

Semi-qualitative Umfragen:

Die Dezimalskala als universell einsetzbares Instrument der Meinungsforschung

Peter Diem (2018)

Die an der Zahl der menschlichen Finger angelehnte Skala 0-10 sowie die auf die allgemeine Vertrautheit mit der Bedeutung der Prozentrechnung und der Celsiugrade bezugnehmende Skala 0-100 haben sich in der Praxis der Umfrageforschung als sehr brauchbare Messinstrumente erwiesen. Als breite Skalen wirken sie der „Flucht zum Mittelwert“ entgegen, die gerade bei der in Österreich üblichen Notenskala 1-5 gerne auftritt.

Während die Zahl 0 (Null) die geschlossenen Fäuste und damit „keinerlei“ Wert symbolisiert, zeigt die Zahl 10 (Zehn) das das Gesamt aller ausgestreckten Finger beider Hände und damit einen Maximalwert an.

Umgekehrt sind „0 Prozent/Grad“ ein leicht verständlicher Minimalwert (etwa beim Alkoholgenuss oder als Gefrierpunkt), während „100 Grad/Prozent“ das höchste erreichbare Maß (etwa beim Kochpunkt des Wassers oder als Inbegriff voller Leistung) symbolisieren.

Während die 100er-Skala die Vorgabe eines Maßstabes oder eines kalibrierten „Sliders“ erfordert, kann die Skala 0-10 leicht als „Punkteskala“ vorgegeben werden (gelernter Effekt: „Der Kandidat hat 10 Punkte“). Sie ist also etwas einfacher in der Anwendung.

In der Folge wird die Brauchbarkeit und Aussagekraft beider Skalen an Hand von realen Ergebnissen der Umfrageforschung aus den verschiedensten Sachgebieten dargestellt – von der Politik- bis zur Konsumforschung.

→ **Beachte:** Bei zahlreichen Auswertungen hat sich gezeigt, dass ein Produkt oder eine Person erst ab einer Punktezahl von **6,6** als „marktreif“ bezeichnet werden kann. Demgemäß kann das Erzielen des arithmetische Mittelwerts von 5,0 Punkten nicht als „durchschnittlicher“ Erfolg gewertet werden, sondern muss als „ungenügend“ bezeichnet werden.

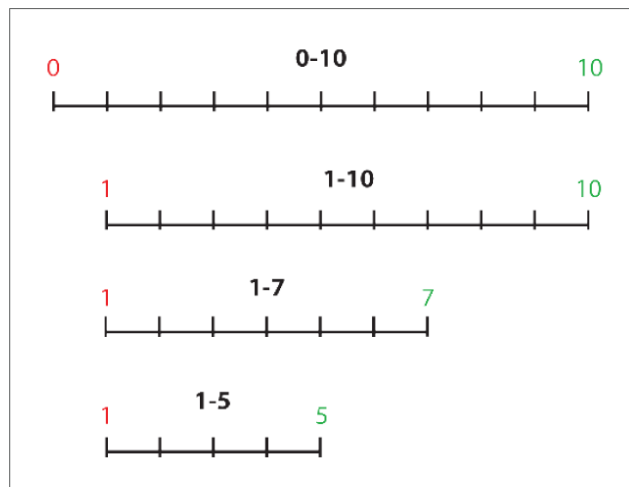
Question Type: Rating Scales

Scales come in many shapes and sizes 1-5, 1-7, 1-10 are common. When measuring customer loyalty, a scale of 0-10 is standard.

Keep in mind that only odd numbered scales allow for a neutral response.

Scales can appear as sliders, which let respondents move them across a spectrum of responses:

Or as a series of horizontal radio buttons that let respondents choose a single option:

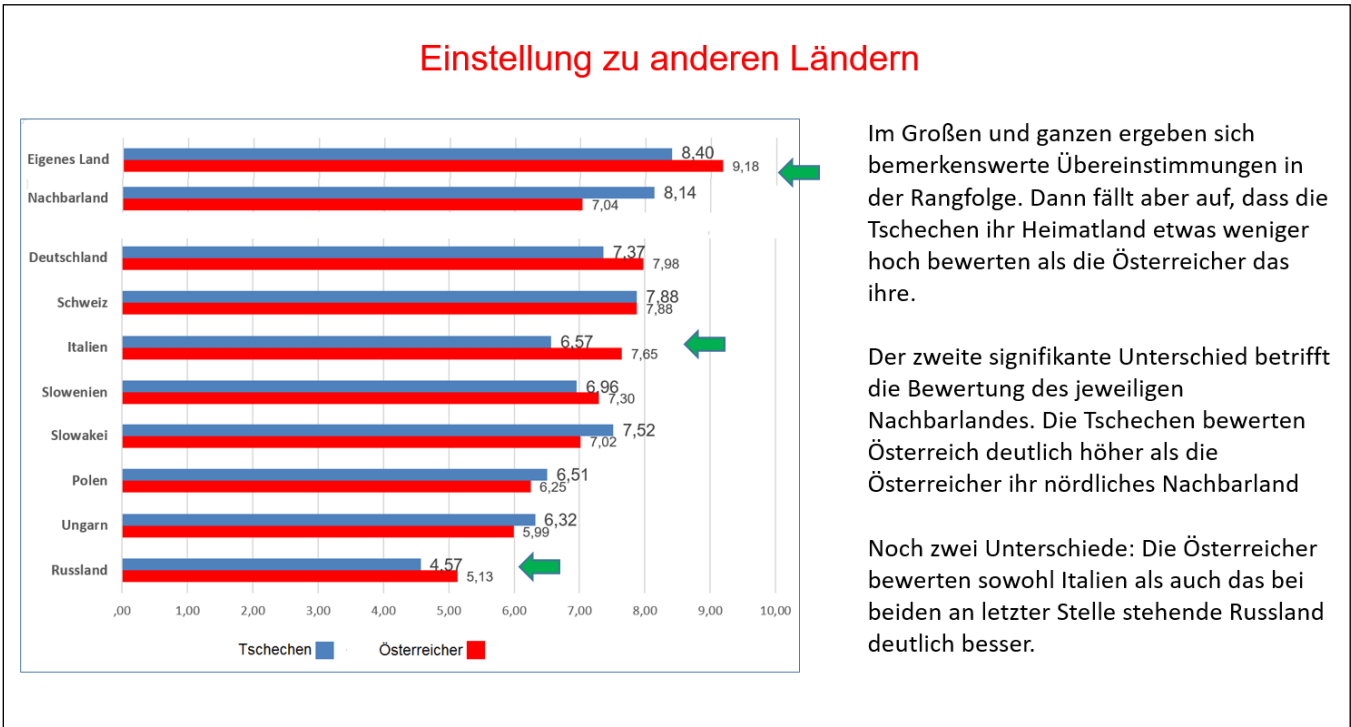


Darstellung aus dem Handbuch der Onlineforschungs-Software SurveyGizmo
<https://www.surveygizmo.com/>

Sympathiemessung (2008)

Wechselseitige Sympathien der österreichischen Bundesländer										
		Wien	NÖ	OÖ	Sbg	Tirol	Vbg	Stk	Ktn	Bgld
	Wien	6,5	7,1	6,7	6,6	6,6	6,5	7,0	7,0	6,6
	NÖ	5,1	8,1	7,8	7,8	8,0	6,7	7,9	7,2	7,0
	OÖ	4,4	6,5	8,4	7,1	7,3	6,5	7,2	7,0	6,2
	Sbg	6,1	6,3	8,2	8,9	8,3	7,6	7,6	7,8	7,5
	Tirol	6,1	6,6	7,4	7,3	9,1	6,4	7,6	6,9	6,4
	Vlbg	4,2	4,8	6,5	6,7	7,7	9,3	5,8	5,8	5,8
	Stk	5,1	6,1	6,7	7,5	7,7	7,1	8,9	7,4	6,5
	Ktn	4,6	6,1	6,8	7,5	7,9	7,5	7,4	8,7	6,2
	Bgld	4,4	6,3	6,8	7,2	7,1	6,9	6,9	6,8	8,8
	Bund	5,3	6,9	7,4	7,4	7,6	6,8	7,6	7,2	6,7
Quelle: Online-Pilot-Untersuchung Juni 2008 (n = 209)										
Skala: 0 = überhaupt nicht sympathisch - 10 = sehr sympathisch										

Die Farben symbolisieren das Hauptergebnis: Natürlich erhält das eigene Bundesland die höchsten Sympathiewerte, wobei Vorarlberg und Tirol mit 9,3 und 9,1 einen sehr hohen Landespatritismus aufweisen, während die Wiener ihrer Heimatstadt mit nur 6,5 Punkten in geradezu spektakulärer Weise kritisch gegenüberstehen. Der starke Mangel an Sympathie, den die Vorarlberger vor allem gegenüber Wien und NÖ. empfinden, tritt bei den anderen Bundesländern etwas gemäßigter auf, ergibt aber insgesamt einen äußerst schlechten Mittelwert von 5,3. Tiroler und Steirer sind in den Augen ihrer Mitbürger die populärsten Österreicher.



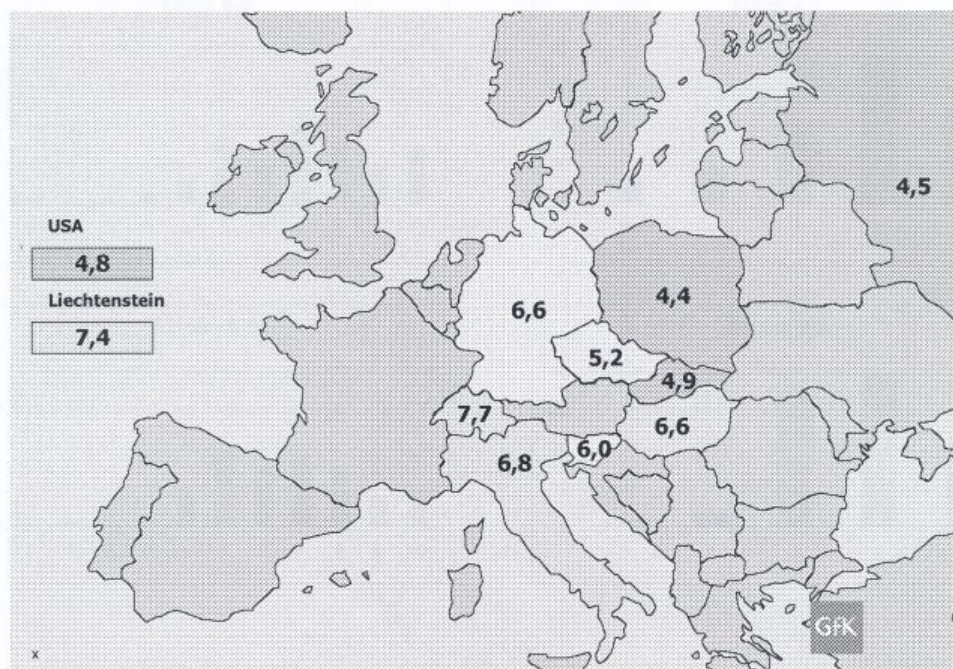
Oben das Ergebnis einer Umfrage unter 165 Österreichern und 68 Tschechen im Februar 2017

Die Sympathie für die Nachbarstaaten Österreichs

16. Juli 2003

Frage: Wie sympathisch sind Ihnen diese Länder?
n = je 200, Juli 2003

Skala	0-10	1-5
Schweiz	7,7	2,0
Liechtenstein	7,4	2,0
Italien	6,8	2,3
Ungarn	6,6	2,3
Deutschland	6,6	2,4
Slowenien	6,0	2,3
Tschechien	5,2	2,9
Slowakei	4,9	3,0
USA	4,8	3,3
Russland	4,5	3,2
Polen	4,4	3,2



Wenn auch die Samples nicht ganz vergleichbar sind (das Sample 2003 war repräsentativ für die österreichische Bevölkerung, während die Stichprobe 2017 praktisch nur Akademiker umfasste), so kann doch festgestellt werden, dass sich die Einstellung der Österreicher zu Deutschland, Italien und Slowenien, vor allem aber zu Polen und der Slowakei im Verlauf von eineinhalb Jahrzehnten offenbar wesentlich verbessert hat. Nur bei Ungarn scheint dies nicht der Fall zu sein.

Messung der Attraktivität unterschiedlicher Begriffe (1998)

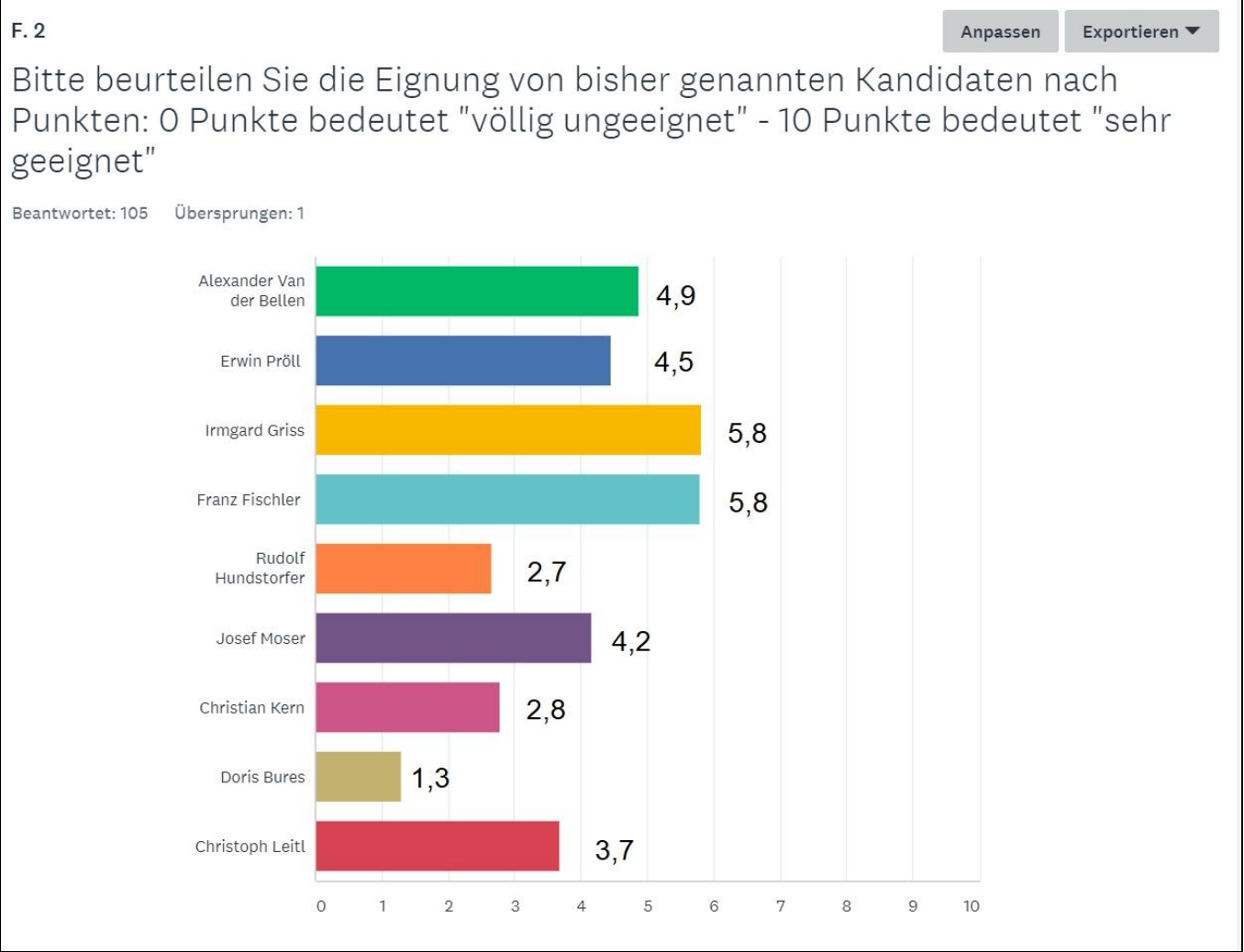
→ Beachte die teilweise großen Unterschiede in der Beurteilung nach dem Alter

Die Attraktivität von Personen, Fernsehprogrammen, Marken und Ideen - gemessen mit der Skala von 0 = "nicht attraktiv" bis 10 = "sehr attraktiv" im Rahmen einer Repräsentativumfrage bei Österreichern ab 14 Jahre (1998)									
	Total	Junge	Alte	Jung:Alt *100		Total	Junge	Alte	Jung:Alt *100
Treue	8,8	8,4	8,8	95,5	Milka	5,7	5,9	5,5	107,3
Liebe	8,6	8,8	8,1	108,6	Palmers	5,6	5,7	5,4	105,6
Glück	8,4	8,6	8,2	104,9	Stronach	5,6	4,5	6,1	73,8
Kinder	8,1	7,3	8,2	89,0	Honda	5,5	5,4	5,4	100,0
BMW	7,5	7,4	7,7	96,1	PTA	5,5	4,9	6,0	81,7
Ferrari	7,5	7,8	7,1	109,9	Spar	5,5	5,0	5,9	84,7
Porsche	7,4	7,5	7,4	101,4	Th. Gottschalk	5,5	5,4	5,5	98,2
Mercedes	7,3	6,8	7,7	88,3	Renault	5,4	5,2	5,7	91,2
Lesen	7,3	6,6	7,7	85,7	Toyota	5,3	5,2	5,6	92,9
Audi	7,2	7,2	7,3	98,6	Kabel 1	5,3	5,0	5,5	90,9
Sex	7,2	8,1	5,6	144,6	ÖBB	5,3	4,6	5,9	78,0
Geld	7,1	7,2	6,9	104,3	B.Stöckl	5,3	3,8	6,3	60,3
Jaguar	7,0	7,4	6,8	108,8	Toni Polster	5,3	4,7	5,9	79,7
Rolls Royce	6,9	6,7	7,2	93,1	Mazda	5,2	5,0	5,6	89,3
Pro 7	6,9	7,5	6,5	115,4	ZDF	5,2	4,3	6,0	71,7
VW	6,8	6,6	7,1	93,0	3sat	5,2	4,2	6,1	68,9
Hermann Mayer	6,8	6,3	7,3	86,3	Mitsubishi	5,1	4,9	5,3	92,5
R. Fendrich	6,7	5,9	7,0	84,3	Super RTL	5,1	4,8	5,4	88,9
RTL	6,5	6,8	6,4	106,3	MTV	5,1	6,6	3,3	200,0
Niki Lauda	6,5	5,6	7,0	80,0	ARD	5,1	4,3	5,8	74,1
ORF 1	6,4	6,3	6,8	92,6	Erika Pluhař	5,0	3,8	5,8	65,5
Sat 1	6,3	6,4	6,4	100,0	Coca-Cola	4,9	5,8	3,9	148,7
Fernreisen	6,2	7,2	5,6	128,6	Fiat	4,7	4,3	5,0	86,0
Sport	6,2	6,6	5,8	113,8	CNN	4,7	4,8	4,3	111,6
Volvo	6,1	5,7	6,4	89,1	Lotto	4,7	4,3	4,7	91,5
RTL 2	6,1	6,3	6,2	101,6	S. Kirchberger	4,7	4,9	4,6	106,5
ORF	6,1	5,7	6,7	85,1	Citroen	4,6	4,0	5,1	78,4
ORF 2	6,0	5,1	7,0	72,9	Arte	4,5	4,2	4,8	87,5
Opel	5,9	5,5	6,4	85,9	Persil	4,5	3,5	5,5	63,6
Haustiere	5,9	5,7	6,0	95,0	Gösser	4,5	4,3	4,9	87,8
Rover	5,8	5,6	5,8	96,6	V. Rußwurm	4,4	3,1	5,4	57,4
Alfa Romeo	5,8	5,7	6,0	95,0	Skoda	4,1	3,9	4,3	90,7
Eurosport	5,8	5,5	6,0	91,7	Sheba	3,3	2,8	3,8	73,7
Ford	5,7	5,3	6,0	88,3	Lada	3,2	3,0	3,6	83,3
Fernsehen	5,7	5,6	6,2	90,3					
					Mittelwert	5,9	5,6	6,0	93,3

An der Spitze stehen Familie und Lebensglück und den Ende der Neunziger Jahre populärsten Automarken. Die Marke Skoda würde heute weit besser abschneiden. Interessant ist die der Programmplanung entsprechende Positionierung von ORF1 und ORF2 nach dem jeweiligen Alter. Beide Kanäle verfehlen den markenstrategischen „Sollwert“ von 6,6 Punkten, was der Konkurrenz (Pro 7 und fast auch RTL) gelingt.

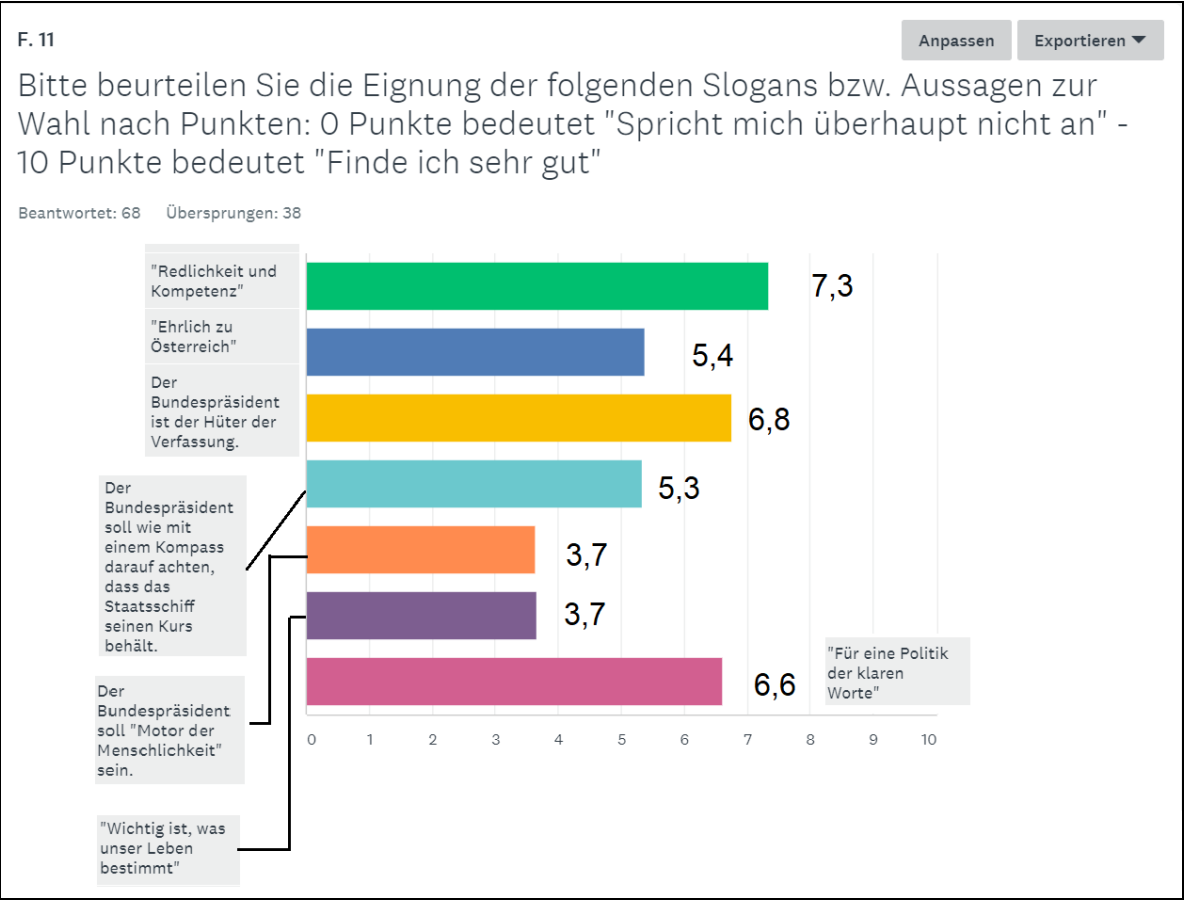
Die Eignung politischer Kandidaten für ihr Amt (2016)

Sehr praktisch ist die Punkteskala 0-10 bei der Beurteilung von politischen Persönlichkeiten. Im folgenden Beispiel aus dem Präsidentschaftswahlkampf 2017 zeigt sich, dass im Grunde keiner der damals genannten Kandidaten als optimal angesehen wurde. Nur Irmgard Griss und Franz Fischler kamen halbwegs in die Nähe der 6,6 Punkte, die für „Marktreife“ erforderlich sind.



Die Umfrage beruhte auf einer Stichprobe von etwa 200 Personen im Raum Wien.

Der Vergleich der Wirksamkeit verschiedener Slogans anlässlich der Bundespräsidentenwahl 2016



Während die Aussage "Hüter der Verfassung" als „stimmig“ angesehen wurde (6,8 Punkte), ohne deshalb als Slogan wirklich geeignet zu sein, zeigt sich ganz deutlich, dass „Redlichkeit und Kompetenz“ ein perfekter Slogan gewesen wäre. Die Aussage „Für eine Politik der klaren Worte“ hatte gerade noch den Sollwert von 6,6 Punkten erreicht.

→ Dieser Umfrage lag zwar keine repräsentatives Bevölkerungsstichprobe zugrunde, doch hat die Erfahrung gelehrt, dass sich ab rund 30 Befragten ermittelte durchschnittliche Skalenwerte auch bei größeren und qualitativ hochwertigen Stichproben nur unwesentlich ändern.

Merkel vs. Schulz - Die Spitzenkandidaten zur Bundestagswahl 2017



1. Oben sehen Sie das Bild Von Frau Angela Merkel, der gegenwärtigen deutschen Bundeskanzlerin. Wenn Sie ganz spontan urteilen: Was sagen Sie zu dieser Politikerin? Drücken Sie Ihre Meinung mit Hilfe von Punkten aus:

> 0 Punkte bedeutet "Das ist eine sehr schlechte Politikerin",
> 10 Punkte bedeutet "Das ist eine sehr gute Politikerin".

Dazwischen können Sie fein abstufen

Durchschnitt: 8,4

Meine Punkte:

2. Wenn Sie nun an die Sympathie für diese Politikerin denken. Wie urteilen Sie da?

> 0 Punkte bedeutet "Das ist eine mir nicht sympathische Politikerin",
> 10 Punkte bedeutet "Das ist eine mir sehr sympathische Politikerin"

Dazwischen können Sie fein abstufen

Durchschnitt: 7,3

Meine Punkte:



9. Oben sehen Sie das Bild von Martin Schulz, dem Spitzenkandidaten der SPD für die Bundestagswahl im September. Wenn Sie ganz spontan urteilen: Was sagen Sie zu diesem Politiker? Drücken Sie Ihre Meinung mit Hilfe von Punkten aus:

> 0 Punkte bedeutet "Das ist ein sehr schlechter Politiker",
> 10 Punkte bedeutet "Das ist ein sehr guter Politiker".

Dazwischen können Sie fein abstufen

Durchschnitt: 7,2

Meine Punkte:

10. Wenn Sie nun an die Sympathie für diesen Politiker denken. Wie urteilen Sie da?

> 0 Punkte bedeutet "Das ist ein mir nicht sympathischer Politiker",
> 10 Punkte bedeutet "Das ist ein mir sehr sympathischer Politiker"

Dazwischen können Sie fein abstufen

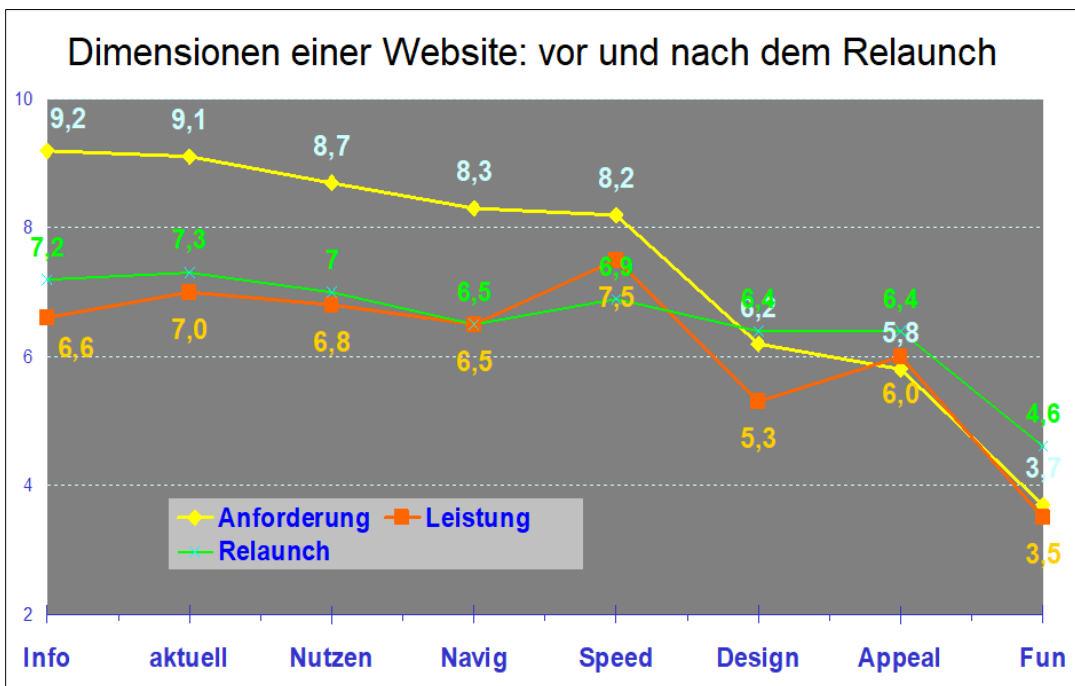
Durchschnitt: 7,1

Meine Punkte:

Angela Merkel führte mit 8,4 Punkten deutlich in der Einschätzung als Politikerin, unterschied sich von Schulz jedoch kaum in der Sympathie.

Die Umfrage wurde im Februar 2017 unter 33 Deutschen durchgeführt, wovon 23 die Frage nach der Beurteilung beantworteten.

Die Skala 0-10 leistet gute Dienste bei der Beurteilung der wichtigsten Dimensionen eines Webauftritts (2000)

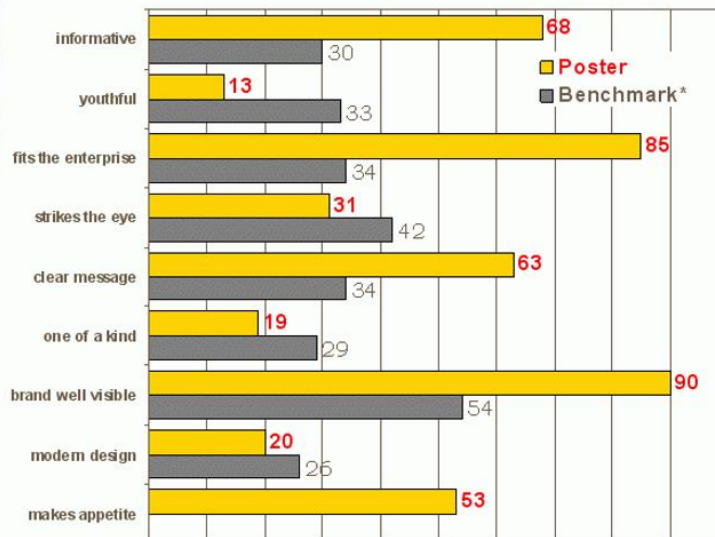


Im gegenständlichen Fall ging es darum, eine relativ „einfache“ Website eines Marktforschungsinstituts durch professionelle Überarbeitung zu verbessern und danach neu ins Netz zu stellen.

Im Rahmen einer Pop-up-Befragung wurde zunächst ermittelt, welche Relevanz die einzelnen Elemente eines Web-Auftritts für die Qualität der Website haben (**Anforderungsprofil** – gelbe Kurve). Es stellte sich heraus, dass, wie zu erwarten, „Informationsgehalt“ (9,2) und „Aktualität“ (9,1) die wichtigsten Dimensionen einer Business-Website sind – gefolgt von „Persönlichem Nutzen“ (8,7), „Benutzerfreundlichkeit“ (8,3) und der zur Zeit der Umfrage noch wichtigen „Ladegeschwindigkeit“ (8,2). Als weit weniger wichtig – typischerweise mit 6,4 knapp unter der „magischen“ Marke von 6,6 Punkten – wurden „Design“ (5,3) und „Anmutung“ (6,0) der Seite beurteilt. Dass eine Informationsseite nur wenig „Unterhaltungswert“ (4,6) haben muss, versteht sich von selbst.

Das Anforderungsprofil der Website wurde ursprünglich besonders in den wichtigsten Dimensionen nur schlecht erfüllt, während die Site in den „unwichtigen“ Dimensionen durchaus punkten konnte.

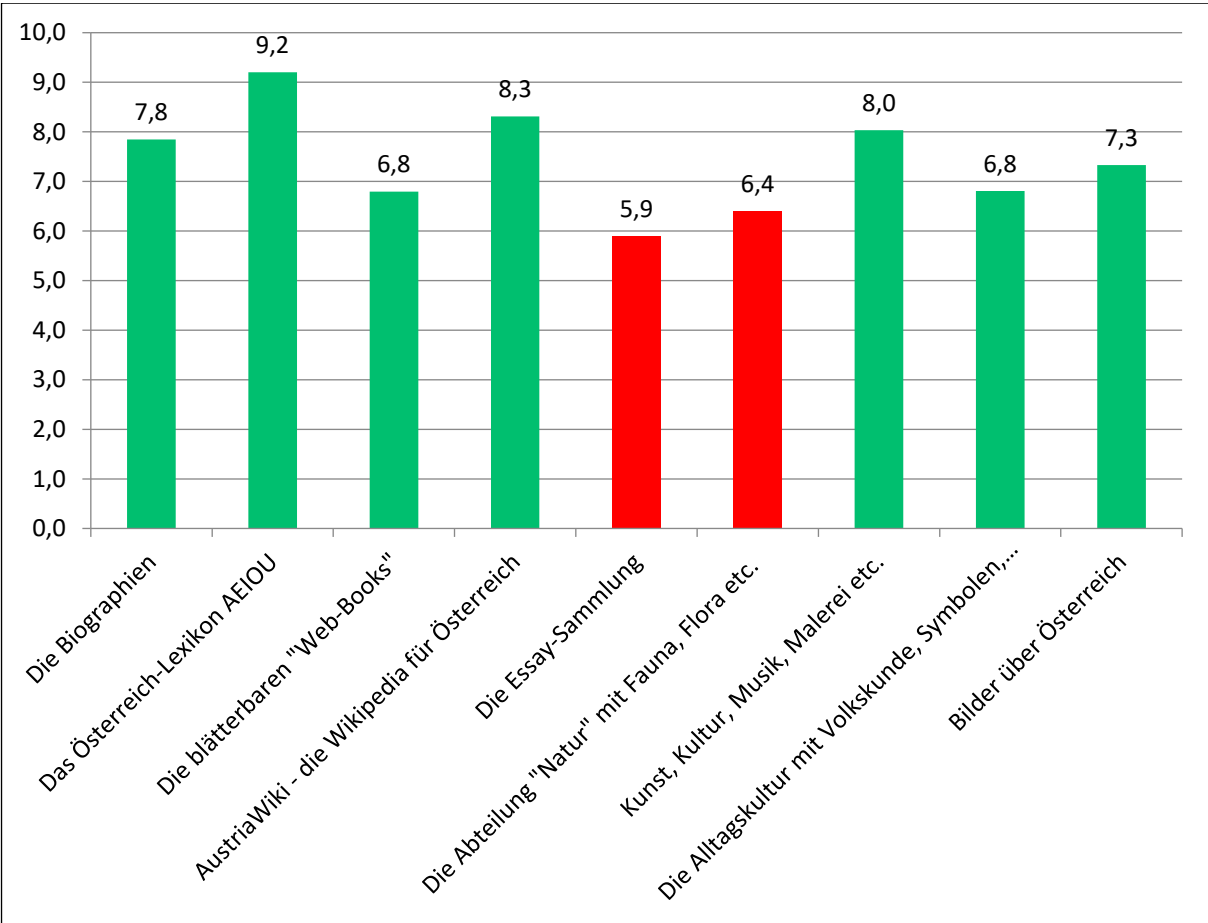
Nach dem **Relaunch** (grüne Kurve) konnte festgestellt werden, dass die Seite insbesondere im Informationsgehalt deutlich besser lag als zuvor (von 6,6 auf 7,2 Punkte gesteigert), was allerdings ein wenig auf Kosten der Ladegeschwindigkeit (6,9 statt 7,5) ging. Als deutlich besser wurde das (professionelle) Design empfunden, das nun den Anforderungswert (6,2) mit 6,4 Punkten sogar übertraf. Auch die weniger wichtigen Dimensionen hatten sich verbessert. Allerdings konnte kein Fortschritt bei der Benutzerfreundlichkeit erzielt werden – die neue Site war ja auch komplexer.



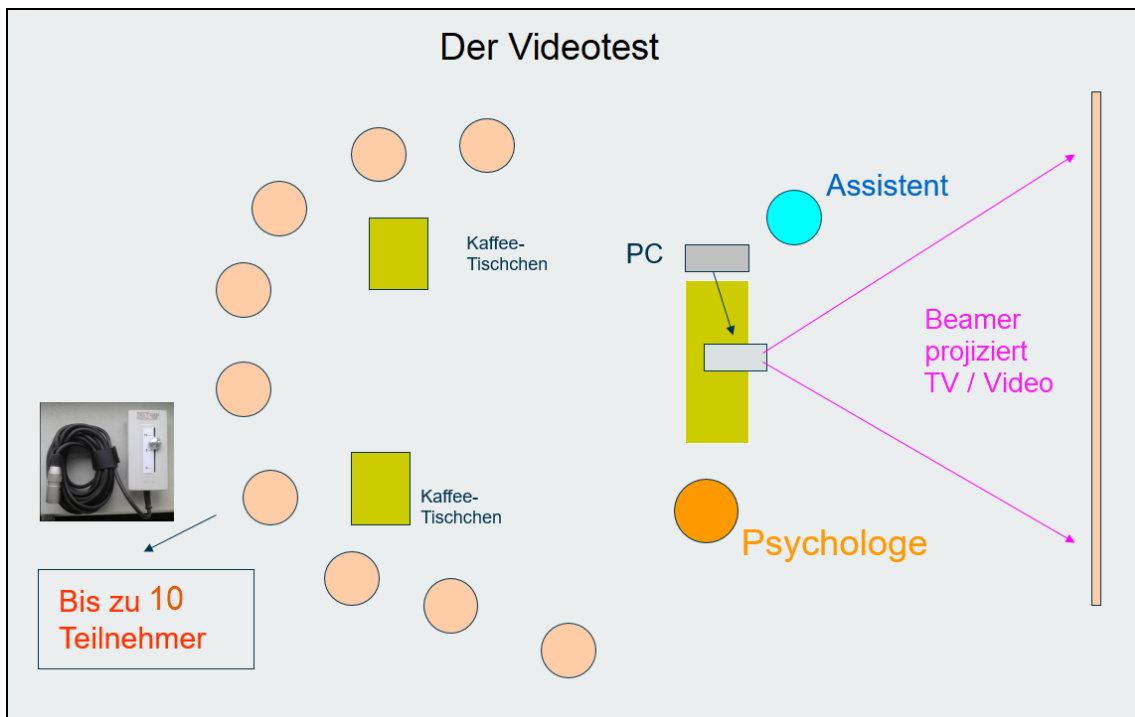
Beurteilung von Webinhalten (2017)

Zum österreichischen Wissensnetz „Austria-Forum“ (<http://austria-forum.org>) wurde 2017 eine Umfrage unter rund 200 Verwendern und Interessenten durchgeführt. Gefragt wurde unter anderem nach dem Interesse an verschiedenen Sachgebieten.

Das Austria-Forum umfasst zahlreiche Sachgebiete. Sie können das jeweilige Sachgebiet durch einen Doppelklick aufschlagen und ansehen. Beurteilen Sie optional bitte die folgenden Sachgebiete nach Punkten. 0 Punkte = "Völlig uninteressant", 10 Punkte = "Sehr interessant "



Skalierung von Videomaterial in Gruppendiskussionen (1998)

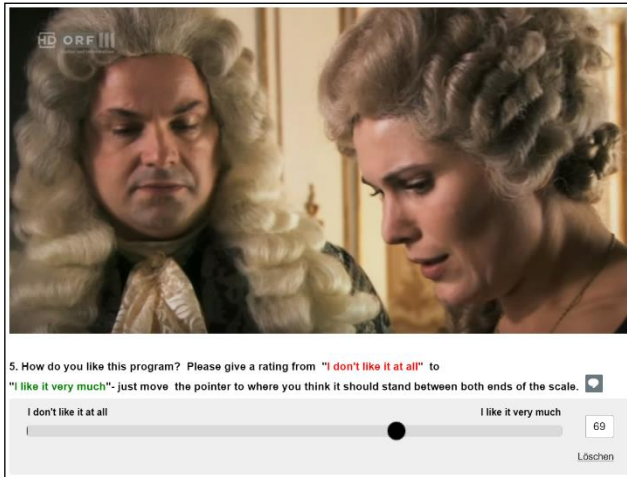


In Kooperation mit dem Marktforschungsinstitut Triconsult <http://triconsult.at> wurde der so genannte "Videotest" entwickelt. Dies ist eine spezielle Form der Gruppendiskussion, in der die Befragten elektromechanische "Slider" bedienen, während sie eine Live-Sendung oder ein Video ansehen. Die mit den Slidern erfassten Daten werden in einen PC eingespeist und auf einer Skala von 0,0 ("sehr schlecht") bis 10,0 ("sehr gut") bewertet. Die Ergebnisse stehen unmittelbar nach der Betrachtung zur Verfügung und können entlang einer Zeitachse angezeigt werden, wodurch die Akzeptanz der Programmteile von Anfang bis Ende angezeigt wird (siehe unten).

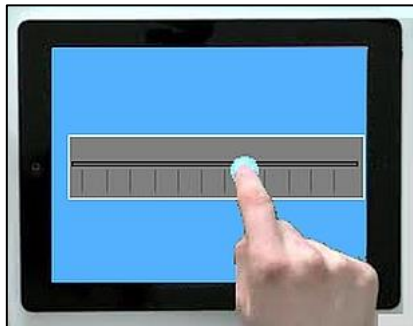
Im Folgenden zunächst Abbildungen eines konventionellen elektromechanischen "Sliders" sowie der einfache Slider, wie er bei „MonkeySurvey“ verwendet wird. Vom marktforscherischen Standpunkt wäre ein etwas lebendigeres Design vorzuziehen.



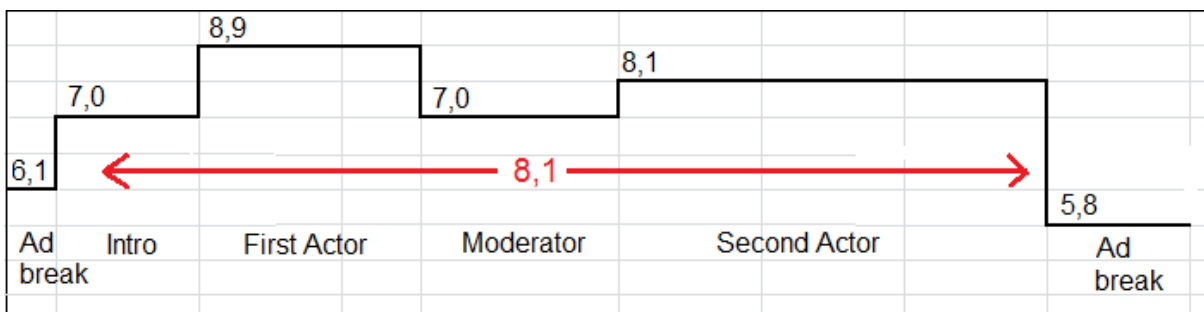
Elektromechanischer Slider



Marktüblicher Slider in der Onlineforschung



Alternative Designs



Fiktives Beispiel für das Ergebnis der Registrierung der Werte einer Anzahl von Schiebereglern, die während der Anzeige eines Fernsehprogramms betätigt wurden. In diesem Fall beträgt der zeitgewichtete Durchschnittswert (oder Median) des Programms (ohne Werbeunterbrechungen) 8,1.

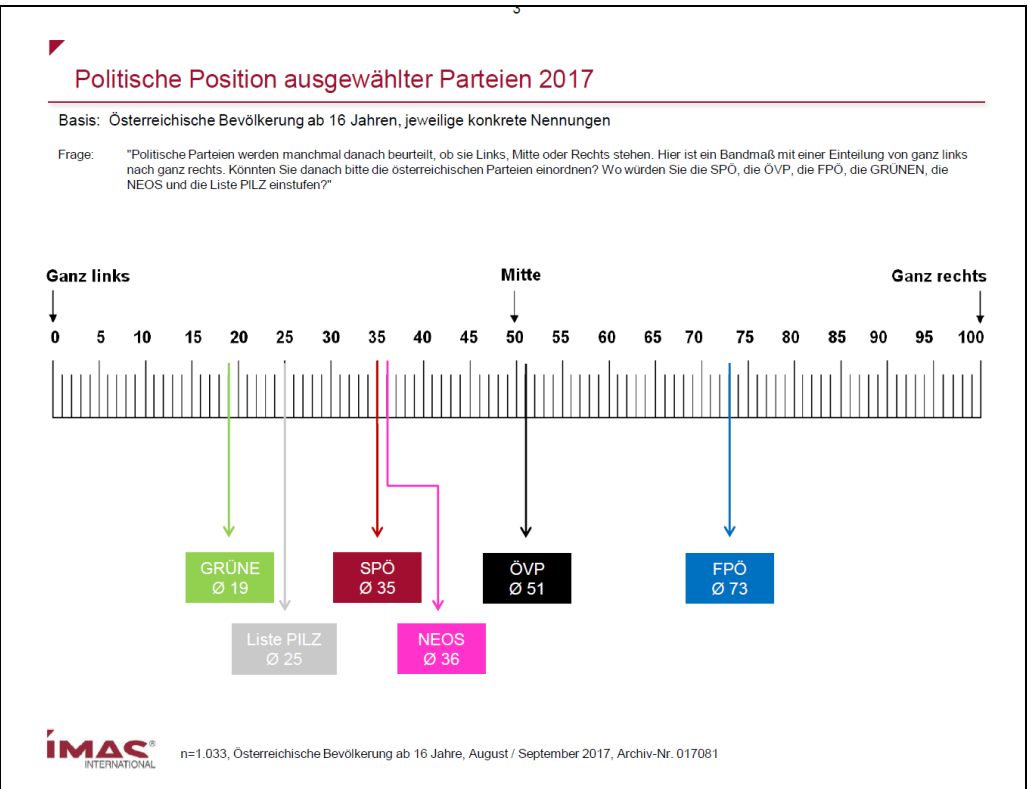
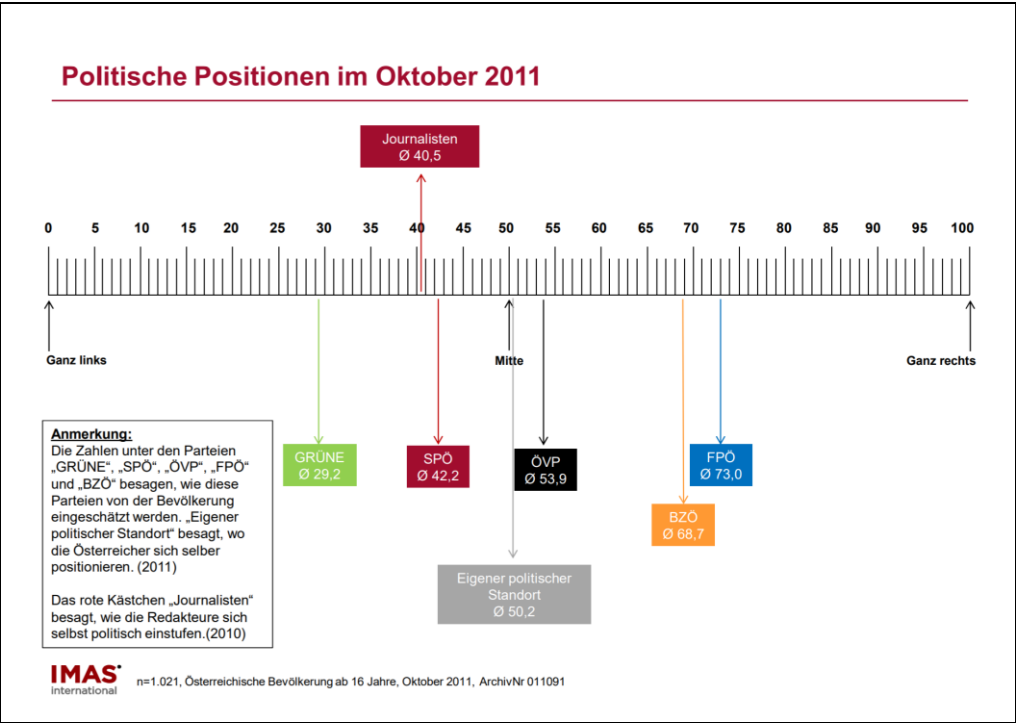
Das Diagramm zeigt, dass der erste Teil des Programms etwas besser empfangen wurde als der zweite und dass die Werbeblöcke die übliche Ablehnung erfahren.

Es sollte in diesem Zusammenhang erwähnt werden, dass der österreichische Soziologe Paul F. Lazarsfeld während seiner Forschungsarbeiten an der Columbia Universität in New York 1937/38 bereits Testpersonen Radioübertragungen durch das Bewegen eines roten und eines grünen Stiftes auf einem laufenden Papierband beurteilen ließ.

→ Eine Würdigung des Programm-Analyzers von Lazarsfeld findet sich im Anhang

Die Verwendung der Skala 0-100 (IMAS/Linz)

Die Denkmuster in Links-Mitte-Rechts sind weiterhin fest im Bewusstsein verankert, die Bevölkerung ordnet die abgefragten Parteien und Themen relativ gleichförmig einer Position auf einer Skala von 0=ganz links bis 100=ganz rechts zu. Entgegen anders lautenden Kommentaren im aktuellen politischen Diskurs hat dieses Denkmuster nicht ausgedient. Die Österreicher positionieren sich selbst im politischen Spektrum mittig (50 auf einer Skala von 0 bis 100), interessanterweise vermutet man aber die Bevölkerung insgesamt deutlich weiter rechts (56). Hier die Entwicklung zwischen 2011 und 2017:

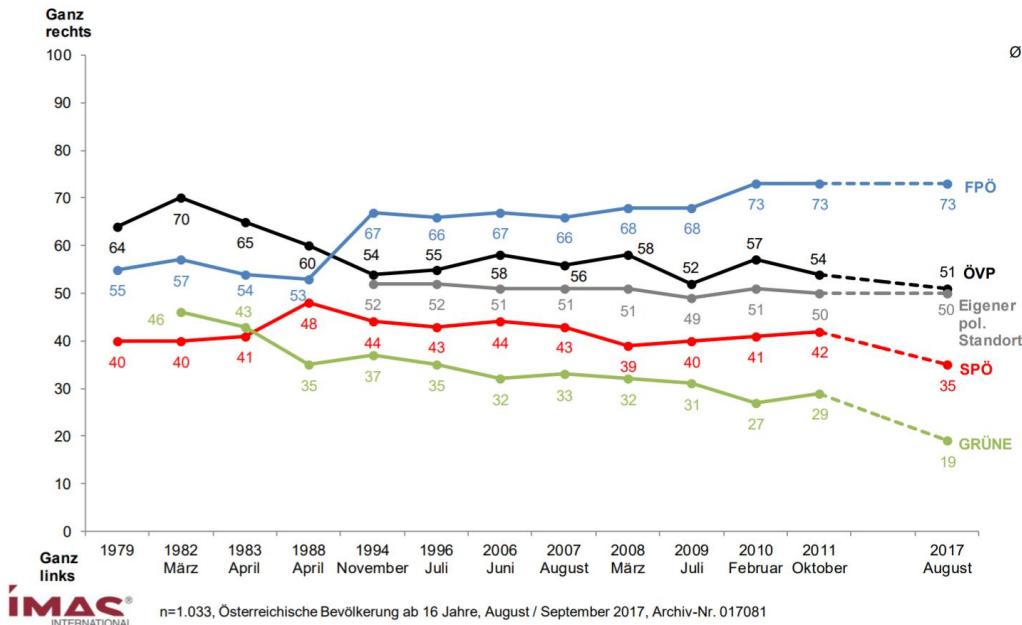


Politische Position ausgewählter Parteien 2017 / eigener politischer Standort – Trend

Basis: Österreichische Bevölkerung ab 16 Jahren, jeweilige konkrete Nennungen

Frage: "Politische Parteien werden manchmal danach beurteilt, ob sie Links, Mitte oder Rechts stehen. Hier ist ein Bandmaß mit einer Einteilung von ganz links nach ganz rechts. Könnten Sie danach bitte die österreichischen Parteien einordnen? Wo würden Sie die SPÖ, die ÖVP, die FPÖ und die GRÜNEN einstufen?"

Frage: "Wie würden Sie Ihren eigenen politischen Standort beschreiben, wo auf diesem Bandmaß würden Sie sich selbst ganz allgemein einstufen?"



Der „eigene“ politische Standort der Österreicher in der „Idealmittte“ hat sich über die Jahrzehnte nicht signifikant geändert. Wie schon in den Neunziger Jahre sehen sich die Österreicher auch 2017 genau in der Mitte des politischen Spektrums.

Während sich die beiden Großparteien und die Grünen in der Sicht der Wählerschaft zunächst nur leicht, in den letzten fünf Jahren aber deutlich „nach links“ bewegt haben, hat die FPÖ ihre seit 2010 bezogene prononciert „rechte“ Position gehalten.

Für die ÖVP scheint dieser langfristige Trend Vorteile zu haben, liegt sie doch im Jahr 2017 mit 51 Punkten am „Idealpunkt“ von 50 Punkten, d.h. genau an der selbstgewählten Position der Wählerschaft.

Für SPÖ und noch viel mehr für die zuletzt ja auch prozentuell abgestürzten Grünen könnte diese Entwicklung bedeuten, dass die Rückkehr an frühere Positionen nur mit einem deutlichen Rechtsschwenk möglich ist – genau das Gegenteil wird aber als sozialdemokratische Wahlstrategie von London bis Wien intern propagiert.

Politische Forderungen im Links-Rechts-Schema

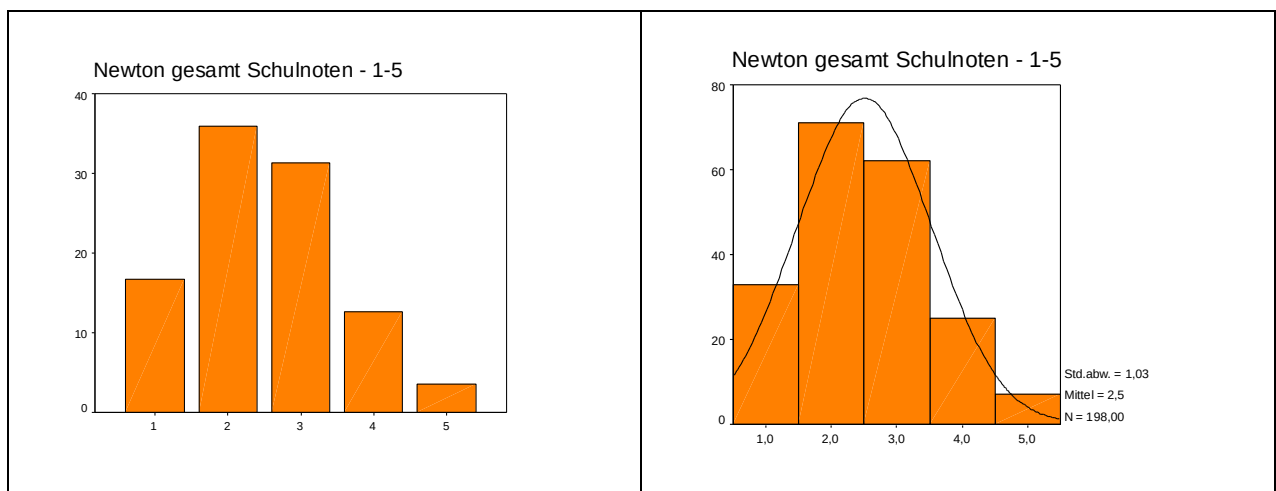
	Positionierung politischer Forderungen auf der Skala 0 = links bis 100 = rechts
0-100	
69	Stopp der Einwanderung
61	Bekämpfung der Kriminalität
53	Wahrung der Interessen Österreichs innerhalb der EU
52	Sparsamer Umgang mit Steuergeld
51	Stärkung des Wettbewerbs in der Wirtschaft
50	Verschärfung des Datenschutzes
50	Senkung der Lohnsteuer
49	Einsatz für den Mittelstand
48	Unterstützung der Interessen der Arbeitnehmer
48	Unterstützung der Interessen der Arbeitgeber
48	Senkung der Mehrwertsteuer
48	Förderung von Familien
47	Förderung der Digitalisierung
47	Erhöhung der Pensionen
46	Verbot der Abtreibung
45	Stärkung des Sozialsystems
45	Privatisierung von Staatsbetrieben
42	Stärkere Zusammenarbeit in der EU
40	Einsatz für Umweltschutz
39	Förderung der Inklusion in den Schulen
39	Einführung einer Ganztagesesshule
37	Senkung der Wochenarbeitszeit
37	Einführung einer Gesamtschule
26	Förderung der Einwanderung nach Österreich
24	Verstärkte Aufnahme von Flüchtlingen
Quelle: IMAS, n=1.033, Österreichische Bevölkerung ab 16 Jahre, August / September 2017	

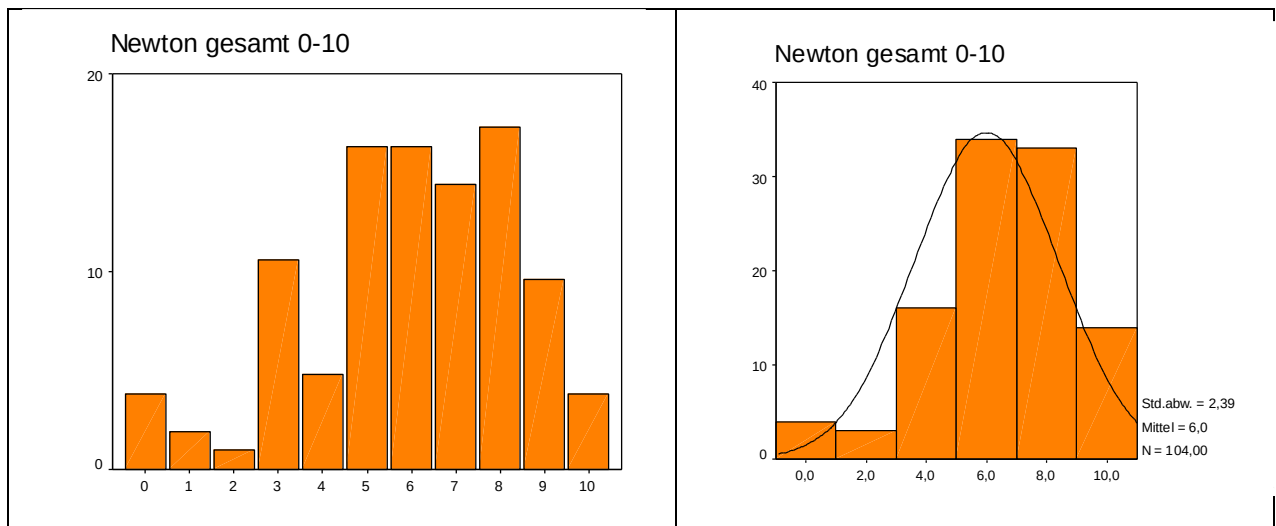
Wie die Einfäbrung der obigen Rangreihe zeigt, hat sich die 2017 gewählte Regierungskoalition genau jene Forderungen auf ihre Fahnen geschrieben, die von der Bevölkerung als „Mitte bis rechts“ eingestuft werden. Umgekehrt kann man die unter der 40er-Marke gelandeten Forderungen durchaus als Programmpunkte der Linksparteien ansehen. Im Falle der Ganztagesesshule dürfte allerdings die Einstufung auf die Nähe des Begriffs zu jenem der Gesamtschule entstanden sein.

Mathematische Beweisführung über den Vorteil der Skala 0-10 gegenüber der Schulnotenskala 1-5.

ewnton gesamt: Schulnoten 1-5			Newton gesamt: Skala 0-10		
	Häufigkeit	Gültige Prozente		Häufigkeit	Gültige Prozente
1	33	16,7	0	4	3,8
2	71	35,9	1	2	1,9
3	62	31,3	2	1	1,0
4	25	12,6	3	11	10,6
5	7	3,5	4	5	4,8
Gesamt	198	100,0	5	17	16,3
			6	17	16,3
			7	15	14,4
			8	18	17,3
			9	10	9,6
			10	4	3,8
			Gesamt	104	100,0

Diagramm

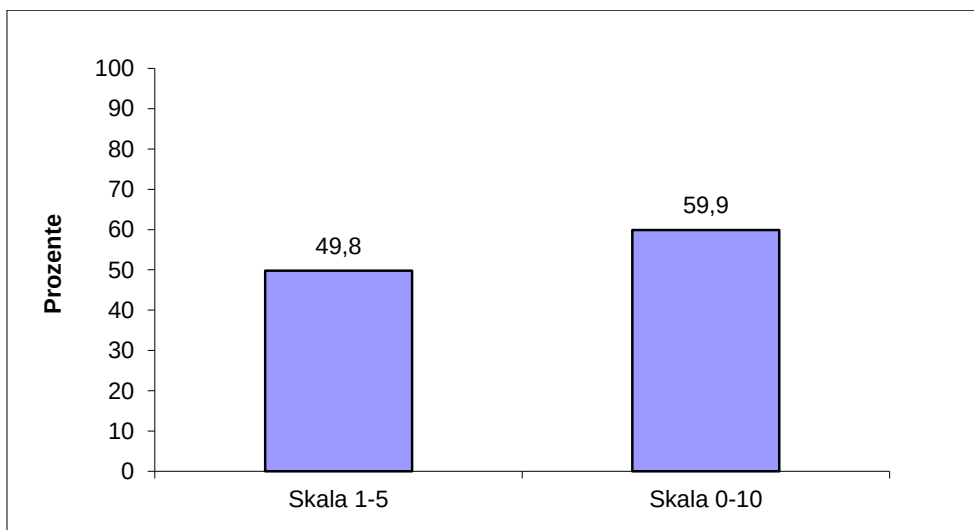




Skalenvergleich:

	Newton gesamt Schulnoten 1-5	Newton gesamt Skala 0-10
N	198	104
Mittelwert	2,51	5,99
Median	2,00	6,00
Modus	2	8
Standardabweichung	1,03	2,39
Varianz	1,05	5,70

Gegenüberstellung der Skalen: wenn 1 bzw.10 = 100%:



Wie man an dieser Darstellung erkennt schlägt die Tendenz zur Mitte bei der breiteren Skala nicht so stark durch. Deshalb werden in der Testtheorie größere und gerade Skalen empfohlen.

Anhang 1 Der Lazarsfeld-Stanton Program Analyzer

Peter Diem

Der **Program Analyzer** ist 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.

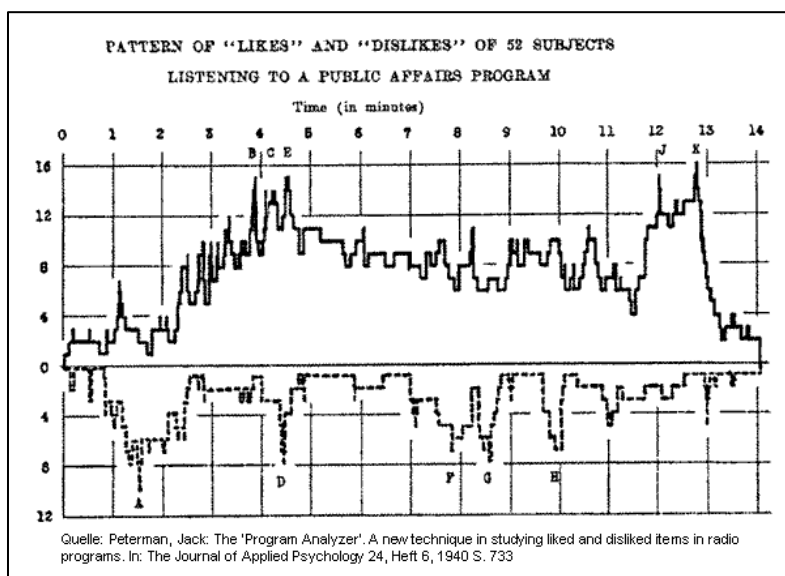
Der Vorteil dieses apparativen Verfahrens liegt im ereignissimultanen Messen, in der Erfassung von Emotionen vor deren bewusster Rationalisierung, sowie in der Vermeidung von Verbalisierungsproblemen.

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.

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

Anhang 2: 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>

ANHANG 3: DER EINSATZ DES PROGRAMMANALYSATORS

21/07/2017

von Flora MESSERKLINGER

Der Blog „Markt- und Mediaforschung“ des Studiengangs Media- und Kommunikationsberatung der FH St. Pölten stellt in regelmäßigen Abständen Seminararbeiten zur Diskussion, die zu ausgewählten und spezifischen Themenbereichen der Markt- und Mediaforschung verfasst wurden.

1. Einleitung

„Der Sammelbegriff Werbewirkungsmessung bezeichnet verschiedenartige Methoden und Verfahren, mit denen die psychische und/oder ökonomische Wirkung einer Werbung empirisch überprüft wird.“ (Leven, 1996, S. 31)

Zur erfolgreichen Werbewirkung von Werbemitteln muss zunächst das Werbeziel formuliert und operationalisiert werden. Im ersten Schritt wird demnach die Frage nach dem „Was“ gestellt. Darauf aufbauend kann das „Wie“ beantwortet werden: *Wie kann ich das Ziel messen?* Hier fällt die Entscheidung für ein geeignetes Testverfahren. (Vgl. Kaiser, 1999, S. 40)

Der Programmanalysator gilt als „non-verbales“ und reaktives Instrument. Häufig wird *diesen Instrumenten* zugeschrieben, sie seien unbeeinflussbar. In der vorliegenden Arbeit wird die als reaktiv und nicht-verbal eingestufte Messmethode *Programmanalysator*, mit dem Schwerpunkt auf dessen Entwicklung seit den 1930er Jahren, genauer betrachtet.

2. Der Programmanalysator

Im Folgenden wird zunächst die Funktionsweise des Programmanalysators erklärt, sowie die Entwicklung und die Einsatzmöglichkeiten. Anschließend wird auf die Messqualität des Programmanalysators eingegangen.

2.1. Funktionsweise

Der Programmanalysator dient der Messung von spontanen Reaktionen, welche Auskunft über Gefallen und Abneigung bzw. Interesse und Akzeptanz geben. (Vgl. Zentes, 1987, S. 132) Die Eindrücke der ProbandInnen zu einem gezeigten Stimulus (z.B. Werbespot) können demnach unmittelbar gemessen werden. (Vgl. Schweiger & Schrattenecker, 2016, S. 392)

Das Kernstück des Programmanalysators ist ein Erfassungsgerät, das entweder über zwei Druckstifte verfügt, die bei positiver oder negativer Reaktion gedrückt werden, oder aus einem Analoghebel, der eine stufenlose Reaktion ermöglicht. Wird der Hebel in keiner Weise betätigt, so wird das als neutrale Haltung gegenüber dem gezeigten Stimulus eingestuft. Die fehlende stufenlose Messung und somit wenig exakte Messung des Gefallens/Missfallens beim Einsatz eines Erfassungsgerätes mit Druckstiften, sowie die farbliche Gestaltung der Druckstifte wird in der Literatur als kritisch betrachtet. Neibecker zitiert Kroeber-Riel beim Verweis, dass unterschiedliche Farbpräferenzen der ProbandInnen bei Programmanalysatoren mit roten und grünen Knöpfen die Messqualität beeinflussen könnten. (Vgl. Neibecker, 1985, S. 100)

2.2. Entwicklung und Einsatzmöglichkeiten

Im nächsten Abschnitt wird die zeitliche Entwicklung des Programmanalysators sowie dessen Einsatzmöglichkeiten vorgestellt.

2.2.1. Die motorisch-manuelle Messung als Vorläufer

Die genaue und kritische Betrachtung des Programmanalysators erfordert eine grobe Analyse der motorisch-manuellen Emotionsmessung. Zu den motorischen, nicht-verbalen Verhaltensweisen zählen Blickkontakt, Körperhaltung und Gestik. In Anlehnung und Orientierung an Scherer und Wallbott (1979) sowie Neibecker (1985) wird die Gestik, insbesondere expressive Handbewegungen, als besonders relevant hinsichtlich der Emotionsmessung betrachtet. Der Programmanalysator versucht, die spontane Einstellung und Haltung einer Person über die motorischen Reaktionen zu messen. Anfang der 1970er Jahre stellte Clynes die Hypothese auf, dass Affektkommunikation über kulturelle Grenzen hinweg funktioniere, weil sie auf wenigen biologisch vorprogrammierten

Mustern des menschlichen Nervensystems beruhe. Diesen Ansatz beschreibt Clynes als „Essentic Forms“, abgeleitet vom Begriff „sentire“.

Das Messinstrument Sentograph besteht aus einem frei beweglichen Knopf, über den der Fingerdruck gemessen wird. ProbandInnen aus unterschiedlichen Kulturen wurden bestimmte Begriffe, die in Zusammenhang mit Emotionen stehen, präsentiert und sie wurden dazu aufgefordert, den Sensor dem Affekt entsprechend lange zu drücken.

Zahlreiche Untersuchungen haben ergeben, dass es für bestimmte Emotionen typische und unterscheidbare Verlaufsformen gibt. Folgende Abbildung zeigt ausgewählte Ergebnisse zur Messung der sentischen Reaktionen. (Vgl. Neibecker, 1985, S. 63ff)

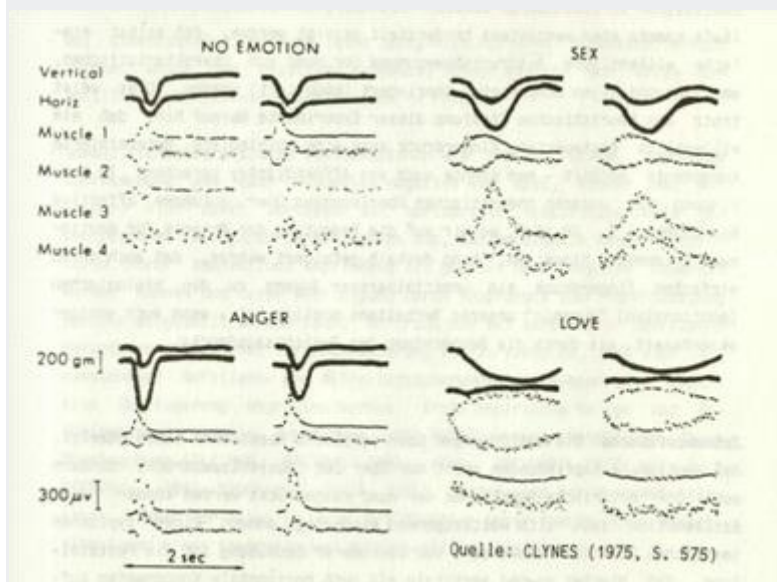


Abbildung 1: Beispielhafte Verlaufsformen des Sentographen (Neibecker, 1985, S. 66)

Die Verlaufsformen bestehen jeweils aus zwei Kurven. Die Obere zeigt die vertikale, die Untere die horizontale Bewegung des Knopfes und die Länge entspricht einer Messdauer von zwei Sekunden. „Keine Emotionen“ ist als Vergleichswert zu betrachten und wurde mit dem Anschlag einer Schreibmaschine simuliert. Auffällig ist, dass je eine Linie ausgeprägter ist als die andere. Von Interesse ist auch, dass die durchschnittlichen Reaktionsverläufe bei etwa einer Sekunde liegen. Die Studie führte zu der Erkenntnis, dass Emotionen, zusätzlich zu dem Gesichtsausdruck über motorische Reaktionen ausgedrückt werden. Das trägt wiederum zur Verifizierung des Einsatzes des Programmanalysators als Instrument zur Messung positiver und

negativer Emotionen mittels einfachen Tastendrucks bei. (Vgl. Neibecker, 1985, S. 66ff.)

2.2.2. Erste Versuche von Lazarsfeld / Stanton und ihre „Big Annie“

Bereits in den 30er Jahren entwickelten Lazarsfeld und Stanton den Programmanalysator, um Rundfunksendungen zu testen. (Vgl. Neibecker, 1985, S. 100)

Mit dem Einsatz des Programmanalysators wollten die beiden Forscher Ursachen für die häufigen Diskrepanzen zwischen der Einstellung und dem tatsächlichen Verhalten von RezipientInnen herausfinden, um schließlich die Frage zu beantworten: Wie werden Radioprogramme bewertet und was sind die Gründe dafür?

Im ersten Schritt wurden die HörerInnen schriftlich befragt. Anschließend bekamen die ProbandInnen einen Radiospot zu hören, welchen sie mittels Knopfdruck simultan bewerteten und abschließend folgte eine Diskussion. (Vgl. Trepte, 2004, S. 8ff)

2.2.3. Die mobile Variante „für Kinder“

Der „Sesame Workshop“, früher „Childrens Television Workshop“ genannt, ist eine amerikanische Non-Profit-Organisation, die hinter der Produktion zahlreicher TV-Bildungsprogramme für Kinder steht. Gemeinsam mit der Ontario Educational Communication Authority entwickelte die NGO das PEAC-System (Program Evaluation Analysis Computer-System), welches eine mobile Form des Programmanalysators darstellt. (Vgl. Gunter, 2000, S. 150)

Bestehend aus 20 tragbaren Erfassungsgeräten und einem Mikrocomputer hat das System vor allem zwei große Vorteile: Es hat lediglich die Größe eines Taschenrechners und kann daher ortsunabhängig eingesetzt werden. (Vgl. Neibecker, 1985, S. 101) Zusätzlich können die 16 Eingabetasten beliebig programmiert werden und somit inhaltspezifische, neue Variablen je nach Anlass hinzugefügt und geändert werden. (Vgl. ebd; Vgl. Gunter, 2000, S. 150)

2.2.4. ASI Marketing Research: Testen im „Theater“

Bei ASI Entertainment in Kalifornien werden in zwei speziellen Theatersälen Werbespots und Fernsehsendungen präsentiert und deren Wirkung auf RezipientInnen untersucht. Das eigens entwickelte

ViewTrac System erlaubt die simultane, sekundengenaue Bewertung durch die ProbandInnen. Für Unternehmen und Werbetreibende hat dies den Vorteil, dass die Reaktionen exakt bestimmten Situationen, Musikstücken, Charakteren und weiteren Elementen des Untersuchungsmaterials zugeordnet werden können. (Vgl. ASI Entertainment, o. J.)

2.2.5. Der Online-Programmanalysator des IKV

Das Institut für Konsum- und Verhaltensforschung der Universität des Saarlandes in Deutschland ist unter anderem Forschungsort des in der vorliegenden Arbeit bereits zitierten Bruno Neibecker. Im Rahmen eines Projektes erarbeitete Neibecker, unterstützt von Werner Kroebl-Riel und weiteren Kollegen, einen Online-Programmanalysator.

Ziel war es, eine Kombination aus Online-Datenerfassung, Mobilität und preiswertem Mikrocomputer zu erschaffen. Zudem ist anstatt des ehemals dichotomen Antwortformats („like“ – „dislike“) eine Messung der Intensität der Reaktion möglich. Die Darbietung und die Beurteilung des Programms werden elektronisch mittels frequenzselektiven Tonschalters synchronisiert, um die Reaktionen sekundengenaue zuordnen zu können. Die Eingaben der ProbandInnen erscheinen sofort am Display und werden extern gespeichert. Gleich im Anschluss an die Messung erhält man Reaktionsprofile, bei welchen das Gefallen und Missfallen entweder in relativen und absoluten Zahlen oder, wie in Abb. 2 gezeigt, geglättet abgebildet werden. Das Beispiel zeigt aus Gründen der Übersichtlichkeit die geglätteten Kurven der Reaktionen auf zwei unterschiedliche Schlagerlieder sowie zusätzlich die Nettowerte (Gefallen minus Missfallen) als strichlierte Linie. (Vgl. Neibecker, 1985, S. 101)

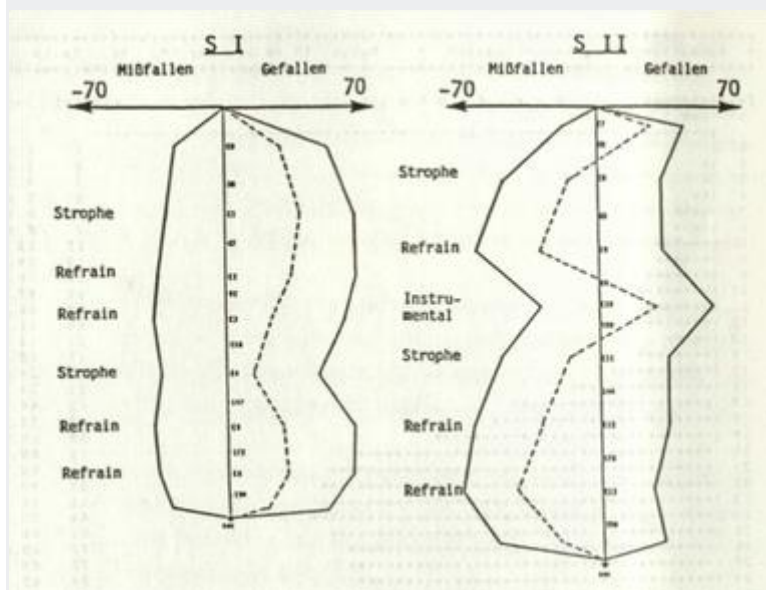


Abbildung 2: Reaktionsprofile des Online-Programmanalysators (Zentes, 1987, S. 136)

Neibecker greift bei der theoretischen Fundierung seiner Arbeit die Meinung seines Kollegen Kroeber-Riel auf, der sich 1983 in „Marktforschung“ die Frage stellt: „Welche Auswirkungen hat der erste, spontane Eindruck eines Meinungsgegenstandes auf die spätere Beurteilung und auf das Verhalten gegenüber diesem Gegenstand?“. (Kroeber-Riel, 1983, zit.n. Zentes, 1987, S. 134)

Als übergeordnetes Ziel der empirischen Forschung Neibeckers könnte die Validierung und Überprüfung der Reliabilität des Online-Programmanalysators auf Individualniveau bezeichnet werden. Dazu wurden drei Hypothesen gebildet, von welchen wiederum zwei belegt werden konnten. Folgendes wird als bestätigt betrachtet: Der Programmanalysator kann dynamische Einzelelemente von Programmen, Musikstücken oder Werbespots als Ursachen für positive bzw. negative Gefallensäußerungen erkennen. Weiters werden die erhaltenen Reaktionsdaten als nützliche und vor allem reliable Ergänzung zu herkömmlichen Befragungen präsentiert. Die Trennschärfenhypothese, welche klar erkennbare Unterschiede zwischen der verbalen (mittels Befragung) und non-verbalen (mittels Programmanalysator) Emotionsmessung erwartete, konnte nicht belegt werden. Neibecker führt das auf die Wechselwirkung(en) zwischen der Neuartigkeit des Präsentationsmaterials bzw. –inhalts und der Bewertung zurück. (Vgl. Neibecker, 1985, S. 152)

2.3. Messqualität des Programmanalysators

Forschungsinstrumente werden anhand von drei Gütekriterien bewertet:

- Objektivität,
- Reliabilität und
- Validität.

Die Messqualität des Programmanalysators wird von zahlreichen AutorInnen kritisch betrachtet. Das folgende Kapitel gibt einen Überblick.

2.3.1. Objektivität

Lienert forderte bereits 1967 Objektivität als eines der drei Hauptgütekriterien für empirische Untersuchungen und verstand

darunter den Grad, „in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind.“ (Lienert, 1967, S. 13)

Lackner (1994, S. 109) bewertet in seiner Dissertation standardisierte Testanweisungen als ausschlaggebend für die Erreichung von Durchführungsobjektivität. Er hält den Programmanalysator für auswertungsobjektiv und begründet das mit der standardisierten Computerauswertung. Die Möglichkeit, die Testergebnisse objektiv zu interpretieren, hängt laut Lackner vor allem von der Anzahl der ProbandInnen und deren Latenzzeiten ab. Aus der gesamten Stichprobe soll schließlich eine einheitliche Latenzzeit quantifiziert werden können.

2.3.2. Reliabilität

Neibecker (1985, S. 82) stützt sich bei seiner Unterscheidung zwischen drei Methoden zur Bestimmung der Reliabilität auf die Ausführungen einiger KollegInnen:

Paralleltest-Reliabilität

Die Ergebnisse von Untersuchungen zweier vergleichbarer Messinstrumente an einer Stichprobe werden verglichen. (Neibecker, 1985, S. 82)

Retest-Reliabilität

Die Testwiederholungsmethode wendet das gleiche Messinstrument an der gleichen Stichprobe zu zwei unterschiedlichen Zeitpunkten an und untersucht die Ergebnisse nach Korrelationen. (Neibecker, 1985, S. 82)

Interne Konsistenz

Die Konsistenzanalyse gilt als Weiterentwicklung der Testhalbierung. Das Messinstrument wird hierbei in zwei gleichwertige Hälften zerlegt – „halbiert“ – und die Ergebnisse werden verglichen. (Neibecker, 1985, S. 82)

Neibecker (1985, S. 151) kommt bei seinen Untersuchungen an Schlagerliedern zu beachtlichen Reliabilitäten, was er als „Indiz für die Genauigkeit des Messverfahrens“ (Neibecker, 1985, S. 152) sieht.

In ähnlicher Weise überprüfte Lackner (1994) die Reliabilität des Programmanalysators mittels Parallel- und Retests. Der Untersuchungsgegenstand, ein TV-Spot von Römerquelle, erreichte hohe Korrelationswerte und nahezu identische Bewertungen bei beiden Versuchen. (Lackner, 1994, S. 242-255)

2.3.3. Validität

Validität gibt „den Grad der Genauigkeit, mit dem ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal gemessen wird“ (Lienert, 1967, S. 12) an.

Lackner (1994, S. 310-314) überprüfte die Validität des Programmanalysators, indem er den ProbandInnen drei sehr unterschiedliche Werbespots präsentierte und diese nacheinander bewerten ließ. Er kam dabei zu dem Schluss, dass ein Großteil der Stichprobe (n=104) kein durchgängiges Antwortverhalten verfolgte.

Neibecker (1985, S. 152) schließt sich der Einschränkung der Validität an. Die Ergebnisse der non-verbalen Reaktionsmessung geben einen Einblick in mögliche Einflussfaktoren auf die Bewertung von dynamischem Testmaterial. Eine vollständige Erhebung des Konstrukts ist somit nicht möglich, die Validität beschränkt sich auf die Erfassung der Valenz der Emotionen. Der Einsatz des Programmanalysators ist demnach als Teil einer dimensionalen Emotionsanalyse zu sehen, um etwaige Umwelteinflüsse nicht außer Acht zu lassen. (Vgl. Neibecker, 1985, S. 153)

3. Conclusio

Im abschließenden Kapitel werden die Vorteile sowie Probleme und Grenzen des Programmanalysators durch Heranziehen unterschiedlicher Literatur erörtert. Die Zusammenfassung gibt schließlich einen Ausblick auf die Zukunft.

3.1. Vorteile des Programmanalysators

Die Vorzüge des Programmanalysators sind vielseitig und wurden mehrfach ausgeführt. In der folgenden Auflistung finden sich ausgewählte Zitate:

- „Die mittels Programmanalysator gemessenen nicht-verbalen Reaktionen der Personen sind für das Marketing deshalb besonders aufschlussreich, weil diese spontanen Gefallenswerte die nachfolgende genauere und rationalere Beurteilung eines Programmes wesentlich beeinflussen.“ (Neibecker, 1985, S. 160)

- „Die eigentlichen Vorzüge unserer Methode [...] kommen jedoch erst in der zweiten Phase eines Copytests richtig zu Geltung, wenn es um die Analyse des Kommunikations- oder Werbeerfolges geht. [...] wie kommt Werbewirkung zustande; welche Defekte sind für das schlechte Abschneiden verantwortlich oder welche strategischen Konzeptionen sind noch nicht berücksichtigt worden?“ (ebd., S. 161)
- Zusätzlich verweist Neibecker (ebd., S. 162) auf den „Zusatznutzen“ des Programmanalysators. Die Analyse ermöglicht Kenntnisse über problematische und gelungene Sequenzen eines Spots, die Bildung von Zielgruppenclustern und die Möglichkeit multivariater Auswertungsverfahren.
- „Der Vorteil dieses Verfahrens liegt in der ereignissimultanen Messung sowie in der Erfassung von Emotionen vor deren bewusster Rationalisierung.“ (Bruhn, 2014, S. 532)

3.2. Probleme und Grenzen des Programmanalysators

Der Einsatz des Programmanalysators kann zu spannenden Erkenntnissen führen, jedoch nicht ohne die Unterstützung durch weitere Methoden und Instrumente sowie der ständigen Weiterentwicklung. Wiederum werden Wissenschaftler zum Thema „Probleme und Grenzen“ zitiert, um einen Überblick über ältere und neuere Meinungen zu erhalten:

- Neibecker verweist darauf, dass empirische Forschung den strategischen Aspekt eines Marketingkonzepts nicht ersetzen kann: „Hier wird deutlich, dass Werbeforschung mehr als das Messen von Werbewirkungen umfasst und letztlich die Entwicklung strategischer Werbekonzepte unumgänglich ist.“ (Neibecker, 1985, S. 161)
- Schwaiger führt den Nachteil der Künstlichkeit der Situation an: „Ein wesentlicher Nachteil ist allerdings, dass die Versuchspersonen sich ihrer Rolle als Testsubjekt bewusst sind und dem Werbemittel daher höhere Aufmerksamkeit schenken als in biotischen Situationen.“ (Schwaiger, 2013, S. 58)
- Kaiser bestätigt die erhöhte Konzentration seitens der ProbandInnen auf den Stimulus. (Vgl. Kaiser, 1999, S. 181)
- Salcher gibt zu bedenken, dass die Reaktionszeiten der Menschen sehr unterschiedlich ausfallen und die Beurteilung somit nicht mit Sicherheit bestimmten Szenen oder Elementen des Spots zugeordnet werden können. (Vgl. Salcher, 1995, S. 123)

3.3. Zusammenfassung und Ausblick

Zusammenfassend lässt sich sagen, dass sich die non-verbale Erlebnisbeschreibung, wie sie das Verfahren des Programmanalysators nutzt, seit beinahe 90 Jahren stetig weiterentwickelt hat und der Prozess vermutlich noch kein Ende erreicht hat. Nimmt man das Beispiel des ASI in der USA zur Hand, bei dem riesige Theatersäle der Forschung dienen, so scheinen den Einsatzmöglichkeiten und technologischen Erweiterungen keine Grenzen gesetzt.

Selbstverständlich ist der Programmanalysator nicht als „allwissende Methode“ und „Alleskönner“ zu verstehen, wie es die aufgezeigten Probleme und Grenzen verdeutlichen. Dennoch ist die non-verbale, apparative Arbeitsweise als fortschrittlich einzustufen. Weitere Überlegungen zur Verbesserung sollten angestellt werden. Betrachtet man die aktuelle Literatur zum Thema, so scheint dies in den letzten Jahren weltweit vernachlässigt worden zu sein.

Literaturverzeichnis

AGSÖ. (kein Datum). *AGSÖ (Archiv für die Geschichte der Soziologie in Österreich)*. Von <http://agso.uni-graz.at/lexikon/klassiker/lazarsfeld/25bio.htm> abgerufen

ASI Entertainment. (kein Datum). *ASI Entertainment*. Von <http://www.asientertainment.com/television-research/> abgerufen

Brennen, B. S. (2012). *Qualitative Research Methods for Media Studies*. London: Routledge.

Bruhn, M. (2014). *Unternehmens- und Marketingkommunikation: Handbuch für ein integriertes Kommunikationsmanagement*. München: Vahlen.

Dentsu Aegis Network. (kein Datum). *Dentsu Aegis Network*. Von <http://dentsu.at/ueber-uns/> abgerufen

Gunter, B. (2000). *Media Research Methods: Measuring Audiences, Reactions and Impact*. New York: SAGE Publications.

Kaiser, C. (1999). *Apparative Werbeforschung*. In *Empirische Marketingforschung* (Bd. 17). Wien: Servicefachverlag.

Koschnick, W. J. (1996). *Standard-Lexikon Werbung, Verkaufsförderung, Öffentlichkeitsarbeit*. München: Saur-Verlag.

Kroeber-Riel, W. (1993). Bildkommunikation: Imagerystrategien für die Werbung. München: Vahlen Verlag.

Lackner, R. (1994). Apparative Werbewirkungsmessung – Die Messqualität der nonverbalen Erlebnisbeschreibung unter besonderer Berücksichtigung von Fernsehwerbung. Wien: Ordinariat für Werbewissenschaft und Marktforschung an der WU Wien.

Leven, W. (1996). Werbewirkungsforschung aus der Sicht der Praxis. Werbeforschung & Praxis (4/96).

Lienert, G. A. (1967). *Testaufbau und Testanalyse*. Weinheim: Beltz Verlag.

Nachman, G. (2012). *Raised on Radio*. New York City: Knopf Double Publishing Group.

Neibecker, B. (1985). Konsumentenemotionen. Messung durch computergestützte Verfahren – Ein empirische Validierung nicht-verbaler Methoden. Würzburg/Wien: Physica Verlag.

Salcher, E. (1995). *Psychologische Marktforschung*. Berlin: Walter de Gruyter Verlag.

Schwaiger, M. (2013). Multivariate Werbewirkungskontrolle: Konzepte zur Auswertung von Werbetests. Wiesbaden: Springer Verlag.

Schweiger, G., & Schrattenecker G.. (2016). *Werbung*. 9. Auflage, München: UVK Verlag Trepte, S. (2004). Zur Geschichte der Medienpsychologie. In G. B. (Hrsg.), *Lehrbuch der Medienpsychologie* (S. 3-25). Göttingen: Hogrefe.

Zentes, J. (1987). EDV-gestütztes Marketing. Ein informations- und kommunikationsorientierter Ansatz. Berlin/Heidelberg/New York: Springer Verlag.

→ Comments welcome: peter.diem (at) gmail.com