

APPROACHES TO MEASURING QUALITY IN BROADCASTING

In 2006 NZ On Air commissioned an extensive research project to examine approaches to measuring quality in broadcasting. The report details and compares measures used in a variety of OECD countries, and the resulting discussion and considerations are contained below. The full report is available from NZ On Air on request.

DISCUSSION AND CONSIDERATIONS

The report has examined an extensive range of literature relating to six dimensions of broadcasting quality (media, production, content, distribution, audience, impact), identified a number of quality sub-categories within each one and outlined different potential approaches to operationalising and measuring each one.

It has also outlined various approaches to broadcasting quality measurement which have been (or are in the process of being) implemented among a range of public broadcasting institutions across seven different countries (Aotearoa/ New Zealand, Australia, Belgium, Canada, Ireland, the Netherlands and the UK).

It has been noted that the conceptualisation of various aspects of quality is to some extent a separate consideration from the operationalisation of those conceptions to render them amenable to measurement and assessment. Accordingly, it was noted that insofar as some aspect of quality in the broadcasting process is identified as desirable, the determination of whether or not that aspect of quality is present or absent from programmes or the schedule may require looking at other aspects of the broadcasting process.

For example, the craft and technical skill in production cannot be validly discerned only by investigating the institution and its creative personnel without any reference to the programmes they actually make. Likewise, examining the informational value of a programme cannot be validly determined by simply analysing the number of words or images it contains without consulting the audience to see if anyone paid attention and, if so, whether it was understood.

Although there are a range of useful and innovative approaches to measuring quality evidenced in the various national examples discussed, it is also apparent that some systems exhibit limited evidence of a coherently theorised and valid articulation between the specific data being collected and the aspect of broadcasting quality ostensibly being measured. It is therefore important to make such links as explicit as possible (while recognising the limitations of whatever methods are adopted).

In this respect, in any endeavour to develop a quality measurement system, it is helpful to ensure congruence between three key elements:

- o *Conception of quality* (i.e. what dimension or sub-category of quality is to be measured/ assessed?)
- o *Focus of measurement* (what data needs to be sourced in order to perform the assessment)
- o *Method of data collection* (how is the data going to be accessed and recorded; e.g. is audience opinion data collected via a rating-scale survey?)

It is also important to consider the *mode of implementation* (i.e. how the measurements of quality are to be interpreted and actually put to practical use), but since this consideration primarily concerns broader policy and institutional issues which are specific to context, this will be discussed separately afterwards.

Attempts to classify some of the various approaches to defining, operationalising and measuring different aspects of broadcasting quality are listed in the table below. This is not a definitive list, and the indication of any particular method should not be taken as an endorsement of its intrinsic validity. In practice, it may be advisable to use parallel methods and triangulate findings, but such possibilities cannot be indicated on this table. Some of the approaches which have been outlined (particularly in the national examples) are bespoke in nature and designed/ adapted to the particular purposes of individual broadcasters or regulators (and therefore not automatically applicable elsewhere).

A typology of broadcasting quality measures by dimension/ conception, focus and method.

Quality Dimension	Quality Conception	Focus of Measurement	Method of Data collection
Media	Total Value	Audience attitude (willingness to pay)	National Voting pricing scale survey (social benefit frame)
	Consumer Value	Audience attitude (willingness to pay)	Gabor Granger pricing scale survey (consumer frame)
	Citizen Value	Audience attitude (willingness to pay)	Total Value minus Consumer Value (as above)
	Plurality	Broadcasting system/ institutional diversity	Identification of number and range of broadcasting operations with distinctive identities and functions
	Benchmarking	Broadcaster operations/ production/scheduling decisions	Interviews with broadcasting managers/ Evidence of correlations (+ or -) between broadcaster activities.
	Efficiency	Accounting records/ Audience ratings Programming outputs	Input-output ratios of cost-to programme or cost-per viewer/ hour (comparison with other broadcasters)
Media		Audience opinion (quality or	Quantitative audience ratings of

	Functional Quality	approval) OR Audience attributions of functional characteristic	approval or quality or delivery of key functions, OR Qualitative audience ascriptions of functions/ key characteristics to broadcaster
Production	Technical standards	Professional broadcaster perceptions/ Audience perceptions	Peer reviews, professional awards/ audience ratings of technical standards in programmes
	Craft/artistry	Professional broadcaster perceptions/ Audience opinion	Peer reviews, professional awards/ Audience identification or ratings of craft/artistry (creativity/ innovation, etc)in programmes
	Ethical Quality	Programme content/ Regulator of broadcasting standards	Absence/ minimisation of formal complaints or breaches of programming standards regulations.
Content	Overall Content Quality-	Audience opinion of schedule/ channel	Quantitative audience ratings-surveys OR Qualitative audience assessment of schedule characteristics (focus groups)
	Individual Programme quality	Audience opinion of programme quality/ Audience perceptions of genre characteristics/ Evidence of audience behaviour/ response/ Mode/ of audience engagement	Quantitative audience surveys to rate quality/ appreciation [see reception]/ Qualitative audience attribution of relevant characteristics to genre (e.g. seriousness/ impartiality in news) Qualitative evidence of relevant responses (e.g. did comedy produce laughter?) [see reception] Assessment of appropriate register/ tenor of programme's engagement with audience.
	Balance	Audience perceptions Content of programme in terms of framing, viewpoints, demographic political representation (either in relation to schedule or programme)	Audience quantitative ratings of balance and representation of their viewpoint/ culture/ ethnicity/ Content analysis of presence of demographic/political groups in relation to census ratios and political spectrum./ Discourse analysis of media framing of issues and portrayal of different demographic groups
	Content Inclusiveness	Relevant audience-segment appeal/ perception/ Programme production process.	Breadth of audience demographic segments who identify themselves as the intended audience for programmes Interviews with programme producers about consideration of minority interests in programmes and schedule.
	Vertical Diversity	Programmes genres within single channel schedule (at different points in time for diachronic analysis)	Quantitative count of different genres in schedule over a given period- possibly including balance and inclusiveness factors. Relative entropy statistical distributions of genres across schedule.

Content	Horizontal Diversity	Programme genres across all schedules (at different points in time for diachronic analysis)	Quantitative count of different genres in schedule over a given period- possibly including balance and inclusiveness factors./ Relative entropy statistical distributions of genres across schedule. Deviation Index (to calculate variation between two channels)
	Synchronous Choice	Programme schedules across all channels broken down into 15-minute segments	Programme Choice Index to identify synchronous choice of between different genres at any point in time.
Transmission	Universal Access	Signal range and demographic/ geographic data/ Statistics about diffusion of signal reception equipment	Calculation of proportion of population who can potentially receive signal/ Calculation of proportion of population who can actually access programmes.
	Signal Quality	Technical measures of signal noise/ distortion	Reception quality across different regions, identification of 'radio shadow' etc.
Audience	Ratings	Viewer/ Listener statistics Demographic statistics	Calculation of the proportion of the total population of the audience listening to or viewing a programme for a minimum period.
	Share	Viewer/ Listener statistics Demographic statistics	Calculation of the proportion of the overall audience across all channels which listening to viewing a programme for a minimum period.
	Reach	Viewer/ Listener statistics Demographic statistics	The proportion of the total population watching or listening to any programme for a minimum period from a particular channel or source at least once over a specified time-period .
	Demographic inclusiveness	Viewer/ Listener statistics Demographic statistics	The range/diversity of different demographic audience segments exhibiting a minimum proportional rating as a result of watching or listening to any programme for a minimum period from a particular channel or source at least once over a specified time-period.
Reception	Audience Enjoyment/ Appreciation/ Satisfaction	Viewer/ Listener statistics Demographic statistics/ Audience perceptions & opinions.	Quantitative surveys of viewers'/ listeners' ratings of various aspects of the schedule or programmes./ Qualitative evidence of programme enjoyment and appreciation in focus groups/ Indirect evidence of relative appeal of programmes by ratings and share.
	Audience Engagement	Viewer/ Listener statistics Demographic statistics/ Audience perceptions & opinions.	Qualitative or Quantitative assessment of the level of audience attention/ engagement while viewing or listening, including absence of channel-switching, reports of

Reception			'appointment viewing', etc.
	Social Function/ Impact	Viewer/ Listener statistics Demographic statistics Audience perceptions, opinions, Evidence of audience attitudinal or behavioural response Shifts in social attitude or behaviour and broadcasting content	Qualitative surveys or focus groups focusing on viewers'/ listeners' reception experiences and responses to programmes, (particularly those involving changes in opinion, attitude or behaviour commensurate with intended social functions). / Number of reported relevant responses in ratio to audience share/ ratings./engagement. Social scientific investigations of changes in public attitude or behaviour over time related to or attributable to broadcasting.

Contextual considerations

It is clear from the literature that there is no way of defining quality in broadcasting in terms of a singular ideal 'Platonic' conception which can be applied unambiguously to all cases and operationalised/measured according to an objective scale. Quality is discerned by applying a set of values or normative criteria to the different aspect of broadcasting in question. Insofar as different stakeholders assume different values and prioritise different aspects of broadcasting, the conceptions and criteria for measuring broadcasting quality are manifold and open to contestation¹. One size does not fit all.

Moreover, the academic exercise of discerning and classifying different notions of broadcasting quality and valid methods of measurement is potentially distinct from the contextual political, economic and institutional pragmatism which may prioritise particular notions or measures for reasons of convenience, affordability, legal obligation, policy legitimisation, or instrumental expediency. What looks like a good idea in the text book may have ramifications that extend beyond the theoretical in practice.

The contextual aspects of different approaches to measuring quality in broadcasting are in some respects separate from the tasks of classification and operationalisation that have been the primary focus of discussion. However, they need to be acknowledged, not only because of what they might imply for the practicalities of policy formulation, but also because they have some bearing on the normative assumptions that underpin the conceptions of quality in question.

Developing a quality measurement system for broadcasting is intrinsically purposive. The complexity and resources involves means that such systems are never developed just for their own sake and an end in their own right, but for ulterior motives. If those motives are only cynical and instrumental, then perhaps conceptual clarity and methodological validity will be a

¹ Indeed, it is worth noting that some interests may be served by avoiding having any system for measuring quality.

secondary concern. However, if the aim is to measure quality in broadcasting in order to improve quality in broadcasting (both by celebrating and emulating what is best and acknowledging and avoiding what is poor) then there is a need to consider how to interpret and make use of such findings.

To digress from broadcasting momentarily, it is worth considering an amusing (but pertinent) literary anecdote. In the late Douglas Adams' well-known science fiction satire, 'The Hitch-Hikers' Guide to the Galaxy', there is a tale of how an ancient civilisation constructed an incredibly powerful computer, 'Deep Thought' to provide the answer to the ultimate question about 'life, the universe and everything'. After several million years' of calculation, the by-now highly expectant descendants of Deep Thought's creators gather in great anticipation only to discover that the epistemological legacy of their ancient ancestors is the rather mundane and wholly unsatisfying answer, 'forty two'.

Trying to define/measure quality in broadcasting is perhaps less grand in its cosmic purpose, but the moral of the story is that there is little point knowing the answer to such questions unless one knows what to do with it. This is a serious point, because for any system of quality measurement to serve a useful function, there must be processes in place for making sense of any data thereby generated and ensuring that it is used to enhance broadcasting practices in the manner desired. For example, if an audience approval- rating survey reveals that two channels receive an average of 5 on an 11-point scale, does that mean they are equally good, equally bad, or both mediocre-to average? If one channel's score of 5 was generated from a cluster of ratings awarding 4, 5 and 6 (low standard deviation) and the other's score of 5 was generated from half the audience awarding 1 or two but the other half awarding 9 or 10 (high standard deviation), then what basis might there be to say that one has outperformed the other? Without benchmarks or an established evaluative framework, such data is largely meaningless.

This also raises many questions about institutional accountability. For example, are the findings collected by an external regulator which compare the performance/quality ratings against formal benchmarks and imposes regulatory or financial sanctions if requirements are not met? Or are the findings subject only to internal, informal review by the broadcaster itself? Bardoel (2003) and Betzel & Ward (2004) differentiate between systems of performance measurement which make broadcasters accountable to i) themselves (i.e. self-assessment), ii) the market (e.g. withdrawal of advertising if ratings fall), iii) the state/government (e.g. ministerial/legal obligation), iv) the public (civil society, e.g. citizen's juries, audience councils), v) other professional media bodies (e.g. press councils), or vi) independent authority (non-government regulators).

Coppens (2005), meanwhile, distinguishes between broadcasting performance/accountability systems which are formal/ binding (e.g. quotas in PSB service contracts) or informal/ non-binding (e.g. qualitative audience data). The nature and magnitude of response is also a key factor. As Bardoel (2003) suggests, if the consequences for under-performance are severe and/or mandatory (such as loss of licence or public funding), then this may preclude using the assessment system to foster progressive change over time.

Moreover, if harsh penalties or great rewards are at stake, then it is incumbent upon the assessment system to be valid, transparent and legally defensible. Even if that is achieved, if too much depends on attaining satisfactory performance (or avoiding underperformance), this may encourage instrumental box-ticking attitudes, whereby the broadcaster pays more attention to operationally maximising its performance in terms of the specific quality measurement criteria than to attaining the aspect of quality performance these are intended to index. If the stakes are high, then assessment systems are liable to promote exactly what they measure. On the other hand, if assessments are merely informal and internal affairs, carrying no potential reward or consequence, then there is a risk of indifference and empty ritualisation² and a lack of incentive to improve.

The detail and number of quality/ performance measures is also a factor here, since over measurement may produce the opposite of the intended results (see Foster et al, 2004); the effort required to ensure a wide variety of indicators/ criteria are satisfied may result in the broadcaster being obliged to engage in instrumental box-ticking and diverting resources away from the production of quality programmes towards compliance with the assessment regime. The additional validity ostensibly gained through the use of multiple indicators may therefore reach a point of 'diminishing returns' if they distract the broadcaster from core business.

The potential end use of quality assessment data therefore needs to be considered as a separate but nevertheless important aspect of any quality measurement typology. The characteristics of the mode of implementation might be described in terms of the following:

- o Agent of Measurement: Which institution(s) conducts the measurement of quality/ performance?
- o Agent of Assessment: Which institution(s) conducts the interpretation of the findings?
- o Agent of Response: Which institution(s) are obliged to respond to the findings and assessment?
- o Access to findings: Which institutions have access to the findings? (e.g. are they private to the broadcaster/ regulator, or made publicly available?)

² Bardoel's term.

- o (For systems using multiple assessment measures) Which assessment criteria/ findings take priority in the overall performance evaluation, and does the prioritisation of one criterion necessitate the neglect of another (zero-sum effect)?
- o Are the findings linked to formal rewards and/or sanctions (and can findings/ rulings be contested/appealed?), or are they informal/ non-consequential?
- o Are any formal rewards/sanctions automatic and final (summative) or are they contingent on response and/or cumulative over time (formative)?
- o Does the system promote optimising (maximising performance, possibly taking some risks) or satisficing (aiming to avoid under-performance, risk aversion)?
- o Does the system have any systemic/ ecological impact on the relations between broadcasting institutions (e.g. increased demand for local minority programmes) and if so is this intentional or undesirable?

It is not the aim of this report to specify either which systems of quality assessment or modes of implementation are most appropriate to specific New Zealand broadcasting institutions. However, by way of observation, it is worth noting that many local media professionals appear sceptical of external assessment either by experts/academics or by officials (and especially by politicians). That said, where public money is made available to support public broadcasting outcomes, it is incumbent on the state to ensure that those outcomes are, in reasonable proportion to the level of input³, realised.

While some media professionals might argue -perhaps rightly- that the existing checks and balances already constitute a sufficiently onerous impediment to accessing public subsidy to deter all but the most dedicated and professional, that is not a reason to object to a review of quality measurement and accountability systems if there is potential to make them more manageable and/or more valid. That does not justify the imposition of any and all accountability measures, but insofar as systems of quality measurement might be introduced which have the potential to increase the likelihood of securing desired policy outcomes, then the grounds for objecting to such measures on principle must be limited⁴.

There is another important factor to consider in the New Zealand context. In the absence of any widely recognised and clearly understood system for gauging broadcasting quality, there is a tendency, particularly within the popular press and some political interests which ideologically oppose public broadcasting, to reduce all matters of broadcasting quality to questions of ratings and audience share: if a programme rates well, it is ostensibly money well spent; if it rates poorly, then questions are raised about why the taxpayer should be funding it⁵.

³ See Hastings, 2004.

⁴ As Cramer (1997) suggests, quality means more than media producers churning out programmes primarily to satisfy themselves or impress their peers.

⁵ See Thompson, 2006.

The lack of a meaningful frame of reference to debate broadcasting qualities beyond ostensible audience appeal indirectly imposes commercial pressures on broadcasters which can suppress their creativity and experimentation in favour of safe, populist formats and also make policy makers reluctant to advance more progressive policies. For that reason, there is a strong argument in favour of exploring alternative approaches to conceptualising and assessing broadcasting quality. Insofar as this report can be of use toward such ends, it might be considered to have fulfilled its objective.

Final Considerations

It was not the objective of this report to make specific recommendations about the particular conceptions, measures or modes of implementation that ought to be adopted by New Zealand On Air or any other broadcasting body. In lieu of a set of firm directives, a list of considerations of the kinds of issue that need to be taken into account in the development of any valid, robust and workable system for measuring/assessing broadcasting quality is provided. The list is not exhaustive, in no special order of priority, and in places it may even include contradictory indications. Insofar as suggestions or indications of preferable uses of quality concepts, methodologies, or courses of action are evident, the aim is to guide the reader away from demonstrable folly and towards potential utility. Other than the stated intention to facilitate a move beyond a reliance on audience ratings as a sole measure, there is no presumption of particular normative preferences for particular conceptions or interpretations of quality.

- o Broadcasting quality cannot be understood as a discrete, singular 'Platonic' conception; it must be understood as a series of relationships between a set of values/norms and different aspects of the broadcasting process
- o In establishing any quality measurement system for broadcasting, validity requires congruence between the conception of quality being assessed, the data needed to assess it, and the method of collecting that data
- o The establishment of a quality measurement system also requires a mode of implementation to be established, including benchmarks/ frameworks of interpretation and institutional arrangements to facilitate desired responses
- o Different stakeholders may have different values and criteria of broadcasting quality. Likewise, different stakeholders may have different, and sometimes unique expertise and ability to assess certain criteria (e.g. the audience is the expert on audience reception; programme producers are the experts on production values)
- o Some stakeholders may have clear views on aspects of broadcasting quality when their actual expertise and qualification to provide valid assessment does not correlate with their own perceptions of that ability (e.g. audience members may know whether or not they enjoyed a programme; they may not perceive superliminal influences)

- o Effective broadcasting quality assessment may require triangulation of different approaches to examine a range of different criteria
- o The implementation of assessment regimes must take account of the potential for broadcasters to engage reflexively and instrumentally with the system so as to optimise the measurement of performance even if that detracts from the delivery of actual quality
- o Assessment regimes which impose sudden or significant sanctions (or rewards) on the basis of the quality measurements deployed are likely to provide a perverse incentive to broadcasters to prioritise measurable indicators over actual performance- if the stakes are high you will tend to get what you measure
- o Assessment regimes which entail no form of sanctions/rewards may not provide sufficient incentive to broadcasters to make improvements to their performance; if the stakes are too low it probably doesn't matter what you measure
- o Even if coherent quality indicators/ criteria are established, these may not be ends in themselves; the menu should not be confused with the meal.
- o It may be useful to develop relative rather than absolute/standard measures of quality attainment in order to promote optimising over satisficing and to maintain high expectations/ ideals
- o Generally speaking, it is not sufficient to define quality in terms of genre or to infer social functions from content; quality within genre and the reception of content cannot be overlooked
- o Genre classifications which are defined solely in terms of their intended audience (children's/minority programming) may need to be expanded to include programme types; genres classifications which are defined solely in terms of content/format may be usefully expanded in relation to audience.
- o Quality measures involving surveys/ interviews or other forms of questioning need to be focused and framed carefully; audiences respond differently from within 'consumer/ enjoyer' frames compared with 'citizen/social benefit' frames
- o The use of multiple measures to triangulate or cross-reference quality measures (e.g. examining the correlations between audience appreciation and content features) may increase validity and facilitate more nuanced understandings of broadcasting quality
- o Employing too many measures and quality criteria is likely to be self-defeating. Validity and practicality may hit a point of 'diminishing returns' if complexity and compliance requirements are too great
- o There is a tension between expensive, complex, rigorous and valid quality measurement systems and cheap, simple and invalid ones; the extremes of either are unlikely to be useful
- o Developing a finite range of key quality criteria is sensible, but care must be taken in constructing 'compound' categories that encompass a range of variables to maintain

discreteness (no overlap) and to avoid including variables which may inversely correlate (contra-indicate) in the same category

- o Avoid compressing too many discrete variables into the same quality category (unless factor analysis indicates strong overlap/correlation), especially if some of those variables indicate different aspects of quality (e.g. audience appreciation does not directly correlate with audience perception of quality or audience enjoyment)
- o Although quality measurement systems may be motivated by a desire to examine factors other than ratings/audience share, this data may still be useful in interpreting other quality measurement findings (e.g. the extent of potential social impact)
- o Data about audience attitudes and perceptions in other countries may not apply to New Zealand, especially in regard to particular ethnic/ cultural demographics
- o Audience perceptions of broadcaster/ programme quality and/or social function may be influenced by extraneous factors which have no bearing on quality or function (e.g. promotional campaign exhorting the quality of particular programmes)
- o Diversity and choice do not on their own ensure a diversity and choice between quality offerings; a choice between two poor-quality programmes isn't worth having
- o Workable quality assessment systems are more likely to emerge if regulators, broadcasters and audiences have some input into their design to ensure 'buy-in'. On the other hand, it is important not to allow any stakeholder interest to dilute or veto measures which are important in the assessment of public functions
- o Technological developments in communications and computing (digital media convergence) and demographic shifts in media usage may mean that quality measures developed today become invalid or misfocused in the future. There is a need to take account of social and technological trends to ensure the quality assessment system is able to perform its intended function.