

Wayne P Eadie

Newsweek
New York, USA

Richard L Lysaker

Audits & Surveys, Inc.
New York, USA

3.6 Developing a magazine readership validating technique

BACKGROUND

In the United States, the 'Recency' or 'Recent Reading' and recognition techniques commonly referred to as 'Through the Book' or 'Editorial Interest' are competing methods for measuring average issue audiences. These two methods exhibit systematic and substantial differences in audience estimates. The differences are particularly pronounced for monthly magazines.

This chronic state of affairs has created substantial interest in developing a validation measure against which the audience data of the two commercial services could be compared. The beginnings of such work were reported in New Orleans in 1981 under the heading, 'The ARF Certitude Tests'. This work, however, did not yield a certitude or validating measure.

In deciding how best to expand upon this initial work, we noted that in New Orleans there was considerable interest in, and discussion about, developing the 'first-time read yesterday' approach as a validating measure. We have pursued this interest with a series of experiments designed to determine if this technique has potential for a larger scale readership study.

The 'first-time read yesterday' technique has a number of virtues.

(1) From an empirical point of view, it appears to put the least strain on one of the weakest elements in readership research — which is memory. Those who read a publication on only one day have only to remember yesterday reading. Those reading on more than one have to remember yesterday reading and whether or not the issue has been read previously.

(2) As a readership model, it measures readership independent of issue age. It does not limit measurable readership of weekly issues to the first five weeks of their lives or of monthlies to the first 12 weeks of their lives, as it is the case with the 'through-the-book' method.

(3) As a readership model, it counts all issues read for the first time and ignores replicate reading. It does not make the assumption that the levels of parallel and replicate reading cancel out each other, as the recency method does.

(4) The first-time read yesterday technique lends itself to fairly direct validation techniques for documenting the existence of both underclaiming and overclaiming.

This paper deals with two methodological studies

and one validation study conducted for *Newsweek Magazine* by Audits & Surveys based on the 'first-time read yesterday' approach. But first, a little background concurrent with *Newsweek's* interest in developing this approach. The Magazine Publishers Association commissioned Audits & Surveys to design and conduct a study to measure magazine page traffic or MPX. This study used the 'Yesterday Recall' method to measure both audience levels and total reading days for a series of 31 magazines. The average issue audience component of the MPX method was 'first-time read yesterday' measurement.

This study yielded a combined audience level for these magazines equivalent to about 18% of the population per day. However, the purpose of the MPX study was to measure not audience level, but page traffic. Therefore, the exact audience measure used was not subjected to systematic experimental study, nor was the sample size of the study large enough to provide a reliable audience measure.

Our work at *Newsweek* pursued the opportunity of building on the MPX work. To date, we have completed experimental studies, a validation study and an actual readership measurement study which Steve Douglas will report on later.

In our experimental work, we chose two operating hypotheses. These hypotheses represent the conventional wisdom in the US that recent reading which estimates higher levels of reading may in fact suffer from too much overclaiming vs. underclaiming. Through-the-book which generates lower audience estimates may suffer from too much underclaiming vs. overclaiming.

The hypotheses were:

(1) For magazines in total, true readership levels in the US are likely to be at or between the levels reported by the two commercial services.

(2) The relationship of monthly to weekly audiences would be at or between those reported by the two services.

Our approach was to experiment with the readership question plus rotation and pick a tentative best fit approach based on its producing audience levels in line with our initial hypotheses. Such a technique and question would then be subjected to a validation study where actually observed reading was to be the standard of truth.

3.6 Developing a magazine readership validating technique

ANALYTICAL VS. ENUMERATIVE STUDIES

The total *Newsweek* research programme on developing the first-time read yesterday method into a validating technique has two basic parts.

Part I Developing work using analytical studies

Part II Audience measurement using an enumerative study

This distinction between an analytical study and an enumerative study is a critical one, as it dictates whether sample estimates lend themselves to across-study comparisons or should be limited only to within study comparisons.

The distinction is probably best discussed by Deming, who has written extensively on the subject (ref. 4,5,6,7). He points out that *enumerative studies* are those which quantitatively describe the characteristics of a universe. The results of such studies lend themselves to cross-study comparisons. All national readership surveys are of this kind.

On the other hand, *analytical studies* are those which deal with the dynamics of different treatments or processes. In such instances, we cannot talk about the size of a universe but we can clearly talk about which experimental treatments in the study yield larger or smaller estimates of readership or any other characteristics. These studies lend themselves to internal analysis but not to comparisons between studies.

Analytical studies tend to be smaller than enumerative ones. For example, the distinguishing feature of an analytical study is that only a subset of the total universe is sampled. In the *Newsweek* development work, we subsampled time by measuring across a three-week interval in the first developmental study and a two-week interval in the second developmental study. In both of the analytical studies, multiple publication intervals were covered for the weeklies but only part of one publication interval for the monthlies. As these two studies were sequential in nature, they were, of course, conducted during different months and thus, to some degree, involved different issues. Therefore, for the developmental work, by design, the analysis must be confined to within study comparisons.

STUDY I: EXPERIMENTING WITH THE FIRST-TIME READERSHIP QUESTION

This study measured the effect of a 2 by 2 matrix of variations on the basic yesterday readership question. The question (with variations tested) was:

'Did you happen to...

...open that issue of (magazine)...

...open that issue of (magazine), however briefly

...read or refer to that issue of (magazine)

...read or refer to that issue of (magazine), however briefly

...before yesterday either at home or away from home?'

Each of the four possible combinations shown above was tested by administering the questionnaire to approximately equal numbers of respondents.

In this study, the 31 magazines used in the MPA/MPX study referred to earlier were randomly ordered. Four orderings were used, across each of the four versions.

A total of 843 interviews were completed with a national sample of adults in telephone households, using random digit dialling. Interviews were evenly distributed over a twenty day period, (approximately 10 interviews were completed per day per treatment). Respondents were randomly selected from within households, and the four samples were then balanced by respondent selection probability, age, and sex.

The questioning sequence was as follows:

— (Introduction questions). Thinking about all the places you happened to be yesterday, including when you were at home or at work, travelling or doing something else...please tell me if you, yourself, happened to watch any television yesterday?...did you happen to read any newspapers yesterday?...did you happen to read any magazines yesterday?

This last question was a filter question: only those who responded yes were asked about yesterday readership of specific titles.

— (Yesterday readership). Now, I'm going to read you the names of some magazines. For each magazine please tell me whether or not you happened to read or look into any issue of that magazine yesterday...please remember...I'm only interested in magazines you happened to read or look into yesterday, (see magazine list at end).

For each magazine read or looked into:

—(Issues read). How many issues of (magazine) did you read or look into yesterday?

— (For each issue read/looked into, ask test question — first time readership).

— Demographics

The question using the phrases, 'read or refer to' and 'however briefly' yielded data which best fitted our initial hypotheses, namely it best fitted the criterion of a total level between that of the two services and a weekly/monthly audience mix most like that of the syndicated services. However, it did show monthlies with a higher audience than weeklies which was contrary to our initial hypothesis.

Results of this test are shown in **Table 1**.

3.6 Developing a magazine readership validating technique

TABLE 1
Comparisons of daily audience estimates

Test combinations	Weeklies	Monthlies	Total
Base (US adults 18+)	(100%)	(100%)	(100%)
'Open'/'...	7.2	13.7	20.5
Open/'however briefly'	10.9	11.9	22.8
'Read or refer to'/'...	7.7	10.1	17.8
Read or refer to/'however briefly'	8.4	9.0	17.4
*Recent reading (MRI)	9.7	9.0	18.7
**Through-the-book (SMRB)	7.4	5.7	13.1



* Recent reading method using a 6-month screen and a card sort and board. Cards were also used for last 7 days reading.

**Also using a card sort, respondent is shown a stripped 5-week old issue of a weekly, a 10-week old issue of monthly.

In order to make the comparisons shown in this table, average issue audience data from the US syndicated services (MRI and SMRB) were derived by taking total reported average issue audience levels and dividing by 30 for the monthlies and 7 for the weeklies. That number is expressed as a percent of the total US population.

In the next study, we used this form of the question as a benchmark for further exploration.

STUDY II: EXPERIMENTING WITH SEPARATING MONTHLIES AND WEEKLIES PLUS FURTHER EXPERIMENTING WITH THE FIRST—TIME READERSHIP QUESTION

In this study, two variables were tested: the wording of first-time reading question; rotation of presentation of magazine titles to the respondent.

Two versions of the first variable were tested (in the first-time readership question):

(1) 'Did you happen to read or refer to that issue of (magazine), however briefly, before yesterday...' this was the 'best fit' question in the first experiment.

(2) 'Did you happen to read or refer to that issue of (magazine), however briefly at any time before yesterday...'

The purpose of testing these two alternative questions was to determine if the new form of the question would yield a relationship between weeklies and monthlies that was more in line with our initial hypotheses.

Three versions of the second variable, *rotation*, were tested (in the yesterday readership question):

(1) One random list — totally random and without reference to publication interval.

'Now, I'm going to read you the names of some magazines...' — where weeklies and monthlies are combined into one randomly ordered list.

Four randomizations of the magazine lists were used.

(2) Weeklies and monthlies grouped, weeklies first.

'Now, I'm going to read you the names of some magazines that come out every week...' — followed by a randomized list of weekly magazines.

'Now, I'm going to read you the names of some magazines that come out once a month...' — followed by a randomized list of monthly magazines.

(3) Weeklies and monthlies grouped, monthlies first.

Same as the preceding, with the monthlies first, followed by the weeklies.

Four rotations were also used when weeklies and monthlies were grouped separately.

Thus, there were a total of six versions of the questionnaire, representing the 3 by 2 matrix of possibilities as shown in **Table 2**.

TABLE 2

Monthlies & Weeklies combined	Monthlies & Weeklies combined
'Did you happen to read or refer to that issue of (magazine), however briefly, before yesterday...'	'Did you happen to read or refer to that issue (magazine), however briefly, at any time before yesterday...'

3.6 Developing a magazine readership validating technique

<i>Weeklies first, then monthlies</i>	<i>Weeklies first, then monthlies</i>
'Did you happen to read or refer to that issue of (magazine), however briefly, before yesterday...'	'Did you happen to read or refer to that issue (magazine), however briefly, at any time before yesterday...'
<i>Monthlies first, then weeklies</i>	<i>Monthlies first, then weeklies</i>
'Did you happen to read or refer to that issue of (magazine), however briefly, before yesterday...'	'Did you happen to read or refer to that issue of (magazine), however briefly, at any time before yesterday...'

The methodology used was otherwise identical to that described for Study I. Twelve hundred and fifty telephone interviews were conducted over a two week period, evenly distributed by day over the six cells in the experimental model.

As the field work was conducted over only a thirteen day period — which is a little less than half the publication interval of a monthly — the readership estimates generated have to be viewed in relative rather than absolute terms.

These question variations had little or no impact on total readership level estimates. The question form without the additional phrase 'at any time' best fitted our initial hypothesis regarding the relative audience levels for weeklies or monthlies. However, we still found, contrary to our initial hypothesis, that the monthlies had higher audiences than the weeklies.

Tables 3 and 4 show the results of this study.

TABLE 3
Issues read for the first-time as a percent of the total sample

<i>First-time readership question variation</i>	<i>% of people reading yesterday</i>		
	<i>Weeklies</i>	<i>Monthlies</i>	<i>Total</i>
However briefly (Original 'best fit' — study I)	9.9	12.6	22.5
However briefly, at any time	7.9	15.9	23.8

We found a marked effect on the weekly/monthly relationship when we varied the rotation, when all magazines were randomized without regard to publication interval. The readership estimate for monthlies was more than twice that for weeklies. However, when magazines of like publication intervals were grouped and then randomized, the monthlies generated an audience estimate about 40% greater than that for weeklies.

TABLE 4
Issues read for the first-time as a percent of the total sample

<i>Presentation order</i>	<i>Weeklies</i>	<i>Monthlies</i>	<i>Total</i>
Weeklies first	10.0	13.3	23.3
Monthlies first	9.8	14.7	24.5
Totally random	7.0	14.8	21.8

This work was followed by a validation study in which the first-time readership question developed through the prior studies was validated by measuring its ability to capture known reading. In many ways, this work duplicates the approaches used in the ARF certitude studies, except that only one magazine was studied and the readership question in the current study was, of course, the preferred version of the first-time read yesterday questions.

STUDY III: PARTIAL VALIDATION BY DIRECT OBSERVATION

In this study, the ability of a respondent to report first-time reading of a new issue of a magazine was tested under two circumstances.

At home reading (subscribers): housewives in households subscribing to *Newsweek* surreptitiously observed and reported on the television viewing and magazine reading behaviour of their husbands on the first day an issue of *Newsweek* arrived in their home. These reports were collected via phone by Audits & Surveys' field staff on Wednesday morning, December 1, 1982. Any household in which the husband had been observed to read or look into that week's issue of *Newsweek* was contacted later that day (Wednesday) for an interview with the husband. If the husband had not yet looked at the issue of *Newsweek*, the procedure was repeated for one more day, with the wife reporting

3.6 Developing a magazine readership validating technique

Thursday morning on her husband's activities on Wednesday.

In order to ensure that virtually all this reading would, indeed, be first-time reading, households were selected from areas in which the normal delivery date of *Newsweek* is Tuesday. Housewives were instructed to tell their husbands nothing of the experiment¹ and were offered a \$10 incentive for their participation. This procedure duplicates that used in the ARF certitude test no. 2.

Public place reading: New issues of *Newsweek* were placed in waiting rooms of doctors' and dentists' offices, beauty parlours, and barber shops². Observers posing as patients/customers observed the behaviour of people in these waiting rooms. Any who picked up and opened the test issue of *Newsweek* were (before they left) recruited for an interview to be conducted the following day, topic unspecified. Again, a \$10 incentive was offered for participation.

The following day the recruited individuals were recontacted and interviewed using the yesterday reading questionnaire³. This approach duplicated the procedure used in the ARF certitude test no. 3.

All observations were conducted on Monday, November 29, to eliminate the possibility of a respondent having seen the test issue of *Newsweek* before the observation day.

Interviewing for both phases of this study was conducted in two areas: New York and Philadelphia. A total of 59 interviews were completed for the "at home" phase, 55 interviews were completed for the "public place" phase.

Based on the results of the first two studies in this

series, it was decided to use the following questioning procedure: ask about weeklies and monthlies separately. Alternate the two versions: ask about weeklies first half the time; ask about monthlies first half the time, — (first-time readership question) 'Did you happen to read or refer to that issue of (magazine), however briefly, before yesterday either at home or away from home?'

This study indicated a high level of capture of first-time reading, as shown in **Table 5**.

It should be noted that the observation procedure greatly reduced but did not completely eliminate levels of underclaiming — twelve percent of the observed at home and eight percent of the observed public place reading was still not claimed.

These results for public place reading are substantially better than those obtained in the comparable ARF certitude test no. 3 as shown in **Table 6**.

¹ Validation: ten of the husbands who were interviewed in the test were recontacted and interviewed to determine whether or not they had been aware of their wives' observation. Only one man reported observing anything unusual. He said his wife had been "acting weird," but had no idea why.

² Two of each, in each of New York and Philadelphia — 16 public places altogether.

³ Validation: 10 respondents were recontacted to determine whether they had any idea of why they had been selected to participate in the study. One woman thought it might have been because she had been reading magazines, but that was based on what had been asked in the interview.

TABLE 5
Yesterday reader

	At Home		Public Places	
Observed <i>Newsweek</i> readers	58	(100%)	52	(100%)
Claimed read	57	(98%)	49	(94%)
Claimed first-time reading	51	(88%)	48	(92%)
Claimed not first-time reading	6*	(10%)	1**	(2%)
Claimed not read	1	(2%)	3	(6%)
Total underclaim	7	12%	4	8%

*one person in the at-home experiment claimed to have read more than one issue of *Newsweek* on the previous day and was *not* included in the calculations.

**Three people in the public place experiment claimed to have read more than one issue of *Newsweek* the previous day and were *not* included in the calculations.

3.6 Developing a magazine readership validating technique

TABLE 6
Comparisons with ARF studies

	Total magazines	Claimed first-time reading	Percent underclaim
This study (Newsweek only)			
At home	100%	88%	12%
Public places	100%	92%	8%
ARF studies			
Study No. 2. (6 magazines)*	100%	82%	18%
At home			
Yesterday recall by day part			
Cover/table of contents prompts			
Study No. 3. (7 magazines)*	100%	54%	46%
Regular waiting room			
Yesterday recall			
Through-the-book method			

*Including both weeklies and monthlies

It is important to note that this validation study was not designed to be an individually conclusive experiment. Rather, it was designed to determine if the basic approach we developed using the 'first-time read yesterday' technique, warranted further work. The reading against which we validated the technique was first-time reading of a new issue. The next steps in this validation work have to involve a look at non-first-time reading of both newer and older issues and an exploration of overclaiming. However, based on these three experiments, we have concluded that the 'first-time read yesterday' technique shows substantial promise and on this basis we used this technique for the much larger scale readership study reported on by Steve Douglas in the paper which follows this.

SUMMARY OF FINDINGS

- The method was affected by rotation
- The method may be affected by wording
- The method reduces but does not eliminate underclaiming
- 'First-time read yesterday' has potential as a validity method because it can measure overclaiming

NEXT STEPS

- A larger-scale readership study using this technique
- A look at non-first-time reading of both newer and older issues
- An exploration of overclaiming

REFERENCES

- 1 Lysaker, Richard L., (1981). The ARF certitude tests; from *Readership Research: Theory and Practice, Proceedings of the first International Symposium, New Orleans*, Harry Henry, ed.
- 2 ARF Technical Staff, (1978). Measures of magazine exposure, yesterday, recall of subscriber at home reading, *Council Report, Certitude Test No. 2*. Dec. Advertising Research Foundation.
- 3 ARF Technical Staff, (1979). Measures of magazine exposure, through-the-book recall of national waiting room reading, *Council Report, Certitude Test No. 3*. Oct. Advertising Research Foundation.
- 4 WE Deming, (1969). Boundaries of statistical inference, chapter 31 in *New Developments in Survey Sampling* by Norman L. Johnson and Harry Smith, (Wiley-Interscience, pp. 652-670).
- 5 WE Deming, *Some theory of sampling*, (Wiley, 1950 Dover, 1966).
- 6 WE Deming, (1975). On probability as a basis for action, *The American Statistician*, vol. 29, pp. 146-151.
- 7 WE Deming, (1975). The logic of evaluation, chapter 4, *Handbook of Evaluation Research*, by Elmer L Struening and Marcia Guttentag (Sage Publications), vol. 1, pp. 53-67.

Note: The authors wish to acknowledge the contribution of Fred Nicholson of Audits & Surveys in the preparation of this document.