



Københavns Universitet



Dialogues between audience research and production

Redvall, Eva Novrup

Published in:
Critical Studies in Television

DOI:
[10.1177/1749602017730262](https://doi.org/10.1177/1749602017730262)

Publication date:
2017

Document Version
Peer reviewed version

Citation for published version (APA):
Redvall, E. N. (2017). Dialogues between audience research and production: The history of testing television drama for the Danish Broadcasting Corporation (DR). *Critical Studies in Television*, 12(4), 346-361.
<https://doi.org/10.1177/1749602017730262>

Dialogues between Audience Research and Production: The History of Testing Television Drama for the Danish Broadcasting Corporation (DR)

Eva Novrup Redvall

University of Copenhagen, Denmark

Abstract

This article offers a historical analysis of the audience testing of television drama from The Danish Broadcasting Corporation (DR) by the in-house DR Media Research Department from the mid-1990s until 2016. The article investigates how the methods for testing have changed from more traditional focus groups and surveys to include biometric methods (e.g. skin conductance) to measure audience arousal. While audience testing is often primarily viewed as a quality measurement tool for executives, the article argues that testing can also be a dialogue-based tool offering creative practitioners the opportunity to learn more about their series and their audiences.

Keywords

audience testing; Danish television drama; Danish Broadcasting Corporation (DR); DR Media Research; DR Fiction

Introduction

In the 2010s, drama series, such as *Forbrydelsen* [*The Killing*] (2007–2010) and *Borgen* (2010–2013), from the in-house drama unit DR Fiction of the Danish Broadcasting Corporation (DR) found remarkable international success. Possible explanations of this new interest in subtitled content have already been offered from many different perspectives. Some scholars have analysed the importance of the Nordic noir ‘brand’ (e.g. Agger, 2011; Gamula and Mikos, 2013; Waade and Jensen, 2013; Creeber, 2015) and investigated the impact of a changed production framework at DR (Redvall, 2013; Nielsen, 2016). Others have discussed the influence of new programming strategies, with for instance BBC4’s scheduling of ‘subtitled oddities’ as a way to frame foreign imports

as a particular kind of public service broadcasting (Ward, 2013; also see Esser's article in this special issue for more on DR series in the United Kingdom).

This article offers a new perspective on the making of the Danish series through a historical analysis of the way in which DR has deliberately worked on learning more about the reception of its series from tests conducted by in-house audience researchers. While audience testing is often primarily viewed as a tool for executives to make more informed decisions about different products in development, this article explores how audience testing can also be regarded as a dialogue-based tool offering creative practitioners the opportunity to learn more about their drama series and their audiences.

The article builds on having had access to in-house audience reports from 1994 to 2015, from cancelled comedy dummies and the one-hour drama series *Taxa* (1997–1999) to recent series such as the inheritance drama *Arvingerne* [*The Legacy*] (2014–2017). These reports have been combined with qualitative interviews with audience researchers at DR and observations of focus group tests of *Bron/Broen* [*The Bridge*] (2011–) and *Bedrag* [*Follow the Money*] (2016–). Based on this material, the article investigates how the methods for testing have changed over time from more traditional focus groups and surveys to biometric methods (e.g. skin conductance) to measure audience arousal. The analysis draws on several examples of how audience feedback has been used to alter or fine-tune elements in, for instance, the storytelling, character development or the use of music, as well as to the shooting of an entirely new episode 'zero' for the Emmy Award-winning crime series *Rejseholdet* [*Unit One*] (2000–2004).

In the 2010s, the DR audience tests are not about whether to greenlight a series or not, but about learning more about the series in production. This can be from a managerial point of view in terms of wanting to explore whether audiences think a series is too violent or whether viewers in the provinces find a series to be too centred on Copenhagen. However, the tests also offer the opportunity for the series' creators to learn more about how the storylines, the main characters or how the look of the series is perceived.

Accordingly, the article discusses how the DR approach has been regarded as 'diagnostic' by pointing to strengths and weaknesses in the productions. The test results are not only intended for learning more about the specific episode at hand, but can also

feed into the production of later episodes and new series. The article concludes by highlighting how this dialogue-based approach can provide a greater understanding among practitioners of their series as well as their viewers that seems conducive when trying to create high-quality public service television drama for the national mainstream audiences.

The in-house audience testing of DR drama series

Among American producers, there are many horror stories about the way in which audience research is used to decide whether to commission series or not. Many of these stories are based on anecdotes of how executives used test results to make what was later judged as the wrong decisions during the so-called pilot season, such as passing on a later hit show like *Seinfeld* (e.g. Polone, 2012; Tomashoff, 2012).

As described in the literature on testing television drama in the United States, there are many different stages of testing (such as synopsis testing, pilot testing and episode testing) and different kinds of research (e.g. Blum, 2013; de Fossard and Riber, 2005; Eastman and Ferguson, 2012). For many years, audience research conducted by ASI Testing in Los Angeles and theatre-style testing in the Television City research centre at the MGM Grand in Las Vegas have been widely used in the television industry. Industry discussions suggest that there are widely differing opinions on the value of these tests, yet they still hold great power in the US television industry.

This article does not attempt to offer an overview of all kinds of audience testing, but focuses on the historical development of testing Danish television drama from the mid-1990s until 2016 and on what DR audience researchers and creative practitioners regard as the main lessons learned in terms of gaining a more nuanced understanding of the series through audience feedback. The analysis draws on qualitative interviews as well as lectures, reports and industry presentations by the DR researchers (the interviews were conducted 2011–2016, and they are listed in the references and referred to by the name of the respondent and the year of the interview). It is of course in the interest of audience researchers to present their work as having value and impact. The analysis aims at ensuring a critical perspective to their perceptions of their work through an awareness of their position as ‘exclusive informants’ with a specific agenda (e.g. Bruun 2014) and

through combining the inside knowledge gained from the audience researcher interviews with the final test reports, the observations of the actual tests and responses from the drama department about this kind of institutionalised initiative to learn more about their productions.

There can be many power struggles and conflicts related to how audience research can, for instance, help justify potentially controversial decisions. This analysis includes examples of highly critical reports and moments of distrust between the audience researchers and drama producers, but it is important to note that all the tests analysed were conducted by the in-house researchers at DR. Contrary to many other commissioning contexts, this is not a case of a broadcaster ordering competing series from external production companies and trying to assess their audience potential before greenlighting only a few, as is often the case during the bloodshed of the American pilot season. In the time period covered by this article, almost all Danish drama series shown on DR's main channel DR1 were produced by the in-house drama department, and, in most cases, at the time of testing one or two seasons had already been ordered. The people producing the series and the people testing the series are thus part of the same broadcasting corporation and it is in everyone's interest to gain a better understanding of the strengths and weaknesses of the upcoming series; it is not about whether to 'axe' a series or not, but about making the most of the material and learning about its potential.

While most tests over the years have been based on screening one or two episodes followed by a relatively broad focus group conversation, each test will normally also address particular issues that the DR management or drama department want to know more about. This can be from a creative point of view in terms of whether the storytelling works and whether the characters have appeal, but the agenda can also include issues of how to position the series. As stated by Ditte Christiansen from DR Fiction (2011), not all series are tested: a series is tested if there is some sort of doubt, but this does not necessarily have to concern the overall quality. As an example, in my observations of the testing of *The Bridge* – which was externally produced by Swedish Filmlance and Danish Nimbus Film for SVT and DR – some of the aims were to see how audiences responded to the violence, to the Swedish spoken language, and to whether the series could be broadcast on Sunday nights at 20:00. People seemed to agree that they had a good show

on their hands, but they needed to know more about the best strategy for finding the right audience.

When giving access to the in-house DR audience reports, Lars Thunø, Head of DR Media Research from 2001–2014 (2012), described the DR tests as ‘a development tool’ rather than ‘a judgement tool’. He argued that the tests are important for understanding what viewers perceive when watching drama series and for creating an awareness of the constantly changing ‘contract’ between the makers of the series and the audience. As he stated, ‘It is healthy for the creators of series to meet the competent viewers. It is an important meeting between the professionals and the audience’ (2012). In his experience, audience tastes and competences change quickly, and media research can give a clearer sense of these changes and bring them ‘into the house’ (Thunø, 2012).

According to DR audience researchers Lene Heiselberg and Jacob Lyng Wieland, this idea of bringing what they describe as users and producers closer to each other is central to their work, and there has been a tradition of insisting that both DR executives and the producers of the series from DR Fiction should be present during audience tests. Rather than only reading the results in the final written report, this quite literally brings the producers closer to the audience by making them hear their spontaneous reactions and interaction (Heiselberg and Wieland, 2011). However, this approach has recently changed, since some creative practitioners find it hard to sit through the actual tests and prefer to gauge the audience response through the researchers’ analysis (Heiselberg, 2016).

In terms of the amount of testing, it is worthwhile noting that drama tests are only a limited part of the work of the DR researchers, who also test a range of other types of programming and track the ongoing viewing figures, among many other tasks. The production of television drama is expensive and DR only produces around two seasons of high-end drama series annually. Normally new series are tested – and sometimes also new seasons of ongoing series – but in some cases not. There were, for instance, never any tests of *Forbrydelsen* [*The Killing*]. According to Heiselberg and Wieland (2011), this was due to a clear sense of the quality and nature of this crime series. However, the decision not to test a series can also be due to a lack of time for testing before the broadcast date.

The process of traditional qualitative focus group tests normally begins with a meeting between the DR management, the DR Fiction drama department and DR Audience Research for the purpose of the test. Based on this meeting, the media researchers develop an interview guide that DR management and DR Fiction agree upon as the contract for the test. There used to be a four-week period from the meeting to the test, since it took about three weeks for Gallup to recruit the participants. In 2016, Gallup recruits members from the so-called DR Panel and this process now only takes about two days, bringing the time of the contract and the test closer together (Heiselberg, 2016).

Having professional recruitment is regarded as essential for the value of the tests. As stated by Wieland, ‘If we don’t know who is in the focus group, the test has no value’ (in Heiselberg and Wieland, 2011). The recruitment depends on the purpose of the test. To find out more about the audience perceptions of the violence in *The Bridge*, for example, the selected target group was families with children, while the test of ‘the depression comedy’ *Lykke [Happy Life]* (2011–2012) focused on finding out what younger versus older audience members thought of this new genre hybrid. In most cases, the tests are conducted at the DR research facilities in the ‘DR Town’ in Copenhagen, but when the aim is to know more about differences in regional responses tests may also be conducted outside of Copenhagen. This was the case with tests of *Borgen*, as there was a concern that the series might be too focused on life in the Danish capital.

Following the test, the researchers only have a few days to produce a presentation for the DR management and DR Fiction in which the main audience reactions are boiled down to insights for further discussion in a feedback meeting. While this kind of qualitative focus group testing was the main tool for testing drama in the 2000s, the 2010s have been marked by DR Audience Research developing new ways to test emotional engagement with drama series. In the following, this will be discussed based on a historical account of how the approach to testing has changed since the 1990s.

‘Rerun tests’, ‘remake tests’ and ‘pre-tests’ in the 1990s

The audience reports from the 1990s show how these years consisted of three different kinds of testing television fiction: ‘rerun tests’, ‘remake tests’ and ‘pre-tests’ of different kinds of comedy dummies as well as tests of proposed episodes for series in production.

Rerun tests were aimed at learning more about whether audiences would be interested in watching previous drama fare or whether these older productions were perceived as dated. One test was of the three-episode mini-series, *Aladdin*, which was based on a play from 1805 (originally broadcast in 1975 and rerun in 1981). In 1994, older audience members were open to another rerun. Younger viewers, however, found the production and dialogue old-fashioned, and it was decided not to repeat the series. Another rerun test showed how even older material could get positive responses if the characters and humour still resonated with viewers. The rerun test of *Huset på Christianshavn* ('*The House in Christianshavn*', 1970–1977, rerun in 1989) found that 92 per cent of viewers would like to see this show again. That kind of feedback is popular with broadcasters who appreciate an opportunity to recycle expensive domestic productions. The rerun tests show the careful considerations that went into deciding whether older productions still had potential or not.

While the rerun tests were based on a managerial attempt to reactivate previous productions, the remake tests can be regarded as an attempt to save money on development and production by adapting popular material from other countries. This is not a common strategy in Danish television drama, and it is interesting to see how the so-called 'pre-test' of the Swedish comedy series, *Svensson, Svensson* (four seasons, 1994–2008), that became the DR series, *Madsen og co./Madsen and Co.* (1996–2000), was not about whether or not to make the series. In the report, the purpose of the test is described as 'providing input for the Danish screenplay adaptation, and the direction and production of the series'. Based on a screening of the first episode of the Swedish series, two focus groups were asked for their opinions. Both the group of 35–50-year-olds and the group of 50–65-year-olds found nothing particularly Swedish in the humour, the family or the setting, and the series was generally described as entertaining.

The report contains a one-page written evaluation of participant descriptions of the episode with three words of their choice and their grade of its overall quality on a scale of 0–10. Following this, the interview themes consisted of more specific questions about the storylines, pacing, use of a live audience, setting and characters. In terms of providing material for the creative practitioners behind the remake, the report thus presented specific feedback from the audience to consider for the adaptation.

In the 2000s, DR moved away from considering remakes, since the new managerial focus was to produce original series. *Madsen and Co.* was the last comedy produced by DR Fiction for many years, maybe because several pre-tests showed that it was challenging to find comedy material with appeal across age groups and audience demographics. In the 1990s, there were ambitions to create national comedy formats that could mirror American sitcoms. DR had a workshop for writers and directors in which American producers taught the craft of creating comedy material. However, the development of new ideas in this genre proved to be difficult since the ambition was to reach the mainstream audiences. Three dummy tests of comedy ideas point to the challenges associated with this kind of material – and to the difficulties of testing material that only exists as screenplays or loosely connected scenes or character studies rather than finished episodes.

In 1994, the DR researchers conducted a pre-test of a possible new comedy. The audience report states that the purpose was to ‘assess the potential of the screenplay from an audience perspective and to test the possibilities of pretesting television fiction already at the screenwriting stage’. There was thus an experimental aspect to the strategy chosen. The material was based on a 12-minute dummy shot in the DR studios based on four scenes from the proposed screenplay. The scenes were not connected, but were linked together by a narrator. The cast consisted of professional actors and a very basic set design had been built for the dummy.

The dummy was tested on three focus groups recruited based on rather broad criteria, since part of the test was to learn more about the ‘insecurity about the target audience’. Firstly, the focus group watched the dummy and each member was then given an individual questionnaire to get everyone’s immediate individual response. Following this, there was a semi-structured discussion based on an interview guide. The test revealed that the proposal got a very mixed response. Among the critical points was the – for comedy – devastating remark that the material was not funny enough, and some respondents said the series was neither crazy nor realistic enough. A later in-house report – summarising the main lessons learned from the testing of comedies in the 1990s – concluded that the audience generally called for a very clear-cut sense of genre if they were to watch Danish comedy.

In 1996, the media researchers tested another potential comedy series that was described as a dummy for a sitcom. The dummy consisted of a presentation of the five main characters based on material developed through improvisation. There was no storyline and no production design. The series' creators presented the sitcom as an eight-part serial targeted at an audience of 25–40-year-olds for the DR 2 channel rather than the main channel DR 1, where the expensive scripted fiction normally premieres.

The test had a dual purpose. First of all, it was meant to explore the potential audience for the series. Secondly, there was a desire to learn more about the possibilities of testing at an early stage in the development phase. Similar to the previous comedy test, part of the goal was to decide whether it makes sense to continue testing along these lines. The report specified that the dummy had not been developed specifically for the pre-test and that 'the pre-test should be regarded, first and foremost, as a part of the development of constructive pre-tests of drama productions'.

The test consisted of two focus groups with 12 people. One group had even numbers of male and female members in the 35–55 age group who watched TV on a daily basis and had a preference for shows such as *Roseanne* (1988–1997), *Cheers* (1982–1993) and *M*A*S*H* (1972–1983). The other group also had a 50/50 gender mix, but with people in the 20–40 age group with a preference for what is described as crazy comedy/satire, such as *Fawlty Towers* (1975–1979) and *The Young Ones* (1982–1984), as well as *Roseanne*. The foreign series mentioned show how the points of reference were primarily US and UK series, partly because there was little national comedy fare to mirror at this point in time. This later changed with TV 2 developing successful low-budget comedy formats – such as *Klovn [Clown]* (2005–2009) – when DR put the funny material aside.

The test audience was very critical of the dummy presentation. The material was hard to place in terms of genre; it was not regarded as funny; and several respondents stated that the material seemed more like theatre than a television production. The report noted that this was partly related to the lack of storyline and production design. The critical response led to a letter to DR from the series producer arguing that this test should not be used as part of deciding whether to greenlight the series or not. The report should be regarded as an internal 'working paper' for the researchers to gain more experience of

pre-testing drama productions. The results might also be used as input in the further development of the series, ‘to the extent that the writer and director found the test of use’. However, the dummy was only a 10 per cent expression of what the writer and director intended the series to be, and it was unfair to ask non-professional audience members to envision the remaining 90 per cent of the series. The letter ended by stating that the report was ‘proof of how a test audience registers everything and puts it all into a context – but since about 90 per cent of the remaining material is not shown, it is not possible to make any general conclusions about the final product based on the audience reactions’. I have not had access to material that reveals whether the report was in fact included in the greenlighting discussions, but the series was never made. And the later reports show how the tests moved away from testing selected scenes from screenplays and unfinished dummy presentations to asking audiences about more finished material.

There is only one other comedy test among the reports from the 1990s. This is a ‘qualitative dummy test’ from 1998 of *Pas på mor* (*Watch out for Mum*, 1998–1999). The test had two mixed-gender focus groups – one in the 35–54 age group and one for people over 55 – who all watched television on a daily basis and were fond of, for instance, *Madsen and Co*. The report never states its purpose clearly, but opens with the conclusion that this sitcom is not suited as ‘broad weekend entertainment’ since it mostly appeals to the 55+ audience members. Again, a main point of criticism was that the series just wasn’t funny enough, and the dummy only scored 2.3 out of 5 among the younger viewers. The report offers advice consisting of specific suggestions on how to improve the material, ranging from warnings against storytelling repetitions to how to make certain characters less irritating and others more interesting. However, the report ends on the note that even if changes are made the series is unlikely to have a significantly broader appeal, and the 12-episode series ended up having limited success on Monday evenings.

The comedy tests of the 1990s and a last attempt with a produced eight-episode comedy series called ‘Bentes verden’ (‘Bente’s World’) that got cancelled following an audience test in 2001 led to DR deciding to focus on high-end one-hour drama series rather than developing both drama and comedy formats. In a news article from 2011 that explores why viewers never got to see the finished series on the world of Bente, Lars

Grarup, head of DR1 in 2001, explained that the quality ‘simply wasn’t good enough’ and that this led to giving up on sitcoms in the 2000s (Nielsen, 2011) (the DR scripted comedy formats have only just returned with a quirky comedy drama such as *Bankerot* (2014–2015) or the comedienne-driven serial *Ditte and Louise* (2015–2016)). Instead, DR decided to focus almost exclusively on creating a new kind of Sunday night drama series. The first attempt at such a series premiered in 1997 and showed how drama was much better than comedy at creating the sense of a quality brand for the DR series and gathering many kinds of viewers.

The early one-hour drama series: *Taxa* and *Unit One*

In 1996, the DR researchers conducted a pre-test of the first episode for the upcoming drama series *Taxa* (*Taxi*, 1997–1999). The series follows the life around a taxi company in Copenhagen and became the first in an impressive streak of popular Sunday night dramas (cf. Degn and Krogager’s article in this special issue). The pre-test of *Taxa* was similar to a pilot test in that the aim was to learn more about the series by asking audiences about the first episode. However, since the 2000s DR has not produced pilots, but moved straight to series. Accordingly, the test was not about whether to commission the series or not, but about gaining more knowledge about the episode before deciding to initiate a second season (before the first had even been aired). The test report also states that the test was intended to provide input for ‘the continued work with form and content’. The test was thus supposed to feed into the work of making future episodes and there were also two explicit questions that DR would like to see answered: whether *Taxa* would appeal to rural audiences and younger viewers, and whether the series was too centred on Copenhagen.

The two mixed-gender focus groups with a daily television diet – one in the 50+ age group and the other in the 25–35 age group – both responded positively to what they saw. They liked the characters and the stories around the taxi drivers. The set-up was perceived as new in Danish television drama, even if younger viewers in particular recognised this use of an arena as a classic trait of long-running US series. Audience members removed from the Danish capital did not find the series to be too Copenhagen-based. And mirroring some of the DR production ideas of creating series containing

aspects of identification as well as fascination, older audience members described identification as the most important aspect of a drama series, while the younger audiences mostly found its fascinating aspects appealing.

The report contained several reasons for continuing the series, which ended up comprising five seasons with a total of 56 episodes. While the series was running in 1998, another qualitative test was conducted, this time based on 1,008 phone interviews, focusing on the audience response to the series' depiction of violence and sex as well as its characters and storytelling pace. The test found that 92 per cent of the audience had never felt offended by the portrayal of violence and only 5 per cent thought that there was too much sex. The report concluded that the audience was generally content with the characters and the storytelling and suggested that it might even soon be time for a rerun (the series was rerun in 2003).

Taxa was the first attempt by DR Fiction to create a long-running series modelled on what was perceived as a Danish approach to US production strategies (Redvall, 2013: 65-66). The next long-running series was the crime drama, *Rejseholdet [Unit One]* (2000–2004), which was intended to follow in the successful footsteps of *Taxa*. Part of the method for doing this was appealing to many audience segments in the so-called 'Minerva model' for lifestyle analysis developed by AC Nielsen AIM, which DR has used as a segmentation tool in several departments for many years.

In 2002, Head of Drama Ingolf Gabold (from 1999 to 2012) told the press that DR had used ideas of audience segmentation during the creation of the series to try to ensure that the end result would appeal to as many segments as possible (Sleiborg, 2002). This was mainly achieved through the character design of the travelling police unit at the core of the series, whose different characters were intended to appeal to certain demographics. DR Fiction has never been shy about the fact that they use audience tests, and when asked whether segmentation and focus groups represent commercial tools from the advertising industry and commercial television, Gabold defended the tests by saying they were most certainly 'public service tools', since it is a public service obligation to make television drama of relevance to the entire nation (Sleiborg, 2002).

The test of *Unit One* is often mentioned in the DR production framework, since it pointed to weaknesses in the material that led to the re-shoot of an entirely new 'episode

zero' to precede the one that was tested. According to Gabold, the test showed that there was plenty of fascination with the police work, but not enough character identification, particularly with the main character, Ingrid Dahl (Charlotte Fich), and her police colleagues. The new episode 0 gave Ingrid a number of challenges in order to put her in a weak position that would make people root for her and want her to succeed in a different way than in the original first episode (in Sleiborg, 2002). The resulting series was very successful with Danish audiences, and *Unit One* ended up with 32 episodes that also managed to find enthusiastic international audiences as a 'classic geo-linguistic export' (Jensen, 2016).

The crucial question of trust: *The Chosen 7* and *Summer*

Unit One won an Emmy for best international drama in 2002, and the next DR drama – the family series *Nikolaj og Julie* [*Nikolaj and Julie*] (2002–2003) – also won the prestigious award, creating a sense that DR Fiction was now on the right track in terms of creating high-end series. However, as repeatedly stated by executives from DR Fiction over the years, there is no recipe for making television drama, and each production has its own challenges (Redvall, 2013: 5).

Two tests illustrated some of these challenges in the early 2000s. I have not had access to the test report of the series *De udvalgte* [*The Chosen 7*] (2001), but according to Heiselberg and Wieland this test was a turning point for the DR researchers, since the series tested positively but performed poorly. The discrepancy between the test results and the viewing figures created a crisis of trust between DR Fiction and DR Audience Research. Heiselberg and Wieland describe how there were several years with 'no collaboration whatsoever' following the test. They were both hired as qualitative audience researchers in 2006 for the first drama test in many years, which according to them took place in 'an atmosphere of mistrust' since DR Fiction had doubts about the value of the findings from the outset (Heiselberg and Wieland, 2011).

However, the test of the family series *Sommer* [*Summer*] (2008–2009) created a new dialogue, since the drama department acknowledged the problems that the researchers highlighted in the material. Part of the audience criticism pointed to storytelling problems, but there were also issues related to the female characters not

appealing to what Heiselberg and Wieland describe as ‘modern women’ (2011). The test led to the making of an entirely new first episode, character alterations and changes in the team behind the scenes. According to Heiselberg and Wieland, the test of *Summer* re-established the trust between DR Fiction and the researchers, which in their opinion has been present ever since. And *Summer* became very popular with national audiences, with three seasons of 10 episodes each.

In the late 2000s, the testing was based on a focus-group framework and did not lead to substantial changes. As mentioned, *The Killing* was never tested. The tests of episode 1 and 2 of *Borgen* showed that audiences liked how the series mixed the political, the press and the private sphere, as well as the stories and characters. The test pointed to some viewers reacting negatively to what they perceived as ‘over-telling’ in certain scenes, such as having a scene with a secret meeting where it is dark, windy and rainy. According to the researchers, some found this clichéd. As described by Heiselberg and Wieland, audience research can thus point to challenges in the material, but – contrary to the earlier comedy reports – the researchers will not suggest solutions. In this case, DR Fiction decided to turn down the sound of the rain. Viewer response can thus influence specific scenes in major as well as minor ways (Heiselberg and Wieland, 2011). Another aspect highlighted in the *Borgen* test was the journalist, Hanne Holm (Benedikte Hansen), who had a strong appeal with several audience segments and thus might be used constructively later in the series. In this way, tests can also provide food for thought in terms of how to use characters in later seasons depending on audience feedback.

As mentioned, the test of *The Bridge* was mostly about addressing concerns about the violence, the use of the Swedish language and finding out when to programme the series. My interviews with screenwriters and producers at DR suggest that there seems to be a general acceptance of the use of tests (Redvall, 2013), but according to the DR researchers it can be a challenge for certain people to understand exactly what is going on in the focus group room. Thus, while the researchers used to encourage the creative practitioners to observe the tests, they are now more careful with this strategy, since it can be a problem if the DR Fiction guests draw their own conclusions based on their immediate observations. The researchers would like to offer their interpretations of what has happened, since observing a test – in their words – can be ‘like looking into a raw

Excel sheet with data'; the inexperienced eye can also get a strange image of the researchers who take on a certain role when trying to get the focus group participants to share their views of the series (Heiselberg and Wieland, 2011). As an example, Heiselberg argues that if audiences respond to a certain character as a wonderful comic relief it does not necessarily mean that one should just add more comic interludes. Rather, this might relate to a general sense of a series being too gloomy, and it might be better to generally lighten the mood instead of trying to add more humour (Heiselberg, 2014). Focus groups can be a good way for makers of series to learn more about their audience, but it can also be complicated to interpret what people actually say.

Can audiences be trusted? New testing methods of the 2010s

This leads to the final part of this article, which focuses on recent DR discussions about whether audiences can actually be trusted when 'self-reporting'. Contrary to many other types of television programming, television drama is very much about emotional engagement. How do you ask audiences to rationally talk about their feelings in a group of strangers, and how can you learn about their subconscious feelings? Questions such as these have led the DR researchers to look into new types of testing that combine qualitative interviewing with methods from cognitive science.

In 2010, DR was part of a research project, financed by Nordvision, together with the Norwegian, Swedish and Finnish public service broadcasters NRK, SVT and YLE, which explored how to better measure emotional engagement with television fiction. The published results focus on a number of 'mentometer tests' in which viewers continuously self-report on a slider (from 0-5) while watching an episode. Afterwards, the results are aggregated and used as the basis for conversations about where they 'peaked' or 'flatlined'. This offers a glimpse into the emotional experience of viewers while they are watching, but it is still self-reporting and asks the viewers to be rational while experiencing the drama (Heiselberg, 2015). DR thus moved on to experiment with a second way of measuring emotional engagement, namely: through EEG measurement (electro-encephalography measurement), an electrophysiological monitoring method to record electrical activity of the brain.

In 2014, Heiselberg and Wieland presented their EEG test of *Arvingerne [The Legacy]* at the Danish TV Festival. They explained that, because the EEG headsets are expensive, only one person would be tested at a time, with electrodes placed along the scalp. Following the test, the company Neurons Inc. computed the data from a total of 28 participants. It showed that viewers generally experienced little arousal during quite a long period in the first episode. As discussed by Heiselberg, the EEG method can measure the intensity of feelings, from high to low, but not their valence. Valence can later be talked about in qualitative interviews: once the emotion is conscious, one can talk about it in a rational way. However, Heiselberg stresses that the DR researchers are generally moving away from analytical questions of ‘what’ and ‘why’ towards asking people to describe or tell more about their emotions. According to her, this ‘keeps viewers on the emotional rather than the rational playing field’, which is what you want when testing television drama driven by emotional engagement (Heiselberg, 2015).

Since EEG testing is both expensive and time-consuming, the latest attempt to measure audience arousal by the DR researchers is to measure skin conductance. Sensors on the fingers of respondents measure their skin conductance through an iPad, and this data is immediately aggregated in a graph showing the average arousal scene by scene. Following this, the results are used as the basis for a group discussion focusing on the scenes with high or low arousal. This method was used for tests of *Bedrag [Follow the Money]* and seems to be the new preferred method for in-house drama tests. As argued by Heiselberg: ‘Self-reports can’t stand alone in audience research; they only tell us half the story’ (2015). Consequently, DR has recently invested in three physiological measuring stations that combine eye tracking and skin conductance, which according to Heiselberg shows the direction of future audience research into audience arousal (see Heiselberg, 2016, for a detailed account of the new tools for audience testing).

Other test methods have also been in use in the 2010s, for instance online focus groups for Ole Bornedal’s historical war series, *1864* (2014), and what the researchers discuss as ‘semi-qualitative testing’, which involves online streaming of an episode followed by an online survey with rather open questions. While the use of online surveys has many advantages – such as getting feedback from viewers all over the country – it does not offer the opportunity to ask follow-up in-depth questions. Depending on what is

to be tested and at what stage in the production, using different methods might thus be the most appropriate, and the DR researchers argue that the exact research design takes careful consideration every time (Heiselberg, Knudsen and Wieland, 2014).

From ‘diagnostic research’ to ‘viewer evaluations’: Conclusions and cliff-hangers

As illustrated by this historical account of DR audience testing, the researchers have worked with several kinds of tests over the years when trying to learn more about the DR television drama productions. While researchers in the 1990s used rerun tests, remake tests and pre-tests of rather unfinished material, the strategy since the 2000s has been to only test material in a finished form that audiences can relate to. There might still be the opportunity to make minor adjustments – or more substantial changes if the tests point to major challenges in the material – but the tests are also fundamentally regarded as a useful tool for keeping creative practitioners aware of audience tastes and competencies. Moreover, the tests can offer input for later episodes or seasons and have led to several different kinds of changes in the finished drama series over the years, from changing small details in the mood of a scene to reshooting entire episodes.

DR Audience Research used to describe their test approach as ‘diagnostic’, but they now use the term ‘viewer evaluations’ to avoid using language that sounds like drama series are sick patients coming in for a check-up: the tests should not sound like the process of diagnosing a disease. Rather, the researchers would like the creative practitioners to think of them as having a more coach-like role by offering assistance through pointing to possible discrepancies between what DR Fiction wants to create and what audiences are experiencing (Heiselberg, Knudsen and Wieland, 2014). According to the current Head of DR Fiction, Piv Bernth (2014), much can be learned from asking audiences about their experiences, but it should always be left to the creative practitioners to decide how to respond to the results. However – as the comedy dummy tests suggested – it can be hard to receive a critical evaluation, and the question is whether the DR researchers, the DR Fiction practitioners and the DR executives all think about the nature of the test and how to use the results in the same way. Audience feedback can help highlight whether they in fact interpret the creative ambitions in the intended way, but as

illustrated in the examples from the past 20 years of testing, test methods – like television drama forms and norms – are also subject to change, scrutiny and improvement.

Just as there is no one recipe for how to make good television drama, there does not seem to be a one-size-fits-all recipe for how to test television drama. That is one of the reasons why it seems beneficial to have a constructive dialogue between executives, audience researchers and drama producers about the specific nature of each test and what can actually be concluded based on its findings. The examples from the DR context show how tests can be conducted for a wide variety of reasons and contribute with many different kinds of feedback, addressing particular issues related to storytelling, characters and content as well as more general responses to audience perceptions of having foreign languages spoken in Danish series or how series are perceived among audiences in certain parts of the country.

Since the high-end DR drama series are targeted at the large mainstream audiences on the national screens, it makes sense for their makers to seek a nuanced sense of how different kinds of viewers relate to their material; for the productions at hand as well as for future productions. It is fair to assume that some of the major points of criticism in the audience feedback that has led to substantial changes of certain series have in fact improved their quality and possibly helped their international circulation. Accordingly, the work of the audience research department can be regarded as one of the contributing elements, albeit a rather hidden one, behind the recent success of Danish series.

It would be interesting with more knowledge about how audience research is carried out in other television drama contexts to compare the strategies chosen in different broadcasting landscapes. As of now – not the least because of challenges in terms of getting access – this sort of research is still very limited. One reason why the DR executives, audience researchers and producers have provided access in the Danish context is related to the national and international popularity of the drama series and the sense that things have generally worked well in the DR production framework in recent years. However, as shown in the historical overview of the testing of DR comedy and drama content this has not always been the case, and there are lessons to be learned from

ideas of best practice as well as the more painful examples of cancelled or criticised productions.

Audience research of television drama can serve many very different purposes, such as whether to greenlight series or how to best schedule or promote certain kinds of material. This article has analysed how one approach is to think of the process of asking audiences as what former Head of DR Media Research Lars Thunø described as a ‘development tool’; one that has a dialogue with audiences about their engagement with the material while their decoding of the series can still influence the series’ final encoding. However, how audiences are asked seems to be currently changing in the DR framework with the move from mostly qualitative focus groups in the 2000s to several new attempts of combining more traditional methods with biometric monitoring. This article has only offered a glimpse into this new cognitive approach of understanding how audiences engage with television drama and more research is definitely needed in terms of understanding the future dialogues between audience research and production.

References

1864 (2014) DR 1. Miso Film/DR.

Agger G (2011) Emotion, Gender and Genre: Investigating *The Killing*. *Northern Lights* 9: 111–25.

Aladdin (1975) DR1. DR.

Arvingerne [The Legacy] (2014–2017) DR 1. DR.

Bankerot (2014–2015) DR 1. DR.

Bedrag [Follow the Money] (2016–) DR 1. DR.

Bernth P (2014) Interview by the author in Copenhagen, 1 October.

Blum RA (2013) *Television and Screenwriting: From Concept to Contract*. 4th ed. Boston, MA: Focal Press.

Bron/Broen [The Bridge] (2011–2018) DR1 and SVT 1. Filmlance/Nimbus Film for SVT/DR.

Bruun H (2014) Eksklusive informanter. *Nordicom-Information* 36(1): 29–43.

Borgen (2010–2013) DR 1. DR.

- Cheers* (1982–1993) NBC. Charles/Burrows/Charles Productions/in association with Paramount Television
- Christiansen D (2011) Telephone interview by the author, 29 April.
- Creeber G (2015) Killing us softly: Investigating the aesthetics, philosophy and influence of Nordic Noir television. *Journal of Popular Television* 3(1): 21–35.
- Eastman S and Ferguson D (2012) *Media Programming: Strategies and Practices*. Andover: Cengage Learning.
- De Fossard E and Riber J (2005) *Writing and Producing for Television and Film*. London: Sage.
- Ditte and Louise* (2015–2016) DR 1. DR.
- Fawlty Towers* (1975–1979) BBC Two. BBC.
- Forbrydelsen* [*The Killing*] (2007–2010) DR 1. DR.
- Gamula L and Mikos L (2014) *Nordic Noir – Skandinaviske Fernsehserien und ihr internationaler Erfolg*. Munich: UVK.
- Heiselberg L (2016) *Seerevaluering af emotionelle oplevelser i fiktionsserier*. PhD thesis, Aalborg University, Denmark.
- Heiselberg L (2015) Guest lecture at The University of Copenhagen, 3 December.
- Heiselberg, L (2016) Interview by the author in Copenhagen, 22 September.
- Heiselberg L, Knudsen HG and Wieland JL (2014) Interview by the author in Copenhagen, 25 August.
- Heiselberg L and Wieland JL (2011) Interview by the author in Copenhagen. 1 April.
- Heiselberg L and Wieland JL (2010) *Emotionelt fokus for kvalitativ metode*. Copenhagen: DR Media Research in collaboration with NRK, SVT and YLE.
- Huset på Christianshavn* [*The House in Christianshavn*] (1970–1977) DR 1. Nordisk Film.
- Jensen PM (2016) Global Impact of Danish Drama Series: A Peripheral, Non-Commercial, Creative Counterflow. *Kosmorama* 263. Available at : <http://www.kosmorama.org/ServiceMenu/05-English/Articles/Global-Impact-of-Danish-Drama-Series.aspx> (accessed 3 August 2017)
- Klovn* [*Clown*] (2005–2009) TV 2 Zulu. Zentropa.
- Lykke* [*Happy Life*] (2011–2012) DR 1. DR.

- Madsen og co.* [*Madsen and Co.*] (1996–2000) DR 1. DR.
- M*A*S*H* (1972–1983) CBS. 20th Century Fox Television.
- Nielsen JI (2016) The Danish Way to do it the American Way. *Kosmorama* 263.
Available at : <http://www.kosmorama.org/ServiceMenu/05-English/Articles/The-Danish-Way.aspx> (accessed 3 August 2017).
- Nielsen S (2011) DR-serie til millioner samler støv. *Tv.tid.tv2.dk*, 26 October. Available at: <http://tv.tid.tv2.dk/nytomtv/article.php/id-44998557%3Adrserie-til-millioner-samler-st%C3%B8v.html> (accessed 22 September 2016).
- Nikolaj og Julie* [*Nikolaj and Julie*] (2002–2003) DR 1. DR.
- Pas på mor* [*Watch out for Mum*] (1998–1999) DR. DR.
- Polone, G (2012) The Folly of Having Focus Groups Judge TV Pilots. *Vulture.com*, 9 May. Available at: <http://www.vulture.com/2012/05/tv-pilot-focus-groups-gavin-polone.html> (accessed 22 September 2016).
- Redvall EN (2013) *Writing and Producing Television Drama in Denmark: From The Kingdom to The Killing*. Basingstoke: Palgrave Macmillan.
- Rejseholdet* [*Unit One*] (2000–2004) DR 1. DR.
- Roseanne* (1988–1997) ABC. Wind Dancer Productions/in association with Carsey-Werner Company/Paramount Television
- Waade AM and Jensen PM (2013) Nordic Noir Production Values: *The Killing* and *The Bridge*. *Akademisk kvarter* 7, Fall:189–201.
- Ward S (2013) Finding ‘Public Purpose’ in ‘Subtitled oddities’: Framing BBC Four’s Danish Imports as public service broadcasting. *Journal of Popular Television* 1(2): 251–258.
- Sleiborg H (2002) Seriesucceser. *Kommunikationsforum*, 20 May. Available at: <http://www.kommunikationsforum.dk/artikler/seriesucceser> (accessed 22 September 2016)
- Sommer* [*Summer*] (2008–2009) DR 1. DR.
- Svensson, Svensson* (1994–2008) DR 1. DR.
- Taxa* [*Taxi*] 1997–1999) DR 1. DR.
- Thunø L (2012) Interview by the author in Copenhagen. 22 November.

Tomashoff C (2012) Programmers Keep Shows' Prospects in Focus (Groups). *The New York Times*, 11 May.

De udvalgte, De [The Chosen 7] (2001) DR 1. DR.

Young Ones, The (1982–1984) BBC Two. BBC.