

**William Belson**  
Survey Research Centre  
London, UK

# 2.12 Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

## PART I: VALIDATION

### ORIGINS AND SCOPE OF THE ENQUIRY

#### Origins

During the 1960s, the Institute of Practitioners in Advertising asked the Survey Research Centre, which I directed, to investigate the validity of their National Readership Survey (NRS), which was the forerunner of Britain's principal readership survey of the present day.

That survey, at that time, used an approach which I have summarised below.

An interviewer asks a respondent to go through a booklet of mastheads (91 at that time), calling out the names of each of them that she had 'looked at' in the past three months. At the end of this sequence, the respondent is asked, of each of the publications she had claimed to have 'looked at', when was the last time she had looked at any copy of it. For monthly publications, 'looking at' in the past four weeks (excluding 'today') counted as 'reading'; for weekly and Sunday publications the qualifying period was the last seven days (excluding 'today'); for dailies, it was yesterday.

The request for a validating study sprang out of evidence of considerable change in the results as elements of the method of elicitation were changed and as the order of presentation of specific publications within the list of 91 was varied through rotation.

This request (for a validity check) came from the Technical sub-Committee of the National Readership Survey Controlling Committee, and that committee stayed in close contact with the study as it progressed.

#### Scope

The validity study was to be concerned with the accuracy of the readership figures when readership was defined as in the National Readership Survey. Thus we set out to determine whether a person who qualified as a reader of a given publication had in fact looked at that publication in the qualifying period. We planned to make this sort of check for each of a number of monthlies, weeklies, Sundays, dailies.

Now the limitation of this objective must be made quite clear. In the first place, readership was being defined in terms of 'looked at', whatever that may mean from the point of view of 'possible exposure to advertising material'. Would 'looked through' have the same

relevance to that ultimate criterion? Or 'read or looked at' or 'paged through'? Or something else again? Moreover, the research strategy of the National Readership Survey of that time took no account of what is called *replicated* or *parallel readership*, or of *intensity* or of *frequency* of reading.

It was against this background that we studied the validity of the National Readership Survey in relation to monthlies, weeklies, Sundays and dailies. The results startled us somewhat and I believe they have important implications for the practice of readership research – the implications that seem to have been strenuously ignored in the years that followed.

### THE CRITERION OF TRUTH OR ACTUALITY THAT WAS USED IN THE ENQUIRY

#### The basic problem

The measurement of the validity of readership data poses a major problem. Ideally, it requires that we compare our readership survey findings with what the respondents *actually* did. However, there is no way in which we can establish the absolute truth of what the respondents actually did. In the first place, no form of observation nor any form of continuous self-reporting could do this without upsetting the normality or the memorability of the behaviour concerned. Nor could observation by another person be continuous enough. Secondly, simulations or mock-ups that depart from normality in any way are likely to be especially well remembered, as well as being non-normal.

#### The criterion of truth or actuality that was adopted

In the circumstances, it seems that the best we can do in establishing a criterion of what actually happened is to conduct intensive interviews which are rigorous, extended and detective-like in character. This detective type operation would also have to be designed to reduce such error-making tendencies as we might suspect would characterise the ordinary readership interview, for example, interviewer deviation from instructions, respondent misunderstanding of questions and instructions, carelessness, memory failure, time distortion, lying. Granted that such an intensive process could be developed, we would regard it as providing evidence that approximated to the truth and we would then compare that evidence, respondent by respondent, with the results

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

from the ordinary readership interview.

### THE BASIC DESIGN OF THE ENQUIRY

The basic strategy of the enquiry was to have the ordinary readership survey interview administered to a population sample and to follow each interview with the intensive procedure, thereby making a comparison possible.

#### The first (ordinary readership) interview

In the first interview, the respondent went through a standard national readership survey interview covering all the 91 publications which were normally included in that survey. In the second interview, conducted very shortly after the first one, the same respondent was interrogated by an intensive interviewer covering up to four of the readership claims (either positive or negative) made by that person in the first interview. A total of 1362 London adults went through this first interview. The interviewers doing this work were changed weekly throughout the 12 weeks of fieldwork\* with a total of approximately 45 interviewers taking part.

#### The intensive interviewing stage

Because the intensive interview is the key to this enquiry, I have described it in some detail, drawing from pages 6–20 of the full report.\*\*

#### SELECTION AND TRAINING OF THE INTENSIVE INTERVIEWERS

These key people were selected from an initial 220 applicants (Press advertisement). There was an initial selective scanning of applications in terms of academic background and previous experience. This was followed by a tape recorded and observed test in the field with selection in terms of ability to interrogate intelligently, purposefully and with gentle tact, ability to see and follow a lead, detective mentality. The 12 or so people passing this test then entered into a training selection sequence lasting for two weeks. Processing during this period took the following forms: (a) there was instruction in the methods of intensive interviewing as planned for the present enquiry; (b) the interviewers went out to practise the technique in the surrounding suburbs, tape recording their interviews; (c) during the afternoon of each day,

each group member played the tape recording that he judged to be the best of his lot to the whole group. He then criticised himself and after this the whole group examined his performance. Between them, they tried to spot weak points in the interview, asking the interviewer why he had not followed up this or that lead, and so on. Next day, there was more practice interviewing in the morning and a further critical session in the afternoon. And so on for the rest of the days in the training period. It soon became evident which of the trainees were sufficiently gifted with intelligent detective mentality, who were self-critical, who could stand being criticised, and could interrupt respondents in a courteous way, and so on. Selection was made on the basis of such criteria. There was an operational team of intensive interviewers and a small group of reservists. From time to time a training session was held again to top up the operational team and the reservist team.

#### THE DIFFERENT STEPS IN THE INTERVIEWING PROCEDURE

The intensive interviewers all worked to appointments made for them by the first stage interviewers. The intensive interview was to be paid for and the respondent had been told that this second interview would last for a little over an hour. The intensive interviewer arrived with an instruction to apply her checking procedure to four publication claims, either two dailies and two Sundays or two weeklies and two monthlies. She carried a sealed envelope containing the questionnaire from the first interview (she did not know if the respondent had qualified or not as a reader with respect to the four test publications). She also carried the last three issues of each of the four publications and a small tape recorder. She took the respondent right through the following stages of her questioning procedure.

(a) to start with, the intensive interviewer explained the purposes of the intensive interview, so that the respondent would see the point of the rigorous and sometimes repetitive nature of the questions that were asked.

(b) after this, she questioned the respondent about her reading situation: what publication she read regularly or ever; where and when she usually did her reading of publication; where else did she occasionally see publications; how she 'comes by' the different publications she looks at; reading over other people's shoulders during travel.

This information was simply stored and would be used as a general background to the questioning that was to follow – just as a detective draws upon his carefully amassed background knowledge of a case in the course of specific questioning. It was also expected that talking about a usual pattern of reading would make it safer to

\* To ensure the use of a wide range of interviewer abilities and tendencies and also to reduce the possibility of the National Readership Survey interviewers realising what the real purpose of the enquiry was.

\*\* Belson, W A (1962). *Studies in Readership Business Publications Ltd, London.*

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

ask respondents, later on, about use of publications in specific periods (ie make them less inclined to answer in terms of the 'usual').

(c) beyond this stage, the intensive interviewer dealt with just one of her four test publications at a time. She began with a familiarisation tactic. The respondent was taken through all three issues of this test publication. She was asked to turn through each, to see if she recognised anything in it or found anything of interest. This was not a test of past exposure. It was meant solely to make it quite clear to the respondent that *this* was a publication that subsequent questions would be about.

(d) now came a vital phase. With the test publication in front of her, the respondent was asked "if she had ever looked at any copy of the publication before the present occasion". If she said "no", she was challenged with information about her own usual and occasional reading opportunities. If still "no", this phase of questioning was discontinued. If "yes", she was asked how she could be sure. If still "yes", she was asked to think carefully back to the very last occasion she looked at it. She was asked to visualise that occasion: who she was with, where she was, any special circumstances. She was asked *not* to try to date it just yet. After this she was challenged with her own usual and occasional reading opportunities on the grounds that this might produce recall of a later 'looking at' event. When the respondent had settled on a last 'looking at' occasion, she was asked to think (again/further) about the detail of that last occasion: where she was at the time, time of day, what else was going on, who was with her, where she was going, what it was she looked at, how she came to be looking at it. Only when this had been done was she asked to put a 'time ago' or a date to that last 'looking at' event. And when she *did* arrive at a date or 'time ago', she was challenged to say how she could be sure of that estimate. Tough challenging about *just when* was relented when the 'time ago' was far outside of the qualifying period.

When a date/time ago was confirmed by the respondent, this was entered on the questionnaire and the same procedure, starting with familiarisation, was then started with the next test publication. And so on for all four.

(e) after that came a stage called 'confrontation'. The intensive interviewer prepared to open the first interview envelope and, with a tape recorder running, said "I am now going to open this envelope with your first questionnaire in it". This done, she got the respondent to help her in comparing the first interview readership claims with the claims for the intensive interview. Where there was a difference, the intensive interviewer asked the respondent to think about which was the nearer to being correct. Almost always the second (intensive) estimate was selected. But either way, the respondent was then

asked to explain how she thought the difference had come about. The respondent's replies were subject to probing by the intensive interviewer, exhaustive probing if necessary. The 'explanations' given were regarded as material for the development of hypotheses about the causes of error. Sometimes this evidence would be backed up with the tape recorded evidence from the first interview. Where the claim related to what the first interviewer had done, it was possible to see if other respondents made the same sort of claim about the same interviewer (as indeed did happen). This procedure was repeated for each discrepancy.

(f) on return to base, the intensive interviewer completed her report from her tape recording. Under her new/final estimate of when the respondent had last looked at the test publication, she entered, firstly, any evidence leading her to doubt the final estimate and, secondly, evidence giving her confidence in it.

### The number of interviews carried out

The total number of respondents going through the first stage interview was 1362. Of these, 963 went also through the intensive interview. The study was focused upon 28 different publications as follows.

<i>Daily Mirror</i>	<i>Woman</i>
<i>Daily Express</i>	<i>Woman's Own</i>
<i>Daily Herald</i>	<i>Radio Times</i>
<i>Daily Sketch</i>	<i>TV Times</i>
<i>Daily Telegraph</i>	<i>Reveille</i>
<i>The Times</i>	<i>Reader's Digest</i>
<i>The Evening News</i>	<i>Practical Householder</i>
<i>News of the World</i>	<i>Do-it-Yourself</i>
<i>Sunday Pictorial</i>	<i>Good Housekeeping</i>
<i>The People</i>	<i>Vogue</i>
<i>Sunday Express</i>	<i>Woman and Home</i>
<i>Sunday Despatch</i>	<i>True Romances</i>
<i>The Observer</i>	<i>True Story</i>
<i>The Sunday Times</i>	<i>Ideal Home</i>

For each of these publications, there were at least 60 double interviews (average = 130). And for each of them the validation procedure was focused on both positive and negative first interview claims.

### FINDINGS

In telling you about the findings from this enquiry, it is necessary to keep it in mind that though some of the findings have application broadly to readership research, others of them relate to the particular research package being investigated. In particular, it would be folly to conclude that the validity findings of *this* 'recent reading'

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

technique apply automatically to any other 'recent reading' technique. In fact the validity findings in this case are much more a function of the particular way the recent reading approach was administered – as you will presently see. Change the mechanics of data collection and you may well change the findings.

### THE PRINCIPAL FINDINGS

#### Error level

The broad character of the findings is usefully indicated by the summary data in **Table 1**. But the content of **Table 1** must first be explained. The NRS type interview had been limited, for the sake of manageability, to just one order of presentation of publications, namely dailies, Sundays, weeklies, monthlies. In normal circumstances, the NRS survey involves systematic rotation of the four publication groups. Whereas a comparison of the intensively derived findings with the NRS findings based upon single order presentation (dailies, Sundays, weeklies, monthlies) is interesting and useful, the important comparison is with NRS type findings based upon the usual systematic rotation of publication orders. There is a way of achieving this. Using the readership data collected through the NRS over the previous year (based on approximately 16,000 cases), it was possible to develop an order-effect conversion factor for each of the 28 publications under study – and hence for each publication group. An example of group conversion factors is given in **Table 6**, but that set is based on all 91 publications, whereas the

present set was based upon the 28 test publications.

If we accept that the intensive interview findings approximate to 'truth', then it appears that the NRS procedure (as it was at the time of the study) was in general slightly underestimating the level of readership (as defined by the IPA) for Sunday and weekly publications, but that it was *underestimating to a marked degree for the monthlies*.

#### Findings for each publication

In **Table 2**, I have presented the same class of findings as in **Table 1** for each of the 28 publications studied.

Whereas the results for the individual dailies are fairly uniformly 'correct', the 'accuracy' levels for the other publication groups are variable and this is particularly the case for the monthlies where there are both 90% accuracy levels and accuracy levels below 40%.

#### The character of the error

The study produced considerable evidence about the character of the error making process and the principal features of this evidence are set out below. The features are presented because they carry clues for the modification of the procedure.

#### BALANCING ERRORS

Each of the findings in **Tables 1** and **2** is highly dependent upon the extent to which overclaims and underclaims (on the NRS system) may balance out. This can be seen from **Tables 3a–d**. There is a sense in which we may say that for the publications with large readership errors, the

**TABLE 1**  
Comparing findings based on the NRS method with findings using the intensive or criterion check

Publication group	First interview (NRS)		Intensive follow-up (check) (c) %	Comparing NRS and intensive follow-up (b)/(c) %	Number of claims checked n
	DSWM order only (a)* %	Full rotation (b)† %			
	Dailies (7)	23.2			
Sundays (7)	26.9	27.3	31.7	86	855
Weeklies (5)	37.6	37.1	42.4	88	929
Monthlies (9)	9.2	9.5	14.8	64	958

\* In calculating the figures in Col (a), weighting was used to normalise the ratio of NRS 'yes' claims to NRS 'no' claims. Details of process are given in **Table 4**. But this weighting does not distort the comparability of the first and the intensive interview findings. This weighting was automatically built into the figures in Col (b) above.

† Using correction for full rotation - see text.

## 2.12 Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

overclaims have been very poorly balanced by the underclaims. This point is important for any correction of the research method because it means that in reducing error we must give attention to both over- and underclaims, seeing that they don't become unbalanced as we go on reducing *overall* error making.

When the study was planned, steps were taken to

build up to a statistically viable level the number of NRS readership claims that would be tested in relation to each of the 28 test publications. This objective related to both negative and positive claims to readership – in that accuracy levels would have to be established for both NRS 'yes' claims and for NRS 'no' claims.

For most publications, this arrangement meant that

**TABLE 2**  
Comparing findings from the NRS and the intensive procedures for all 28 test publications

Publication	First interview (NRS)		Intensive follow-up (check) (c) %	Comparing NRS and intensive follow-up	
	DSWM order only (a)* %	Full rotation (b)† %		intensive follow-up (b)/(c) %	Number of claims checked n
	Daily Mirror	48		48	48
Daily Express	29	29	30	97	197
Daily Sketch	18	21	19	111	125
Daily Telegraph	14	13	13	100	80
Daily Herald	12	14	11	127	119
The Times	5	4	5	80	61
Evening News	37	34	40	85	157
News of the World	49	50	51	98	155
Sunday Pictorial	45	46	47	98	160
The People	36	37	41	90	146
Sunday Express	26	26	33	79	157
Sunday Despatch	13	14	19	74	95
Sunday Times	10	9	17	53	70
The Observer	9	10	14	71	72
Radio Times	55	55	59	93	197
TV Times	43	42	46	91	206
Woman	38	37	42	88	185
Woman's Own	34	36	41	83	173
Reveille	18	17	24	71	168
Reader's Digest	23	24	26	92	150
Do It Yourself	12	11	19	58	112
Practical Householder	10	9	13	69	120
Vogue	10	11	12	92	115
Good Housekeeping	9	10	8	125	110
Ideal Home	7	7	22	32	101
Woman and Home	6	8	21	38	88
True Story	4	4	5	80	79
True Romances	2	2	7	29	73

\* The figures in this table have been weighted to preserve normality of NRS 'yes' and NRS 'no' ratios. However this does not distort the comparability of the NRS and the intensive findings. See footnote on page 198 for details of the weighting system.

† Taking rotation effects into account. See text on page 195.

# 2.12 Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

**TABLE 3a-d**  
Showing the extent of balance of underclaims and overclaims

(a) Dailies	<i>Intensive check</i>				<i>All claims</i> <i>n</i>	(b) Sundays	<i>Intensive check</i>				<i>All claims</i> <i>n</i>
	<i>No</i> <i>%</i>	<i>Yes</i> <i>%</i>	<i>?</i> <i>%</i>	<i>All</i> <i>%</i>			<i>No</i> <i>%</i>	<i>Yes</i> <i>%</i>	<i>?</i> <i>%</i>	<i>All</i> <i>%</i>	
NRS No	70.9	5.3	0.6	76.8	517	NRS No	64.0	9.0	0.1	73.1	444
NRS Yes	4.7	18.4	0.1	23.2	382	NRS Yes	4.2	22.7	0.0	26.9	411
All	75.6	23.7	0.7	100.0	899	All	68.2	31.7	0.1	100.0	855

  

(c) Weeklies	<i>Intensive check</i>				<i>All claims</i> <i>n</i>	(d) Monthlies	<i>Intensive check</i>				<i>All claims</i> <i>n</i>
	<i>No</i> <i>%</i>	<i>Yes</i> <i>%</i>	<i>?</i> <i>%</i>	<i>All</i> <i>%</i>			<i>No</i> <i>%</i>	<i>Yes</i> <i>%</i>	<i>?</i> <i>%</i>	<i>All</i> <i>%</i>	
NRS No	51.2	10.2	1.0	62.4	520	NRS No	79.2	9.6	2.0	90.8	644
NRS Yes	5.0	32.2	0.4	37.6	409	NRS Yes	3.6	5.2	0.4	9.2	304
All	56.2	42.4	1.4	100.0	929	All	82.8	14.8	2.4	100.0	948

**TABLE 4a**  
Before weighting

	<i>Intensive check</i>			
	<i>No</i> <i>Number</i>	<i>Yes</i> <i>Number</i>	<i>?</i> <i>Number</i>	<i>All</i> <i>Number</i>
NRS No	58	12	0	70
NRS Yes	15	75	1	91
All	73	87	1	161

**TABLE 4b**  
After weighting

	<i>Intensive check</i>			
	<i>No</i> <i>Number</i>	<i>Yes</i> <i>Number</i>	<i>?</i> <i>Number</i>	<i>All</i> <i>Number</i>
NRS No	82	17	0	99
NRS Yes	15	75	1	91
All	97	92	1	190

**TABLE 5**  
Frequency of errors related to claimed frequency of reading

<b>Claimed frequency reading</b>	<i>Publication group</i>				
	<i>Dailies</i> <i>%</i>	<i>Sundays</i> <i>%</i>	<i>Weeklies</i> <i>%</i>	<i>Monthlies</i> <i>%</i>	<i>All</i> <i>%</i>
Every copy	11	17	9	27	16
Most copies	39	34	37	43	38
Now and then	12	29	23	31	24
Hardly ever	7	10	8	14	10
Never	2	1	1	1	1

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

the ratio of NRS 'yes's' to NRS 'no's' was higher than the ratio ordinarily found in NRS surveys. Accordingly the 'no' group was weighted up to the level ordinarily found in NRS surveys. An example of this weighting system follows. Suppose the validity check for the *Daily Mirror* gave the results shown in **Table 4a**.

Let the NRS 'yes' group, numbering 91 cases, remain unchanged in size. But weight the NRS 'no' group up to the usual NRS level of that period. Since the readership figure for the *Daily Mail* at that time was 48%, the NRS 'no' group would have a total  $52/48 \times 91 = 98.6$  (say 99) and the top line in **Table 4b** would have to be shown. This conversion or weighting process was applied for each publication group and the summary tables (**Tables 3a-d**) are simply the arithmetical means of the weighted figures.

### THE 3-MONTH SCREENING QUESTION

(1) The three-month screening question excluded many reading events that the intensive interviewer identified as qualifying for readership: 7% for the dailies, 9% for the Sunday papers, 13% for the weeklies, 9% for the monthlies. The screening question is obviously a liability that must be attended to in any correction of the procedure. (2) On the other hand, reading events that occurred more than three months ago could get past both the three-month screening question and question 2: this occurred to a minor degree for the dailies, Sundays and weeklies (2%, 4% and 1% of the NRS 'yes' claims) but to a major degree for the monthlies (22% of the NRS 'yes' claims).

### ACCURACY IN RELATION TO FREQUENCY OF READING

Accuracy level of the NRS claims was substantially associated with the claimed frequency of reading of the publications concerned (**Table 5**).

It is very clear from this table that a useful screening device might turn out to be a reply of 'never' to a question about frequency of reading.

### CORRELATES OF ERROR

Error making was analysed by age, sex, social class, occupational level, age of leaving full-time school, further education and approximately 15 more variables. Though there was variation from one sub-group to another, the evidence did not support a conclusion of any strong and general relationship between error making and characteristics of respondents. The exceptions were 'number of adults in the household' (positive association with error making), 'number of household members in a full-time job' (positive association), 'extent of commercial viewing' (positive), 'publication delivered to home' (positive), 'publication kept round the house for a while'

**TABLE 6**  
Showing the effect on readership level of rotating the order of presentation of the different publication groups (all 91 publications)

	1st order	Second	Third	Fourth
Dailies	1.00	1.07	1.02	1.03
Sundays	1.00	1.05	1.06	1.07
Weeklies	1.00	0.99	0.93	0.88
Monthlies	1.00	0.83	0.77	0.74

(positive), 'respondent cares about seeing the publication' (positive), 'the reading of publications over an extended period' (positive).

### ORDER EFFECTS

A study of the 1958 NRS results makes it clear that there is a considerable variation in results as the order of presentation of the four groups of publications is changed.\* In **Table 6**, the first order of presentation of any one of the four publication groups has been converted to 1.00 so that the effect of position change can be more clearly seen.

For economy reasons, the present validity check on the NRS method had been based on only 28 of the 91 publications of the NRS and upon only one order of presentation of publication groups, namely dailies, Sundays, weeklies, monthlies. However, with a knowledge of the rotation effects for those 28 publications\*\*, it is possible to estimate the readership figure that would have emerged for any other order of presentation of their publication groups. Such estimates indicate that the best order (for securing accuracy) of the four publication groups would have been monthlies, weeklies, Sundays, dailies\*\*\* (**Table 7**). This order would not have produced an accurate result, but a result less inaccurate than any other order. Accordingly that is the order that should be used in any modified NRS procedure.

\* See page 23 of Studies in Readership based on approximately 16,000 cases.

\*\* Not to be confused with the data in **Table 6**, which relates to all 91 publications.

\*\*\* These findings relate to the presentation order: dailies, Sundays, weeklies, monthlies. But it may be assumed (from the data already presented that broadly similar findings would emerge for the NRS with full rotation.

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

**TABLE 7**  
**Estimation of error levels when publications are presented in the order**  
**monthlies, weeklies, Sundays, dailies**

Monthlies first		Weeklies second		Sundays third		Dailies fourth	
Modified NRS %	Intensive check %	Modified NRS %	Intensive check %	Modified NRS %	Intensive check %	Modified NRS %	Intensive check %
10.9	14.8	38.3	42.4	26.9	31.7	23.7	23.7

### INTERVIEWER PERFORMANCE

A third of the NRS interviews were tape recorded and also were observed. An analysis of the evidence from these sources indicated serious deficiencies in the interviewing situation.

(a) interviewers frequently deviated from the required wording of questions and from related instructions. In this respect it should be noted that questions 1 and 2 were required to be as follows:

**Q1** "I want you to go through this booklet with me, and tell me, for each paper, whether you happened to have looked at any copy of it in the past three months, it doesn't matter where."

NOW EXPLAIN: 'looked at', 'any copy'; 'past three months' (ie, 'since ...'); 'doesn't matter where'.

**Q2** "When was the last time you looked at a copy of ...?"

DISREGARD ANY READING ON THE DAY OF INTERVIEW: IF THE ANSWER IS 'Today', ASK: "When did you last look at a copy apart from today?"

The deviations of the interviewers took the following principal forms: omission of the terms 'any copy', 'it doesn't matter where', 'past three months'; substitution of the term 'looked at' by words like 'see', 'look or glance at', 'read', 'take', 'regularly see'; failure to explain the terms 'looked at', 'any copy', 'it doesn't matter where', 'past three months'; the offering of explanations of the above terms in jumbled form and prior to the asking of question 1; the interviewer asks question 2 ("When was the last time you looked at a copy of ...?") just once and thereafter reads through the publications (that had qualified through Q1) without reminding the respondent of what was wanted through question 2 (ie, 'when she last looked at a copy ...').

Several illustrations of deviant deliveries are given in the footnote to this page\*.

(b) following question 1, interviewers often took to reading through the list of 91 publications as the respondent turned through the 91 mastheads. This raised the possibility of the interviewer calling out the name of one publication while the respondent was looking at the

masthead for another. This was observed to happen.

(c) respondents were also observed to be turning more than one masthead at a time, especially towards the end of the booklet when dealing with the monthly publications.

(d) in dealing with question 2, the interviewer was to ask "When was the last time you looked at a copy of ...?" When the reply was in any way ambiguous (eg "We always take *The Times*"), the interviewer was required to probe for a clear and unambiguous reply. In such cases there was almost always some degree of breakdown of the probing procedure.

(e) occasionally the interviewer seemed to be in a great hurry, pushing ahead irrespective of apparent respondent difficulty (observer and respondent claims). On other occasions, the interviewer caught the respondent in the midst of some preoccupying activity or preparation, the most error-producing of which appears to have been 'about to go out'. See **Table 8**.

\* "I want you to go through this booklet with me and tell me the ones you think you've looked at in the last three months. Can you call them out - the ones you have looked at in the last three months"; "I want you to go through this and tell me which ones you have seen in the last three months. By 'seeing them' it means anything. You don't have to have read them or bought them. You might have seen them at the doctors or dentists or barbers"; "We're ... making a survey of the daily papers, Sunday papers, weekly and monthly publications looked at by the public. Would you tell me with the aid of my little booklet. Just flip through them and tell me which it is you see - that you've looked at. Would you start at the beginning dear. Flip the first one over and say 'News Chronicle, yes or no', and say the same to each as you go along." The interviewer suddenly remembered when she got to the Herald and said "Within the last three months". She did not go back.



# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

### INDICATIONS FROM THE CONFRONTATION SEQUENCE

The confrontation sequence at the end of the interview was particularly useful in developing clues about causes of error. The more frequently given explanations (of errors) are given in summary form in **Table 9**.

### Summing up on the character of the error

A study of the character of the errors made and what seems to lie behind them provides many clues for any wide-awake researcher interested in modifying the NRS procedure towards greater validity. These were regarded as vital in taking the enquiry into its second phase, namely the development of a modified procedure of greater validity.

**TABLE 8**  
Respondent activity at time of the interviewer's call

Activity	Number of estimates tested <i>n</i>	Percentage of these in error <i>%</i>
Cooking/preparing meal/eating meal	734	15
Looking after children	96	11
Bathing/washing/shaving/dressing	122	12
About to go out	157	25
Housework	647	16
Decorating/carpentry	99	21
Working at full-time job/at work	151	18
Gardening/cleaning car/sewing	176	13
Just came in	209	17
Relaxing	884	17
Watching TV/listening to radio	239	13
Social activities	101	10

**TABLE 9**  
Reasons offered by respondents for error in the NRS interview

Reasons given for error	Number of cases
Poor memory/mind wandered/too long ago/confused over dates	275
Occasion was borderline (re definition or date)	35
Respondent sees too many to be accurate/confused papers asked about	96
Paper not cared for/not admitted	19
Wrong interpretation of 'looked at'/thought interviewer meant 'looked at regularly, bought or ordered'/claims interviewer said 'regularly seen', 'bought or ordered'	213
Misunderstood question/misheard	33
Said 'no' because it was someone else's copy/said 'yes' because it was in the house (but had not looked at it)	26
Answered in terms of the 'usual'/interviewer asked for ones 'usually' read/isolated event overlooked	323
No real effort by respondent/interviewer did not make it clear that effort was needed/easier to say 'yes' than 'no'/guessed	230
Respondent in a hurry at first interview/interviewer insisted in spite of time being inconvenient/respondent was distracted at first interview	147
Interviewer allowed two respondents to answer together/lack of control over respondent	15
Book of mastheads not used at all/book not properly used	58
Interviewer did not deal with all the publications/the 'when' question was not asked	74
Interviewer did not date start of the three-month period	21
Interviewer seemed in a hurry	26
Interviewer misheard/misunderstood reply	75
Interviewer decided when respondent doubtful	16

### PART II: MODIFYING THE READERSHIP RESEARCH TECHNIQUE

#### ORIGINS AND METHODS USED

Once the results of the validity study were placed before the Technical sub-Committee of the NRS Controlling Committee, the Survey Research Centre was asked to develop a procedure capable of producing accurate readership figures. I am presenting here a short statement about the development procedure and about the method that was so developed.

#### The general form of the development procedure used

The procedure used for developing an alternative method for measuring readership was *Progressive Modification*. With this procedure, the following steps are typically involved.

- (a) *Cycle 1* A study is made of the problems that stand in the way of collecting accurate information and of the circumstances that facilitate it. In the present case, the validity study of Part I provided such information.
- (b) *Cycle 2* On the basis of such information, a first

approximation to the new procedure is made or constructed. This first and tentative form of the procedure is then subjected to as rigorous a form of testing as seems feasible. In the present class of work, such testing would yield information about the incidence both of overclaiming and of underclaiming and the apparent causes of such erroneous estimates.

(c) *Cycle 3* On the basis of *this* evidence, the readership research procedure would be further modified and then tested once more.

(d) *Further cycles* Further cycles of modification and test would follow until the test result indicated that an acceptable level of accuracy had been reached.

(e) *A reliability/validity check* At this point a reliability/validity check based on the full intensive procedure was used to challenge further the final form of the new procedure.

#### Some features of the progressive modification method as used in the present enquiry

(a) in the present study there were five cycles of modification test and, following this, the standard reliability/validity check.

(b) the tests of 'accuracy' were, for certain of the cycles, extended to all 91 publications of the ordinary NRS. This large coverage necessitated some reduction in the intensity of the tests made, but nonetheless they were intensive and full details about them are available in Part II of the full report of the enquiry.

**TABLE 10**  
Showing the sequence of cycles in the Progressive Modification procedure

Cycle number	Publications dealt with in		
	New procedure <i>n</i>	Intensive interview <i>n</i>	Number of cases <i>n</i>
1 modification of original NRS procedure, followed by intensive tests for all 91 publications	91	91	174
2 second modification with the changes and tests focused on the 30 monthlies	91	30	29
3 third modification with the changes and tests focused on the 30 monthlies	91	30	30
4 fourth modification with the change and tests focused on the 30 monthlies	91	30	53
5 fifth modification with the changes and tests focused on all 91 publications	91	91	172
6 fifth modification in first interview, with extended intensive checks on the 28 publications featured in Part I	91	28	153

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

(c) in Cycle 1, all four publication groups were subject to test. At this point it became clear that the monthlies were still a major source of error and so the next three cycles of the Progressive Modification process were devoted solely to securing accurate information about readership of the *monthlies*. This objective was achieved by the end of the fourth cycle and for Cycle 5, the Progressive Modification procedure was applied to all four publication groups. By the end of that cycle, the results yielded by the new procedure were broadly accurate for three of the publication groups, publication by publication. For the dailies there was now a moderate overestimate. The full intensive check was subsequently applied.

(d) the number of cases tested in each cycle of the construction procedure is given in **Table 10**.

(e) the whole of the Progressive Modification procedure was carried through in approximately seven weeks, with cycle analyses and modifications going on through each weekend.

(f) for each testing of a modification, a new team of first stage interviewers was brought into the operation. A team would be briefed on Monday mornings and would work throughout that week.

(g) from the very first modification, the order of presentation of the publications in the first interview was: monthlies, weeklies, Sundays, dailies – this being the best order from the point of view of accuracy (see Part I for grounds for this conclusion).

### Some features of the intensive checks used in the Progressive Modification operation

In Cycle 6, the intensive checking process used was the one employed in Part I of this enquiry. In Cycles 1 and 5, the intensive check was of necessity abbreviated to allow us to test claims made with respect to all 91 publications. Nonetheless the check made remained a tough challenging one and some of its features were as follows.

(a) intensive interviewers used a specialised book of mastheads into which certain instructions and investigatory procedures had been built (on interlacing cards and on the backs of the masthead pages). These steps included special emphasis on the name of each publication being checked, repeat definition of 'looked at', repeat specification that 'any copy'/'any situation' count, instructions for administering challenges and probes, instructions for confrontation over errors.

(b) for each publication, the interviewers went early to the question "When was the last time you looked at any copy of . . . . .?" On getting a reply, the intensive interviewer asked "What about since then?" and took the respondent through a range of possible 'looking at' situations to draw out forgotten occasions of more recent occurrence.

(c) on getting the respondent to think about the last

occasion, the intensive interviewer made a major effort to set the respondent thinking about the detail of that occasion – of what was going on then. She challenged that detail and that recall of an event. She then tried to get the respondent to date that set of details of the occasion. If the last occasion was far outside of the qualifying period, attempts to date it fairly precisely were not made. A special effort was made to stop respondents answering in terms of what they actually did rather than in terms of 'the usual'.

(d) the intensive interviewer was to make no reference to specific qualifying periods, concentrating solely upon locating and then dating the last occasion.

See the report for the full procedure.

### FINDINGS

The principal findings for Part II of the enquiry are set out in **Tables 11** and **12**.

#### The sequence of results from the different forms of the developing new technique

In **Table 11** are presented 'accuracy indices' for (a) the original NRS procedure (figures based upon full rotation of publication groups) and (b) each of the different forms of the new procedure as these emerged through Cycles 1–5. In each case, the index presented is the ratio of the figure for the readership research procedure to the figure for the intensive (check) interview.

(a) the fifth modification produced a major improvement in the 'accuracy' levels for monthlies and an increase for both weekly and Sunday publications. But there was some reduction in accuracy level for the dailies. Presumably the techniques that lifted the other groups towards accuracy worked also to move the dailies towards overclaim. At the same time, enough is known about the workings of the fifth modification to allow effective further modification of that procedure as it applies to dailies.

(b) the first modification of the old NRS procedure turned a major understatement (0.64) for the monthlies into a major overstatement (1.36). This switch arose almost totally out of the reduction of underclaims whilst the overclaims remained much the same. However, as attention was switched to reducing the overclaims, the ratio fell towards unity (1.03 by the fourth modification).

(c) the same phenomenon also occurred in relation to the full set of 30 monthlies carried in the old NRS survey.

#### Comparing 'accuracy' levels for specific publications in going from the old NRS method to the new procedure

**Table 12** compares 'accuracy' levels stemming from the old NRS method and the new procedure.

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

The main indications of **Table 12** are as follows.

(a) the new method produces a much more even (and a more accurate) set of findings for monthlies than did the old NRS method. The findings for the weeklies and the Sunday publications are also more even and more accurate.

(b) for the dailies, on the other hand, the fifth modification is in general somewhat less accurate. Improvements could be made through focusing the Progressive Modification method on that publication group.

### THE NATURE OF THE FINAL FORM OF THE PROCEDURE: ITS PRINCIPAL FEATURES

The final form of the procedure is fully described and illustrated in Part II of the full report. Some of its salient features are set out below.

(a) the three-month screening question (with its demonstrated tendency to exclude genuine readership) was eliminated and the respondent was taken straight to the question "Did you look at any copy of . . . . . in the last four weeks/seven days/yesterday?" Thus the process became one-stage instead of two-stage.

(b) a modified book of mastheads was used.

(i) it was attached to the interviewer's board and so could not be passed to the respondent for page turning. It remained in the control of the interviewer.

(ii) the order of presentation of the mastheads was fixed in the sequence monthlies, weeklies, Sundays, dailies - because of the demonstrated superiority of that order.

(iii) on the backs of the different mastheads were specific instructions. Some were meant to help the respondent differentiate between similar sounding publications and some were spaced reminders of the basic rules and arrangements for reacting to the mastheads (eg, repeat definitions of 'looked at' and a test to see if the respondent understood that key term, reminders of the qualifying period and of when it started, reminders that 'any copy' (old or new) could count and that the 'looking at' event could be anywhere). This was done because these instructions were frequently omitted by interviewers and tended not to be repeated by the interviewer in the course of the NRS procedure.

(iv) in the fifth modification, the qualifying period for the group of publications being presented at any time was displayed on an extended card that was

**TABLE 11**  
Showing 'accuracy' levels achieved in going from the original NRS method through the different modifications of the new readership research technique

*Comparing results from 1st and intensive checks*

Readership procedure being tested	Based on 28 test publications				Based on all 91 publications			
	Monthlies 9	Weeklies 5	Sundays 7	Dailies 7	Monthlies 9	Weeklies 5	Sundays 7	Dailies 7
Standard NRS (full rotation)	0.64	0.88	0.86	0.99	—	—	—	—
First modification (MWSD)	1.36	1.19	1.12	1.19	1.52	1.30	1.14	1.19
Second modification (MWSD)	1.24	—	—	—	1.28	—	—	—
Third modification (MWSD)	0.91	—	—	—	1.00	—	—	—
Fourth modification (MWSD)	1.03	—	—	—	0.98	—	—	—
Fifth modification (MWSD)	1.01	0.98	1.00	1.12	0.95	0.96	0.99	1.07
Fifth modification reliability/validity check	0.95	0.92	0.98	1.08	—	—	—	—

# 2.12

## Measuring and then increasing the accuracy of Britain's National Readership Survey: a validation project

part of the book of mastheads (eg, any copy in the LAST FOUR WEEKS) – to reduce forgetting of the key readership periods.

(v) for each claim of 'yes' to the readership question, the interviewer was required to deliver specific challenges and probes, working to a set of printed instructions that were clearly and constantly on display (on an extended page in the masthead pack) throughout the interview. This display would have been hard to ignore, with its three challenges to any 'yes' claim.

### LOOKING TO THE FUTURE

#### The fifth form of the procedure

The fifth modification requires that a little more work be done on it to make more accurate the figures for the dailies. This should be a fairly simple matter, involving a further modification stage geared to the ascertainable causes of error for the publications concerned.

#### Other readership methods and objectives

Whatever the purpose of a readership survey, certain steps in the construction of its procedures are vital.

(a) in the first place, the variable used to measure or indicate readership (eg 'looked at', 'paged through') must be established, empirically, as the most powerful of the available correlates of whatever the readership survey is intended to measure (eg extent of exposure to the advertising content of publications). Possibly a small *composite* of such correlates is necessary, these to be regarded as the indicators of whatever is to be measured.

(b) once this is done, the Progressive Modification method should be applied for the development of a valid research procedure or package.

(c) where a method already exists, its validity should be assessed. This calls for much more than a comparative study (eg the comparison of two competing methods). The intensive technique (or an improvement of it) should be used. If the tested method is found to be wanting, that test can be made to reveal enough about the problems of collecting the required information for the formulation of the first modification in a Progressive Modification sequence.

The methods described in this paper, and in the full report on which they are based, cannot be expected to

provide actual certainty or proof. No methodology is likely to give us that. But if soundly used, they can take us very usefully in that direction. Certainly they can give us something much better than mere comparisons of one operating readership research procedure with another.

**TABLE 12**  
Comparing 'accuracy' levels for the old and new methods, publication by publication

Publication	'Accuracy' ratio*	
	NRS	New
Reader's Digest	0.92	1.00
Do It Yourself	0.58	1.00
Practical Householder	0.69	1.00
Vogue	0.92	1.13
Good Housekeeping	1.25	1.11
Ideal Home	0.32	1.00
Woman & Home	0.38	0.88
True Story	0.80	1.30
True Romances	0.29	1.00
Radio Times	0.93	0.96
TV Times	0.91	0.93
Woman	0.88	0.94
Woman's Own	0.83	1.08
Reveille	0.71	1.05
News of the World	0.98	1.00
Sunday Pictorial	0.98	1.00
The People	0.90	1.02
Sunday Express	0.79	1.04
Sunday Despatch	0.74	0.86
Sunday Times	0.53	0.90
The Observer	0.71	1.25
Daily Mirror	1.00	1.16
Daily Express	0.97	0.96
Daily Sketch	1.11	1.09
Daily Telegraph	1.00	1.20
Daily Herald	1.27	1.20
The Times	0.80	0.83
Evening News	0.85	1.18

\* First survey figure/intensive survey figure. Ratios all based on rounded percentages. (See pages 192–4 of full report for further details.)