

Qualitative Methods in TV Research - Appreciation

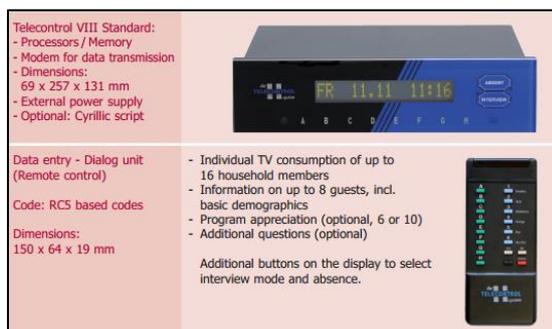
Peter Diem

There are two basic approaches to practically-oriented qualitative television research. One is based on data delivered as the result of appreciation scaling. Such data can be collected by meter, diary, phone or with the help of online surveys. The second approach is based on the use of in-depth interviews, group discussions or auditorium tests for data collection. These studies can be conducted both offline and online.

Appreciation Measurement by Meter

In the Austrian Public Service TV (ORF) use was made of both methods. Austria is probably one of the last countries to run a meter system with integrated program appreciation. Years ago such systems were employed by Belgium, Denmark, Finland, The Netherlands, Norway, Portugal, Spain and Switzerland, while in the UK a supplementary appreciation diary was used which provided Appreciation Indices (AI) for all programs. It is interesting to note that at that time the three major European markets Germany, France, and Italy did not employ such a system. Countries formerly relying on diary collected data, such as the Czech Republic and Slovakia, also used appreciation. The University of Hong Kong in its reports on TV program quality uses the UK-type Appreciation Index (AI).

In the meantime, practically all of the countries mentioned above, have „lost“ appreciation by meter. As TAM contracts were renewed, important customers (chiefly media and advertizing agencies) said they were not interested in appreciation. But public service stations were – due to their „public value“ obligation which calls for the proof of „quality“.



Former Telecontrol VIII People Meter



Traditional Nielsen PeopleMeter

So in the course of time, a number of nations ruefully paddled back to appreciation – albeit on the basis of a completely new system: the „TV Appreciation Panel“. This type of research was invented and developed by GfK Intomart in the Netherlands (its TV system was formerly very fond of appreciation by meter.) The next to sign up was the UK (BBC and ITV). While Germany (ZDF) and half of UK (ITV) dropped out after a few years, the following countries continue to run a TV Appreciation Panel:

TV Appreciation Panels



Client	Panel	Local GfK, hosting	Since
 Dutch Public Broadcasters	8,000	Intomart GfK	2002
 BBC UK Television and radio Cross media	21,500 650 per week	GfK NOP, Intomart GfK	2004
 RTE, Ireland	3,000	GfK NOP, Intomart GfK	2007
 TV3 Catalonia, Spain	1,500	GfK Emer	2008
 VRT, Flanders, Belgium	4,200	GfK Significant, Intomart GfK	2012

→ For more information and a discussion about the use of TV Appreciation Panels cf. [Appendix 1 and 2](#)

Rating vs. Appreciation

Note: There is no fixed correlation between ratings and appreciation.

Large TV audiences are by no means always satisfied audiences. Large audiences are composite audiences, which means that a considerable number of viewers is sitting before the screen without really having personally chosen to watch this particular program. And sometimes it even seems that viewers hate what they have just watched. This is not only the case when the national soccer team suffers a painful loss against an 'inferior' opponent or when ski champions miss a victory by milliseconds. There are also other instances in which rather large audiences may react with disgust. Current affairs programs with violent pictures from a theater of war or discussion programs about unpopular topics are of this type. Thus in many instances, *high ratings may go along with low appreciation*. Relatively low appreciation often occurs with typical Friday or Sunday night thrillers – one has the impression, viewers regret that they „wasted“ time again on some murder case...

The second category is *low appreciation combined with low ratings*. In such cases, a minority program has failed to draw enough satisfied viewers within the target audience. This occurs quite often with programs about modern art or contemporary music which do not appeal to many viewers. Such broadcasts tend to draw small audiences, who on top express their criticism about what they have been offered by low appreciation scores.

A third category comprises programs in which *high appreciation is combined with low ratings*. Good public service broadcasters produce quite a lot of such programs.

These programs are aimed at small target groups - opera fans are a typical example - who take pleasure in special offerings. Another examples are religious broadcasts which tend to be highly appreciated by small audiences.

Program makers are naturally most satisfied with the fourth category: *high ratings accompanied by high appreciation*. This type is by no means an exception. There are some program categories which tend to score well in both respects. These include nature/animal programs or expeditions, top-flight cinema films, special folk music offerings etc. Here is a general overview:

Typical Genres - with exceptions	Rating	Appreciation
Classical Concert, Romantic Films, Nature/Science Documentaries, Sports	high	high
Religious Broadcasts, Classical Opera, Ballet, Operetta, Literature	low	high
Reality TV, Comedy, Quiz Shows, Sometimes Current Affairs, Thrillers	high	low
Modern Opera, Avantgarde Films, Reality TV, Horror Films, Thrillers	low	low

What are the advantages of appreciation?

First of all, appreciation is used by public service broadcasters to document the *public value* of their program. Appreciation is a major performance indicator. For details on the concept of „performance indicators“ cf. [Appendix 2](#).

Besides helping public broadcasters to justify license fees, an important function of appreciation scores is their *predictive value*. Many cases have shown that a first episode which is well accepted – not only according to rating but also with respect to appreciation – will result in good ratings for subsequent episodes. Conversely – if a series of episodes starts with a good or satisfactory rating but with low appreciation, it can be assumed that the following episodes will not perform well – except if substantial improvements are being made in the program.

A third interesting feature of high appreciation scores is their *value for the acceptance of TV advertizing*. In contrast to the opinion of most advertizing agents/media agencies, a TV program with high appreciation has a positive influence on ads placed immediately before or after such a program. According to empirical studies performed in the Netherlands (Ster/Intomart/MarketResponse) , ads after well-accepted programs produce a higher buying intention than ads placed in other program environments. According to these findings, even a slight increase in appreciation has a measurable effect on the propensity to buy.

Finally, it should be noted that appreciation is an „*incentive*“ for the members of a TAM panel. Respondents who are invited to judge the programs they watch by having to enter an appreciation score, have more fun in participating and are therefore more disciplined in operating the meter than panelists who do not have a chance to express their opinion in the form of appreciation.

Some methodological considerations

There are some methodological objections against the use of appreciation as an 'one-dimensional' measurement. It is sometimes contended that only by means of a *cluster of qualitative metrics* (measuring a variety of program properties), is it possible to determine the „quality“ of a program. Still, experience with more than 12,000 ORF TV programs per year over a period of almost 15 years has shown that the interpretation of appreciation scores in the context of the respective program genre by experienced researchers and program makers allows for unbiased insight into audience satisfaction.

Of course, *additional qualitative questioning* will lead to a still deeper understanding of program quality. Therefore, in all TV Appreciation Panels, questions are included to measure such factors as viewing attention, personal enrichment, willingness to recommend a program to others etc. In particular, *answers to open questions* (about „likes“ and „dislikes“) collected in online panels give insight into the different factors and motives constituting program „quality“.

The „viewing experience („Sendungserlebnis“)

Note: The general appreciation score measures the *individual viewing experience*, not the artistic value of a program or the journalistic quality of its content. The motives underlying program appreciation may differ from case to case. As we said, the transmission of a sports event may be of high technical quality – but if the national team is defeated, appreciation will be low.

Nevertheless, the average numerical appreciation score (which also needs no further processing) together with the respective program category is a most practical instrument of measuring program quality – simple, cheap and effective.

Other qualitative methods of TV research

In the sphere of radio research the in-depth interview and call-out testing are well established techniques. But also in the field of television, program pretesting by means of focus groups has proved to be a very good solution. In order to collect as much information as possible, a special setting for the use of a focus group was developed by the ORF media research department.

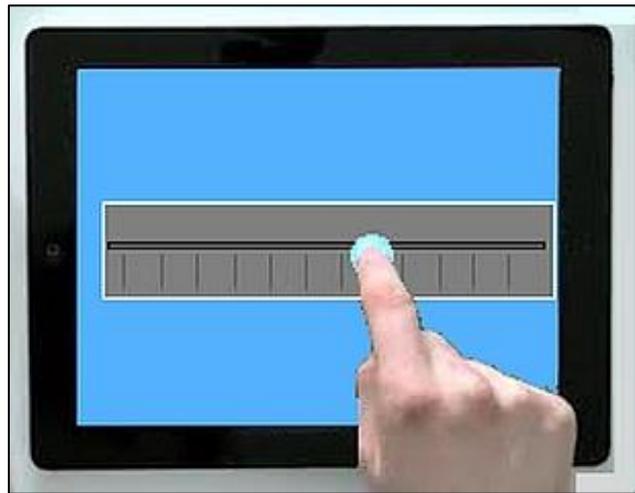
The Videotest

In cooperation with the market research institute Triconsult (<http://triconsult.at>), the so-called „Videotest“ was used until the year 2000. This was a special form of group discussion in which the respondents operated electro-mechanical „sliders“ while viewing a live broadcast or a video tape. The data recorded by means of the sliders were fed into a PC and evaluated on a scale from 0,0 („very bad“) to 10,0 („very good“). The results were available immediately after the viewing and could be

displayed along a time axis, thus showing the acceptance of the program parts from beginning to end (see below).



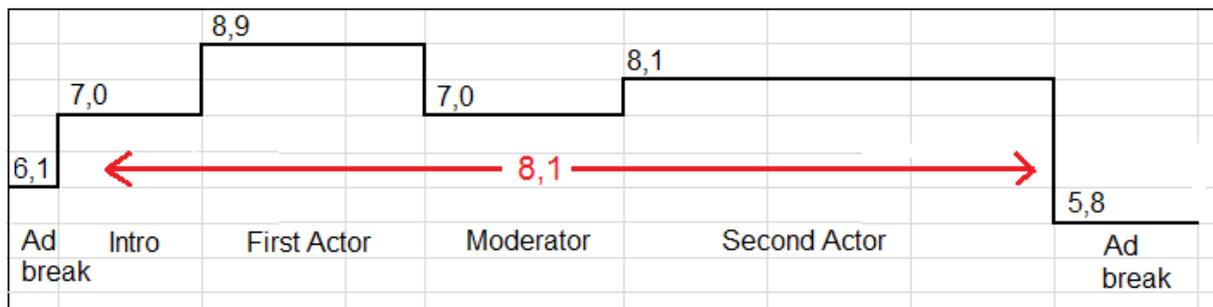
Old-fashioned slider device



Slider scale designed for use on tablet



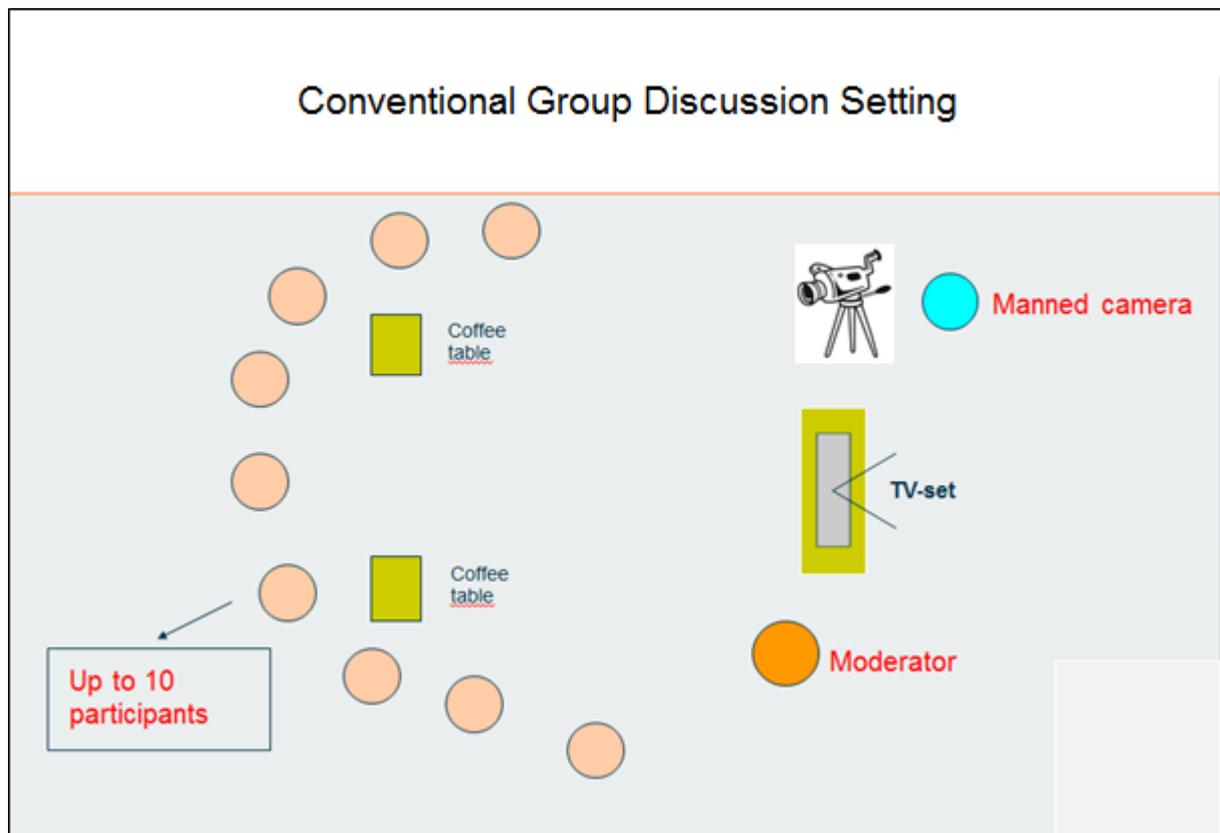
Slider programmed as an „App“ to be used on a smartphone



Sample read-out of a number of sliders operated during the viewing of a TV program. In this case, the time-weighted average (or median) appreciation score of the program (without ad breaks) is 8,1. It shows that the first part of the program was better received than the second one.

It should be mentioned in this connection that the Austrian sociologist Paul F. Lazarsfeld while doing research at Columbia University in New York in 1937/38, already had test persons judge radio broadcasts by operating a red and a green pen on a moving paper band to record their appreciation of what they were listening to (cf. *The Lazarsfeld-Stanton Program Analyzer Appendix 5*)

After having viewed the program, the Videotest group was led into a one-hour session to discuss the *pros* and *cons* of the program. To produce a perfect videotape of this discussion – including body language and facial expression – the group was seated in a semicircle and was being filmed by a semi-professional camera man. Emphasis was led on good lighting and the use of a high quality directional microphone in order to produce a videotape which professional program makers would be willing to view. Body language and facial expression could thus be recorded. Here is a sketch of the set-up.



If electronic sliders or other sentiment recording equipment is not at hand, the Appreciation Index can also be produced by paper and pencil work. With regard to the scale form 0,0 to 10,0, long-term benchmarking has taught us that products, services or media content (such as TV programs) which receive a score lower than 6.5 are serious marketing risks. We have proposed to GEAR, the professional Group of European Audience Researchers, that this scale - which is widely used in the US, England and Germany - be accepted as a kind of 'harmonized' international measure. This would enable researchers to make simple comparisons in the field of globalized television.

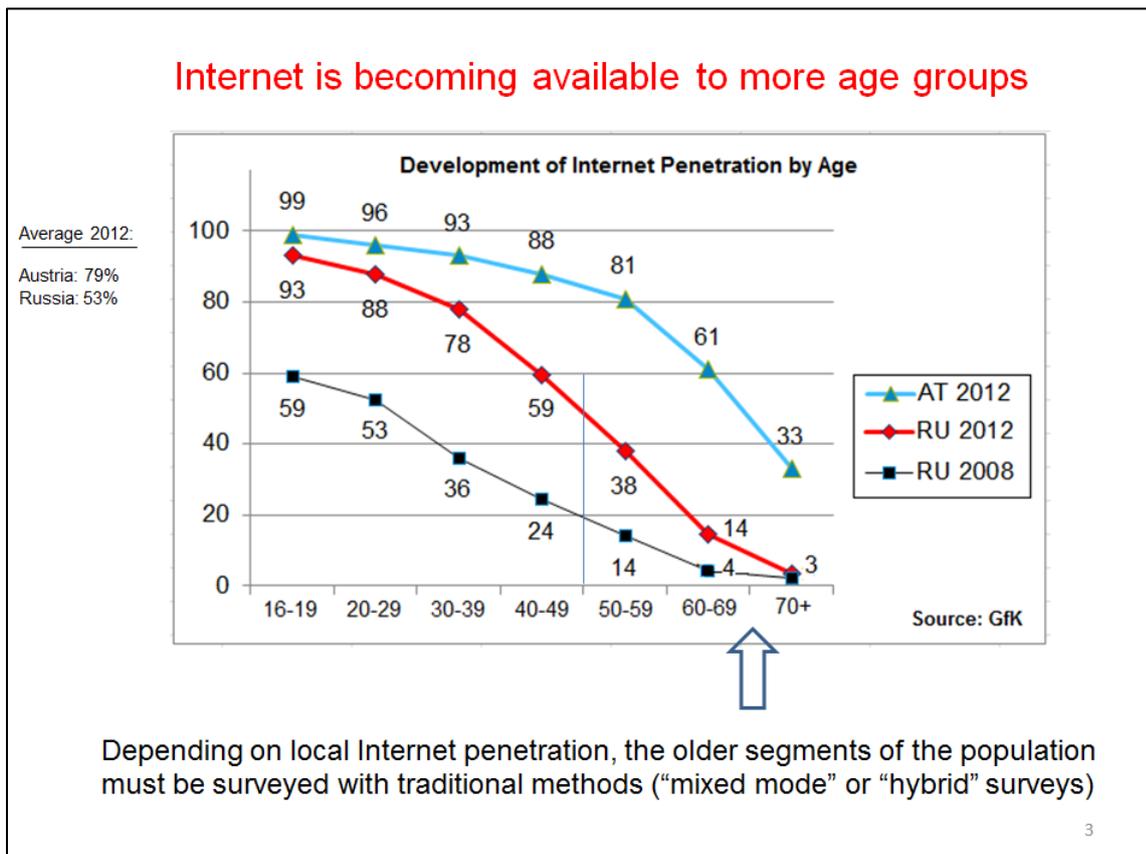
Since quite a number of years, the Internet has opened the possibility of doing qualitative TV research online. One possibility is the *Online Focus Group (OFG)*, a

discussion held online among up to 8 participants. For practical reasons (formalities, limited time), simultaneous online discussions tend to produce rather „thin“ results.

Another method, the *Online Bulletin Board (OBB)* has established itself as the most practical and most effective method of qualitative TV research. Engaging a number of 12 to 20 participants and lasting a series of days – up to a fortnight – the OBB produces very rich results. This is because the participants can follow up at their convenience at home and take their time to express themselves without stress. There is also enough time to view sample footage shown online or delivered to the home on disk, as participants can enter their opinion at any time during the exercise. There are a few other possibilities of qualitative online research for TV - such as diaries, web logs or the creation of „viewer communities“. The number of participants can be higher in these cases. For logistic reasons, these methods are rather complex and therefore also expensive.

Note: Due to the rapid spread of the internet also in developing markets, qualitative online TV research is no longer restricted to young and up-market segments of the population. The small numbers of participants required for a targeted *Online Bulletin Board* can be easily recruited by professional market research institutes.

The following table shows the penetration of Internet in different age groups in Austria and Russia.



Source: This presentation is based on a paper published in „Media-Perspektiven“ 2/94, an article published in „Communications“ - the „European Journal of Communication Research“ # 21 (1996), and a presentation given in at the Higher School of Economics in Moscow (April 2013).

Appendix 4 is a German-language article on the same subject by the Dutch TV researchers Lex van Meurs and Erica Simpson

Appendix 1

TV Appreciation Panels

As can be seen in the graphs below, appreciation panels use different types of „additional“ questions after „general“ measurement by the „Appreciation Index“.

1. Selection of programm(s) viewed „yesterday“

BBC ONE	BBC TWO	ITV 1	Channel 4	five	BBC THREE	BBC FOUR	SKYONE	4
18:00 BBC News at Six	18:00 Breab	18:00 STV News at Six	18:00 The Simpsons	18:00 Home and Away	18:00 MOTD Live: Italy v Egypt	18:00 World News Today	18:00 Futurama	18:00 Big Brother's Little Brother
18:30 Reporting Scotland	19:00 Earranan	18:30 ITV Evening News and Weather	18:30 Hollyoaks	18:30 Zoo Days	21:30 Bizarre ER	19:30 Return To... Airport	18:30 Futurama	18:30 Friends
19:00 The One Show	19:30 'S Ann An Ile	19:00 Emmerdale	19:00 Channel 4 News	19:00 Five News at 7	22:00 EastEnders	20:00 Hidden Histories	19:00 The Simpsons	19:00 Hollyoaks
19:30 EastEnders	20:00 Springwatch Home Movies	19:30 Countrywise	19:55 3 Minute Wonder: On the Ward	19:30 Highland Emergency	22:30 The Real Hustle on Holiday	20:30 Moon Landing: Days That Shook the World	19:30 The Simpsons	19:30 My Name Is Earl
20:00 Celebrity MasterChef	21:00 Krok Mandoon and the Flaming Sword of Fire	20:00 The Bill: To Die For	20:00 The Home Show	20:00 Paul Merton in India	23:05 Family Guy	21:00 Crude Britannia: The Story of North Sea Oil	20:00 The Real A & E	20:00 Friends
21:00 Occupation	21:30 That Mitchell and Webb Look	21:00 Wild Islands	21:00 Undercover Boss	21:00 The Mentalist	23:25 Family Guy	22:00 The Secret Life of the Airport	21:00 Bones	20:30 Friends
22:00 BBC News at Ten	22:00 Psychoville	22:00 News at Ten and Weather	22:00 Big Brother Show	22:00 Law and Order: Special Victims Unit	23:50 Ideal	23:00 The High Life	22:00 Lie to Me	21:00 My Name Is Earl
22:25 Reporting Scotland	22:30 Newsnight	22:35 Cops with Cameras	23:10 The TNT Show	23:00 Law and Order: Criminal Intent		23:30 Crude Britannia: The Story of	23:00 Bones	21:30 The Big Bang Theory
22:35 Question Time	23:00 Newsnight Scotland	23:35 Politics Now	23:40 Derren Brown: Trick or Treat				23:30 Fonejacker	
23:35 This Week	23:20 Cricket: World Twenty20							

Programme selection screen, programme files uploaded on day of broadcast

2. Adding appreciation index (1-10), effort to watch, amount of viewing a program









In the list below are all the programmes that you watched yesterday for at least 5 minutes. Could you please rate each of these programmes with a mark out of 10, where 10 is the highest score

	1	2	3	4	5	6	7	8	9	10
EastEnders (BBC 1/19:30)	<input type="radio"/>									
Wild Islands (ITV 1/21:00)	<input type="radio"/>									
Krod Mandoon and the Flaming Sword of Fire (BBC 2/21:00)	<input type="radio"/>									

And how much effort did you make to watch each of the programmes?

	Watched because the TV was on	Made some effort to watch	Made a special effort to watch
EastEnders (BBC 1/19:30)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wild Islands (ITV 1/21:00)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Krod Mandoon and the Flaming Sword of Fire (BBC 2/21:00)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Standard questions asked for all programmes seen

And how much of each programme did you watch yesterday? Please mark on the scale below, where 1 means that you watched hardly any and 10 means that you watch all of it.

	1, Watched hardly any	2	3	4	5	6	7	8	9	10, Watched all of it
EastEnders (BBC 1/19:30)	<input type="radio"/>									
Wild Islands (ITV 1/21:00)	<input type="radio"/>									
Krod Mandoon and the Flaming Sword of Fire (BBC 2/21:00)	<input type="radio"/>									




2. Respondents agree/disagree to descriptive statements, enter likes and dislikes.

The following questions are about the programme **EastEnders** on BBC 1 at 19:30

	Agree strongly	Agree slightly	Disagree slightly	Disagree strongly	No strong views
It is the kind of programme I would talk to other people about...	<input type="radio"/>				
This programme felt original and different from most other TV programmes I've seen	<input type="radio"/>				
This programme was inspiring?	<input type="radio"/>				
This programme encouraged me to do something new?	<input type="radio"/>				

In-depth questions for selected programmes
 Questions per channel, genre or specific title
 Optional: programme or broadcast specific questions

What was your opinion of the storyline?

This programme had a very good storyline This programme had a fairly good storyline This programme had a fairly bad storyline This programme had a very bad storyline Does not apply/don't know

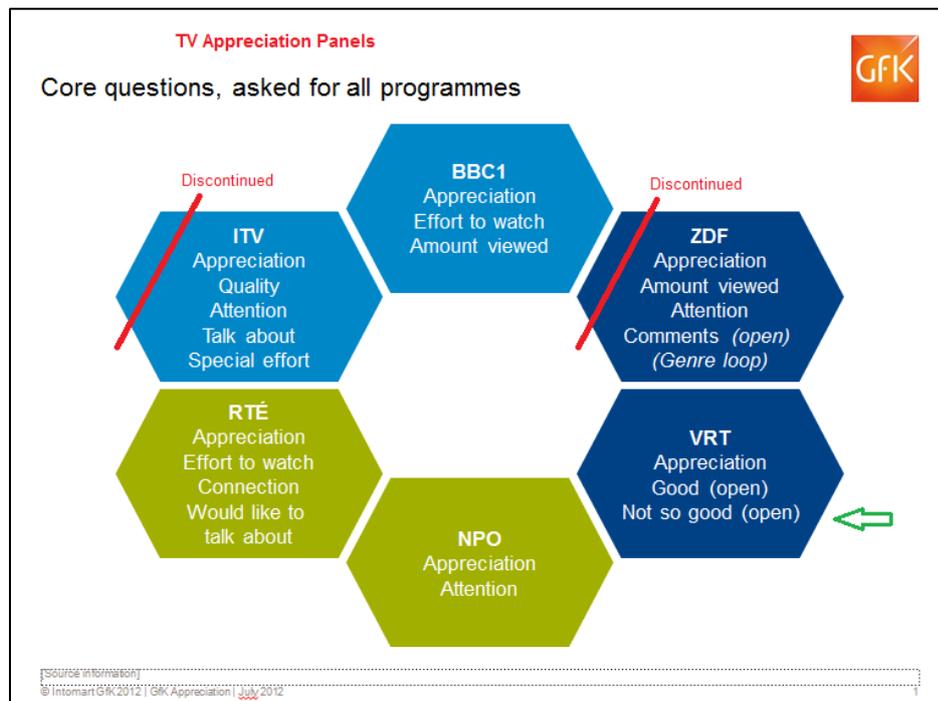
What was your opinion of the characters in it?

It had very good characters It had fairly good characters It had fairly bad characters It had very bad characters Does not apply/don't know

What did you think about this programme? Please write in what, if anything, you liked or didn't like about it.

Open ended question

Note: For expressing „likes“ and „dislikes“ (the so-called „*verbatim*“) it is advisable to use *two* separate text fields. This has already been realized by VRT/ Belgium.



Here are a few examples of general and genre-specific questions used in TV appreciation panels:

All genres except film & children

- This was a high quality programme.
- It is the kind of programme I would talk to other people about.
- Did you feel you learnt anything from watching this programme?
- This programme was thought-provoking.
- Would you recommend this programme to a friend? Please give a mark out of 10, where 10 indicates 'definitely yes', and 1 indicates 'definitely not'.

All genres except film, children & news

- This programme felt original and different from most other TV programmes I've seen.
- This programme was inspiring.

Drama (no film)

- What is your opinion of the storyline?
- What is your opinion of the characters in it?

Entertainment

- How entertaining did you think this programme was?
- And how much did you like the presenter(s), if there were any?

News

- How trustworthy did you think this programme was?
- How impartial did you think this programme was?
- What did you think about the way the programme was presented?

Discussion

The author contends that it would be more efficient to include appreciation in the people meter itself instead of collecting all sorts of „qualitative“ information by a special TV Appreciation Panel. Day-by-day „qualitative“ questionnaires are bound to produce lots of redundant data (which – by the way – become available only two days after transmission). It is much cheaper to conduct *ad hoc* qualitative online research to find out details about viewers' attitudes towards certain programs.

At the moment, there is need for a completely new generation of TAM meters because of digital signal transmission plus internet-based and mobile TV reception. „Sound matching“ and „watermarking“ are the catchwords for the new technologies. Portable recording devices such as „media watches“ and „pagers“ are among the instruments required to meet the new challenges. Such devices are already being used in Kasachstan; and a Moscow based firm is also trying to produce such cutting-edge apparatus. In the view of the author it would be a real pity if these new meters would not provide for appreciation!

The need for a TV Appreciation Panel

If, however, a TAM system does not provide (or has de-activated) the feature of appreciation in the people meters, the TV industry should certainly decide to commission a TV Appreciation Panel. Such a system has several advantages.

First of all, a TV Appreciation Panel delivers quality data in the form of *numerical and verbal TV metrics*. Secondly, it enables researchers to include *ad hoc questions* from time to time - such as testing TV personalities. The third advantage of an online operated panel is the possibility to extend quality research also to *radio and internet*.

When negotiating an Appreciation Panel, a few modifications should be considered. As shown below, it is suggested to use simple „checkmark“ questions for determining programm „properties“ („entertaining“, „informative“, „practical“ „suspenseful“, „well presented“, „high quality“ „none of these“ etc.) The proportion of such properties will explain *why* the respective Appreciation Index was given. In contrast to the evaluation of the „*verbatim*“ which requires additional effort (e.g. text analysis software) these data are delivered automatically by the system. Furthermore, if the *same* „properties“ are being asked across *all* program types, this will facilitate the creation of a *reception quality catalogue* or „*genre mapping*“ which in turn could be used to optimize program schedules. Sometimes program makers are not fully aware of factual audience tastes. For example, TV programs about animals are not only considered „informative“, but also „entertaining“ and „useful“.

Basic Function of a TV Appreciation Panel

1. Presenting yesterday's programs
 2. Highlighting the programs watched
 3. Rating the programs watched (from 0 to 10 or along any other scale)

In the list below are all the programmes that you watched yesterday for at least 5 minutes. Could you please rate each of these programmes with a mark out of 10, where 10 is the highest score.

	1	2	3	4	5	6	7	8	9	10
neighbours BBC 1 / 19:15 (3)	<input type="radio"/>									
BBC News and Weather BBC 1 / 19:00 (3)	<input type="radio"/>									
Masterchef Goes Large BBC 1 / 19:00 (3)	<input type="radio"/>									
The Singing Detective (Nov 7 19:15 (3)	<input type="radio"/>									
EastEnders BBC 1 / 19:00 (3)	<input type="radio"/>									

Please rate the programs you have watched yesterday by giving points from 0 = very bad to 10 = very good

Evening News	6
Society Magazine	4
Primavera Health Magazine	7
King of Queens	3

Alternative: Numerical fields

Possibility: Use of a slider

Move the slider to indicate what you thought of the program

entertaining	<input type="range"/>	boring
good presenter	<input type="range"/>	bad presenter
nice pictures	<input type="range"/>	dreary
informative	<input type="range"/>	not informative

TV Appreciation Metrics

Basically, a TV Appreciation Panel offers:

- an appreciation score for each program viewed,
- „likes“ and „dislikes“ in the form of text entries, and
- a number of special properties such as „attention to a program“ or judgement of certain properties like „informative“, „entertaining“ etc.

To this end, a program schedule is offered in which panelists highlight the program(s) viewed „yesterday“. They then enter appreciation and other judgements.

Only a large panel of Internet users makes it possible to ask a daily sample to judge their use of media (TV, radio, Web)

Program Properties, Likes and Dislikes

Please check the properties which describe the programs you viewed.

	well moderated	informative	entertaining	highbrow	practical	suspenseful	none of these
Evening News	<input type="checkbox"/>						
Society Magazine	<input type="checkbox"/>						
Health Magazine	<input type="checkbox"/>						
King of Queens	<input type="checkbox"/>						

Applying properties stated by simple check

What did you think about this program?
 Likes in what you liked: [text area]
 And what didn't you like?: [text area]

Likes and dislikes filled in separately for easier evaluation

As shown above, three types of questions can be used to record appreciation. In spite of the possibility of „straight-lining“ (entering the same score in all lines), the matrix question with radio buttons is the most popular one.

Appendix 2

BBC TV Quality Measures (2011)

Appreciation Index (AI)

The AI (or Appreciation Index) is the main currency inside the BBC for measuring how much the audience appreciates and enjoys BBC programmes. The Appreciation scores are collected via the 'Pulse' panel; an online panel of 19,000 adults aged 16+ (representative of the UK) run by an independent research agency GfK. The panel enables the BBC to collect over 5,000 responses a day on TV and Radio programmes.

AI Data

The BBC now holds over five years of AI data from this panel (and has experience of this currency going back to the early days of television) which has enabled the organisation to build up robust benchmarks along with a good understanding of what makes for a good or bad score. Following an extensive study undertaken by the BBC into how people score programmes and what variables help to make up an AI score, it was found that one of the leading factors in determining an AI score was whether a programme was deemed to be "high quality". Other factors such as the amount of effort people take to view the programme or whether viewers found it entertaining were also important. The AI score is now the lead metric in assessing programme quality within the BBC. This said, although it is the lead metric it is not the only benchmark for quality, the BBC uses a number of other tools alongside AI such as whether the programmes are different from other TV programmes in the market, whether they are fresh and new or, in the case of factual programmes, whether the viewers learnt anything from it. AI scores are also judged in the context of the type of programme it is and whether it is a new or established programme - in fact, there are many factors that can influence AI scores and it is important to understand the context of these scores in order to interpret them confidently.

The main factors that affect the AI score

- o **Type of programme** - different types of programmes achieve different levels of scores. For example a score for a sport programme on TV can sometimes be dependent on whether or not you support the team that won or lost. In addition, scores can vary according to whether the programme is new or long running, for example, new comedy typically starts with a lower average score as viewers become accustomed to the characters/sketches and this then tends to increase over time.

- o **Changes in the TV market** - the introduction of expanded choice via digital TV, introduction of HD and the ability to watch what you want when you want via PVRs such as Sky Plus and catch-up services such as BBC iPlayer, has been helping to drive up the quality of the experience and hence, AI scores. For example, people

typically give programmes that they have recorded or have watched in HD a higher score than the same programme watched live or in standard definition. This explains why BBC HD has a higher average AI score than other channels.

o **Size of Audience** - niche, targeted programming, even though attracting a limited number of viewers, can often score very highly, as it is actively chosen and appreciated by the viewer; programming aimed at a wider audience, attracting incidental viewers as well as those keen to see the programme, may score lower as it is scored by a wider, more varied group of people. This explains why BBC Three and Four have higher average AI scores than BBC One.

The Method:

Daily Pulse (Pulse) is a continuous BBC survey designed to assess what audiences think of the programmes they have seen on TV and listened to on the radio on a daily basis. It is an online survey, recruited and administered online by GfK and has been running since 2005. Key Pulse measures include the Appreciation Index and programme ratings for quality and distinctiveness.

The Pulse is a panel of around 20,000 people who are invited to complete a survey every day to say what they have watched and listened to and what they thought of each programme. Pulse measures a wide range of BBC and competitor stations.

The daily reporting panel is weighted for age, social grade, sex, presence of children, region and the household digital type on a daily basis to ensure it is representative of the UK as a whole.

The main Pulse measure is the AI or Appreciation Index – respondents are asked to mark out of 10 each of the programmes they watched or listened to the previous day, where 10 is the highest score and 1 the lowest. The average of all these marks out of 10 for a programme is then multiplied by 10 to give an Appreciation score (AI). For example, when aggregating scores for all programmes together for BBC TV we get an average score of 8.2, which gives us an AI of 82.

Additional measures include ratings for the quality and distinctiveness of the programmes.

Source: BBC Audience Information January – March 2011

Appendix 3

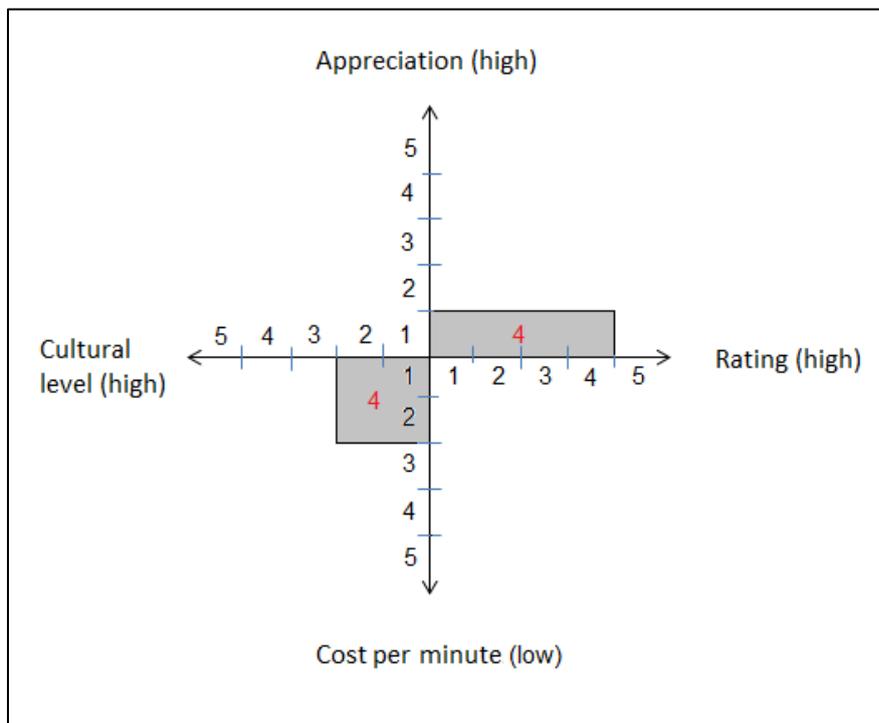
Performance Indicators for (Public Service) TV

In an article published in the German media research journal „Media Perspektiven“ 2(1994), the author suggested the combination of four dimensions of quality control and performance measurement for TV programs:

1. *Rating* (percentage of adult viewers)
2. *Appreciation* (measured in the form of an index and recorded by meter, diary, online panel, or ad-hoc research)
3. *Cost per minute* (preliminary internal and external expenses, indexed over the average program cost)
4. Cultural/intellectual level (from a purely commercial to a high cultural standard)

Without going into details, the idea of this system is the following: TV programs are being classified according to their audience size (rating), their perceived quality (appreciation), their cost efficiency (total cost per minute) and their fulfilment of cultural standards. The latter is based on a collective assessment by program makers. (Other authors have suggested to measure this dimension by means of an analysis of program reviews by public critics.)

Each factor is expressed by a range of 5 classes (very high, high, medium, low, very low). Here is an example:



In spite of a satisfactory rating, the program in this example has a **low overall performance** due to its low appreciation, its low esteem and its relatively high cost per minute. The larger the total area, the better the performance.

Die Programmbewertungspanel der GfK. Von Dr. Lex van Meurs und Erica Simpson, Intomart GfK, Niederlande

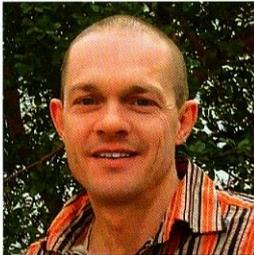
Noten statt Quoten

Reichweitenmessung und Quoten liefern Fernsehmachern seit Langem wichtige Informationen darüber, welche Sendungen beim Zuschauer ankommen und wo die breite Masse abschaltet. Doch jenseits der Quoten bleiben viele Fragen offen. Warum entscheiden sich die Zuschauer für ein bestimmtes Fernsehprogramm? Warum lieben sie die eine Daily Soap und hassen die andere? Und was wünschen sie sich von den Programmverantwortlichen? Antworten auf die Frage nach dem Warum geben die Appreciation Panels, die Programmbewertungspanel der GfK.

Das Prinzip der onlinebasierten Appreciation Panels ist schnell erklärt: Die Panelteilnehmer wählen aus einer von der GfK zur Verfügung gestellten, virtuellen Fernsehzeitschrift die Sendungen aus, die sie am Vortag gesehen haben. Für jede dieser Sendungen füllen sie einen Online-Fragebogen aus und vergeben abschließend eine Note. In Anlehnung an das Schulsystem der Niederlande – dem

Ursprungsland des Panel-Bewertungssystems fürs Fernsehen – können die Zuschauer zwischen den Noten 1 (sehr schlecht) und 10 (sehr gut) wählen und so ihr Gesamturteil abgeben.

Die GfK betreibt Appreciation Panels in den Niederlanden, Großbritannien, Deutschland, Irland und Spanien mit jeweils 500 bis 5.000 Teilnehmern. Die Intomart GfK in den



Dr. Lex van Meurs

arbeitet seit 1990 für die Intomart GfK in den Niederlanden. Mittlerweile ist er Research Director und Mitglied des GfK Cross Media Competence Center. Van Meurs befasst sich unter anderem mit der Fernsehreichweitenmessung, mit Prognosen für Sendungsbewertungen, Plakatwerbung und den Appreciation Panels.

Lex van Meurs studierte Politik- und Kommunikationswissenschaft an der Universität Amsterdam und promovierte zum Thema „Zappen in der Werbepause“. Zudem hat er eine außerordentliche Professur an der Amsterdam School of Communications Research (ASCoR) der Universität Amsterdam, Niederlande.



Erica Simpson

arbeitet seit 2008 als Senior Media Researcher in der Abteilung Appreciation Measurement der Intomart GfK. Erfahrung in der Medientorschung sammelte sie zuvor als Kundenbetreuerin bei AGB Nielsen Media Research Südafrika, wo sie verantwortlich für die Inbetriebnahme der Auswertungssoftware zur Fernsehreichweitenmessung war.

Erica Simpson studierte Marketing an der University of Pretoria, Südafrika, und Projektmanagement an der School of Business Leadership der University of South Africa.



Die Erkenntnisse, die Fernsehprogrammbewertungen liefern, sind für zahlreiche Abteilungen der Fernsehanstalten von hohem Wert

Niederlande befragt täglich rund 3.500 Zuschauer zu ihrem Fernsehgeschmack und kommt somit auf rund 1,3 Millionen Bewertungen pro Jahr.

Darf's noch ein bisschen genauer sein?

Der Online-Fragebogen liest sich in etwa so: „War das Programm Ihrer Ansicht nach von hoher Qualität?“ „Fanden Sie die Sendung unterhaltsam?“ „Fanden Sie den Film spannend?“ Die Antworten geben Aufschluss darüber, wie hoch eine Sendung innerhalb ihres Genres in der Gunst des Zuschauers steht. Schließlich werden die Fragen genau an die Sendungsgattungen angepasst. Beurteilt der Zuschauer einen Krimi, wird er nach dessen Grad an Spannung gefragt. Gibt er seine Meinung zu einer Soap ab, bewertet er den Unterhaltungsfaktor – ein Faktor, der nach der Übertragung eines Promi-Begräbnisses natürlich nicht abgefragt würde. Komplexe Filter steuern den Aufbau der vertiefenden Fragen: Vom Programm-Genre über den Namen der Sendung hin zu speziellen Themen und Inhalten

wird der Panelteilnehmer immer detaillierter befragt. Das Online-Fragebogensystem bietet den Vorteil, dass die Fragen noch in letzter Minute dem aktuellen Programm angepasst werden können. Die Vergleichbarkeit wird dadurch gesichert, dass innerhalb eines Genres jeweils die gleichen Kriterien abgefragt werden. Die Zahl der vertiefenden Fragen ist pro Befragten begrenzt – und die Panelteilnehmer bleiben bei Laune, selbst wenn sie viel gesehen und zu bewerten haben.

Die GfK betreibt Appreciation Panels in den Niederlanden, Großbritannien, Deutschland, Irland und Spanien mit jeweils 500 bis 5.000 Teilnehmern.

Am Ende des Fragebogens werden einige Teilnehmer danach gefragt, ob sie „fünf Minuten extra“ haben, um einen zusätzlichen Fragenkatalog zu beantworten. Die Aufforde-

rung, diese Studien zu unterstützen, geht an Teilnehmer, die bestimmten Kriterien entsprechen. So werden beispielsweise Zuschauer ausgewählt, die einer bestimmten Altersgruppe angehören oder solche, die ein bestimmtes Programm gesehen und es mit einem Wert unter 6 beurteilt haben. Dieser Extra-Fragebogen bleibt für sieben Tage abrufbar und kann von den Panelteilnehmern innerhalb dieser Zeit ergänzt werden.

Abb. 1: Mit Dashboard bewerten Zuschauer die Qualität einzelner Sendungen

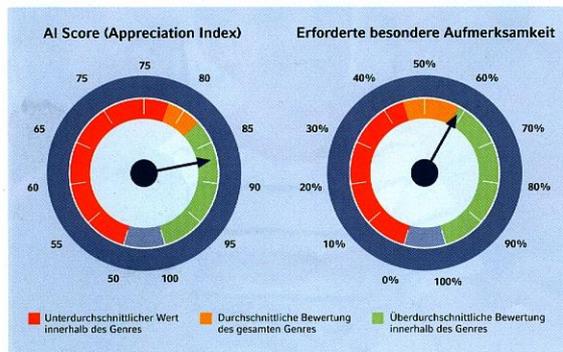


Abb. 2: Programmbewertung im Zeitverlauf



Bewertung früher und heute

Die Idee, Programme vom Zuschauer bewerten zu lassen, ist nicht neu. In den Niederlanden dachte man bereits seit 1965 darüber nach, wie zufrieden die Menschen mit den Programmen tatsächlich sind – und warum. Zu dieser Zeit nutzte man Tagebücher, in denen die Zuschauer ihre Eindrücke festhalten konnten. Dieser qualitative Ansatz floss 1987 in das People Meter mit ein – ein elektronisches System zur Reichweitenmessung, das mit einem „Bewertungsknopf“ für einfachste qualitative Einschätzungen ausgestattet wurde. Im Jahr 2001 entwickelte die Intomart GfK

für den öffentlich-rechtlichen niederländischen Sender NPO das Appreciation Panel System (APS) und setzte damit neue Standards. Seit 2002 fließen diese täglichen Sendungsbewertungen direkt in dessen Programmgestaltung mit ein.

Blick auf die Bewertungseinheiten

Die Erkenntnisse, die Fernsehprogrammbewertungen liefern, sind für zahlreiche Abteilungen der Fernsehanstalten von hohem Wert. So erhalten Redaktionen von Nachrichten- oder Magazinsendungen Informationen zur Qualität ihrer Sendungen – aufgeschlüsselt bis hin zu einzelnen Beiträgen innerhalb einer Sendung. Programmierer können den Erfolg einer Sendung oder die Beliebtheit einer Person oder einer Handlung messen. Und Werbe- und Marketingexperten erhalten Aufschluss darüber, welche Sendungen ihre relevanten Zielgruppen ansprechen und warum. Die Auswertungssoftware der GfK, Dashboard, liefert ihnen zügig eine Momentaufnahme der Bewertungen.

Abbildung 1 zeigt auf der linken Skalenuhr ein Beispiel für eine Sendung, die in ihrem Genre eine überdurchschnittliche Punktzahl (Appreciation Index) erreicht hat. Wäre der Zeiger im roten Bereich, hätte die Sendung im Genre-Vergleich unterdurchschnittlich abgeschnitten. Nach demselben Prinzip zeigt die Uhr rechts an, ob der Zuschauer es als mühsam empfand, einer Sendung zu folgen oder nicht. Mit Dashboard lässt sich auch schnell und einfach sagen, wie sich eine Sendung entwickelt. Abbildung 2 zeigt eine Sendung, die innerhalb von drei Jahren jeweils einige Wochen lief. Die Grafik präsentiert den wöchentlichen Appreciation Index im Verlauf, verglichen mit der Entwicklung des gesamten Genres von 2006 bis 2008.

Noch mehr Informationen dank offener Fragen

Spezielles Highlight der Appreciation Panels sind die offene Fragen. Die Antworten darauf bringen Themen ans Licht, an die die Fernsehforscher vielleicht gar nicht gedacht hätten und durch die die Programmplaner die qualitative Dimension dessen erforschen können, was die Aufmerksamkeit des Zuschauers tatsächlich auf ein Programm lenkt.

Auf die offene Frage „Wie fanden Sie die Sendung xy?“ antworten Zuschauer beispielsweise mit: „Großartiger Moderator, interessante Themen, originell und faszinierend, präsentiert auf eine sehr neutrale, nicht wertende Art. Der Moderator war präsent, ohne die Show zu dominieren.“ Oder: „Nicht ganz so gut wie die vorhergehenden Folgen. Es ist schade, dass die Untertitel nicht besser hervorgehoben waren. Das hätte es leichter gemacht, sie zu lesen.“ Solche Antworten zeigen Redaktionen und Programmverantwortlichen mehr als eindeutig, wo Verbesserungspotenziale liegen. Sie geben auch Aufschluss darüber, wie eine neue Sendung beim Zuschauer ankommt und ob es sich lohnt, sie im Programm zu behalten.

Appendix 5

Lazarsfeld-Stanton Program Analyzer

Ein Gerät, das es Testpersonen in der **Werbeforschung** ermöglicht, ihre spontanen Reaktionen während des Betrachtens eines Objekts wie z.B. eines Werbespots entweder über zwei Druckstifte oder über einen stufenlosen Analoghebel bekannt zu geben.

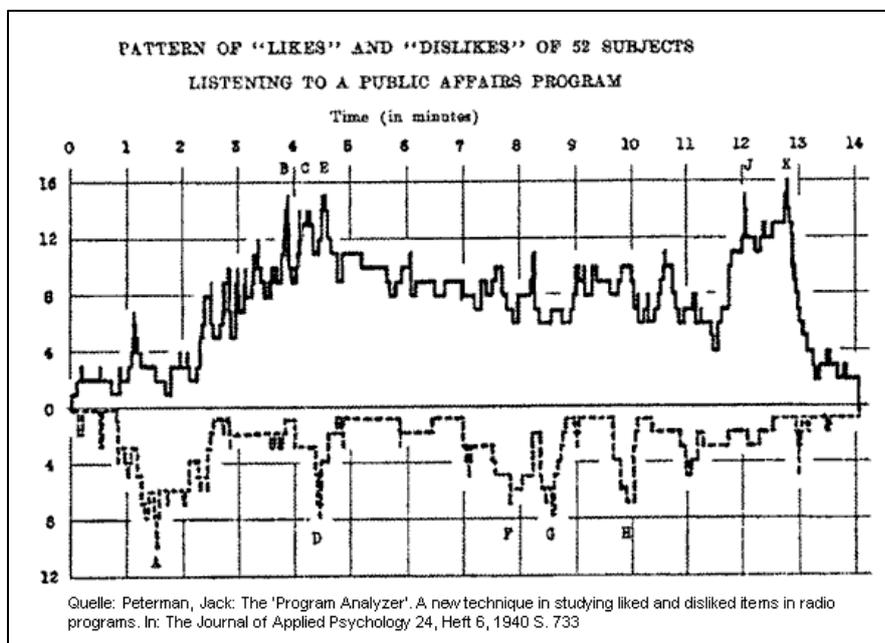
Der Vorteil dieses **apparativen Verfahrens** liegt im ereignissimultanen **Messen**, in der Erfassung von **Emotionen** vor deren bewusster Rationalisierung, sowie in der Vermeidung von Verbalisierungsproblemen.

Der Program Analyzer wurde 1937/1938 im Rahmen des Radio Research Program an der Columbia University von den amerikanischen Hörfunkforschern Paul F. Lazarsfeld und Frank N. Stanton entwickelt. Dabei ging es von vornherein nicht allein darum zu wissen, welche Wirkungen bestimmte Sendungen haben, sondern darum, „*warum* sie sie haben, bezogen auf die Eigenschaften des Programms“. (Lazarsfeld 1940, S. 661)

In der Originalvariante verfügten die Testpersonen über einen grünen und einen roten Knopf. Mit dem grünen Knopf in der rechten Hand drückten sie aus, dass ihnen ein Programm gefällt, mit dem roten in der linken, dass es ihnen nicht gefällt. Die Reaktionen der Testpersonen wurden unmittelbar in eine Grafik übertragen, die im Zeitverlauf eine Kurve des Gefallens bzw. Nichtgefallens zeichnete.

Ähnlich wie der österreichischen RAVAG-Studie liegt dem Einsatz des Program Analyzers das Bemühen zu Grunde, die unüberschaubaren individuellen Rezeptionsvorgänge mit Hilfe statistischer Korrelation überschaubar zu machen.

Abbildung: Der Lazarsfeld-Stanton-Program-Analyzer



Dabei geht es um subjektive Hörerfahrungen, die auf quantifizierbare Lust-Unlust-Empfindungen auf einer Zeitachse reduziert und in ihrer Häufung gewichtet werden. Die Gründe für die positiven oder negativen Reaktionen werden erst auf der Grundlage der Aufzeichnung von den Versuchspersonen erfragt. Nicht die Individuen werden dabei getestet, sondern das Programm hinsichtlich seiner optimierbaren Korrelation mit den Reaktionsmustern der Probanden.

Die Programmwirkung wird dabei auf ein quantifizierbares statistisches Feld bezogen. Quantifizierbare Individualität und technisch geschaltete, radiophone Subjekteffekte werden so in Konstellation gebracht.

Das durch die radiophonen Subjekteffekte konstituierte Publikum ist prinzipiell unüberschaubar und als Publikum erst mit Hilfe der quantifizierenden Verfahren der Hörerforschung überschaubar zu machen.

Der Lazarsfeld-Stanton Program Analyzer wurde bis in die 1950er-Jahre in der **Hörerforschung** eingesetzt. In den letzten Jahren haben sich mit den Real-Time-Response-Verfahren (RTR) in der **Kommunikationsforschung** verschiedene prozessbegleitende, kontinuierliche Messverfahren durchgesetzt.

[zuletzt überarbeitet am 24.11.2008]

Literatur:

- Giessauer, Maria: Die Wirkung des Traubensaft-Fernsehwerbespots der österreichischen Weinmarketinggesellschaft. Diplomarbeit an der Abteilung Werbewissenschaft und Marktforschung. Wirtschaftsuniversität Wien 1999
- Giessauer, Maria: Beurteilung eines Traubensaft-Fernsehwerbespots mittels Programmanalysator. In: transfer – Werbeforschung & Praxis, Heft 4/1999, S. 42-44
- Hollonquist, Tore/Suchman, Edward A.: Listening to the Listener. Experiences with the Lazarsfeld-Stanton Program Analyzer. In: Lazarsfeld, Paul F./Stanton, Frank N. (Hrsg.): Radio Research 1942-43. 2. Auflage. New York 1979, S. 265–334 (1. Auflage 1944)
- Lazarsfeld, Paul F.: Audience research in the movie field. In: Annals of the American Academy of Political and Social Science, 254, 1947, S. 160-168
- Lazarsfeld, Paul F.: Introduction by the Guest Editor. In: The Journal of Applied Psychology 24, 6, 1940, S. 661-664
- Lazarsfeld, Paul F./Field, H.: The people look at radio. Chapel Hill, NC, 1946
- Lazarsfeld, Paul F./Stanton, Frank N. (Hrsg.): Radio Research 1942-43. 2. Auflage. New York 1979, S. 265–334 (1. Auflage 1944)
- Levy, Mark R.: The Lazarsfeld-Stanton Program Analyzer: An Historical Note. In: Journal of Communication Autumn 1982, S. 30-38
- Mark, Desmond (Hrsg.): Paul Lazarsfelds Wiener RAVAG-Studie 1932. Wien 1996
- Peterman, Jack N.: The 'Program Analyzer'. A new technique in studying liked and disliked items in radio programs. In: The Journal of Applied Psychology 24, Heft 6, 1940, S. 728-741
- Schrage, Dominik: Psychotechnik und Radiophonie. Subjektkonstruktionen in artifiziellen Wirklichkeiten 1918-1932. München 2001, S. 309-312

→ verweist auf Synonyme oder ähnliche Begriffe
vgl. verweist auf Antonyme, d.h. Begriffe mit gegenteiliger Bedeutung

Quelle: <http://www.medialine.de/deutsch/wissen/medialexikon.php?snr=4508>

Methods for Studying Mass Media

Written by Peter Simonson and Lauren Archer

Paul Lazarsfeld was known as a methodologist and the Office of Radio Research/Bureau of Applied Social Research used or played a part in developing several new methods for studying mass media. Below is more information on those mentioned in Out of the Question or elsewhere on this website.

CONTENT ANALYSIS:

Content analysis is a systematic method for reading, categorizing elements of, and analyzing some text. The term "content analysis" seems to have been coined by the University of Chicago's Bernard Berelson and Douglas Waples around 1940, though sociologists and political scientists had done systematic studies of newspaper content for at least three decades before. As it developed in the 1930s and during World War II, it was a method for analyzing propaganda, which might aid critics, social scientists, and producers alike. It involved a trained researcher undertaking a careful reading of a text, searching for particular themes, linkages, phrases, or particular words, and then subjecting

them to qualitative or quantitative analysis. Though originally designed for print, it came to be applied to radio, motion pictures, and other communications media as well.

In the film, Joan Goldhamer recalls doing content analyses of popular magazine fiction and discovering the villains were often ethnic minorities. When testing propaganda and morale programs during World War II, researchers would give the film or broadcast a content analysis before bringing in test audiences, who were then asked questions about how they responded—the origins of what today is known as the focus group. Content analysis fed interview guides and helped researchers determine whether government-sponsored communications were having their intended effects on target audiences during the war.

Further Reading:

Klaus Krippendorff, *Content Analysis: An Introduction to Its Methodology*. 2nd ed. (Sage, 2003), Ch. 1: "History"

Thelma McCormack. "Content Analysis: The Social History of a Method," *Studies in Communication* 2 (1982), 143-178.

LAZARSFELD-STANTON PROGRAM ANALYZER:

Paul Lazarsfeld and Frank Stanton developed the Program Analyzer as a way to record viewer reactions as they were happening. Lazarsfeld first experimented with recording reactions in real time back in Vienna. Stanton, as head of audience research for CBS, had created a machine for recording the size of radio audiences. During a conversation together in the late 1930s Lazarsfeld and Stanton swapped stories and decided to pursue inventing a mechanical recording device for gauging audience reactions. Although it went through several versions, the final one accommodated twelve people, each of whom was placed in front of two buttons, a green and a red one. They were instructed to push the green button for positive reactions and the red for negative reactions. Each button was connected to a pen which marked the reactions on recording paper, which flowed in time with the program. The unwieldy machine sometimes gave an electrical shock to those who operated it but allowed researchers to determine what specific moment in a program listeners liked or disliked. With that information in hand, researchers would then follow up with focused interview questions to probe more deeply. This was the origin of the focused group interview, which would diffuse from social science into advertising and become known as the focus group.

The Program Analyzer was put to use at CBS (where it was known as the Stanton-Lazarsfeld Program Analyzer, and by Lazarsfeld's Office of Radio Research (ORR). Critical theorist Theodor Adorno's disgust for the Analyzer in 1939 fueled his disgust for Lazarsfeld's ORR. In the summer of 1942 the ORR shipped an Analyzer to U.S. Army bases, where Robert K. Merton and stenographic assistants were conducting tests for the Office of War Information (OWI), whose Hollywood-produced "Why We Fight" series (directed by Frank Capra) presented a rousing case for American involvement. After the war, the advertising giant McCann-Erickson bought exclusive commercial rights to the machine, where it was enthusiastically supported by former ORR associate Herta Herzog, who had been hired into McCann's research department (and who sent machines to South America and London). By 1954, fifteen different variations on the Analyzer were in use.

Source/Further Reading:

Mark Levy, "The Lazarsfeld-Stanton Program Analyzer: An Historical Note," *Journal of Communication* 32.4 (1982), 30-38.

IN-DEPTH/FOCUSED INTERVIEWING:

In-depth (or focused) interviewing was a method developed at the Office of Radio Research as a way to probe the subjective experiences of audience members exposed to a particular text, campaign, or

event. It grew from a collaboration between Herta Herzog and Princeton psychologist Hadley Cantril. She had been asking questions of radio listeners since the early 1930s when she was a graduate student in Vienna busy writing one of the earliest dissertations on radio listening. In November of 1938, she sought out listeners who had been set into a panic by Orson Wells' famous War of the Worlds broadcast and asked open-ended questions that helped establish the research plan for a classic audience study, *The Invasion from Mars*. Afterwards, Herzog would cultivate the technique further in pioneering studies of women who listened to soap operas as well as in war-related work. Robert Merton also used the technique during the war and codified the method in a 1945 article co-written with Patricia Kendall.

In-depth, qualitative interviewing required highly skilled interviewers who could put respondents at ease, make decisions on the fly, and sensitively probe deeper. It was far more complicated than a simple multiple-choice survey interview and so the people who conducted them-usually women at the ORR-needed to be both trained and talented. The method proved useful as a way to probe how test or real-life audiences responded to programs and campaigns, how particular elements of them might have generated "boomerang effects," and what meanings and "gratifications" radio listeners found in particular programs. Herta Herzog helped transplant the method into the advertising industry when she took a job in 1943 at the McCann Erickson agency and became a director of research. It was a method that could be practiced one-on-one (as in the Kate Smith study, which Joan Goldhamer describes in the film) or in groups-where it was often used in conjunction with the Lazarsfeld-Stanton Program Analyzer. The group form of the interview would later become known as the "focus group," which in the 1970s and '80s became a popular method for marketing research on consumers and political campaigns.

Further Reading:

Susan J. Douglas, *Listening In* (University of Minnesota Press, 2004). Pgs 102, 130, 137, 139-141, 144-48, 165.

Robert K. Merton, "The Focussed Interview and Focus Groups: Continuities and Discontinuities," *Public Opinion Quarterly* 51 (1987), 550-566.

Robert K. Merton and Patricia L. Kendall. "The Focused Interview," *American Journal of Sociology* 51 (1946), 541-557.

David Morrison, *The Search for a Method: Focus Groups and the Development of Mass Communication Research* (University of Luton Press, 1998).

PANEL STUDIES:

The panel was a method developed in the 1930s that involved repeated interviews conducted of the same individuals over time. Paul Lazarsfeld and his colleagues used and refined the method in marketing and media research in the 1930s and '40s, and he and Marjorie Fiske introduced it to public opinion researchers in a 1938 article. Working with a smaller sample of people than public opinion polls typically did, the panel was an economically favorable alternative that also gave investigators a richer array of information, including how opinions might change over time. Lazarsfeld and his women interviewers had used it in a study conducted for *Woman's Home Companion* magazine launched in late 1935, and continuing over several years. Six interviewers recruited 250 women (which would grow to 1,500), whom they spoke with about the magazine and the women's lives and opinions. The magazine made adjustments to its content as a result, and cultivated a relationship with the 1,500 "Reader-Editors" who made up the panel. The method was also utilized in the 1930s by the National Board of Motion Picture Review, which enlisted 250 volunteers (most of them women) to make recommendations for those films most suitable for children under 12. The Office of Radio Research also used it as a tool for studying the effectiveness of propaganda and promotional campaigns, and made it a central part of their classic 1940 and '48 election studies, where they tracked voters' opinions over time. In comparison to focused interviews, panel studies tended to rely more on survey-type questions, though open-ended queries also found a place in them.

Further Reading:

Paul F. Lazarsfeld and Marjorie Fiske, "The 'Panel' as a New Tool for Measuring Opinion," *Public Opinion Quarterly* 2 (1938) 596-612.

Paul F. Lazarsfeld, "'Panel' Studies," *Public Opinion Quarterly* 4 (1940) 122-28.

Paul F., Lazarsfeld, Bernard Berelson, and Hazel Gaudet. *The People's Choice: How the Voter makes Up His Mind in a Presidential Campaign* (Duell Sloan, 1944). [Based on panel interviews in Erie County, Ohio during the 1940 presidential campaign]

Bernard Berelson, Paul F. Lazarsfeld, and William N. McPhee, *Voting: A Study of Opinion Formation in a Presidential Campaign* (University of Chicago Press, 1954). [Based on panel interviews in Elmira, New York during the 1948 presidential campaign]

Quelle: <http://www.outofthequestion.org/Media-Research-of-the-1940s/Methods.aspx#LSPA>