

Dr. Rolf Schoch

Skalierungsverfahren – Wie wirken sich Änderungen in der Konstruktion von Skalen aus?

„Wenn man messen kann, worüber man spricht und es in Zahlen ausdrücken kann, dann weiss man etwas darüber. Wenn man es nicht in Zahlen ausdrücken kann, dann ist das Wissen dürftig und unzureichend.“ Lord Kelvin, britischer Physiker

Dieses Motto von Lord Kelvin lässt sich sinngemäss und grundsätzlich auch für die Sozialwissenschaften und die Marketingforschung anwenden. Wichtig für uns ist auch der gelegentlich gehörte Zusatz: „Was man nicht messen kann, kann man nicht verbessern“¹. Oder einfach ausgedrückt: Wir brauchen Zahlen und Fakten, um vernünftig über eine Sache sprechen zu können. Unser gefühltes Wissen über Dinge reicht nicht aus, um Probleme zu erkennen und nachfolgend Lösungsansätze zu finden.

Skalierungsverfahren und Messskalen sind wichtige methodische Instrumente für die Erhebung standardisierter numerischer Daten. „Eine Messung im Sinne der Messtheorie liegt vor, wenn ein Isomorphismus oder Homomorphismus zwischen einem empirischen und einem numerischen Relativ existiert“ (Diekman, 1995, 247). Diese Definition präzisiert die ursprüngliche Begriffsumschreibung von Stevens (1951), der Messung als „Zuordnung von Zahlen zu Objekten nach bestimmten Regeln“, aufgefasst hatte. Skalierungsverfahren sind Messverfahren zur Erfassung qualitativer Aspekte, insbesondere von Einstellungen und Motiven (Attleslander, 1975, 243).

Bei der Konstruktion von Skalen hat der Marketing- oder Sozialforscher eine ziemlich grosse Gestaltungsfreiheit. Es stehen viele Auswahlmöglichkeiten zur Verfügung, z.B. betreffend Skalentyp, Anzahl und Art der Skalenstufen, Startpunkt der Nummerierung, statistische Auswertung der Skalenwerte etc. Er muss bezüglich aller diese Optionen zweckmässige Entscheide treffen. Angeregt durch einen

¹ <http://de.wikipedia.org/wiki/Kelvin>. Das Kelvin (Einheitenzeichen: K) ist die SI-Basiseinheit der thermodynamischen Temperatur und zugleich gesetzliche Temperatureinheit; es wird auch zur Angabe von Temperaturdifferenzen verwendet.

neuen Fachartikel in einer Marketingzeitschrift werden verschiedene Gestaltungsmöglichkeiten für interaktive Online-Skalen diskutiert.

In der deutschsprachigen Fachliteratur konnte man dazu kaum praktisch anwendbare Anleitungen oder Ratschläge für die konkrete empirische Marketing- und Sozialforschung finden. Zwar beschreiben die klassischen Methodenbücher auf dem hohem Abstraktionsniveau der Messtheorie Grundbegriffe des Messens, ferner die Mess- oder Skalenniveaus – Nominal-, Ordinal-, Intervall-, Ratio-Skalen etc. –, sodann verschiedene Arten von Skalen, z.B. für die Messung speziell von Einstellungen, wie Likert-Skalen, Semantische Differential-Skalen (Polaritätsprofil), Guttman-Skalen, Stapel-Skalen, Thurstone Skalen, sowie Qualitätskriterien für die Beurteilung von Skalen, Validität, Reliabilität, Objektivität (Atteslander, 1991; Diekman, 1995; Friedrichs, 1980; Hammann & Erichson, 1990; Holm, 1976; Kromrey, 1991; Miller, 1970; Scheuch & Zehnpfennig, 1974; Selltiz et al., 1972; Tull & Albaum, 1973). Jedoch enthalten diese Lehrbücher keine Aussagen zu den konkreten *Detailfragen*, die den Praktiker interessieren. Zu diesen technischen *Details der Skalenkonstruktion* gehören beispielsweise die folgenden Fragen, die hier behandelt werden:

- *Wie viele Skalenstufen* sollen die zu verwendenden Skalen enthalten, d.h. wie stark unterteilt sollen mit anderen Worten die Skalen sein?
- Soll eine *gerade oder eine ungerade Zahl* von Skalenstufen verwendet werden, mit anderen Worten mit oder ohne neutralen Skalenmittelpunkt?
- Nummerierung des *Startpunkts* der Skala?
- Statistische *Auswertung* der Daten?

Zusätzlich gibt es eine Fülle von weiteren Optionen für die Gestaltung der physischen Form der Skala (Malhotra, 1996, 300 ff). Zu entscheiden sind die Fragen:

- Sollen die Skalenstufen vertikal oder horizontal präsentiert werden?
- Sollen die Skalenstufen durch verbale Beschreibungen (Labels) oder ohne Text bloss durch Kästchen zum Ankreuzen, durch Linien oder Einheiten auf einem Kontinuum dargestellt werden?
- Sollen den Skalenstufen numerische Werte zugeordnet werden, positive oder negative oder beides, und sollen nur die Endpunkte der Skala oder alle Skalenstufen so angeschrieben werden?

Zwei typische Skalen in der Marketingforschung sind im Weiteren die Temperaturskalen mit Gradeinteilung zwischen 0° und 100° und 7-stufige Gesichtsskalen („smiley faces“; nach Kunin). Bei der ersteren bedeuten höhere Temperaturen, bei der zweiten glücklichere Gesichter vorteilhaftere Einstellung zum Beurteilungsgegenstand (Malhotra, 1996, 301).

Einige U.S.-amerikanische Marketingforschungslehrbücher behandeln diese Details ebenfalls nur in allgemeiner Form, darunter ausgerechnet Standardwerke der American Marketing Association über „Marketing Research“ (Chuckrapani, 2000). Andererseits findet man solche forschungspragmatische Aussagen in anderen Werken, wie z.B. Churchill (1995), Green & Tull (1975) oder Malhotra (1996).

Die Handbücher über Skalierungsverfahren im Marketing gehen einen Schritt weiter. Die 1. Auflage des „Handbook of Marketing Scales“ (Bearden et al., 1993) präsentiert insgesamt 88 verschiedene Skalen mit bibliografischen Angaben aus dem Bereich der Marketingforschung. Kriterien für die Aufnahme einer bestimmten Skala in das Handbuch waren:

- Die Skala hatte eine vernünftige theoretische Basis und/oder Definition der Grundbegriffe
- Die Skala bestand aus mindestens drei Items oder Fragen
- Die Skala wurde innerhalb des Sachgebiets Marketing oder Konsumentenverhalten entwickelt oder allgemein angewandt
- Mindestens einige Skalierungsverfahren wurden in der Entwicklung der Skala angewandt
- Es liegen Angaben vor über Reliabilität und/oder Validität

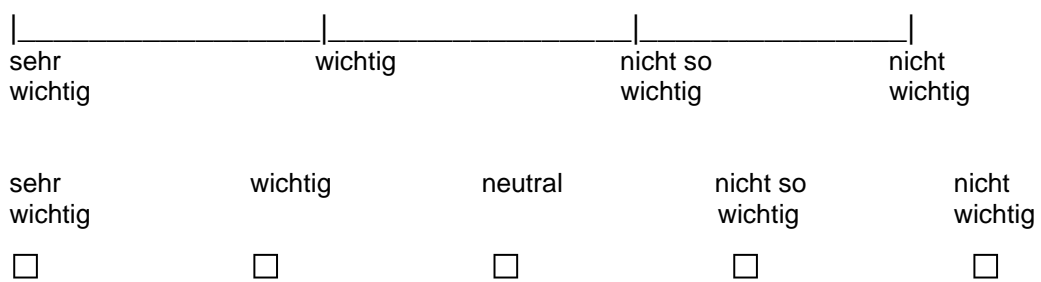
Das „*Marketing Scales Handbook*“ in der dritten Auflage ist eine umfangreichere Kompilation von insgesamt 941 Skalen, die in Marketingforschungsstudien tatsächlich verwendet und zwischen 1994 und 1997 in wissenschaftlichen Periodika publiziert worden sind (Bruner et al., 2001). Die erfassten Sachgebiete sind „Konsumentenverhalten“ (402 Skalen), „Werbung“ (83 Skalen) und „Organisationales Verhalten“ (456 Skalen). Über jede einzelne der aufgeführten 941 Skalen werden nach einem einheitlichen Raster Angaben geliefert über Name, Beschreibung/Skalentyp, Herkunft, Stichproben, Reliabilität und Validität, Anwendung, wichtigste Forschungsergebnisse, Literatur, Skalenitems und psychometrische Eigenschaften. Eine inhaltliche Auseinandersetzung mit den Skalen oder eine me-

thodische Beurteilung und Bewertung, eine Analyse der jeweiligen Vor- und Nachteile, geben die Herausgeber des Handbuchs in ihrer Auflistung zwar nicht. Jedoch präsentieren sie beispielhaft verbale und grafische Darstellungen der Skalen sowie des Wortlauts aller Items. So kann sich der Leser doch eine Vorstellung von den Vorzügen oder allenfalls Mängeln der dargestellten Skalen (alle auf Englisch) machen.

Über die „Do's“ und „Don'ts“ der Skalenkonstruktion in der Marketingforschung seien schon Bände geschrieben worden, meinen Green & Tull (1975, 181). Dessen ungeachtet besteht auch heute in Theorie und Praxis noch keine völlige Einigkeit über die jeweils beste Lösung.

1. Gerade oder ungerade Zahl von Skalenstufen?

Der Möglichkeiten für die Wahl der Anzahl Skalenstufen in einem Messinstrument gibt es sehr viele; so z.B. die Nummerierung von 1 – 10, entsprechend 10 Skalenstufen oder 0 – 11 (11 Skalenstufen) oder 0 – 7 (acht Skalenstufen) oder 1 – 7 (sieben Skalenstufen). Andere Optionen sind 0 – 5 (sechs Skalenstufen) oder auch 1 – 5 (fünf Skalenstufen). Wie diese Möglichkeiten zeigen, geht es dabei vor allem auch um die Auswahl einer *geraden* vs. *einer ungeraden* Anzahl von Skalenpunkten. Die folgenden zwei Beispiele von grafischen Rating-Skalen illustrieren den Unterschied (Tull & Albaum, 1973, 110):



Die Argumente für die erstere Variante, *gerade Zahl*, werden darin gesehen, dass darin *kein neutraler Skalenmittelpunkt* existiert. Das bedeutet: Die Befragten haben sich zu entscheiden, ob sie mehr dem einen oder dem anderen Extrem oder Endpunkt der Skala zuneigen, z.B. „wichtig“ vs. „unwichtig“ bzw. „positiv vs. negativ“ oder „gut vs. schlecht“ etc. Für den Marketingforscher (oder den Auftraggeber einer Studie) ist es vermutlich wünschenswert, z.B. bei einer Konsumentenumfra-

ge eine klare, eindeutige Stellungnahme der Befragten für die eine oder andere Seite einer Frage zu bekommen.

Bei einer *ungeraden Zahl* von Skalenstufen gibt es dagegen zwischen den Extremen, d.h. den Endpunkten der Skala, eine *mittlere Skalenstufe*, die als „neutral“ interpretiert wird. Dies scheint insofern realistisch, als in einer Umfrage manche Befragte schlicht keine ausgeprägten Meinungen oder Einstellungen zum Thema haben. Bei vielen Befragungsgegenständen gibt es zweifellos eine *neutrale Mitte* (wie z.B. betreffend Temperatur weder „heiss“ noch „kalt“). Drängt man die Befragten zu einer Stellungnahme indem man die mittlere Antwortmöglichkeit eliminiert („forced choice“), wird möglicherweise eine Verzerrung in die Resultate eingeführt. Denn dadurch würde die Fähigkeit oder Bereitschaft der Versuchspersonen zu Beantwortung der Frage reduziert und die Validität der Ergebnisse beeinträchtigt. Unsere Auffassung in diesem Punkt deckt sich mit derjenigen des eingangs erwähnten Artikelverfassers (Cook, 2013, 22). Dieser stellt unzweideutig fest, es sei sehr wichtig, Skalen mit *ungerader Zahl* an Antwortalternativen zu verwenden, welche die Option haben, eine neutrale Mitte zu wählen. Allerdings kann man, um eine Konfundierung der Mittelposition mit alternativer Interpretation wie z.B. „weiss nicht“ oder „keine Meinung“ zu vermeiden, diese Optionen als zusätzliche separate Antwortmöglichkeiten explizit in das Befragungsinstrument aufnehmen (Churchill, 1995, 414).

Die Frage gerade oder ungerade Anzahl Skalenstufen hat einen grossen Einfluss auf die *Antworthäufigkeiten*. Die explizite Nennung einer neutralen Mitte bzw. der Verzicht darauf kann deutlich unterschiedliche Häufigkeiten produzieren. Churchill (1995, 414-415) präsentiert aus einer Studie, welche sowohl 4- wie auch 5-stufige Skalen zur Messung der Kaufabsicht für drei verschiedene Konsumgüter mit drei Zeitperioden (kurz-, mittel- und längerfristige Kaufabsichten) verwendete, die folgenden Antwortverteilungen:

	5 Stufen	4 Stufen
1. werde definitiv kaufen	13 %	14 %
2. werde wahrscheinlich kaufen	15 %	20 %
3. (würde kaufen oder auch nicht kaufen)	25 %	---
4. werde wahrscheinlich nicht kaufen	21 %	39 %
5. werde definitiv nicht kaufen	26 %	27 %

Die Skalen waren jeweils dieselben, ausser dass in der 5-stufigen Kaufabsicht-Skala eine neutrale Mittelposition (3.) enthalten war. Die obige Tabelle enthält als

Beispiel für 9 Messungen jene für „Kaufabsicht für Kinotickets in den nächsten 7 Tagen“ für die beiden Varianten mit 5 bzw. 4 Skalenstufen.

Welche von beiden Varianten misst nun die Kaufabsichten der Befragten „richtiger“, d.h. gültiger? Hier stellt sich offenbar die Frage der *Validität* der Skalen. Aufgrund der vorliegenden Daten lässt sich diese Frage nicht beantworten. Dazu wären Angaben z.B. über die Prognosegültigkeit, durch Korrelation mit einem externen Validitätskriterium – Kaufabsicht mit tatsächlichem Kauf der Produkte –, erforderlich. Churchill (1995) beschränkt sich auf die folgenden, pragmatischen Aussagen: Das Ergebnis zeige wenig Unterschiede in den beiden extremen Skalenstufen 1 und 5. Wenn also das Forschungsinteresse sich auf diese beiden Endpunkte – „definitiv kaufen bzw. nicht kaufen“ – konzentriert, können offensichtlich beide Arten von Skalen verwendet werden. Interessiert man sich jedoch eher für allgemeine Tendenzaussagen – im Sinne von „sicher oder wahrscheinlich kaufen“ bzw. „sicher oder wahrscheinlich nicht kaufen“, d.h. Antworten 1 und 2 bzw. 3 und 4 jeweils zusammengenommen – so liefern die beiden Skalen unterschiedliche Antworten, die 4-stufige Skala massiv häufiger negative (66 % vs. 47 %)(vgl. Churchill, 1995, 415-416).

2. Anzahl Skalenstufen?

Soll eine Skala drei, fünf, neun, elf oder mehr Stufen umfassen, wenn es denn wie gesagt eine ungerade Zahl sein soll? Beispiele für alle diese Varianten finden sich in Literatur und Praxis. Dazu einige Beispiele:

Die Skala von King & Summers (1970) zur Messung der Meinungsführerschaft (Opinion Leadership) von Personen bei Produktinnovationen beispielsweise kommt mit zwei *dreiteiligen* (trichotomen) Skalen aus.:

„Würden Sie sagen, dass Sie sehr wenige Informationen, ein mittleres Mass an Informationen oder eine grosse Menge an Informationen über ...(Name der Innovation) an Ihre Freunde weitergeben?“

- | | |
|---|---|
| • Sehr wenig Information | 1 |
| • Eine durchschnittliche Menge an Informationen | 2 |
| • Eine grosse Menge an Informationen | 3 |

Ein anderes Instrument zur Messung der gleichen Variable „Opinion Leadership“ bei Neuerungen verwendet dagegen eine *fünfstufige* Skala mit 7 Items (Childers, 1986). Ein Beispiel:

„In einer Diskussion über Kabelfernsehen würden Sie am ehesten:

den Ideen Ihrer Freunde zuhören					Ihre Freunde von Ihren Ideen überzeugen
	1	2	3	4	5

Hartline & Ferrell (1996) verwendeten eine *siebenstufige* Likert-Skala mit sechs Items um die Anpassungsfähigkeit im Verhalten von Verkaufspersonal gegenüber Kunden zu messen. Beispiel:

stimme völlig zu											stimme gar nicht zu
	1	2	3	4	5	6	7				

- Wenn ich das Gefühl habe, dass mein Vorgehen nicht funktioniert, kann ich leicht auf einen anderen Ansatz wechseln
- Ich versuche zu verstehen, wie ein Kunde sich vom anderen unterscheidet.

Churchill (1995) druckt in seinem Lehrbuch eine Serie von *10-stufigen* Ratingskalen mit vier Beurteilungskriterien für Kaffee ab. Nur die Endpunkte sind angeschrieben und zwar mit „ganz und gar nicht wichtig“ (01) und „ausserordentlich wichtig“ (10). Für alle Skalenstufen sind Kästchen zum Ankreuzen vorgesehen. Musterbeispiele daraus sind:

	ganz und gar nicht wichtig											ausserordentlich wichtig
Aroma	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10	<input type="checkbox"/> 11	
Frische	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10	<input type="checkbox"/> 11	

Welche von diesen Alternativen für die Skalenkonstruktion werden empfohlen? Eine grössere Zahl von Skalenstufen ist offenbar von Vorteil für die statistische Datenanalyse. Churchill & Peter (1984) haben eine Meta-Analyse von total über 150 empirischen Marketingstudien (Zeitschriftenartikel) aus der Periode 1964-1982 durchgeführt. Das Ziel war, Einflussfaktoren auf die *Reliabilität* der Skalen zu identifizieren. Von insgesamt 23 möglichen und geprüften Einflussfaktoren aus den Bereichen Sampling, Skaleneigenschaften und Skalenentwicklung erwiesen sich nur zwei als von Bedeutung: die *Anzahl Items* in der Skala und die *Anzahl Skalenstufen*. Beide waren positiv korreliert mit der Reliabilität der Skala (die Vali-

dität wurde von den Autoren nicht untersucht). Alle anderen geprüften Einflussfaktoren erwiesen sich als unbedeutend für die Reliabilität der Skalen.

Obwohl es also keine einzelne optimale Anzahl gibt, wird in der Marktforschungspraxis traditionellerweise häufig mit 5- bis 9-stufigen Skalen gearbeitet. Und es wird empfohlen, eine Zahl von 11 Skalenstufen nicht zu überschreiten (Cook, 2013) – vorzugsweise mit dem Bereich 0 bis 10. Dies scheint für die Befragten plausibel, da die Zahl „Null“ im Allgemeinen als „schlecht“ definiert und die Zahl „10“ mit der Bestnote assoziiert wird. Weniger geeignet erscheint die Bezeichnung „1“ für die erste Stufe der Skala, da eine „Eins“ manchmal als beste, manchmal als schlechteste Stufe betrachtet wird (wie z.B. bei Schulnoten).

3. Nummerierung des Startpunkts der Skala

Soll die Skala auf der *ersten Stufe* mit der Zahl „0“ oder mit der Zahl „1“ *beginnen*? Wenn die Skalenstufen nummeriert werden, ist die Wahl der Zahlen mehr oder weniger Geschmacksache, vorausgesetzt, dass diese klar definiert und deutlich voneinander verschieden sind. Eine 1 – 10-Skala ist empirisch gleichbedeutend und gleichwertig mit einer 0 – 9-Skala und sogar mit einer 2 – 11-Skala, denn in jeder Variante existieren 10 Skalenstufen. Die Regeln für die Transformation der Skalenwerte erlauben bei einer *Ordinalskala* jede beliebige Umformung der Ursprungswerte, solange die grundlegende *Rangordnung* der Messobjekte erhalten bleibt. Zulässig sind somit alle *monotonen positiven Transformationen* der zugeordneten Zahlen. Statt 0, 1, 2, 3 ... 9 wären somit auch die Werte 0, 5, 10, 15 ... 45 gültig, nach der Regel $Y = bX$. Wenn die erste Stufe, der Anfangspunkt, mit Worten wie z.B. „gar nicht“ oder „keine“ oder „nie“ oder anderen Bezeichnungen die auf „Null“ hindeuten umschrieben wird, dann leuchtet es ein, dass die Zahl „0“ die geeignetste ist, um die Skala zu starten. Bevorzugt wird bei einer 10-stufigen Skala somit die Variante 0 – 9; für die anderen Varianten fehlt es an überzeugenden Begründungen. Überdies ist hier die Umrechnung in Prozentwerte am einfachsten.

Bei Online-Fragebogen im Internet werden mit Vorliebe optische *Schieberegler-Skalen* eingesetzt (mit 10, 11 oder mehr Skalenpunkten). Bei diesem Schieberegler-Messverfahren sind oft nur die Extrempunkte der Skala beschriftet – z.B. „0 gar nicht wichtig“ bis „10 ausserordentlich wichtig“. Die Versuchsperson verschiebt

Aufrunden oder Nichtaufrunden von Mittelwerten?

Auch das Runden von Skalen-Mittelwerten wird nicht als „gute Idee“ bezeichnet, wenn mit den Durchschnitten beispielsweise zwei Subpopulationen verglichen werden sollen. Dies zeigt das folgende Zahlenbeispiel mit unterschiedlichen Mittelwerten für 6 verschiedene Stichproben aus einer Skala 0 – 10 (Cook, 2013, 25, Tabelle 4):

		Mittelwerte aus einer 10-stufigen Skala		
		tatsächlich	gerundet	tatsächlich vs. gerundet
1.	Gruppe A	4.00	4.00	0.00 %
	Gruppe B	5.00	5.00	0.00 %
	B vs. A	25.00 %	25.00 %	0.00 %
2.	Gruppe C	4.49	4.00	12.25 %
	Gruppe D	4.50	5.00	-10.00 %
	D vs. C	0.22 %	25.00 %	-99.11 %
3.	Gruppe E	3.50	4.00	-12.50 %
	Gruppe F	5.49	5.00	9.80 %
	F vs. E	56.86 %	25.00 %	127.43 %

Solange die Messwerte in zwei Vergleichsgruppen ganzzahlig sind (wie im Beispiel 1. oben Gruppen A und B) gibt es keine Probleme – tatsächliche und gerundete Werte sind ja identisch; die Mittelwerte der Gruppe B sind und bleiben 25 % höher, als jene der Gruppe A.

Anders ist es, wenn Mittelwerte mit Dezimalstellen auf ganze Zahlen auf- oder abgerundet werden (wie im Beispiel 3. oben), z.B. von 3.50 auf 4.00 und von 5.49 auf 5.00 (Gruppen E und F). Die Unterschiede zum Beispiel 1. mit nichtgerundeten Messwerten (Gruppen A und B) sind frappant.

Gruppierte oder nichtgruppierte Daten?

Cook zeigt an einem Zahlenbeispiel mit verschiedenen Varianten (2013, 26, Tabelle 5/6) den Effekt, den die *Zusammenfassung* von Prozentzahlen aus einer Skalenanalyse zu *Gruppen* auf die Mittelwerte ausüben und wie diese zu Fehl-

terpretationen führen kann. Das Beispiel umfasst 6 verschiedene Stichproben aus 0 – 10-Skalen, wobei jedes Mal gleichermassen 25 % der Skalenwerte von 7 bis 10 in eine Gruppe (mit unterschiedlichen Antwortverteilungen) zusammengefasst wurden. Die Mittelwerte der 6 Samples differieren deutlich; sie variieren zwischen 4.63 und 1.75 im Extremfall, je nach Antwortverteilungen im gruppierten Bereich von 7 oder höher.

Zusammenfassung und Schlussfolgerungen

Nicht alle Skalenanalysen sind gleichartig und gleichwertig. Tatsächlich können einige richtiggehend täuschend ausfallen. Die Empfehlung des Artikelverfassers Cook (2013, 27) ist klar: Wenn immer möglich sollten 1 – 10-Skalen verwendet werden; Mittelwertanalysen sollten nichtgerundete Daten verwenden; die Rohdaten können, wenn nötig, in Prozentzahlen umgewandelt werden.

LITERATUR

- Atteslander, Peter* (1975): Methoden der empirischen Sozialforschung. 4. erw. Aufl. Sammlung Göschen. Berlin/New York: Walter de Gruyter
- Bearden, William O., Netemeyer, Richard G. & Mobley Mary F.* (Eds.): Handbook of Marketing Scales. Multi-Item Measures for Marketing and Consumer Behavior Research. Published in Cooperation With the Association for Consumer Research. Newbury Park, CA: Sage Publications Inc.
- Bruner, Gordon C., James, Karen E. & Hensel, Paul* (Eds.): Marketing Scales Handbook. A Compilation of Multi-Item Measures, Vol. III. Chicago, IL: American Marketing Association
- Childers, Terry L.* (1986): Assessment of the Psychometric Properties of An Opinion Leadership Scale. Journal of Marketing Research, Vol. 23, No. 2, May, 184-188
- Churchill, Gilbert A.* (1995): Marketing Research. Methodological Foundations, 6th Edition. Forth Worth TX: Dryden Press
- Churchill, Gilbert A. & Peter, Paul* (1982): Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. In: Journal of Marketing Research, Vol. 16, No. 4, November, 360-375
- Cook, Adam S.* (2013): An analysis of the impact of survey scales. In: Quirk's Marketing Research Review, November, 23-27
- Diekmann, Andreas* (1995): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek b. Hamburg: Rowohlt Taschenbuch Verlag
- Friedrich, Jürgen* (1980): Methoden empirischer Sozialforschung, 13. Aufl. Opladen: Westdeutscher Verlag
- Green, Paul E. & Tull, Donald S.* (1975): Research for Marketing Decisions. 3rd edition. Englewood Cliffs NJ: Prentice-Hall
- Hammann, Peter & Erichson, Bernd* (1990): Marktforschung, 2. Aufl. UTB Stuttgart/New York: Gustav Fischer
- Hartley, Michael D. & Ferrell, O.C.* (1996): The Management of Customer Contact Service Employees. An Empirical Investigation. In: Journal of Marketing, Vol. 60, No. 3, October, 52-70
- Holm, K.* (1976) (Hrsg): Die Befragung 4. München: UTB Franke
- King, Charles W. & Summers, John O.* (1970): Overlap of Opinion Leadership Across Product Categories. Journal of Marketing Research, Vol. 7, No. 1, February, 43-50
- Kromrey, Helmut* (1991): Empirische Sozialforschung, 5. Aufl. Opladen: Uni-Taschenbücher 1040
- Malhotra, Naresh K.* (1996): Marketing Research. Applied Orientation. 2nd Edition. Upper Saddle River N.J.: Prentice Hall
- Miller, Delbert C.* (1970): Handbook of Research Design and Social Measurement. Second Edition. New York: David McKay
- Scheuch, Erwin K. & Zehnpfennig, Helmut* (1974): Skalierungsverfahren in der Sozialforschung. In: Handbuch der empirischen Sozialforschung, Band 3a: Grundlegende Methoden und Techniken, zweiter Teil, 3. Aufl. Hrsg. René König. Stuttgart: Ferdinand Enke

Selltiz, Claire, Jahoda, Marie, Deutsch, Morton & Cook, Stuart E. (1972): Untersuchungsmethoden der Sozialforschung, Teil I. Neuwied/Darmstadt: Luchterhand

Stevens, Stanley S. (1951): Mathematics, Measurement and Psychophysics. In: Stanley S. Stevens (Ed.): Handbook of Experimental Psychology. New York: Wiley

Tull, Donald s. & Albaum, Gerald S. (1973): Survey Research. A Decisional Approach. Aylesbury: International Textbook