frequency distribution of the answers was typically reversely J-

shaped for dailies, while they were more L-shaped for monthlies. Additionally, the results of focus group discussions strongly supported our view of how the scale should be interpreted.

It should be mentioned here that a very similar scale is used in Canada, as Hastings Withers⁴ reported in his 1995 Berlin paper. The main difference is, that their scale starts with the "all or almost all" alternative, and that these alternatives are collapsed into one.

Withers⁴ had an interesting further discussion in his paper: "... One of the criticisms of the frequency method is that the scale can be misunderstood by the respondent to mean a variety of other things⁵. For example, some people may not understand we are asking about the number of issues read out of four issues of the publication, and reply instead in terms of their attachment to the publication. Should this be a strongly significant effect then there is a risk that even if the overall readership numbers might be close between RR and frequency, their demographic make-up might be quite different. ... The test did not show that kind of tendency. The profile by age, sex, and income for the irregular publications shows no significant differences between the RR AIR and the frequency AIR."

Another interesting support for the frequency of reading approach was delivered by Dr. Scott C. McDonald³ under the intriguing heading: "For Comparative Purposes, Ask for General Estimates of Frequency". He continues: "...The Seattle HMO study ... found that when asked to provide a general estimate of the number of visits made to a doctor in the past year, they reported 87% of the actual visits. However, when frequency is estimated from the recollection of specific visits, only 39% of the actual visits are remembered. This under-reporting occurs, even if lax rules are applied to the dating of visits. If people have a pretty good idea of how many times certain events happened in their lives, even if they cannot specifically recall each instance, then the more general estimates may have value."

8. The importance of the shape of frequency scales to media accumulation results

The way we create our media analysis computer programs and algorithms is strongly influencing the shape of the curves of accumulation and of the reach/frequency estimates that we obtain.

An additional dimension in this context, not frequently considered or discussed, is the strong impact caused by the shape and construction of the frequency scales themselves. It is my contention that the closer the frequency scale is matching the 'natural' distribution of the frequency of reading, the more it should be able to reflect the 'real' accumulation.

9. The quality of reading dimension - heavily related to reading frequency

Some readership surveys include questions about proportion read, quality of reading and/or attitude to the publication ("how much would you miss it if it ceased to be published") etc. These dimensions are of course of interest to ad sales people, to be able to show that their advertising vehicle is actually read and liked.

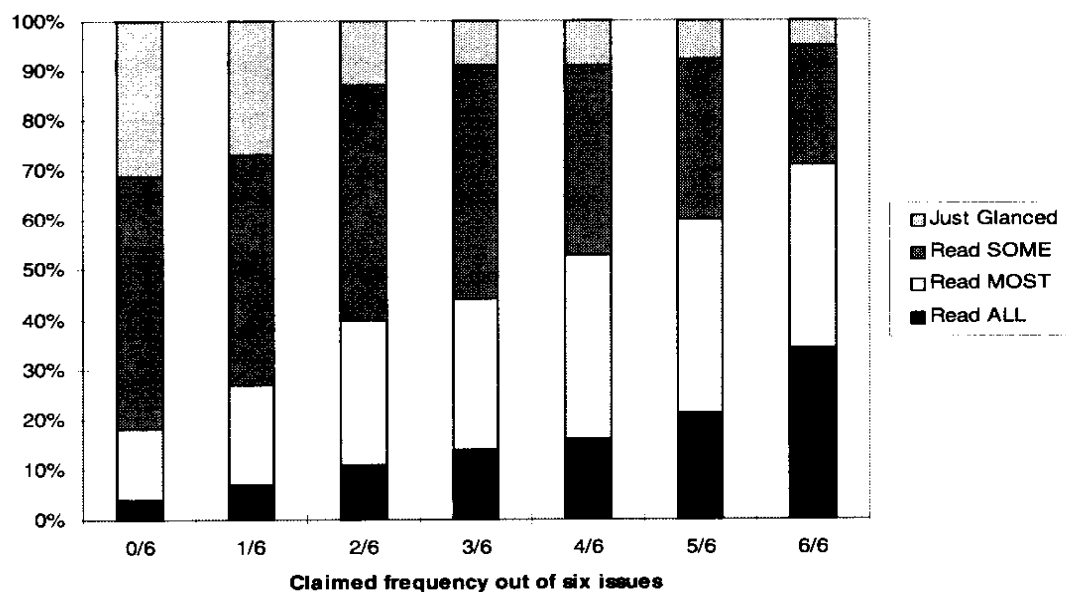
The late Wally Langschmidt⁵ has an interesting illustration to this, where he "...shows the marked differences between thoroughness of reading of the regular readers and the occasional readers. The chart (see below) is based on a cross-tabulation of claimed thoroughness of reading within the various frequency groups. The figures are for a well-known South African magazine and are taken from the AMPS '77 survey."

³ McDonald, Scott C.: as above

⁴ Withers, Hastings: as above

⁵ Dear Reader Michael Brown - p116 quoting Wally Langschmidt and Pym Cornish

Graph showing how the thoroughness of reading improves as the frequency of reading increases



This chart (and other sources as well) show the strong correlation between frequency of reading and relation to a publication, and is another reason for emphasising the importance of using and developing good frequency scales. In the further computer analysis of the data, it is then i.e. easy to amend AIR (if one would want to), by excluding infrequent reading alternatives, or add further weights if found necessary.

Then there is, of course, the question if there is any correlation between amount of the editorial read and ad exposure. Peter Masson⁷ discusses this in one of his 1995 Berlin papers (referring to the works of Cerha¹). The point he makes is, that ad exposure is related to the target market person, rather than to the advertising medium: "...when we add into the target definition an 'interest in private cars' requirement (very/rather interested) we find not only a higher selectivity by these titles (business titles and car magazines) but also a strong discrimination between them and in favour of the special interest titles...". "...It is our contention therefore, that the media researcher can better, and more cost effectively, serve the planner by including purchase intentions or usage/loyalty data along with product and topic interest - rather than pursuing within mainstream media studies, very laborious 'proxy' measures of vehicle effect...".

It goes without saying that an inclusion of such target market data, in a true single source survey concept, will additionally increase the accuracy of the analysis. In the end the availability of relevant target market information is very crucial for the end result of the analysis. However correct readership figures may be, they cannot be used and analysed properly, unless they can be related to relevant target markets.

5. Concluding remarks

According to Chandon & Fagot²: "...Overall, it appears that, for print audience measurement, self-administered surveys mailed to panel members are a reliable alternative to traditional face-to-face surveys. They have excellent return rates, lower costs and they can eventually be combined with purchase data to develop behaviour-based media planning."

These and other references are very strong arguments for a self-administered methodology. The respondent not only decides when is the best time to fill in the questionnaire, but also can take the time necessary to look at and understand the questions and get an overview of what it is all about.

¹ Cerha, Jarko: as above

² Chandon & Fagot: as above

⁷ Masson, Peter "Towards a more cost effective route for evaluating media in terms of communication effect"
Worldwide Readership Research Symposium VII, 1995

From the questionnaires we get back it is obvious, that the self-administered procedure works very well. This includes the function of the composite standard concept, in that we see answers being frequently changed by the respondent while going through the questionnaire (much to the annoyance of our OCR data entry people).

It should be complemented with the observation that it works best in combination with a good frequency of reading scale.

Obviously our procedures and experiences are not necessarily applicable to other countries and cultures, with different possibilities and conditions. There is, however, a lot that leads us to expect that future development will inevitably go in the self-administered direction, be it the new German touch-screen technique, questionnaires via Internet or in other forms. This also opens up for major methodological changes, including development of the so far unexplored potential of the frequency of reading approach.

There is a lot to be gained, not just economically.

