

Planung und Analyse 06/2005

Schwerpunktthema: Marktforschung zwischen Pflicht und Kür

Artikel von: ForschungsWerk GmbH

Thema: Skalenoptimierung bei telefonischen Kundenzufriedenheitsstudien (CATI)

In Abhängigkeit vom Studiendesign, der Befragungsart und eingesetzter Methode steht der Marktforscher regelmäßig vor der Wahl: Welche Skalenform bietet sich an? Eigentlich sollte man meinen, dass es im Laufe der „Entwicklungsgeschichte der Marktforschung“ schon genügend Untersuchungen hierzu gegeben hat, so dass sich eine Best Practice herauskristallisieren konnte. Mitnichten, wie man bei näherem Studium der Literatur feststellt. Häufig wird eine in der täglichen Praxis bewährte Skalenform herangezogen. In der wissenschaftlichen Literatur findet sich eine Vielzahl von Veröffentlichungen, die sich mehrheitlich jedoch mit schriftlichen Befragungen auseinandersetzen. Da CATI-Befragungen nach wie vor eine große Bedeutung in der kommerziellen Marktforschung haben, verwundert der geringe Anteil an Studien zu dieser Methodik. ForschungsWerk untersuchte deshalb im Rahmen einer Studie in Zusammenarbeit mit der Universität Erlangen-Nürnberg, welche Skalen für eine telefonische Kundenzufriedenheitsstudie die größte Aussagekraft besitzen.

Generelle Problematik der Skalenverwendung

In aller Regel kommen bei quantitativen Kundenzufriedenheitsstudien Ratingskalen zur Anwendung. Die Gestaltungsvielfalt fängt bei numerischer oder verbaler Verankerung der Skala an und reicht von der Anzahl der Kategorien bis zur Frage nach schiefer oder symmetrischer Skalenform. Generell gilt, dass mit steigender Kategorienanzahl auch eine höhere Informationsübertragung möglich ist. Einschränkend ist jedoch zu sehen, dass die Erhöhung der Kategorien auch ein erhöhtes Abstraktionsvermögen des Befragten erfordert. Überforderung führt aber zu einer Ergebnisverzerrung, die nicht steuerbar ist.

Ein weiteres Problem, das bei der Verwendung von Skalen auftritt, liegt in Antworttendenzen. In der Literatur werden zwei wesentliche Antworttendenzen unterschieden: Zum Einen die Akzentuierungstendenz, bei der sich die Befragten unabhängig vom Frageinhalt zustimmend äußern, sowie die Zentrale Tendenz, bei der die Befragten keine eindeutige Position beziehen möchten und immer die mittlere Kategorie der Skala wählen. Weitere Effekte sind der Primacy und der Recency Effekt. Während in schriftlichen Befragungen die Probanden eher zur Auswahl der ersten Antwortalternative neigen (Primacy Effekt), treten bei Telefoninterviews Recency Effekte auf, bei welchen die Befragten die letztgenannte Alternative wählen.

Anforderungen an Skalen bei telefonischen Kundenzufriedenheitsbefragungen

Um zu untersuchen, welche Effekte von Skalen in einem CATI-Interview hervorgerufen werden, wurde im

Rahmen der erwähnten Studie mit sechs unterschiedlichen Skalen die Zufriedenheit mit den Leistungen eines bekannten Produktes abgefragt. Einerseits sollte eruiert werden, ob in einem Telefoninterview auch Skalen mit einer relativ hohen Anzahl an Kategorien eingesetzt werden können. Einen zweiten Schwerpunkt bildete die Frage, ob die Verwendung bipolarer, numerischer Anker (bspw. -5; +5) zu einer Antwortverlagerung gegenüber unipolaren Ankern führt. Frühere Untersuchungen hatten gezeigt, dass Befragte verbalen Bezeichnungen in Kombination mit negativen Nummern eine noch negativere Bedeutung beimessen. Dies führt, im Vergleich mit unipolaren Skalen, zu positiveren Antworten. Getestet wurden symmetrische Skalen mit 4, 5 und 7 Kategorien, zwei Skalen mit 11 Kategorien und jeweils unterschiedlicher Verankerung [-5;+5] bzw. [0;10] sowie eine asymmetrische 5-stufige Skala. Letztgenannte wurde als einzige vollständig verbalisiert, bei den anderen Skalen wurden nur die Endpunkte verbalisiert.

Beurteilung der symmetrischen 4-, 5- und 7-Punkt-Skalen

Insgesamt kann bei den symmetrischen Skalen mit 4 und 5 Kategorien festgestellt werden, dass sie den Befragten nicht genügend Differenzierungspotenzial im positiven Bereich bieten. Dass graduiertere Urteile getroffen werden können, zeigt ein Vergleich der Verteilungen der 4-, 5- und 7-Punkt-Skalen. Es ergeben sich signifikante Unterschiede beim Median sowie im oberen Bereich der Skala. Die Urteile fallen bei den 4- und 5- Punkt-Skalen „besser“ aus, da der Extrempunkt bei der 7-stufigen Skala seltener verwendet wird. Bei einer Erhöhung der Skalenpunkte von 5 auf 7 ist eine

signifikante Abnahme der mittleren Kategorie festzustellen. Es zeigte sich, dass die Probanden das höhere Differenzierungspotenzial der 7-Punkt-Skala im positiven Bereich der Skala nutzen, wobei es hier keinerlei Anzeichen einer kognitiven Überforderung gab.

Insgesamt bestätigte sich bei allen drei Skalen die Tendenz, den negativen Bereich der Skala zu meiden - ein Ergebnis, das aufgrund einer prinzipiell erwarteten hohen Zufriedenheit sicherlich häufig in Kundenzufriedenheitsstudien auftreten dürfte.

Für die Beurteilung der Reliabilität wurde Cronbachs α für jede Skala berechnet. Es zeigte sich, dass neben der bipolaren 11-Punkt-Skala die 7-Punkt-Skala die höchste Reliabilität aufweist.

Beurteilung der asymmetrischen, verbalisierten 5-Punkt-Skala

Sowohl bei der Reliabilität als auch bei der Validität hat die asymmetrische, verbalisierte 5-Punkt-Skala im Vergleich zu allen anderen Skalen schlecht abgeschnitten und weist die geringsten Koeffizienten auf. Der Einsatz einer asymmetrischen, verbalisierten Skala ist bei einem CATI-Interview deshalb auch nicht zu empfehlen, da hier leicht eine Überforderung der Befragten auftreten kann. Dies zeigte sich bei den Antwortverweigerungen, die höher ausfielen als bei allen anderen Skalen. Des Weiteren ist der Zeitaufwand durch das Vorlesen der einzelnen Skalenpunkte höher. Als entscheidender Nachteil ist hier jedoch zu sehen, dass bei dieser Skala keine auch nur annähernd metrischen Daten generiert werden, was aber Voraussetzung für eine Reihe multivariater

Auswertungsverfahren wie bspw. Regressionsanalyse ist. Ein weiterer Nachteil dieser Skala liegt in dem nicht vorhandenen Differenzierungspotenzial des negativen Bereichs. Sollten Befragte doch einmal verstärkt negative Urteile abgeben, kann dies nicht ausreichend abgebildet werden.

Positiv ist bei dieser Skala zu sehen, dass durch die Asymmetrie tatsächlich eine größere Differenzierung hinsichtlich des Zufriedenheitsgrades im oberen Bereich erfolgt, der auch von den Befragten angenommen wird.

Beurteilung der 11-Punkt-Skala

Obwohl sich die 11-Punkt-Skalen durch gute Reliabilitäts- und Validitätswerte auszeichnen, sind sie u.E. weniger geeignet für einen Einsatz in einem CATI-Interview. Grund ist das unzureichend ausgenutzte Differenzierungspotenzial der Skala, so dass de facto nur maximal eine 8er-Skala verwendet wurde. Zudem ist von einem Einsatz bipolarer, numerischer Anker abzuraten. Die Ergebnisse bestätigten Erkenntnisse früherer Studien, wonach die Befragten numerische Anker nutzen, um sich die Bedeutung der verbalen Anker zu erschließen. Negative Zahlen bei einer bipolaren Skala verstärken noch die Bedeutung des verbalen Labels, was offensichtlich dazu führt, dass Befragte negative Werte nicht verwenden. Es zeigt sich, dass die Probanden bei der bipolaren 11-Punkt-Skala ein signifikant positiveres Urteil abgaben.

Empfehlung für telefonische Kundenzufriedenheitsstudien

Auf Basis der Ergebnisse kann der Einsatz einer 7-Punkt-Skala mit unipolarer, numerischer Verankerung

empfohlen werden. Dies gilt ebenso für Befragungen, bei denen mittels Erfahrungswerten eine starke Zufriedenheit erwartet wird, wie auch für Befragungen, in denen vorab keine Informationen über den Zufriedenheitsgrad vorhanden sind.

Kurzfassung

Welche Skala für welche Fragestellung? Der vorliegende Artikel beschäftigt sich mit der Frage, welche Skalenform besonders für telefonische Kundenzufriedenheitsbefragungen geeignet ist. In einer Studie wurden 4-, 5-, 7- und 11-Punkt-Skalen (uni- und bipolar, symmetrisch und asymmetrisch) miteinander verglichen und hinsichtlich ihrer Einsatzfähigkeit und Aussagekraft bewertet. Auf Basis der Studienergebnisse kann der Einsatz einer unipolaren, numerischen 7-Punkt-Skala empfohlen werden, unabhängig davon, ob vorab Informationen über den erwarteten Zufriedenheitsgrad vorliegen. Denn diese Skala bietet ein hohes Differenzierungspotenzial sowohl im positiven wie auch im negativen Bereich, ohne den Befragten zu überfordern. Außerdem lassen sich annähernd metrisch skalierte Daten generieren – Voraussetzung für viele multivariate Auswertungsverfahren.

Abstract

This article discusses which kind of scale is the most preferable for customer satisfaction studies carried through by telephone (CATI). Different scales (with 4, 5, and 7 grades, unipolar and bipolar, symmetrical and asymmetrical) had been compared and evaluated regarding their utilisability and significance in a study.

Based on this study's results it is recommended to use an unipolar, numeric scale with 7 grades. This scale gives the possibility to differentiate the test person's statements without overextending. Furthermore, metrical data can be recorded which is a precondition for multivariate data analysis.

Autorinnen:

Dr. Hariet Köstner, ForschungsWerk GmbH
Sandra Bauer, Lehrstuhl für Statistik und Ökonometrie,
Friedrich-Alexander-Universität Erlangen-Nürnberg

Email: koestner@forschungswerk.de

sandra.i.bauer@web.de

Telefon: 0911-99 08-105

Fax: 0911-99 08-111