Abstract

The much touted transformations of television not only towards more television but more ways of accessing television presents particular challenges for ‘business as usual’ in audience measurement—particularly with respect to the trade-offs involved in establishing and maintaining broadcast ratings conventions. 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 both becoming more complex and needing to recalibrate its instruments to meet the new challenges of covering not only more but 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 pay-TV, DVD and downloads of programs is entering a new and more mixed phase. What the authors want to do in this paper is to follow the debates and controversies within audience measurement in Australia and the USA, especially as they relate to the future of exposure as the standard measure.

Introduction

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, 2008).

Helen Crossley, an icon of the American Association for Public Opinion Research (APPOR), was personally present at the first revolution of audience ratings — its creation. Her father Archibald Crossley created them at the beginning of last century for the analysis of radio audiences (Beville 1985). Helen Crossley and her father went on to help Gallup establish public opinion polling.

Archibald Crossley measured exposure in his radio ratings analysis and exposure has become the standard way — the convention — for measuring broadcasting ratings. Before audience ratings there was little systematic research on audiences. Exposure measures whether an audience has been exposed to a program and for how long. It does not measure whether the audience likes a program or indeed whether it is engaged.
But with a many media and fragmented audience market there appears to be an erosion in confidence in a ‘single number’ exposure measure based on a random sample (Napoli 2003). To what extent can the broadcast ratings retain its currency as a single standard? Will the interests of the different parties to the audience ratings convention become sufficiently divergent as to require different conventions co-existing in the same market?

To address these questions the authors draw from their Australian Research Council (ARC) project *The Emergence, Development and Transformation of Media Ratings Conventions and Methodologies, 1930–2008*. This research involves interviewing those involved in the historical development of audience ratings, those involved in its contemporary formations, as well as archival research at audience ratings company or national archives in Australia and overseas.

**Decline of the audience product**

Media ratings systems provide an economic foundation for advertiser-supported media. Consequently the nature of the audience measurement process affects the structure and behaviour of media companies and regulators alike. So when the techniques and technologies of the ratings change these changes can as Napoli (2003) observes have ‘a significant effect on the economics of media industries (because these changes can affect advertiser behaviour), the relative economic health of various segments of the media industry, and the nature of the content that media organisations provide’ (Napoli 2003, 65).

Although changes to the ‘ratings convention’ governing audience measurement can be disruptive, these changes are driven by the inevitable gap between the measured audience and the actual audience for a service or programs. With the advent of a more diverse and fragmented media environment and fractionated audiences increasingly demographically defined this gap has become even more evident with the validity of ratings as currency for buying and selling media being challenged in the USA. Napoli (2003) suggests that this is leading to a decline in quality and value of the ‘audience product’ — data on who is watching when — because of changes in technology and audiences. The provision of reliable third-party syndicated and customised audience measurement technology for the production of ratings, however, remains essential to good media management nationally and internationally. As Australia moves into a more fragmented and differentiated multichannel broadcasting environment these issues will become more important.

Napoli (2003, 181) identifies in his analysis an empirical trend of separation of audience and content markets, which historically have been connected. The audience market and ratings have traditionally informed programmers and advertisers on what content to provide and where to target advertising. With the emergence of media technologies such as satellite radio, interactive television, digital cable, and digital television broadcasting there is a focus on audience payment for content rather than advertising support (2003, 180). In the new media environment, more personal information is becoming available about individuals but less information is becoming available about audience preferences. It becomes more expensive, therefore, for media producers and distributors to find audiences, to discover their viewing or listening preferences and to deliver content across a range of different media.

In order to understand the decline in the audience product it is important to understand how the statistics and standards of audience ratings have developed over time.

**The Old Regime**

The early ratings methodologies in what the authors will call ‘the old regime’ used families and social class definitions to segment the audience. McNair and Anderson 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 Andersons did not use age breakdown in reports until 1961 and then
reported on the age ranges 1–15, 16–24, 25–39, 40–54, 55–99. McNair also did not use age categories until 1963 and then reported on the age categories 0–11, 12–19, 20–35, 36+.

By 1970, however, McNair and Anderson age categories for radio and television became uniform, with McNair and Andersons using the same age categories 10–17, 18–24, 25–39, 40–54, 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 began in 1959 in one of McNair’s January/February reports when he started using Class AB (without separating them, as the company had done in previous reports). In modern marketing circles there appears to be a visceral reaction to continuation of the use of these 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)

What Joseph is touching on, of course, is not the statistics of audiences but on the standards by which we agree to measure audiences and what counts as an audience. The old regime in audience ratings may appear to be a simple model, but it is in fact a complex agreement between many often conflicting parties on what should be counted and how frequently it should be counted. The need for standards or conventions governing measurement is not knew, as Stigler (1999) has pointed out.

Stigler (1999) 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 was a 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 made individual weighing 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).

This history of statistics is not the same as the history of standards. By extension, the history of audience ratings is not simply a history of the statistics of ratings. The relatively simple definitions of audiences in the McNair and Anderson surveys are not simple because early audience ratings
experts had no idea on how to create more complex definitions of demographics. The measures of two competing ratings systems were accepted because both were, in fact, a default audit, one system acting as a reference to the other. By the 1940s modern statistical, probability, sampling was also in place. The definitions of the audience also had, underpinning them, assumptions about how many people might be surveyed or indeed included in the sweeps to represent whole nations.

The importance of the difference between measurement (statistics) and standards (convention) in audience ratings becomes particularly noticeable when there are controversies over ‘accuracy’. For example, the 1963 Congressional committee into broadcast ratings established in the US came about because what was presented as results was not matched by an understanding of the underlying standards governing their use.

The hearings suggested that the illusion of exact accuracy was necessary to the ratings industry in order to heighten the confidence of their clients in the validity of the data they sell. This myth was sustained by the practice of reporting audience ratings down to the decimal point, even when the sampling tolerances ranged over several percentage points. It was reinforced by keeping as a closely guarded secret the elaborate weighting procedures which were used to translate interviews into published projections of audience size. It was manifested in the monolithic self-assurance with which the statistical uncertainties of survey data were transformed into beautiful, solid, clean-looking bar charts’ (Bogart, 1966, 50)

The US Congressional committee also found that, ironically, some of the audience ratings companies were completely bogus and not even conducting surveys. The US committee was the first serious independent analysis of audience ratings methodology and conventions. It is also an important cultural marker that shows us that *more measurement* does not necessarily mean a better standard. Like the barons watching the weighing of the coins in the Trial of the Pyx, it is in audience ratings the agreements on the approximations that matter more than finer and finer modifications to decimal points or having more information about audiences.

The hearings had a dramatic effect on the audience ratings industry in the United States. Gale Metzger was a young Nielsen employee when the hearings started and later went on to set up his own company and a competing system to Nielsen. He recounted the fallout:

And in those Congressional hearings it was determined that some people, some measurement companies were bogus, were making up numbers in effect, and others were not doing that but they weren’t quite all they represented themselves to be. And while there were a number of companies that were looked at, maybe 10 or 12, some went out of business after that. But Nielsen was clearly a focus, because Nielsen was the primary game in town. And they really went over Nielsen in a very harsh way and determined to the embarrassment of many of their clients, that Nielsen was not being totally forthright in their disclosure of their methods. Not that they were doing anything evil but they were not telling the clients what they were doing or where they made compromises (Gale Metzger 2008).

The Congressional hearings led to the establishment of an independent ratings council in recognition that audience ratings were not just a commercial matter but in the *public interest*. The hearing of course also led to a media industry much more sensitive to how audience ratings worked. Australia, interestingly, avoided the kinds of controversies that arose in the United States precisely because the ‘standard’ was more robust than in the United States (Balnaves and O’Regan 2008). The Anderson and McNair systems were transparent. The stability of the ratings convention has also, of course, a direct impact on business models for the delivery of ratings.

While measurements and standards are different things, changes in the technology for the collection of audience ratings can affect the audience ratings convention — the standard. The Peoplemeter was a new method for the measurement and collection of data.
The Transitional Regime

In the old regime advertisers or broadcasters could monitor simple audience categories within a narrow range of media. With the emergence of television there was an additional demand for the provision of audience ratings on a more frequent basis. The Peoplemeter emerged as a way of capturing from households viewing information nearly all the time and transferring it directly from a set top box on the television to a computer. Diaries are still used for the Australian radio market but Peoplemeters for the most part dominate television. Ironically, the Peoplemeter did not radically alter the classifications of the audience but it did introduce greater expectations of speed of delivery of the results of audience ratings. The old regime in radio and television employed ‘sweeps’ — maybe a one week sweep in Perth and a three or four week sweep across the year for Sydney, for example. Electronic audience ratings potentially allowed collection of the data all the time and for users of ratings to manipulate the results using software packages created for this purpose.

But it was not the Peoplemeter as measurement that has created problems with the modern audience ratings convention. Advertisers want exposure to audiences through 30 second messages because it is assumed that there is an audience that is big enough to make it worthwhile spending large amounts of money. But the major free to air television networks in the world, and especially in the United States, have experienced a dramatic downturn in audiences. There is a range of explanations for this:

- **Conduit multiplication and audience fragmentation**: With the advent of more television channels, more radio channels, thousands and if not millions of internet sites, mobiles phones, and so on, there is now an assumption of fragmentation of audiences. People have far more options available to them.

- **Time shifting**: People can use their time-shifting technology, like the Personal Video Recorder (PRV) or TiVO to by-pass advertisements and make monitoring of their exposure difficult.

- **New uses for television**: DVDs, game stations and other television-attached-media reduce attention to traditional broadcast television.

- **Ambience**: Traditional broadcast television like radio is becoming a background medium with people doing other tasks while watching television.

- **Shift to below-the-line advertising**: There is a trend towards a more accountable media spend because of conduit multiplication and fragmentation. Advertisers are demanding that there is evidence that their audiences are in fact exposed to advertising in a fractured media market.

Exposure in the transitional regime appeared to have lost or reduced its value. Some exposure like advertisements during the Super Bowl still deliver critical mass. But how do other advertisers gain affordable access to audiences in the new media environment? This is what Napoli (2003) means by decline in quality and value of the ‘audience product’. There appears to be no guarantee that a critical mass of audiences can necessarily be delivered or that audience data can effectively capture who is watching or reading or listening and when.

The Contemporary Moment

The old regime measured very few media and in the transitional regime audience ratings found itself with many media and many audiences. The solution in the contemporary moment has been to measure absolutely everything that people do. Nielsen in 2007 used the term A2/M2 to describe this new strategy. A2/M2 literally means anytime anywhere measurement. A2/M2 collects information everything from Web surfing through to mobile phone use, all day and every day (Story 2008).

Nielsen created 17 panels to cover the full range of media activities. Table 1 in the appendix shows how extensive this process is in the United States and internationally. Each household in a panel can participate for a number of years. The television panel, for example, can participate for two years.
In 2008 Nielsen cancelled a joint Arbitron venture *Project Apollo* that surveyed the radio and television behaviour of 5,000 households. The joint venturers found that their clients could not afford the data and participants in the panel withdrew as the number of things monitored in their household grew people were less likely to participate as the number of media monitored grew.

Nielsen of course is not the only audience measurement company. Internationally, TNS offers different types of ratings information, for example summarizing preferences of DirecTV’s satellite subscribers. comScore competes with Nielsen on evaluating websites and M:Metrics competes with Nielsen on mobile phone tracking. In Australia AGB Italia collects the television audience ratings; Nielsen collects the radio ratings data. Foxtel and regional pay-TV group Austar, in conjunction with MCn, announced in 2008 the launch of a new digital television audience measurement system (AMS) that would, the group argued, be the largest measurement system in Australia, providing viewing results from a panel of 10,000 Australian subscription TV homes (Bodey 2008). The system is designed to give the Pay-TV group information on how Australians are adapting to the digital TV environment, the acceptance of the new standard definition and high definition multi-channels and trends in time-shifted viewing. ‘No one was willing to admit it was a competitive service to OzTAM’s, which measures free-to-air and pay-TV viewing. Rather the new system is seen as complementary.’ (Bodey 2008). Leaving aside how the AMS actually works — how it measures and what it measures — AMS almost seems to represent a return to the days of McNair and Anderson, with competing systems providing information to a skeptical market.

The interesting thing, however, about the contemporary moment is that there is now measurement of many media, from use of mobiles through to television, but there is no agreement on a currency that covers all the measurements. In the old regime audiences were relatively stable and there were few media. An audience ‘rating’ reflected an audience that had critical mass and was, hopefully, for example watching or listening to the medium. What has occurred is that the media industry has decided to do more measurement without a corresponding standard — statistics without a convention. We have, therefore, a revolution in measurement — more measurement of human media behaviour than ever undertaken before — but no revolution in the audience ratings convention (a currency that can conveniently summarize all audiences in all media).

The other major trend in the contemporary moment is to buy large audiences. Google bought YouTube in 2006 for $1.65 billion. At that time YouTube had over 100 million videos viewed every day and over 72 million individual visitors each month. Google has made its money from small text advertisements that you see next to Google search results. These advertisements deliver most of Google’s US$16.6 billion in revenue. This is a different model from traditional audience measurement and delivery of advertising to audiences. Google also plans to sell advertisements that appear inside videos on. It has purchased smaller companies that provide different types of advertising services or software for use in an online video environment precisely to build this market.

Google has also started partnering with video content providers. AdSense for Video, a Google purchase, for example, helps Internet publishers to make money from their own videos. AdSense for Video provides advertisers with a choice between a video or a text advertisement on a portion of the video viewer. But Google has gone further and calls these providers ‘participating user-partners’. ‘Participating user-partners will be treated as other content partners and will have the ability to control the monetization of the videos they create. Once they’ve selected a video to be monetized, we’ll place advertising adjacent to their content so participating user-partners can reap the rewards from their work. Before you start fantasising about the hundreds of pennies you’ll surely make from that video of Granny’s teeth falling out, you should know that the program is not open to all, just a select invited few’ (Google, 2007).

AdSense for Video has started with 20 partners including video aggregator sites Revver and Blip.tv and video technology companies Brightcove and YuMe. These companies deliver video content and advertisements to their own networks of Web publishers. The long term strategy is of course to
include as partner the *individual user* who wants to make money from the video that they produce for the Internet. Google is anticipating what has been called co-creation — where users and producers of media content become much closer (Banks 2002). Co-creation has most commonly been associated with games fans and their participation with games content creators (and the role of mods, skinners and mappers for example).

John Banks, who worked for Auran, a now defunct multimedia games company, reflected on the practice.

To assume that gamers and the corporation stand in an exterior relationship — on one side the producer and on the other side the consumer or audience — would be a mistake. For a start, many — if not most — of the designers, artists and programmers who work at Auran are also themselves game fans. Many of them participate actively in the fan networks that form around particular game titles and genres on the Internet. These gamer-employees bring with them skills, practices and cultures that are very much influenced by their activities as gamers. These practices in turn influence and impact upon Auran at the most basic level of workplace practices. (Banks, 2002, 195)

Google's approach is to buy massive audiences and to experiment with them. Google does not need panels or samples because it has either a record of what its users do, or a site like YouTube, where the audience gathers. That audience in itself has economic value.

At first sight, it looks like there is indeed a revolution in audience ratings — exposure as a metric has been eroded and alternative means of assessing and delivering audiences is underway. However, this might also be far from the truth. The demand for a more open view of what was happening with audiences has complex origins. For example, in 1908 in Britain the *Observer* proceeded for damages against the Advertisers’ Protection Society, formed in 1900 to ensure that manufacturers who paid for advertising would get better value for their money. Its first action was to invite newspaper owners to divulge circulation figures and when they refused they provided its members with private estimates for leading newspapers and journals. The Society had published an estimate of 5,000 for the *Observer*’s circulation. The *Observer* could show that its net sales were over 80,000, but the Society was acquitted, ‘because, in the struggle for circulation which then consumed Fleet Street, there was no basis for more accurate estimates.’ (Harris & Seldon, 1954 p. 14.) After this court case newspapers began publishing their circulation figures. The Society then embarked on a campaign for independently audited figures that led to the establishment of the Audit Bureau of Circulation in 1931. The Audit Bureau is itself a convention formalising the need for public and market accessibility to audience data.

The need for a standard or a social convention governing the measurement or reporting on audiences has never simply been a matter of private organisations better measuring audiences or keeping proprietary control of their own audiences. It has, interestingly, been a demand for transparency for both the commercial and the public good. The US Congressional hearings on audience ratings were extraordinary not only because of the pressure from commercial owners who woke up to their importance, but also because of the recognition that the conventions governing audience ratings were also in the public interest.

**Conclusion**

In the United States there has been extensive historical study of the development of audience ratings (Beville, 1985; Webster, Phalen & Lichty, 2000). However, there has been little historical research on the audience ratings conventions. The audience ratings convention — or standard — the authors argue — has had several important components. The convention:

1. Has exposure as the key measurement.
2. There must be an 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.
7. Is expected to work in the public interest (that is, accurately represent the public as an audience).

Audience ratings involve a social convention where all parties have a common interest in a coordination rule, none has a conflicting interest, and none will deviate lest the coordination is lost (Miller 1994). There can be little doubt that the traditional audience convention or standard is under scrutiny with no real agreement on what happens next. In the absence of any agreement on new standards, the traditional measures remain in place and other methods and means are being either trialled or adopted. This does mean, of course, the end of exposure as the standard way of measure audiences but it does mean that we will see a new audience convention emerging in the near future.

APPENDIX

Table 1: Nielsen panels

<table>
<thead>
<tr>
<th>Panel Description</th>
<th>Year</th>
<th>Households in panel</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Peoplemeter</td>
<td>1987</td>
<td>14,000</td>
<td>US</td>
</tr>
<tr>
<td>Local Peoplemeter</td>
<td>2002</td>
<td>600–800</td>
<td>US in 13 markets</td>
</tr>
<tr>
<td>Set Metered Markets (Electronic boxes that track viewing but information about the view is in a diary)</td>
<td>1959</td>
<td>21,000</td>
<td>US</td>
</tr>
<tr>
<td>Hispanic People Meter Supplement</td>
<td>1994</td>
<td>270</td>
<td>US</td>
</tr>
<tr>
<td>Out of Home (Measures TV viewing at work, bars, airports, and so on, using sounds from the programmes that are recorded automatically by special mobile phones)</td>
<td>2008</td>
<td>4,700</td>
<td>US</td>
</tr>
<tr>
<td>Homescan Global (Purchasing behaviour)</td>
<td>1988</td>
<td>135,000</td>
<td>27 countries</td>
</tr>
<tr>
<td>Homescan US consumers</td>
<td>1988</td>
<td>125,000</td>
<td>US</td>
</tr>
<tr>
<td>Homescan Hispanic</td>
<td>2007</td>
<td>2,500</td>
<td>US/Latin America</td>
</tr>
<tr>
<td>FANLinks (Cross-references Homescan with their fan interests)</td>
<td>2005</td>
<td>50,000</td>
<td>US</td>
</tr>
<tr>
<td>Project Apollo (Multimedia consumption and purchasing – now cancelled)</td>
<td>2006</td>
<td>5,000</td>
<td>US</td>
</tr>
<tr>
<td>Nielsen BookScan (US book industry data)</td>
<td>2001</td>
<td>12,000</td>
<td>US Booksellers</td>
</tr>
<tr>
<td>Your Voice (Online community for opinions)</td>
<td>2000</td>
<td>500,000+</td>
<td>Global</td>
</tr>
<tr>
<td>Nielsen Mobile Bill Panel (Activity on mobiles)</td>
<td>2005</td>
<td>20,000</td>
<td>US Mobile bills</td>
</tr>
<tr>
<td>Hey! Nielsen (Website where users rate TV shows, movies, and so on)</td>
<td>2007</td>
<td>30,000</td>
<td>US</td>
</tr>
<tr>
<td>NetView Panel and MegaPanel (Offline and online audience and market research)</td>
<td>1997</td>
<td>475,000</td>
<td>US and 9 other countries</td>
</tr>
<tr>
<td>Pine Cone Research (Product and concept surveys)</td>
<td>1999</td>
<td>173,000</td>
<td>Global</td>
</tr>
<tr>
<td>The Hub (Former members of other panels who allow Nielsen to track them)</td>
<td>2008</td>
<td>1,000</td>
<td>US</td>
</tr>
</tbody>
</table>

Source: Story (2008).


Acknowledgements: The authors are grateful for the assistance from Helen Crossley, Ian Garland and Gale Metzger and their insights into the history of audience ratings. The authors also wish to thank Mark Neely, part of a family involved in the development of early Australian ratings system, for access to Nielsen’s Australian archives. The interviews and research were made possible by funding from the Australian Research Council (ARC) project *The Emergence, Development and Transformation of Media Ratings Conventions and Methodologies, 1930–2008*. 

---

**REFERENCES**


Acknowledgements: The authors are grateful for the assistance from Helen Crossley, Ian Garland and Gale Metzger and their insights into the history of audience ratings. The authors also wish to thank Mark Neely, part of a family involved in the development of early Australian ratings system, for access to Nielsen’s Australian archives. The interviews and research were made possible by funding from the Australian Research Council (ARC) project *The Emergence, Development and Transformation of Media Ratings Conventions and Methodologies, 1930–2008*. 

---