Balnaves, Mark, Tom O'Regan, and Ben Goldsmith. "The Convention." *Rating the Audience: The Business of Media*. London: Bloomsbury Academic, 2011. 21–48. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781849664622.ch-002>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 12:10 UTC.

Copyright © Mark Balnaves, Tom O'Regan and Ben Goldsmith 2011. You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

2 The Convention

In Chapter 1 we saw how a new FM station, when FM was itself new as a technology, needed to demonstrate the existence of its audience. If W71NY had already been recognized in the radio ratings of the time, then the work of demonstration of an audience would have already been done. However, FM was new and the popular stereotype was that it appealed to a very narrow, classical music-listening demographic. The customized, commissioned study by WOR showed that its audience was both more diverse and more valuable than this. The need to measure audiences and to agree on what counts as an audience has continued unabated since the early days of broadcasting media. In this chapter the authors provide a detailed introduction to the rise of the audience-ratings convention and the rules, established very early on, that have grown to define expectations among the different players in each market.

‘The Crossleys’ – Archibald Crossley

My father would bring them home, interview cards, and he trained me how to do tabulate data from them – 1, 2, 3, 4 across, 1, 2, 3, 4 across, to make bunches of fives that could be added up by hand. I got into that by the time I was 10. I remember that there were four radio networks across the top of the sheet, and you put your check mark under whichever network the listener was reporting, so when you added them all up and counted them in piles of five, you knew how many listeners you had out of 20 calls. Helen Crossley (Crossley 2008)

Helen Crossley is here recalling working for her father as a child, in the first weeks after Archibald Crossley set up his system for measuring the radio audiences of national radio networks. He had decided to measure ‘exposure’ in his radio ratings analysis – who listens, for how long and with what regularity. His calculus at first glance might look simple, but the environment in which it was agreed on was not. The alternative possible measure was ‘engagement’ – how involved people are in their radio programmes. However, there was no agreement on how engagement might be used let alone charted as a universal measure. For example, one person’s ‘like it very much’ might be another’s ‘like it’. Exposure, on the other hand, delivered a more comparable objective measure. A person listening to the radio for ten

minutes was directly comparable to another person listening to the radio for five minutes. This did not mean that audience researchers did not collect data on whether people did or did not like the radio programmes. But for the purposes of buying and selling radio airtime, or programmes, a metric that showed the fact of tuning in to a programme and the amount of time listening to a programme had a simplicity that was essential for bargaining in highly competitive environments. All competitors, though, had to agree to the measure being used.

The first ratings convention was not a broadcast station initiative but an advertiser initiative. Archibald Crossley was hired in 1929 by an organization of radio advertisers, the Cooperative Analysis of Broadcasting, Inc. (CAB), to measure the 'unseen audience'. The radio advertisers had an interest in a single rule and an authoritative audience measure covering those radio stations, particularly the national radio networks, that were relevant to the larger mostly national advertisers. They wanted a measure to enable them to pick and choose among broadcast outlets and, since this was an era where advertisers were also programme producers, an instrument to gauge whether they were getting their target audiences.

Crossley was a methodologist in his own right, working with George Gallup on public opinion polling, and experienced in the development and application of statistical sampling methods which he adopted to the audience ratings. With no prior methodology to work from, he created basic measurements for CAB that covered over 80 US cities for 16 years. His employees thumbed through telephone directories, called subscribers, and asked them what programme they had been listening to – a technique called telephone recall. Crossley's ratings estimated the number of telephone subscribers tuned in to any show. This approach provided timely survey results ensuring that they could be used iteratively by advertisers and their programme producers. But this relative immediacy came at a cost. As telephone ownership was limited to those who could afford it these ratings were skewed by its sample with the result that particular audiences, most especially regional, minority and African American audiences, were under-represented. (This skewing gave rise to criticism, which persists to this day, that the ratings discriminates and disenfranchises minority audiences.) The CAB ratings became known as the 'Crossleys' and were the currency for determining the popularity of programmes. 'Even top stars like Jack Benny and Edgar Bergen worried more about their "Crossleys" than their hairlines,' *Time* magazine reported in 1946 (*Time* 1946).

Archibald Crossley is the acknowledged founder of broadcast ratings (Beville 1988). His exposure measure quickly became the *standard* for measuring ratings. Crossley soon found that the demand for his radio ratings was extraordinary. Initially only made available to the advertisers participating in CAB, a bootleg market for

radio ratings emerged among radio networks, stations and advertising agencies. Warren Dygert, in his 1939 advice to ‘radio men’ in *Radio as an Advertising Medium*, shows why:

The old-time advertising man ... knows if he buys advertising space in *Redbook*, *True Story*, *Business Week*, *Country Life*, he buys a ready-made type of reader. The publisher can tell him the reader’s buying habits, income bracket, even the group age and sex. This solves a big problem in periodical advertising. In radio, with the exception of foreign-language stations, the advertiser has no such help. Rich and poor, employed and unemployed, educated and uneducated, moron and Solon [the ancient Greek lawyer] listen to the radio and are no respecters of stations. The radio advertiser has the added task here of picking out the group he wants and building his program to attract and hold that group. If he wants both mass and class, he may do as the Ford Motor Company did in building a special custom-made program for each group. Or he may attempt to build one program for all, a difficult feat for, as already pointed out, no one program can possibly suit everyone. (Dygert 1939: 11)

In a broadcast environment where audiences could not be easily classified in the same way that subscribers to print could be, radio ratings provided the stability and certainty by providing authoritative measures of audiences for time segments and programmes. What Archibald Crossley set up was the core of the modern audience-ratings convention:

- 1 Exposure is the key measurement;
- 2 The inherent correctness of that measurement must appeal to all parties;
- 3 The ratings deliver a ‘single number’;
- 4 A probability, statistical, sample is used for data collection;
- 5 Distortion of the ratings by the research provider or by subscribers is unacceptable and requires monitoring and control.

Both formal and informal arrangements emerged to deal with how this agreed measure was collected, used and distributed. Conflict over accuracy and outright cheating had to be dealt with. However, one of the biggest problems was *hypoing*, where a systematic attempt is made to distort ratings figures. This might involve simple means like running competitions during ratings sweeps. C.E. Hooper, whose ratings later replaced those of Crossley and CAB, banned radio stations or others that attempted to distort ratings from participating in the survey. This threat

to exclude organizations that were blatantly hypoing was remarkably successful, although, as we shall see later, more complex means of distorting the ratings numbers in favour of providers emerged.

Hooper replaced Crossley as the major ratings provider in the 1940s. CAB was seen as less reliable in its methodology and approach and, because of its close association with advertisers, less independent. Hooper's innovation was the 'telephone coincidental' method. It involved telephoning the household at the same time that the programme was running. By the time Crossley decided to adopt Hooper's technique it was too late, and subscribers had moved to Hooper's service. This did not change the industry agreement that exposure should be the core measure for audience ratings. After all, only the means of collection and the group that collected the ratings had changed. While methodological questions would continue to be raised, the ratings were now mostly accepted as a convention – an industry-wide agreement on how audiences would be measured and valued. On a day-to-day basis, radio ratings were used in a variety of contexts. They were combined with other information to help organizations make decisions on everything from programming to sponsorship and talent.

If Crossley instigated the first revolution – audience ratings themselves and exposure as the key measure – then it was Art Nielsen who embedded technology into audience ratings. Crossley, like Hooper, used self-reporting by audiences as the key technique for collecting the data on ratings. Nielsen thought that there was another step to take which would remove self-reporting altogether. He developed mechanical devices to register whether people indeed had the radio set turned on and tuned in to a particular station.

Arthur C. Nielsen (and the 'Black Box')

Arthur C. Nielsen set the scene for Nielsen's dominance of national audience ratings in the United States when he tied technology to the measurement of exposure. The legacy in the United States of this move has been to create, to use the economists' phrase, a path dependency in American audience measurement. Technology became with Nielsen's introduction of the 'black box', the Audimeter, an intimate part of the data collection process and of the convention itself, in the 1940s. In 1936, Nielsen acquired the Audimeter from two inventors, Bob Elder and Louis Woodruff, who had developed it to pick up and record frequencies tuned on a radio (Beville 1988: 20). The compiled information was stored on a wax drum then transferred to film and ultimately solid state memory. Nielsen then spent the next few years developing the technology and proof of concept for its use as an alternative ratings technology to measure radio and later television set on/off status

and channel/station tuning. Nielsen began providing audience estimates to the radio industry in December 1942; by 1946 he had coverage of 63 per cent of all US homes, and by 1949 the company had 97 per cent (Beville 1988: 21–2).

Prior to his involvement in broadcast ratings Arthur Nielsen ran his own market research company specializing in testing the viability of new products and determining the market shares of products in retail stores. Nielsen is credited with formulating, popularizing and institutionalizing the concept of market share through his Drug Index and Food Index, established in 1933 and 1934 respectively. A.C. Nielsen was a well-established market research firm before it moved into broadcast ratings. The company's move into ratings was a natural extension of Arthur Nielsen's longstanding concern to measure market share in its various manifestations and to measure, as Gale Metzger recalls, 'the whole marketing process, step-by-step'. Metzger reckons that Nielsen was one of the first who 'wanted to report on all elements of the marketing process in a single measurement'. He also points out that Nielsen came at the ratings from a different position. Unlike the other notable audience researchers Nielsen was not an attitude researcher:

In the early days, there were attitude researchers who became involved with media audience measurement. George Gallop is one name that comes to mind. The social research approach involved asking people what they did; what stations or programs they listened to. That was in contrast to an engineering perspective where a meter was used to record what was happening in the home. Art believe the metered measurement was more precise and a superior technique. He sold that point of view effectively.

For Metzger the turning point for Nielsen's metered approach came with the Advertising Research Foundation's review of audience measurement, which concluded that the meter measurement of set tuning was a superior way to estimate audience size:

They affirmed that audiences should be defined by set tuning. That conclusion created great turmoil. There were very angry exchanges. Some in the industry objected strongly to the ARF taking that position. They felt it was highly prejudicial and wrong. Their feelings were expressed in angry terms. The appendix to the ARF report includes a strongly expressed minority opinion.

For his part Beville reckons Nielsen won 'handily' because of a superior product offering greater coverage: 'he had a continuous meter service that operated around the clock, 48 weeks a year' (Beville 1988: 22). Whatever the sources of 'industry endorsement', Nielsen was able to expand its media services in size and scope after 1950. Its national sample size was increased and with the development of TV as a

mass medium it extended its coverage to television eventually moving from radio to television as that medium came to dominate broadcast advertising revenues. Nielsen was able to use its radio sample to leverage its television coverage, as Gale Metzger observes:

When television came in the early 50s, Nielsen had a national panel of radio homes. At the time, they were measuring in-home use of radio. Television was measured by metering television sets that were purchased by the radio sample homes. As 10 per cent of the homes got a television set, they measured those 10 per cent. The sample size grew as television ownership grew.

But Nielsen was not able to organize the convention in quite the way he wanted to: he, like Metzger, felt that the major users of the ratings – the media companies, the advertising agencies and the advertisers – should equally contribute to its support. But this is not what emerged. It is worth quoting Metzger in full here:

The history has been that the rating services' revenues have been at least 80 per cent from the media. Media often pay 90 per cent or more. For network radio, they paid about 95 per cent of the bill. Others paid what amount to token fees. The agency/advertiser perspective is that they pay for audience measurement when they buy the media. But their voice at the table is effectively a secondary voice rather than a primary voice. I've always felt and Art Nielsen advocated originally, that the bill should be split up a third, a third, a third – a third by the media, a third by the agencies and a third by the advertisers. That was never achieved. The Nielsen national service had relatively more revenue from the advertisers and agencies than other services. For print, publishers paid virtually the entire cost for audience measurement.

For Metzger this financing structure privileged the media companies' voice at the expense of the advertiser and advertising agency/market research voice. He sees this as the conceptual flaw in this financing structure:

It puzzled me because of the interest in the return on investment for advertising. If one is spending hundreds of millions of dollars on advertising, it seems obvious to me that better, more accurate measurement would pay for itself. If advertisers could get an extra 5 per cent or 10 per cent efficiency out of advertising budgets, it would be an extraordinary return.

I believe many advertiser research personnel see that potential also, but they were never able to sell that to the management. Art Nielsen worked very hard to achieve closer advertiser involvement. Whenever he had something new or different, he would always get on a train and go

to Cincinnati. That was the home of Proctor & Gamble. They were the industry bellwether – in his mind. If you could sell Proctor & Gamble you could sell anybody. They would set the pace. Proctor was more active in the earlier days and, at the same time, very guarded because they didn't want to 'throw their power around' and be exposed to accusations of anti-trust activities. So anyway the funding issue was a significant focus. We had advertisers who wanted to put up money, but instead of millions of dollars they were putting up tens or hundreds of thousands of dollars. While that is different dimension, it was moving in the direction of getting them more involved. (Metzger 2008)

Nielsen's Audimeter did not just suddenly appear without debate among those who had an interest in its use. Audience ratings are not just numbers and data collection techniques but a complex convention. C.E. Hooper sold his methodology in academic papers and in the commercial market. His telephone coincidental technique was, he argued, a better way to capture exposure than Crossley's telephone recall method. Nielsen did not change exposure as the core metric – he changed the technology for counting it. Nielsen funded his own innovation out of his own pocket and had a very different relationship to advertisers and other clients compared with Archibald Crossley. Crossley was employed by CAB but did not own CAB. Nielsen, however, had to convince parties to the convention that this new way of measuring exposure was superior to previous methods. This did not mean that Nielsen did not use diaries or other techniques to capture audience engagement with programmes or their demographics. It did mean that he set up the machine as an independent error-free source. Measuring radio exposure in a diary or by telephone appeared by contrast to be unreliable, because people might not fill out the diary at the time of listening to programmes or they might misremember or mis-record which stations they were listening to at particular times.

Nielsen emerged as a near monopoly player in the metropolitan television markets in the United States along with Arbitron in the radio markets. While there continued to be a range of smaller audience ratings providers, the United States became characterized by monopoly provision in broadcast ratings with the ratings and Nielsen becoming synonymous. The historical experience of Australia with audience-ratings methodology, technology and ratings firm provision was significantly different. Exposure was still the core measure, but technology was not seen as the answer to error. While Nielsen was consolidating its market power as a near monopoly provider of television ratings Australia supported two competing ratings firms operating in the same market.

Bill McNair and George Anderson

Rather, as commodities themselves, the ratings were constructed in response to market pressures, including competition and monopolization as well as continuities and discontinuities in demand. The ratings producer was no scientist motivated by curiosity, but rather a company seeking its self-interest through the profitable manipulation of demand. E.R. Meehan (Meehan 1984: 201)

Meehan gives the impression that audience ratings emerged in an uncontested way, part of a capitalist engine that discovered research and measurement and simply adapted them as a way of increasing profit. However, audience ratings and survey forms of media research undertaken for advertisers, producers and broadcasters have been an integral part of our broadcast system for so long that it is difficult to see how much work had to go into constructing them, and then having them accepted and utilized by the different industry players. In the case of Australia, compared with Archibald Crossley's experience, early promoters of ratings and media survey research more generally had to invent, and then sell, the very concept of ratings and survey research to reluctant radio stations, advertisers and advertising agencies. Like Crossley, Hooper and Nielsen, what they were selling were new forms of knowledge, practices and techniques associated with the application of social survey research techniques as integral to the very management and orientation of the businesses associated with broadcasting. Typically what was being promoted and accepted was the general concept of 'scientific research' – market research.

Bill McNair, the founder of the McNair ratings system, gave his account of the need for audience ratings in his 1937 work, *Radio Advertising in Australia*. McNair lamented that in Australia 'systematic research has hardly been tried. The agencies with competently staffed research departments are in the minority; and on matters affecting newspaper and magazine circulations and radio owners' listening habits very little information has been collected' (McNair 1937: 44). Of those who tried to survey audiences, 'in most cases the results have not been published, this has been of little use to the great body of advertisers. The few surveys on which information can be obtained have differed widely in method and scope' (McNair 1937: 248). At the same time, McNair noted the inappropriate uptake and problems of market research in the United States, where market and audience research had just taken off.

In America when consumer research first obtained recognition, the new technique was quickly abused. Business executives with no statistical training

would run off long questionnaires abounding in irrelevant and ambiguous questions, and have them taken by untrained interviewers from door to door through all sorts of unrepresentative localities. Completed questionnaires which showed unfavourable results were freely discarded in order to make the final percentages more impressive. Such methods could not but cast discredit on the whole practice of consumer research. (McNair 1937: 248)

Bill McNair saw himself as promoting audience and market research that was independent and rigorous. His book is a report on his own surveys of radio audiences and listening habits of the time, pitched to convince the media industry of the need for sustained independent research. McNair attempted to establish the academic credentials of audience ratings by submitting his 1937 book as a PhD to a university in the United Kingdom. His application was knocked back but his academic interest in audience ratings remained, like that of his competitive counterpart George Anderson. Bill McNair's study was the first to touch on specifically 'audience' issues. But there were attempts at audience segmentation before McNair's survey.

J. Walter Thompson (JWT), McNair's employer at this time, hired psychologists A.H. Martin and Rudolph Simmat to measure consumer attitudes towards advertising. Simmat was research manager for JWT in Sydney in 1929 and began some of the earliest research into the segmentation of audiences. Simmat divided Australian society into four market segments based on income and housewives. Classes A and B were high income housewives; C and D were average or below average income housewives. Class D was 'barely sufficient or even insufficient income to provide itself with the necessities of life. Normally Class D is not greatly important except to the manufacturer of low price, necessary commodities' (Simmat 1933: 12). Simmat standardized interviewing techniques because experience had shown him that women were usually more effective as fieldworkers than men: 'Experiments have indicated that persons with a very high grade of intelligence are unsatisfactory for interviewing housewives ... usually a higher grade of intelligence is required to interview the higher class of housewife than is required to interview the lower grade housewife' (Simmat 1933: 13). Simmat and his team had interviewed 32,000 Australian housewives by 1932. Advertising was then targeted to specific audiences, with sophistication 'definitely soft-pedaled' for Classes C and D: 'We believe that farce will be more popular with our *Rinso* [detergent] market than too much subtlety.' Soap and detergent were the major advertising markets during the 1920s and crossed all market segments.

McNair does not mention Martin or Simmat's work in his book, even though he was familiar with it as JWT also supported the McNair radio survey. McNair

used similar stratification variables, with the home as the key focus. Gwen Nelson, arguably the first Australian woman in a senior capacity within audience ratings research, managed Bill McNair's sampling, fieldwork and interviews. Nelson spoke to us about her experience of audience measurement. McNair used interviews and recall methods to ascertain audience ratings:

The household was a household, it was a family. They weren't scattered here, there and everywhere. There was one radio set and the family sat around and listened to the radio. The advertiser thought it was ridiculous, nobody would sit around and listen to a voice. They'll read it when they see it in print, in the newspaper, and they'll believe it, but they won't take any notice of an odd voice talking to them over the radio. And also, the housewife was always at home, or mostly at home, we had no problem with what we would do with all the outs we found. We were able to get a very reasonable cross section of all the groups and all social structures, which was ideal for personal interviews.

For Nelson at that time the interview had several advantages over its Australian alternative of the diary:

I was brought up on personal interviews, and I left on personal interviews. I didn't like diaries, I never liked diaries. Though I must admit, today you couldn't possibly work as we worked. It would be an impossibility to do house to house interviews. You wouldn't find the people at home. As for phones, we didn't use them at all. The phone ownership wasn't very high, and it was so impersonal that your interview had to get the confidence of the person you were interviewing, and make them feel that they were really contributing to something. Which I think they did. And they're the early days. They were the days ... I staggered for 28 years, but I mostly ... was in Sydney quite a lot, but I did mostly interstate work. Every month I went somewhere. We were doing three surveys a year in Melbourne, two in Adelaide, two in Brisbane and one in Perth. I went to Perth every year for twenty years. And we had teams of interviewers. We did quite a lot of New Zealand. But I suppose we had two hundred casual interviewers throughout Australia and New Zealand, and we just called on them whenever we needed them, and we paid them at a casual rate and they were thrilled to work for a week or so. (Nelson 2000)

At a time when the United States was using telephone interviews and telephone coincidentals as the standard, McNair was using door-to-door interviews. In a way they had to: telephone penetration in Australia was much lower than in

the United States ruling out as Bill McNair observed that method for Australian conditions. At the same time the cost of these interviews meant that ratings sweeps would be more limited than in the United States. Finally, for decades in the twentieth century, Australia was not able to be grasped as a 'national audience' in any media or in any media organization conceptualization. There are various historical reasons for this arrangement. First, there was an industry resistance to measurement of audiences; ironically from the media itself, like radio. Second, there were, over time, limitations in what ratings could do in their classifications or which areas, geographically, they could reasonably cover in frequency. This meant that particular cities and regional markets were always the targets with the consequence that a national ratings figure was not the focus of the research as in the United States; the focus was instead on individual markets – the domain, in the United States, of 'local ratings'.

Bill McNair, however, never gained the ascendancy over the Australian ratings market that Art Nielsen achieved in the United States. McNair's chief rival was George Anderson.

George Anderson had a background in regional and then Sydney radio where he acted as station manager of 2GB and sales manager of the Macquarie Radio Network. Don Neely, one of his executives, recalls that Anderson became interested in media research, because he wanted to know 'who was on the other end of the microphone, listening':

He corresponded, I believe, with some of the Americans who were active at that time, including Hooper, who was doing his Hooper ratings, and so George struck out on his own with a capital of fifty pounds. He founded the company and started measuring radio audiences. In between times, he'd found time to be a Colonel in the army, so he was a busy man and a real character. A very determined man. (Neely 2000)

George Anderson started with personal interview and recall, but after talking to colleagues in the United States and reading the research of the National Association of Broadcasters and C.E. Hooper, he settled on diaries as the preferred method of data collection. In 1947 Anderson wrote a confidential report to the Australian Association of National Advertisers, *Report on the Effectiveness of the Radio Listener Diary in Measuring Radio Audiences in Australia*. The extract at Figure 2.1 shows that he was familiar with all the major methods, including Nielsen's Audimeter.

Anderson concluded in his report to the advertisers that the Audimeter solution was too capital-intensive, and that the most reliable method for Australia was the Radio Listener Diary, used by C.E. Hooper as an adjunct to the coincidental

2. When, in June 1944, The Anderson Analysis of Broadcasting commenced to measure radio audiences in Australia, there were ~~two~~ methods in general use, viz:-
  - (a) The Telephone Method, by which people with telephones were contacted by telephone and asked what radio station their radio set was tuned to:-
    - (i) at the time of the call (the Co-incidental Telephone Method) and/or
    - (ii) during a short period prior to the call (the Recall Telephone Method.)
  - (b) The Personal Interview Method, by which people were interviewed in their homes to find out:-
    - (i) The station being listened to at the time of the interview (the Co-incidental Personal Interview Method) and/or
    - (ii) The stations which were listened to during a period prior to the interview (the Recall Personal Interview Method.)
3. At the same time (June 1944) two other methods were being developed as follows:-
  - (a) The Mechanical Recorder Method, by which a mechanical recording device was attached to a radio set and, by means of a moving tape, recorded the times when the set was switched on and off and the time when each station was tuned in and tuned out.
  - (b) The Radio Listener Diary Method, by which people in their homes kept a record, in diary form, of their listening during a number of consecutive days.

Figure 2.1 Extract showing description of methods

Source: Nielsen Historical Archive

telephone method. Anderson attached a sample diary for the advertisers, shown at Figure 2.2.

As we have seen, Bill McNair, by contrast, came to a completely different conclusion and settled on personal interview and recall. No doubt Anderson hoped that his report to the advertisers would knock McNair out of the race. However, quite the opposite happened. Both methods, personal interview and recall and diary, ran side by side as audience measurement technique until the two firms merged in the 1970s. For the whole period from 1944 through to the merger, the debate about the two methodologies was both personal and public. Bill McNair's son Ian McNair, himself a major figure in Australian audience research, takes up the story:

It was an argument that went on between the two research companies. It was a very public argument, because my dad liked sending out circulars and

letters, and quotations and methodology arguments and George responded in his way, so the argument was going on all the time.

Figure 2.2 Early historical example of Anderson Analysis Radio Ratings Diary
Source: Nielsen Historical Archive

HOW TO KEEP THIS RADIO DIARY

- INTRODUCTION**
In keeping this record of home listening, you are taking part in a new scientific survey, designed to find out when people listen and what they listen to.
- IMPORTANT**
Just because you are keeping a record of your radio listening, please do not let it affect the use of your radio set in any way.
Providing all members of your household observe the following simple rules, your diary will be just as valuable whether you use your radio every day, or whether you use it for only a few minutes during the whole eight days of the survey.
- SPECIMEN PAGE**
The Specimen Page opposite demonstrates the simple rules set out below, and clearly shows how easy it is to keep an accurate record of your listening.
- JUST A SIMPLE RECORD**
There are but three things to record:—
(a) The times when your radio set is switched on
(b) The call signs of the stations to which it is tuned
(c) The number of people who actually listen
- WHEN YOUR RADIO SET IS SWITCHED ON**
Put a tick in the appropriate square (thus ☒) when the radio is switched on. Tick every quarter-hour when it is on. If any square has no tick in it, we will know that the radio was not switched on at that time.
Notes—Tick as "ON" if radio is switched on for five minutes or longer in any quarter-hour.
- RADIO STATION CALL SIGNS**
When you switch on your radio set, write opposite the tick you have placed in the appropriate square, the call sign of the station tuned in. When the radio is switched to another station, write the new station's call sign.
Where you continue to listen to a station for more than a quarter-hour, place ditto marks under the call sign for every quarter-hour concerned.
It is most important to write the call signs at the time you switch on your set, or when you change from one station to another, and also, when you switch off your set to see that no ticks or ditto marks have been omitted.
- NUMBER OF PEOPLE ACTUALLY LISTENING**
Write in the actual number who listen each quarter-hour. In this connection "M" stands for Men, "W" for Women and "C" for Children—only boys and girls under 16 years of age are classed as children.
If the set is switched on but no one is listening, place crosses opposite the times concerned.
Notes—Record the number of Men, Women and Children listening if someone was listening for 5 minutes or longer in each quarter-hour concerned.
- NO LISTENING**
If, on any day, your radio is not switched on at all between 6.00 a.m. and midnight, write across the form for that day, in large letters, "No listening" or words to that effect. This is most important.
- QUERIES**
If there is any doubt about how to keep this Radio Diary, please 'phone us as follows:
9.00 a.m. to 6.00 p.m. Mondays to Friday and 9.00 a.m. to 12 noon on Saturdays Phone BX 1645
At all other times Phone JX 3373 or XL 2171
- FINALLY**
Broadcasting Stations and Advertisers are waiting for the information we are obtaining from this survey. It will be of great help, therefore, if you will post the completed Diary to us on Monday, 10th February, using the attached stamped, addressed envelope.
On behalf of Broadcasting Stations and Radio Advertisers, and on our own behalf, please accept our most sincere thanks for your co-operation.
We can assure you that the time spent in making this record will be of great assistance to people responsible for planning programmes, and thus will be to the benefit of all listeners in Sydney.

SPECIMEN PAGE

MORNING 6:00 A.M. - 12:00 NOON				AFTERNOON 12:00 NOON - 5:00 P.M.				EVENING 5:00 P.M. - 12:00 MID.			
TICK	WHEN RADIO SET ON	STATIONS	NUMBER ACTUALLY LISTENING	TICK	WHEN RADIO SET ON	STATIONS	NUMBER ACTUALLY LISTENING	TICK	WHEN RADIO SET ON	STATIONS	NUMBER ACTUALLY LISTENING
☐	6:00			☐	12:00			☐	5:00		
☐	6:15			☐	12:15			☐	5:15		
☐	6:30			☐	12:30			☐	5:30		
☐	6:45			☐	12:45			☐	5:45		
☐	7:00			☐	1:00			☐	6:00		
☐	7:15			☐	1:15			☐	6:15		
☐	7:30			☐	1:30			☐	6:30		
☐	7:45			☐	1:45			☐	6:45		
☐	8:00			☐	2:00			☐	7:00		
☐	8:15			☐	2:15			☐	7:15		
☐	8:30			☐	2:30			☐	7:30		
☐	8:45			☐	2:45			☐	7:45		
☐	9:00			☐	3:00			☐	8:00		
☐	9:15			☐	3:15			☐	8:15		
☐	9:30			☐	3:30			☐	8:30		
☐	9:45			☐	3:45			☐	8:45		
☐	10:00			☐	4:00			☐	9:00		
☐	10:15			☐	4:15			☐	9:15		
☐	10:30			☐	4:30			☐	9:30		
☐	10:45			☐	4:45			☐	9:45		
☐	11:00			☐	5:00			☐	10:00		
☐	11:15			☐				☐	10:15		
☐	11:30			☐				☐	10:30		
☐	11:45			☐				☐	10:45		
☐	12:00			☐				☐	11:00		
☐	12:15			☐				☐	11:15		
☐	12:30			☐				☐	11:30		
☐	12:45			☐				☐	11:45		
☐	1:00			☐				☐	12:00		
☐	1:15			☐				☐	12:15		
☐	1:30			☐				☐	12:30		
☐	1:45			☐				☐	12:45		
☐	2:00			☐				☐	1:00		
☐	2:15			☐				☐	1:15		
☐	2:30			☐				☐	1:30		
☐	2:45			☐				☐	1:45		
☐	3:00			☐				☐	2:00		
☐	3:15			☐				☐	2:15		
☐	3:30			☐				☐	2:30		
☐	3:45			☐				☐	2:45		
☐	4:00			☐				☐	3:00		
☐	4:15			☐				☐	3:15		
☐	4:30			☐				☐	3:30		
☐	4:45			☐				☐	3:45		
☐	5:00			☐				☐	4:00		
☐	5:15			☐				☐	4:15		
☐	5:30			☐				☐	4:30		
☐	5:45			☐				☐	4:45		
☐	6:00			☐				☐	5:00		
☐	6:15			☐				☐	5:15		
☐	6:30			☐				☐	5:30		
☐	6:45			☐				☐	5:45		
☐	7:00			☐				☐	6:00		
☐	7:15			☐				☐	6:15		
☐	7:30			☐				☐	6:30		
☐	7:45			☐				☐	6:45		
☐	8:00			☐				☐	7:00		
☐	8:15			☐				☐	7:15		
☐	8:30			☐				☐	7:30		
☐	8:45			☐				☐	7:45		
☐	9:00			☐				☐	8:00		
☐	9:15			☐				☐	8:15		
☐	9:30			☐				☐	8:30		
☐	9:45			☐				☐	8:45		
☐	10:00			☐				☐	9:00		
☐	10:15			☐				☐	9:15		
☐	10:30			☐				☐	9:30		
☐	10:45			☐				☐	9:45		
☐	11:00			☐				☐	10:00		
☐	11:15			☐				☐	10:15		
☐	11:30			☐				☐	10:30		
☐	11:45			☐				☐	10:45		
☐	12:00			☐				☐	11:00		
☐	12:15			☐				☐	11:15		
☐	12:30			☐				☐	11:30		
☐	12:45			☐				☐	11:45		
☐	1:00			☐				☐	12:00		
☐	1:15			☐				☐	12:15		
☐	1:30			☐				☐	12:30		
☐	1:45			☐				☐	12:45		
☐	2:00			☐				☐	1:00		
☐	2:15			☐				☐	1:15		
☐	2:30			☐				☐	1:30		
☐	2:45			☐				☐	1:45		
☐	3:00			☐				☐	2:00		
☐	3:15			☐				☐	2:15		
☐	3:30			☐				☐	2:30		
☐	3:45			☐				☐	2:45		
☐	4:00			☐				☐	3:00		
☐	4:15			☐				☐	3:15		
☐	4:30			☐				☐	3:30		
☐	4:45			☐				☐	3:45		
☐	5:00			☐				☐	4:00		
☐	5:15			☐				☐	4:15		
☐	5:30			☐				☐	4:30		
☐	5:45			☐				☐	4:45		
☐	6:00			☐				☐	5:00		
☐	6:15			☐				☐	5:15		
☐	6:30			☐				☐	5:30		
☐	6:45			☐				☐	5:45		
☐	7:00			☐				☐	6:00		
☐	7:15			☐				☐	6:15		
☐	7:30			☐				☐	6:30		
☐	7:45			☐				☐	6:45		
☐	8:00			☐				☐	7:00		
☐	8:15			☐				☐	7:15		
☐	8:30			☐				☐	7:30		
☐	8:45			☐				☐	7:45		
☐	9:00			☐				☐	8:00		
☐	9:15			☐				☐	8:15		
☐	9:30			☐				☐	8:30		
☐	9:45			☐				☐	8:45		
☐	10:00			☐				☐	9:00		
☐	10:15			☐				☐	9:15		
☐	10:30			☐				☐	9:30		
☐	10:45			☐				☐	9:45		
☐	11:00			☐				☐	10:00		
☐	11:15			☐				☐	10:15		
☐	11:30			☐				☐	10:30		
☐	11:45			☐				☐	10:45		
☐	12:00			☐				☐	11:00		
☐	12:15			☐				☐	11:15		
☐	12:30			☐				☐	11:30		
☐	12:45			☐				☐	11:45		
☐	1:00			☐				☐	12:00		
☐	1:15			☐				☐	12:15		
☐	1:30			☐				☐	12:30		
☐	1:45			☐				☐	12:45		
☐	2:00			☐				☐	1:00		
☐	2:15			☐				☐	1:15		
☐	2:30			☐				☐	1:30		
☐	2:45			☐				☐	1:45		
☐	3:00			☐				☐	2:00		
☐	3:15			☐				☐	2:15		
☐	3:30			☐				☐	2:30		
☐	3:45			☐				☐	2:45		
☐	4:00			☐				☐	3:00		
☐	4:15			☐				☐	3:15		
☐	4:30			☐				☐	3:30		
☐	4:45			☐				☐	3:45		</

was that ours was more reliable because it was what people remembered yesterday, and it wasn't just the people who filled out a diary. (McNair 2000)

The motives for collecting audience ratings were also personal, which often left Anderson especially on a financial tightrope because of George Anderson's commitment to methodology and service. Don Neely recalls Anderson over-providing to country TV stations 'for the reason George said, "it's their livelihood. Forget the agencies, it's the station that's got to live or die by these numbers, and I want them to be right".'

The fact that we've quoted a price and are providing a service which was more expensive than we got was quite irrelevant to George. As long as he was right. And that's why the company was never really financially flush. McNair's on the other hand tended to go the other way. And I noticed this when we got together, for example in regional television, they only did it one year, and the spread of diaries they used was abysmal. They might have used twenty drop points in a market, ours would have used sixty. So you can imagine the cost differential. (Don Neely 2000)

By the mid-1960s McNair moved, under industry pressure, to adopt the diary system. For nearly the next decade there were two providers operating two systems based on the same ratings instrument. This dual system of audience ratings and its ongoing public display, including the arguments over methodology that were presented in many public fora, had practical purpose for clients. One system was a check on the other. No client bought one set of ratings without buying the other. The ratings could be used to present a case to advertisers as to why they should advertise on a station or be used in programming decisions. Only a minority of stations would ignore the ratings surveys, as Des Foster, general manager of 2GB at the time, notes:

They weren't cheap, and some stations, I recall, who didn't do well in the ratings and were not doing well, revenue wise, would say 'we don't want them', they didn't have the surveys, they didn't want to know about them. My recollection was that 2KY might have been one of those stations at the time, and they took that view that surveys only damaged them. They're far better off going to clients and talking to them and selling them advertising and not raising the question of surveys. The big problem with surveys – there's always a debate in the industry, of course – not a problem with surveys, a problem with the way you deal with them. That is, when you have a good survey from the stations point of view, the station's sales people all want to race around the advertisers and say 'look what a good survey we've got'. And then, three

months later the positions reverse, they're telling you something else and perhaps one of the negative effects of that over the years is that there has been not enough selling of the radio ... the merits of radio as a medium ... there has been a much greater emphasis, unfortunately, in selling the merits of *my* radio station as a medium. And I think the industry has suffered to some extent because of that. (Foster 2000)

As the Australian market grew, including the introduction of new media like television, so did the cost of collecting audience ratings. Who should bear this cost, as mentioned by Gale Metzger in Chapter 1, is an ongoing debate in the ratings convention. McNair and Anderson finally merged in 1973 because of industry pressure to do so and affordability. The diary method then became the main method of data collection.

The main single reason was that being seven days, from the same respondent, we're able to work out reach frequencies much better than we could with the single day samples. And it wasn't so much personal interviews versus anything else, it was the fact that you got data over a whole week from a diary, that really persuaded us to continue on with that method for both radio and television. That's what we did, right from 1973 onwards. Anderson's were in town and they came across to the same building as us, 40 Miller Street North Sydney, and no one was sacked or anything because of the merger because McNair's continued doing their market research as well, and so we're able to employ everyone in the merger without laying off anyone. We're very proud of that. A lot of the market research that McNair's were doing was able to continue throughout the seventies. (McNair 2000)

Interestingly, the diary system ran in Australia up until 1991 when peplemeters were finally introduced for the television market. Diaries, however, are still used for many remote parts of Australia and in Darwin, and remain the only means of collecting audience data for radio broadcasting markets. Commercial Radio Australia's *Global Overview & Australian Industry Position* emphatically stated in 2007 that 'Australia currently uses the diary system, the only proven and reliable ratings method used around the world at present. Electronic measurement is currently not in use country-wide anywhere in the world as the sole radio audience measurement system, nor is it regarded anywhere as industry currency, except in a few areas in Switzerland and one city, Houston, in the USA' (Commercial Radio Australia 2007).

Arbitron, the monopoly provider for radio ratings in the United States, has, by contrast, for many years been trying to replace the paper diary with a Personal People Meter (PPM). The PPM is like a pager. It can pick up digital codes from

different media and send the information to a database. Arbitron has used paper and pencil diary systems for radio since 1965. But the Australian radio broadcasting market continues to reject the technology push. Australia's Commercial Radio Australia, however, argues that the PPM:

- Costs 2 to 3 times the cost of the diary system
- Has no real evidence of long-term compliance across all demographics – e.g. once listening is lost, it is lost and cannot be regained
- Drops in breakfast listening have been recorded in trials with no valid evidence provided for device purveyors' explanation that 'people must have been over-reporting breakfast listening'! And industry does not accept this explanation as no other daypart has been supposedly over-reported. We do not believe that tens of millions of diary keepers across the world over the last 50 years have all over-reported breakfast and no other daypart. (Commercial Radio Australia 2007)

The Australian situation gives us some tantalizing insights into the complexity of the audience-ratings convention and how formal and informal agreements work within this compact. While both the US and Australian markets agreed on exposure as the core measure there were fundamental differences between them in the *speed* by which new innovations or technologies were adopted and arguments about the validity of different technologies. In the United States the machine is still seen as the best solution to capturing exposure. However, competitive pressure in the United States – with Arbitron seemingly under-reporting radio audiences – has led to radio networks putting pressure on Arbitron by inviting Nielsen to introduce a paper diary system. The entry of Nielsen into the diary radio market, however, is more likely to be a stalking-horse attempt to force Arbitron to change its methods. This type of intervention itself is also not new, and often functions as an 'informal' way of creating change, as we will see later.

While the recall and diary methodologies continued to run side by side in Australia until the merger of McNair and Anderson in 1973, the underlying conceptualizations of the audience had changed during the 1950s. Programming formats and attention to demographics became interlinked. There was, by the time television became standard, no radio or television manager that was not looking carefully at the demographics of their audience and measuring their performance by the audience ratings. Australia also differed from the United States in its methodology and approach. As you can see from the accounts of senior participants in the creation of the audience ratings system in Australia, there was a dedication to the art of audience ratings in and of itself, to the point where a company was willing to lose

money to ensure accuracy. Why did Australia keep two competing ratings systems, both of which collected data from the same audiences? Why would hard-nosed station managers and media magnates use two methodologies that produced different figures and were expensive?

The answer lies in the character of the competition between McNair and Anderson. Both thought that their methods were the best. Both were respected in the media industry and both were trusted. When the audience became more fragmented and there were more stations to measure, the economics of keeping two ratings agencies changed. But embedded in that older system was a perception of checks and balances, even if it was costly to run two methods of audience measurement. Those checks and balances, interestingly, did not disappear when the two companies merged in the 1970s. Don Neely recalls:

The merger happened at the end of 1973, which saved the problem for the industry, because there was still an amount of sympathy for McNair, and because he'd been in the industry for so long, having been dumped in the tendering situation. So they were delighted when we announced that we were getting together. But they insisted that the methodology be the Anderson methodology, the format was the Anderson format, in fact it was an Anderson show. Except that McNair was part of the deal. And that really did help, having the two got together. There was never really enough market for two companies in Australia. One company made a lot more sense. So the industry was happy. They had a research committee, they and we agreed to have an auditor, Dr Arthur Meadows was the first auditor, and he would come in and go through all the diaries and all the paperwork to see if we did what we said we'd do. And I believe today there still is an auditor. (Neely 2000)

The shift to an audited regime in radio audience ratings in Australia, the authors would argue, is a direct consequence of the way that McNair and Anderson developed over time. The dual system of audience ratings acted as a default audit and indeed as an integral part of the audience-ratings convention.

New Forms of Knowledge about Audiences

The measurement of audiences led to the creation of a whole new language and form of knowledge centred on media buying as a speciality. Once all the results could be received in a form that was easy to manipulate by slide rule and then computer other kinds of pictures of the audience could be created. As we have seen, decisions about what counts as correct measurements in the world of audience ratings are negotiated. A method such as telephone recall might be argued as valid at one point

in time and invalid at another. Ironically, C.E. Hooper, for example, returned to recall later in his career, despite arguing strenuously against it when Crossley ran recall and Hooper ran telephone coincidental. Arguments about methodology, of course, are also arguments about what to accept as methodology and what trade-offs to make in the process. Choosing exposure itself is a trade-off in that it is an indirect, not direct, measure of audience like or dislike of a programme or station. It is automatically assumed in exposure that ten minutes of listening is ten minutes of listening, when in fact people may have done other things or, indeed, very little listening at all.

The audience ratings numbers, however, did not only affect commercial decisions on media buying. The whole process of using audience ratings created a new type of knowledge to deal with the complexities of translating audiences as raw figures into analytic categories.

On the media side, for example, media analysis tools were created to assess the reach and frequency associated with different media and the costs to be charged against them. Reach refers to unduplicated exposures or gross impressions and the number of different people exposed to the message. Frequency refers to the average number of exposures and how many times an audience is exposed to a message. Outdoor, newspapers and magazines tend to be the best media for frequency. Broadcast advertising and magazines tend to be the best for reach. The best combination is found in radio.

A cost per thousand (CPM) calculation is made on how much it costs to deliver 1,000 gross impressions or to reach 1,000 listeners, viewers, readers, households and so on. CPM allows media planners to compare media based on audience and cost. The lowest cost per thousand medium therefore is the most efficient.

For print, when the cost of the advertisement is known: the CPM is the cost of advertisement x 1,000 circulation; when audience data are known: the CPM is the cost of 1 advertisement x 1,000 readers reached; with broadcast media: the CPM is the cost of 1 advertisement x 1,000 homes reached. Suppose that an advertiser can buy a television commercial minute for US\$1,200 during a situation comedy that reaches four million people and a commercial minute for US\$1,000 during a football telecast that reaches only one million people. When the costs per minute are compared the situation comedy seems to be a better buy:

Situation Comedy:	Football Game:
\$1200 = \$0.3 CPM	\$1000 = \$1 CPM
4000 (000) total audience	1000 (000) total audience

But if an advertiser's target is men aged 18–49, the situation comedy may not be so attractive. If only one million men watch the comedy and all the football viewers are men, the football game is a better buy.

Situation Comedy:

\$1200 = \$1.2 CPM

1000 (000) men (18–49)

Football Game:

\$1000 = \$1 CPM

1000 (000) men (18–49)

The comedy costs US\$1.2 CPM and the game only US\$1. Thus, football is a better buy for the advertiser because it reaches the target audience at a lower unit cost. Lest this hypothetical example prove misleading, it should be emphasized that it is deliberately simplistic. Many other factors may be taken into account when measuring an audience and in general prime-time Australian television audiences don't normally subdivide so markedly. In prime time stations are competing for the whole family. They normally attract the middle to lower income groups, as those on higher incomes tend to watch the Australian Broadcasting Corporation (ABC) or be relatively 'light' television viewers. The most desirable age group to reach may be the 25–39 demographic as it delivers the best combination of viewing numbers, spending power, flexibility of attitudes and predictability of tastes.

Again speaking generally, a show must score high ratings to survive in prime time – between 6 p.m. and 10 p.m. If it is rating less than 'twenties', or 20 per cent of homes with televisions, it probably won't survive. If it is screened on a commercial station and scores ratings of between 20 and 30 it is probably getting a fairly even share of the market in Sydney, Melbourne, Adelaide and Brisbane. If it is getting 'thirties' it is probably doing very well indeed as the ABC typically accounts for between 10 and 20 per cent of the homes using television and Special Broadcasting Service (SBS) for around 5 per cent. When an ABC public service show goes into 'thirties', commercial station managers often complain that the corporation is not serving its minority audiences. Ratings of programmes outside prime time are another matter. Off-peak ratings of only 2–3 per cent of sets are common and 10 per cent might be considered good. Sets in use vary according to the time of day, the day of the week and the time of the year (higher in the winter, lower in summer). They also vary by city.

All stations seek high rating shows because they tend not only to attract viewers to that particular programme but also to hold them for other programmes. Many stations, for example, often spend more than they reap on the main evening news bulletin because it marks the beginning of prime time and may lock in its audience for the evening. Thus, the sequence of programmes is important because a show following a high-rating programme tends to do better than it otherwise might have done. A series rating thirties can lift a channel out of the economic doldrums by attracting viewers to the channel itself. The station management may then bargain with advertisers, offering them good 'positioning' in the highly rated shows if they will agree to fill several less favoured time spots in slack periods.

Series and serials are the dominant programmes on most prime-time commercial television. But because of the spectacular successes of a handful of drama programmes, many of them produced overseas, it is easy to overrate the viewing loyalties of series. A series is a drama programme in which the same characters appear in successive episodes, each with a self-contained plot. A serial differs in that it has a continuing story line. Such programmes are normally shown for a certain number of episodes at the same time each week, with relatively few changes throughout a TV season. Such repetitive programming gives television viewing a firm structure, attracting viewers back night after night or week after week at the same time. They also know what to expect when they return. Producers in particular like serials because they are cheaper to make than 'episode' or one-off dramas, as they can use the same sets, cast and crew.

But it does not necessarily follow that the viewers of one episode of a series or serial will automatically watch the next, even though they may like it very much and even though they may say they watch every episode. In *Television Audience: Patterns of Viewing* Goodhardt, Ehrenberg and Collins showed that at that time only about 55 per cent of viewers of one episode of a programme would watch the following episode, no matter what the type of programme or content or even when completely different programmes (with similar ratings) were shown in the same timeslot each week. 'Repeat viewing,' they argued, 'therefore appears to be more a function of social habits (i.e. people's availability) than of programme content' (Goodhardt, Ehrenberg and Collins 1975: 125). They also found that the extent to which different programmes share the same viewers also follows a simple pattern. They called this 'the duplicate of viewing law': 'for any two programmes the level of duplication in their audiences depends on the *ratings* of the programme and not on their *content*'. One pair of programmes generally has the same degree of audience duplication as any other pair of programmes with the same ratings' (Goodhardt, Ehrenberg and Collins 1975: 126). They also worked out a multiplier constant for audience overlap in the second programme of a series. For example, if the constant or 'proportionality factor' for Nine Network programmes is 1.4 and the rating of the second programme is 20, then the duplication of viewing law states that about $20 \times 1.4 = 28$ per cent of the audience of the *first* programme will have watched the second programme. That is, the audience overlap generally depends only on the proportionality coefficient (here 1.4) and the programme's rating (here 20) and *not* on programme content. Summing up their findings, the authors say very few viewers of any series see all or most of its episodes. Even self-proclaimed 'regular viewers' are not all that regular.

With the introduction of computers a range of permutations of individual variables is possible. Moreover, with the introduction of peplemeters, near

IN-HOUSE	ETAM ANALYZER AdQuest eTAM Analyst OzTAM	ETAM SCHEDULIZER AdQuest eTAM Scheduler OzTAM	ETAM MAXIMIZER AdQuest eTAM Maximizer OzTAM	VIEWERGRAPHICS AdQuest eTAM Viewergraphics OzTAM	SCRUTINIZER AdQuest eTAM Scrutinizer OzTAM	TSS – MAX MEDIA MAX MEDIA
	RADIO ADVISOR AC Nielsen Radio Advisor Nielsen Media Research	AD VIEWS REPORTER AdViews Reporter	ETAM MILLENIUM AdQuest Millennium OzTAM	ETAM EBBINIZER AdQuest eTAM OzTAM	AARDS aards.	ROY MORGAN Roy Morgan Research
FEE BASED	OMD CURVE curve © OMD	OMD SNAPSHOTS snapshots © OMD	OMD DATAPILOT DataPilot © OMD	OMD PHASER phaser © OMD	OMD ROAR ROAR © OMD	COMMUNIGRAPHICS communigraphics
	OMD FLIGHTPLAN flightplan	OMD METRICS Metrics © OMD	OMD INTEGRAL Integral	OMD ALLOCATOR allcator	OMD PALETTE palette	

Figure 2.3 Sample of in-house and fee-based analysis services

instantaneous data collection became possible as the time of exposure can be transmitted directly to databases. A large range of software packages quickly became available to analyse ratings and other audience research data on users' desktops. In an earlier era there was no intermediary between the ratings firm and their clients. But now media buying, whether in-house or as a separate agency, provides a raft of analytical services that interpret the ratings. The people who do this interpretation are no longer, necessarily, methodologists: they are knowledge brokers in the real sense of the word. Nor is the manipulation of audience data limited to simple classifications.

The early ratings methodologies used families and social class definitions to segment the audience. McNair and Anderson, surprisingly, ran different definitions of the audience up until 1963. McNair used A,B,C,D,E and Anderson used A,B,C (upper, middle, industrial). The age categories for television were also different. For example, in reporting on audience ratings for television Anderson did not use age breakdown in reports until 1961 and then reported on the age ranges 1–15, 16–24, 25–39, 40–54 and 55–99. McNair also did not use age categories until 1963 and then reported on the age categories 0–11, 12–19, 20–35 and 36+. By 1970, however, both McNair and Anderson were using the same age categories for radio and television: 10–17, 18–24, 25–39, 40–54 and 55+. Housewives were their own category in this period, together with reporting on 'housewife lifecycles'. The use of social class definitions also affected the 'argot' or local language for reporting audience ratings. The term 'AB', for example, was first used in one of McNair's reports in 1959; previously the categories had been separate.

Once the audience ratings structure was in place and once it was established by McNair and Anderson and accepted as the currency, changes in the machineries of knowledge became intimately linked to changes in demographics or technology or the needs of stations and programmers. The first big change in the 'simple model' came with the introduction of the transistor radio in the 1950s, and, of course, with television:

Radio was a medium for groups. That's peak time radio. Very few car radios and virtually no portables. So it was all done in the home. Then there was the invention of the transistor. The transistor suddenly made it possible to produce a light little tiny radio set that you could carry around with you. So the technology gave us very very quickly the capability to transform the medium from a group medium, a medium for a family, into a medium for a person, one person. And so that was highly significant. Portable, you could have it in your car, you could take it to the beach, you could do what you like. Portable, personal medium. You don't have one radio in the household. Whereas in the household you might have had a big one in the living room, perhaps one in the kitchen for mum. Now junior's got his own radio, and sis has got her radio and they could all be doing their homework in their different rooms, listening to different programs. This mobility allowed a switch of focus from nighttime radio and radio was never ever to dominate at night again. So what they were able to offer in the daytime was programs for individuals in the daytime and for people on the move, and for people at the beach and for people at leisure, mobility, individuals in their own rooms, in the garden. Suddenly the medium began to transform. And the major emphasis was on music formats.

I think that gave a shove along on the concentration on demographics, also. Suddenly the 12–18 year-old group became a very identifiable segment, 12–18, whatever you like. But I mean all these age demographics became very very relevant and it was possible to tailor programs for them. And instead of thinking in mass terms, mass audience, it was no longer quite so important to be number one in the ratings if you could be number one in 10–17. (Foster 2000)

The whole practice of radio, in this case, changes because of both a new technology and a newly defined and differently valued demographic. At the same time, more data are collected as categories expand and time of collection expands. Diaries collect data in 15 minute segments. Peoplemeters collect it by the second. Agreements on the valuable demographics, however, are not accidents of history, accidents of technology or accidents in measurement. The agreements are a part of ongoing debates within the audience-ratings conventions.

Theorizing the Convention

In modern marketing circles there appears to be a visceral reaction to continuation of the use of A,B,C,D classifications:

As far as marketing jargon goes, nothing bothers me more than use of the term AB to refer to a demographic target audience. As far as I'm concerned, all it stands for is Absolute Bollocks. The term originates from the social-economic classifications of the English class system. It could never have originated here as Aussies have no class ... When was the last time you heard a consumer self-refer as an AB? Just imagine the scene in a restaurant: 'Waiter, why is my main course taking so long you lowly D, and tell the chef he's a useless FG. This is outrageous – I'm an AB don't you know!' (Joseph 2008: 17)

Joseph is not disputing exposure as the key measure in ratings, but rather aspects of its use. The early regime in audience ratings may appear to be a simple model, but in fact it was and remains a complex agreement between many often conflicting parties on what should be counted and how frequently, even if there are different ways of 'cutting up' the demographics of exposure. The need for standards or conventions governing measurement is not new, as Stigler has pointed out (Stigler 1999). Stigler uses the 'Trial of the Pyx' to show the differences between 'standards' and 'statistics' and the role of sampling. The Trial of the Pyx system was created in 1150. The trial was where the mints' coinage was put to test: there were standard and statistical methods. A standard was needed for comparison to tell if a newly minted coin was as promised. For example, as a standard of fineness, a bar of gold was retained as a reference. Statistical methods were needed because the sheer volume of the coinage makes the individual weighing of a coin impossible. Tests of fineness were destructive and made tests of each coin impossible. Sampling, as a consequence, was absolutely necessary.

The earliest documents are not specific about how the samples would be drawn, but it is impossible to believe that the different (and very suspicious) parties would have been satisfied with a noticeably biased selection, and the documents support this. But sampling was not the only statistical method born of necessity in this trial; there were two others of note. One of these was the mean. In order to avoid having vaguely understood uncertainties of weighing mask major variations in the weight of coins, the coins (like Kobel's feet) were measured in the aggregate, say 100 at a time – essentially it was the average weight of the tested coins that was compared with the standard. Of course from the point of view of experimental design, this was admirable – the aggregate was subject to one measurement error rather than 100. And

there was yet another statistical method employed, an allowance for variability: because mint technology was not perfected, it was granted by all that some allowance had to be made for variability. If the coins weighed too little, the barons were being cheated and the acceptance of the coins in circulation was jeopardized. If the coins weighed too much, the larger coins would be culled from circulation, melted, and recoined, with the profit going to the merchant. Neither situation was tolerable. The allowance that the contract specific was called the 'remedy,' since measures outside these limits would need to be remedied by the master of the Mint. (Stigler 1999: 367)

'Allowance for variability' that was made in the case of the Mint is no different from the 'allowance for error' in modern audience surveys, which will be discussed in the next chapter. What is important from Stigler's history of sampling is the recognition that agreement on measures and standards may not only be a scientific matter. The negotiations around standards are often complex and who is accepted as arbiter equally complex.

Exposure as the core measure in modern audience ratings has often been debated by the ratings methodologists and its weaknesses betrayed, yet the parties have agreed to accept the weaknesses. Where improvements have been proposed, like Art Nielsen's Audimeter, the 'superior' method of measuring exposure has to be argued and sold. New categories of audience, likewise, are not static or immediately universally accepted, and agreement is needed on any changes to the convention. The measure and the standard remained much the same for many years because more complex analytics were not required, and not because early audience ratings experts had no idea how to create more complex definitions of demographics based on exposure data.

There are two levels at which exposure, the standard, can become contested:

- 1 the method by which exposure data is collected (the measuring instruments, like telephone coincidental versus recall); and,
- 2 the analysis and classification of that data (demographics, sizes of audiences, and so on).

The dynamism and ongoing fact of this contestation is little talked about and only sometimes comes to the surface in key debates about audience ratings and their validity. What has to be recognized, however, is that this dynamism has an inherently democratic aspect to it and is built-in to the convention. In particular, the adoption of modern audience ratings means that the data collection methods and analytics need to be transparent. Hypoing is still not acceptable in the convention. What this means in practice is that countries like China are adopting measures and standards

that will affect the presentation of their audiences to the global market. If China seeks to distort its audience for national purposes or to influence the international market, then it will be seen to be breaching the ratings convention.

China is quickly becoming an established ratings user. China originally teamed up with one of the largest ratings providers, Taylor Nelson Sofres (TNS). Chinese TV viewers are measured by peplemeters and the radio users by diary (as they are in western countries). There is one panel (sample) of 4,000 households to represent 1,199,564,000 people in China, as the national panel. There is also measurement at city and province level and where peplemeters are not possible for television measurement diaries are used (in Australia cities like Darwin are still measured by diary for television).

It is not at all clear how a global audience-ratings convention might work, but there can be little doubt that audience ratings are expanding on all continents. What we do know is that audience ratings as they have developed in western industrial countries have a *dual identity*. Audience ratings are not like other forms of social science survey research because they require a higher level of scrutiny than other types of research. They are at one moment a standard by which media markets

Table 2.1   Structure of China's TV panels

	Number of Panels	Sample Size	Universe '000
National Meter	1	4,000	1,199,564
City Meters	43	9,950	103,042
Provincial Meters	4	2,050	155,969
City Diaries	125	15,500	106,612
Provincial Diaries	21	12,400	917,077
Total	194	43,900	

Source: CSM Media Research

Table 2.2   Structure of China's radio panels

	Number of Panels	Sample Size	Universe '000
City Diaries (constant)	13	3,900	32,216
City Diaries (3 sweeps)	6	1,800	13,386
City Diaries (4 sweeps)	13	3,900	14,758
Total	32	9,600	

Source: CSM Media Research

make financial decisions and at another moment a public agreement, a convention, on what counts as an audience.

Summary

Broadcasters, advertisers, universities, media, ratings firms, government and equipment manufacturers, among others, are all involved in both the creation of audience ratings and their use. In Chapter 3 we will see in more detail how ratings intellectuals crossed over from university to industry. The broadcast audience-ratings convention, or standard, on the authors' historical analysis, has had several important components. The convention:

- 1 is based on exposure as the key measurement;
- 2 must appeal to the inherent correctness of the measurement;
- 3 uses a probability, statistical, sample;
- 4 delivers a 'single number';
- 5 is syndicated to reduce costs to subscribers;
- 6 has generally been third-party (the ratings firm, except in the case of Crossley, has been separate from the advertisers or the stations);
- 7 is expected to work in the public interest (that is, accurately to represent the public audience).

Unlike McNair and Anderson, Archibald Crossley did not have the same control over his destiny even though broadcasters and advertisers could and did put pressure on the two Australian broadcast ratings companies over time. By the 1940s the principle of 'current opinion/view/actions' based on probability sampling principles as the basis for future estimates was established in the United States and Australia. The broadcast ratings built by Crossley drew on the credibility of opinion polling and came to occupy a similar position as regular repeatable information to be used predictively (Beville 1988).

The early ratings methodologies used families and social class definitions to segment the audience. McNair and Anderson even ran different definitions of the audience up until 1963. The age categories for television were also different. The aim of this methodology was to produce relatively simple data for the production of figures for buying and selling data.

The United States and Australia did not differ significantly on this side of the methodology – the 'single number' principle governed media buying. However, as

discussed, the audience ratings differed significantly. The history of audience ratings is not therefore simply a history of the statistics of ratings. We have seen that there are significant differences between the American and Australian audience-ratings conventions in practice even though they share a very similar overarching convention. The United States for much of its history has had a 'competitive' model in which Nielsen came to dominate as a free-standing company, now owned by private equity shareholders. Australia has had a succession of models: first competing providers with different methodologies pursued with integrity; and then a broadcaster-defined ratings contract which ratings companies tender for and in which broadcasters call the shots, with ratings owned by the networks but subject to auditing. In Australia the broadcast ratings companies themselves went out of their way to make public the 'black box' of ratings and how the methodologies work.

We can see glimpses of the convention in operation when the audience and media technology context changes or when the methodology is questioned and the measure – exposure – or sampling become an issue. We can see it when the currency appears to be failing or when proprietary control of audiences and their information is queried or problematic. The gradual erosion of the standard broadcast model and the evolution of niche broadcasters alongside a variety of platforms including YouTube demonstrate an increasingly more complex and mixed set of televisual arrangements and business models. In this context audience measurement is becoming more complex and recalibrating its instruments to meet the new challenges of covering more and also different kinds of viewing. At the same time the relation between the audience market defined by advertising and the content market defined by sales of media product including cable TV, DVD and downloads of programmes is entering a new and radical phase.