

Servicequalität messen, bewerten und auszeichnen: Standardisierte Methodik bei Servicetests

Erstellt in Zusammenarbeit mit Prof. Dr. Thomas Liebetruh,
Hochschule Regensburg

Version: 2.3
Stand: Januar 2022

Inhalt:

| | |
|---|-----------|
| 1. Das Marktforschungsinstitut | 2 |
| 2. Methodische Grundlagen der Servicemessung..... | 3 |
| 2.1 Wissenschaftlich fundierte Servicequalität aus Kundensicht..... | 3 |
| 2.2 Objektive Erhebung der Servicequalität: Mystery-Tests..... | 8 |
| 3. Testmethodik Deutsches Institut für Service-Qualität | 11 |
| 3.1 Testdesign: vielschichtiger kanal- und branchenspezifischer Analyseansatz..... | 11 |
| 3.2 Datengewinnung und -Analyse: Durchführung, Rollenspiele und Fragebogen..... | 12 |
| 3.3 Umfassende Qualitätssicherung | 13 |
| 4. Detaillierte Testkriterien | 14 |
| 4.1 Filial-Test..... | 15 |
| 4.2 Telefonische Kontaktqualität..... | 16 |
| 4.3 E-Mail-Beantwortung..... | 16 |
| 4.4 Online-Chat-Gespräche | 16 |
| 4.5 Video-Call..... | 17 |
| 4.6 Internet-Test..... | 17 |
| 4.7 App-Test..... | 17 |
| 4.8 Versandqualität | 17 |
| 5. Literatur | 18 |

Geschäftsführung:
Markus Hamer
Beirat:
Brigitte Zypries (Vorsitz),
Jochen Dietrich, Marianne Voigt
Ust.-Id.-Nr. DE249603922
Amtsgericht Hamburg HRA 104009
Persönlich haftende Gesellschafterin:
buscha Verwaltungsgesellschaft mbH,

1. Das Marktforschungsinstitut

Ausgangssituation – Informationsflut, Angebotsvielfalt, Wettbewerbsdruck: In einer zunehmend komplexen Welt wird Servicequalität immer wichtiger – für Unternehmen und Kunden gleichermaßen. Der Verbraucher benötigt häufig neben den physischen Produkten eine optimale Beratung und einen schnellen, individuellen Service. Für den Anbieter ist eine hohe Kundenzufriedenheit entscheidend, um im Markt zu bestehen – Servicequalität ist ein Wettbewerbsvorteil.

Mission – Das Deutsche Institut für Service-Qualität ist der beste Qualitätskompass für smarte Verbraucher. Mit unseren Werten Objektivität, Unabhängigkeit und Transparenz bietet wir Orientierung und fördern fairen Wettbewerb. Das Institut verfolgt seit seiner Gründung im Juni 2006 das Ziel, die Servicequalität in Deutschland zu verbessern und die Kundenzufriedenheit zu erhöhen. Dem Verbraucher wollen wir Entscheidungshilfe bieten, dem Unternehmen Servicedefizite und Optimierungspotenziale aufzeigen.

Leistung – Das privatwirtschaftliche Marktforschungsinstitut führt als Mediendienstleister unabhängige und wissenschaftlich fundierte Servicestudien von Unternehmen und Branchen durch. Die Studien werden von renommierten Fach- und Publikumsmedien beauftragt und veröffentlicht. Das Hamburger Institut gibt dem Verbraucher damit Orientierung im Bereich Service und bietet Anhaltspunkte für Kaufentscheidungen. Unternehmen können Studien nicht in Auftrag geben.

Methode – Untersucht werden durch objektive, mehrdimensionale Analysen Serviceleistungen in unterschiedlichen Kanälen der Kommunikation zwischen Kunden und Anbietern (z.B. Filiale, Telefon, E-Mail, Website und App). Die Beurteilung erfolgt anhand von Beratungstests und Testkäufen (Mystery-Tests). Insgesamt sind rund 2.000 geschulte Tester in ganz Deutschland im Einsatz. Ein Team aus Ökonomen, Soziologen und Psychologen entwickelt die Testkonzeptionen und wertet die Ergebnisse aus. Um die Servicewünsche der Verbraucher zu kennen, werden regelmäßig Gruppendiskussionen geführt.

Nutzen – Auch getestete Unternehmen gewinnen durch die Tests des Instituts wertvolle Informationen für ihr Qualitätsmanagement. Sie erfahren, in welchen Bereichen sie im Vergleich zur Konkurrenz gut aufgestellt sind und wo sie ihren Service noch verbessern müssen.

Die Auftraggeber der Studien:



2. Methodische Grundlagen der Servicemessung

Die Durchführung von Studien zur Messung der Servicequalität erfolgt in einem standardisierten, wissenschaftlich fundierten und qualitätsgesicherten Marktforschungsprozess, der spezifisch auf die jeweilige Branche zugeschnitten wird.

Die besondere Kombination aus einer wissenschaftlich fundierten und anerkannten Operationalisierung des Konstrukts der Servicequalität einerseits und einer objektiven Erhebungsmethode andererseits zeichnet den Marktforschungsansatz des Deutschen Instituts für Service-Qualität aus.

2.1 Wissenschaftlich fundierte Servicequalität aus Kundensicht

Unter Service werden im deutschen und im englischen Sprachgebrauch teilweise unterschiedliche Inhalte verstanden. Im Gegensatz zur englischen Literatur wird in deutschsprachigen Texten in einigen Fällen zwischen Dienstleistungen und Service unterschieden. Die Dienstleistung wird als kostenpflichtiges Kernprodukt betrachtet, die über den kostenlosen Service in Form von Zusatzleistungen ausgestaltet wird (z. B. Bruhn 2008, S. 24.). Nach unserem Verständnis werden die Begriffe jedoch – in Einklang mit dem englischen Sprachgebrauch – synonym verwendet. Unter Service wird daher die Gesamtheit aller selbstständigen und produktbegleitenden Dienstleistungen verstanden.

Die Messung und damit auch die Definition von Servicequalität ist der Gegenstand einer intensiven Diskussion in Wissenschaft und Praxis. Die Herausforderungen bei Definition und Messung der Servicequalität liegen dabei in besonderen Eigenschaften von Dienstleistungen im Vergleich zu physischen Produkten (Homburg 2012, S. 953):

- Viele Eigenschaften von Dienstleistungen können erst nach Inanspruchnahme „erfahren“ werden, etwa die optische Wirkung eines Haarschnittes, oder die Beurteilung der Dienstleistung würde, wie bei einem Arztbesuch, ein hohes Fachwissen voraussetzen. Dagegen können Form, Farbe oder technische Eigenschaften von physischen Produkten bereits vor dem Kauf einfach festgestellt oder getestet werden.
- Dienstleistungen erfordern häufig, dass der Kunde selbst oder ein Objekt aus seinem Besitz bei der Erbringung der Dienstleistung unmittelbar betroffen ist.
- Dienstleistungen sind in der Regel nicht speicherbar, was für die Anbieter die Herausforderung der Kapazitätsauslastung birgt und sich für den Verbraucher häufig in einer schwer nachvollziehbaren Preisgestaltung bemerkbar macht.

Es hat sich – ebenso wie die Vielfalt und Heterogenität der Dienstleistungsarten (Bruhn 2010, S. 34 f.) – eine Vielzahl von Abgrenzungen und Definitionsansätzen herausgebildet (Bruhn 2008, S. 19 ff.). In der Literatur ist deshalb eine Vielzahl verschiedener Ansätze zur

Operationalisierung des Konstrukts Servicequalität zu finden (Seth/Deshmukh/ Vrat 2005, S. 913 ff.). Diese Ansätze stellen im Wesentlichen auf verschiedene Dimensionen von Dienstleistungen ab, die in den Ansätzen unterschiedlich stark berücksichtigt werden (Homburg 2012, S. 955):

- Potenzialqualität: Leistungsvoraussetzungen, die für die Erbringung von Dienstleistungen notwendig sind (z. B. Anzahl und Qualifikation der zur Verfügung stehenden Mitarbeiter)
- Prozessqualität: Aktivitäten, die während des Prozesses der Erstellung der Dienstleistung stattfinden (Freundlichkeit der Mitarbeiter, Fehlerfreiheit)
- Ergebnisqualität: Ergebnis des Dienstleistungsprozesses im Sinne der Kundenanforderung (Funktionsfähigkeit eines reparierten Fahrzeugs)

Der bekannteste wissenschaftlich fundierte, vielfach praxiserprobte und gleichzeitig branchenunabhängige Ansatz ist der SERVQUAL- bzw. SERVPERF-Ansatz (Homburg 2012, S: 953). Er eignet sich darüber hinaus insbesondere, da er auf dem Einnehmen einer Kundenperspektive basiert.

Im Rahmen des SERVQUAL-Ansatzes (Parasuraman/Zeithaml/Berry 1988), ein Kunstwort aus Service und Qualität, wurde eine standardisierte Skala zur Messung der Qualität von Service und der daraus folgenden Kundenzufriedenheit entwickelt. Es wird einerseits die Kundenerwartung und andererseits das Kundenerlebnis abgefragt. Über die Differenzbildung kann eine Aussage über Leistungsdefizite getroffen werden. Der Ansatz gehört zu den in der Wissenschaft meistzitierten und in der Praxis meistbenutzten Verfahren zur Messung der Servicequalität.

Der Ansatz definiert fünf Dimensionen, die sich aus einer Vielzahl von einzelnen Kriterien (insgesamt 22 Einzelkriterien) zusammensetzen (Parasuraman/Zeithaml/Berry 1988, S. 23):

- Materielles Umfeld (Räumlichkeiten, Ausstattung, Personal)
- Zuverlässigkeit (Fähigkeit die avisierte Leistung akkurat und verlässlich zu erfüllen)
- Reaktionsfähigkeit (Fähigkeit und Bereitschaft schnell und flexibel auf Kundenanfragen und -wünsche zu reagieren)
- Leistungskompetenz (Wissen, Höflichkeit, Sicherheit, Vertrauen)
- Einfühlungsvermögen (emotionale Kompetenz, individuelle Kundenbetreuung)

Der SERVPERF-Ansatz ist eine Weiterentwicklung des SERVQUAL-Ansatzes (Cronin/Taylor 1994). Der einstellungsorientierte Ansatz geht davon aus, dass Probanden, die die Servicequalität eines Anbieters

bewerten, automatisch einen Vergleich zwischen wahrgenommener und erwarteter Servicequalität vornehmen bzw. die Servicequalität über eine relativ dauerhafte, gelernte innere Haltung gegenüber einem Objekt beurteilen. Der klassische SERVQUAL-Ansatz stellt dagegen explizit auf den Unterschied zwischen der erwarteten Dienstleistung und der wahrgenommenen Dienstleistung ab.

Die separate Erhebung von Kundenerwartungen wird aufgrund der Überlegungen des SERVPERF-Ansatzes als überflüssig erachtet; es kann somit allein auf Basis der Wahrnehmungskomponente von SERVQUAL eine Qualitätsmessung vorgenommen werden. Der dadurch deutlich reduzierte Erhebungsaufwand trägt zu einer Erhöhung der Reliabilität der Messung bei (Brown/Churchill/Peter 1993, S. 127 ff.; Bruhn 2008, S. 137 ff., Schmidt 2007, S. 109 f.).

Zusammenfassend können als Vorteile von SERVQUAL bzw. SERVPERF insbesondere die Ganzheitlichkeit und Einfachheit der Rangfolgenbildung genannt werden. Diese Merkmale stellen einen hohen Praxisbezug sicher und zeigen die nach wie vor hohe empirische Relevanz des Konstrukts.

Der hohe Praxisbezug und damit gleichzeitig eine branchenübergreifende, universelle Einsetzbarkeit des SERVQUAL/SERVPERF-Ansatzes werden durch eine Vielzahl durchgeführter wissenschaftlicher Studien in verschiedenen Branchen dokumentiert (Auswahl):

- Banken, Versicherungen (Parasuraman/Berry/Zeithaml 1991, Zhou 2004)
- Handel (Meng/Summey/Herndon/Kwong 2009)
- Hotels (Getty/Getty 2003, Ingram/Daskalakis 1999; Lau/Akbar/Yong Gun Fie 2005, Saleh/Ryan 1991, Ting-Kwong Luk/Layton 2004)
- Dienstleistungsqualität im Internet und von IT-Systemen (Ding/Hu/Sheng 2011, Nambisan/Watt 2011, van Dyke/Kap-pelman/Prybutok 1997)
- Gesundheitswesen (Cronin/Taylor 1994, Dagger/Sweeney/Johnson 2007, McAlexander/Kaldenburg/Koenig 1994)
- Bildungswesen (Arambewela/Hall 2006)

Einige empirisch nachgewiesene Erkenntnisse, die hauptsächlich aus dem Forschungsfeld der **Verhaltensökonomie**¹ stammen, führen dazu, dass bei der Bewertung des Konstrukts Servicequalität weitere Aspekte berücksichtigt werden müssen.

¹ Die Verhaltensökonomie basiert auf der Ablehnung bzw. der Erweiterung des Modells des „Homo-Öconomicus“, also einem Menschen, der vollständig rational ist, seinen Eigennutz maximiert, frei von Emotionen ist und keine Fehler in der Informationsaufnahme und -Verarbeitung macht (vgl. Beck 2014, S. 1 ff.)

Kundenzufriedenheit, Wirtschaftlichkeit und Kundenaufwand

Das Angebot einer nachhaltigen und guten Servicequalität kann – neben sozialen und ggf. ökologischen Aspekten – nur einhergehen mit einer wirtschaftlich profitablen Erstellung und einer möglichst einfachen Inanspruchnahme durch den Nutzer. Eine gute Servicequalität und hohe Kundenzufriedenheit sollte also immer eine Win-Win-Situation sein. Dies kann sich beispielsweise in Form von Wiederholungskäufen, positiven öffentlichen Äußerungen oder ganz allgemein Anbieterloyalität materialisieren.

Obwohl in empirischen Studien ein starker Einfluss zwischen Unzufriedenheit mit einer Dienstleistung und dem Kundenverhalten erkennbar ist (Finkelmann/Goland 1990, S. 3 ff.), kann ein Zusammenhang zwischen hoher Kundenzufriedenheit und Kundenloyalität jedoch vielfach nicht bestätigt werden (Stum/Thiry 1991, S. 34 f. oder Gierl 1993, S. 90 ff.). Statt eines Zusammenhangs zwischen Kundenzufriedenheit und Kundenloyalität lassen empirische Studien den Schluss zu, dass Kundenloyalität eher mit dem Kundenaufwand korreliert (Dixon/Freeman/Toman 2010, S. 116 ff.). So wurde ein stärkerer positiver Zusammenhang zwischen dem Customer Effort Score² und der Wiederkaufswahrscheinlichkeit ermittelt, als mit dem Net Promoter Score und ein wesentlich höherer als mit dem Zufriedenheitswert. Je einfacher und weniger aufwändig der Servicekontakt für den Kunden ist, desto eher wird er diesen Anbieter auch beim nächsten Mal wählen.

Diese Erkenntnisse decken sich mit der Transaktionskostentheorie, die besagt, dass Individuen unter der Annahme begrenzter Ressourcen Transaktionen nach deren Effizienz bewerten und gestalten. Wenn man sich also beispielsweise einmal aufwändig durch die Anmelde- masken einer Buchungsplattform gequält hat und die Buchung ansonsten auch akzeptabel war, wird man das nächste Mal wahrscheinlich auch wieder dort buchen.

Betrachtung kritischer Erlebnismomente

Kritische Erlebnismomente spielen bei der Analyse der Dienstleistungsqualität eine große Rolle. So ist die Methode der kritischen Ereignisse (Critical Incident Technique) eine gängige Vorgehensweise, um Hintergründe für eine Bewertung der Dienstleistungsqualität zu erfahren. Dabei werden Kunden in systematischen Befragungen gebeten, sich an Kontaktsituationen (Moments of Truth) zu erinnern, die für sie besonders zufriedenstellend oder besonders unzufriedenstellend waren (vgl. Homburg 2012, S. 961).

Besonders positive oder besonders negative Ereignisse sowie das Erleben am Ende des Kontakts spielen bei der rückblickenden Bewertung eines Serviceerlebnisses häufig die größte Rolle. Nach der sogenannten Peak-End-Regel sind für eine rückblickende Bewertung eines

² Der Customer Effort Score wird dabei auf einer Skala von 1 bis 5 mit der Frage ermittelt, wie viel Anstrengung der Kunde für die Befriedigung seines Bedürfnisses aufwenden musste.

Erlebnisses der intensivste Moment und das Ende am wichtigsten (vgl. Kahnemann et al. 1993). Hintergrund ist, dass das Gehirn das Erleben einer Situation anders bewertet als die Erinnerung daran. Bei der rückblickenden Bewertung wird zudem auch die Dauer der Situation ignoriert (duration neglect).

Diese Erkenntnisse wurden in verschiedenen Studien zur Erinnerungswirkung von Werbung (vgl. Baumgartner/Sujan/Padgett 1997) oder der Bewertung von Restaurants (vgl. Just/Sigirici/Wansink 2015) bestätigt.

Einfluss positiver Emotionen beim Erleben der Dienstleistung

In vielen Branchen sind die Leistungen anhand objektiver Kriterien nur noch wenig differenzierbar. Das gilt insbesondere für Dienstleistungen im digitalen Umfeld, wo die materiellen objektiven Rahmenbedingungen wie der Look der Webseite sehr schnell verändert werden können. Wettbewerber können sich hier nur schwer einen nachhaltigen Vorsprung erarbeiten.

Gleichzeitig haben Studien ergeben, dass Kunden Entscheidungen häufig intuitiv treffen und die Customer-Experience³, die aus einer Vielzahl von bewussten und unbewussten Komponenten besteht, auf der Basis von erfahrenen Emotionen bewerten. Faktoren wie die Einfachheit der Problemlösung und besonders intensiv erlebte schöne Momente können auch eine positive emotionale Wirkung haben. Anhand von positiven Emotionen und emotionaler Nähe lässt sich zukünftiges Kundenverhalten gut vorhersagen (vgl. Shaw/Hamilton 2016).

Ähnliche Ansätze lassen sich auch in der (Marketing-)Psychologie finden: Das Konzept der Reduzierung der kognitiven Dissonanz versucht beispielsweise zu erklären, warum Menschen eine emotionale getroffene Kaufentscheidung durch rationale Argumente nachträglich vor sich selbst rechtfertigen. Auch im Relationship-Marketing greifen Konzepte wie Customer Engagement (vgl. dazu Vivek/Beatty/Morgan 2012 oder Brodie et al. 2011) diese Erkenntnisse auf und bewerten, wie stark Kunden im Erstellungsprozess (cocreation) emotional involviert sind.

Zusammenfassend haben diese drei Aspekte nun folgende Implikationen für die Erweiterung des SERVQUAL/SERVPERF-Modells bei der Bewertung der Servicequalität:

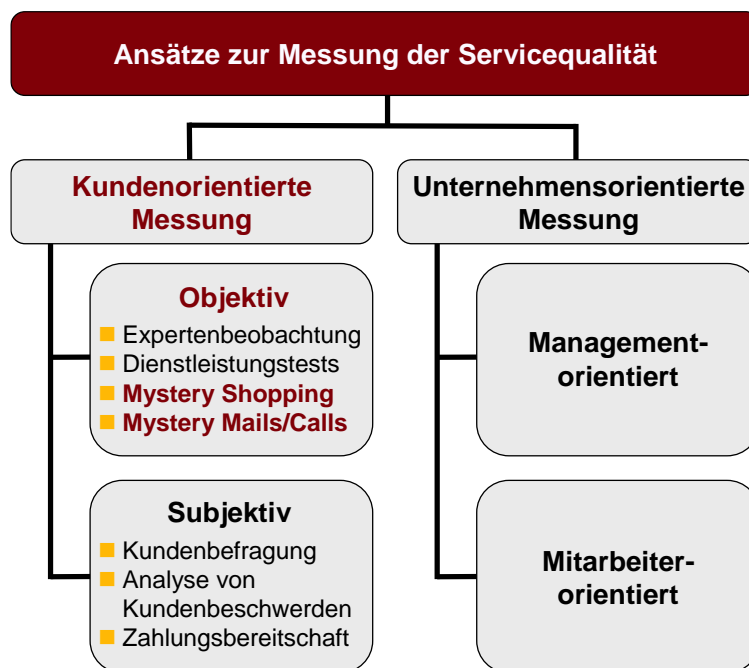
- Die emotionale Wirkung der Dienstleistung auf den Nutzer sollte ergänzend erhoben werden
- Kritischen Erlebnismomenten sowie den finalen Momenten der Dienstleistung sollte eine höhere Bedeutung eingeräumt werden

³ Shaw/Hamilton (2016, S. 8) definieren Customer Experience als "a customer's perception of their rational, physical, emotional, subconscious, and psychological interaction with any part of an organization. These perceptions influence customer behaviors and build memories, which drive customer loyalty and thereby affect the economic value an organization generates."

- Der Aufwand bei der Inanspruchnahme der Dienstleistung durch den Kunden sollte bewertet werden

2.2 Objektive Erhebung der Servicequalität: Mystery-Tests

Auch für die Erhebung bzw. Messung der Servicequalität existieren je nach Zielsetzung unterschiedliche Verfahren. Gemäß der Perspektive wird hierbei zwischen kunden- und unternehmensorientierten Ansätzen unterschieden. Unternehmensorientierte Verfahren eignen sich für die betriebswirtschaftliche Ergebniskontrolle bzw. sind auf den Prozess der Dienstleistungserstellung gerichtet.⁴ Liegt das Ziel dagegen in der optimalen Erfüllung der kundenseitigen Erwartungshaltung, sind die Qualitätsanforderungen sowie deren Erfüllung auf kundenorientierter Basis zu evaluieren (Bruhn 2008, S. 129 ff.).



Innerhalb der kundenorientierten Verfahren lassen sich objektive und subjektive Ansätze unterscheiden. Zu den subjektiven Ansätzen zählen zum Beispiel Kundenbefragungen oder Fokusgruppengespräche. Hierbei wird die individuelle Kundenmeinung abgefragt. Zu den objektiven Verfahren gehören Expertenbeobachtungen, Dienstleistungs- und Warentests sowie Mystery-Aktivitäten. Objektive Verfahren dienen einer nachprüfbaren, globalen Messung der Dienstleistungsqualität (Bruhn 2008, S. 131 ff.).⁵

⁴ Diese Verfahren, die die Qualität aus Sicht des Unternehmens (Management oder Mitarbeiter) beurteilen sind relevant die Ermittlung der verschiedenen Abweichungen im Rahmen der Dienstleistungserstellung und -kommunikation sowie der Vorgaben des Managements innerhalb des Unternehmens (Bruhn 2008, S. 191 ff.).

⁵ Die beiden Ansätze Mystery Shopping und Kundenbefragung unterscheiden sich grundlegend voneinander. Bei einer Kundenbefragung werden die Probanden nach ihrer persönlichen Meinung befragt. Diese subjektiven Kundenmeinungen können durch Phänomene wie z. B. Vergessen oder sozial erwünschtes Verhalten verzerrt sein

Die objektiven Verfahren lassen sich im Wesentlichen der Beobachtung zuordnen. Bei der wissenschaftlichen Durchführung von Beobachtungen im Rahmen der Marktforschung sind einige Gestaltungsaspekte relevant, anhand derer auch eine Beurteilung der Eignung der objektiven Verfahren vorgenommen werden kann (Homburg 2012, S. 267 f; Berekoven/Eckert/Ellenrieder 2006, S. 149 ff.):

- Transparenz der Untersuchungssituation – Grad, indem dem Beobachteten die Beobachtung sowie der Zweck bewusst ist: Es können offene, nicht-durchschaubare und verdeckte Untersuchungen unterschieden werden.
- Partizipationsgrad des Beobachters – Grad, indem der Beobachter an der Situation teilnimmt: Es können teilnehmende und nicht-teilnehmende Beobachtungen unterschieden werden.
- Standardisierungsgrad – Grad der Strukturierung der Beobachtung: Hier können standardisierte und nicht-standardisierte Erhebungen unterschieden werden.
- Beobachtungsort – Natürlichkeit des Umfelds, in dem die Beobachtung stattfindet: Hier können Laborexperimente und Feldbeobachtungen unterschieden werden.

Als besonders geeignete Erhebungsformen gelten teilnehmende verdeckte Beobachtungen in einer natürlichen Situation, da sogenannte Reaktivitätseffekte der beobachteten Personen vermieden werden können und die untersuchte Situation damit so realistisch wie möglich abgebildet werden kann. Eine strukturierte Erhebung hilft darüber hinaus die Vergleichbarkeit und somit die Zuverlässigkeit der Beobachtungsergebnisse zu erhöhen (Schmidt 2007, S. 86).

Ein etabliertes Verfahren der teilnehmenden, verdeckten und strukturierten Beobachtung im Rahmen der Marktforschung ist Mystery Shopping bzw. Silent-Shopping. Darunter werden Verfahren zur Erhebung von Servicequalität verstanden, bei denen geschulte Tester als normale Kunden auftreten und reale Kundensituationen wahrnehmen. Neben Testkäufen gibt es weitere Arten von Servicetests, welche analog zum Mystery Shopping durchgeführt werden. Als Mystery Call bezeichnet man das Durchführen von Testanrufen. Ein Mystery Guest testet die Dienstleistungen in Hotels oder Restaurants. Auch das Versenden von Test-E-Mails ist eine Form von Servicetest. Ebenso können Internetauftritte aus Nutzersicht untersucht und analysiert werden (Finn/Kayande 1999, S. 196 f.).

Das Serviceerlebnis wird dabei nach einem zuvor festgelegten und standardisierten Kriterienkatalog bewertet. Branchenspezifika können durch explorative Experteninterviews oder Gruppendiskussionen in den Kriterienkatalog eingebracht werden. Eine weitere Möglichkeit ist

(Berekoven/Eckert/Ellenrieder 2006, S. 98 ff., Schmidt 2007, S. 82 ff.). Beim Mystery Shopping beobachten geschulte Testkunden dagegen Verhalten und Tatbestände und bewerten diese anhand vorgegebener Antwortmöglichkeiten.

die Identifikation von kritischen Punkten, indem der Dienstleistungserstellungsprozess intensiv durchleuchtet wird (Bruhn 2008, S. 134).

Einflussfaktoren des Ergebnisses von Mystery-Shopping-Studien sind neben externen Elementen wie der zu untersuchenden Dienstleistung oder der Kettenzugehörigkeit⁶ auch beeinflussbare Erfolgsfaktoren wie eine intensive Schulung und Evaluierung der Tester, eine für die normale Kundenpopulation repräsentative Zusammensetzung des Testerspanels, eine standardisierte, repräsentative Prüfung der Dienstleistung und eine möglichst mehrfache Begutachtung des zu untersuchenden Ortes der Dienstleistungserstellung zu unterschiedlichen Zeitpunkten.

Je nach Vorliegen der oben genannten Erfolgsfaktoren werden in der Literatur als eine ausreichende Anzahl Werte von 4 bis hin zu 40 bis 60 Testkontakten je Unternehmen genannt. (Dawson/Hillier 1995, S. 417 ff.; Finn/Kayande 1999, S. 212 ff.; Finn 2001, S. 316 ff.; Schmidt 2007, S. 127 ff.). Finanztest wertet bei einzelnen Studien sieben Testkontakte aus (Finanztest 2010, S. 26).

Zusammenfassend ist nicht die subjektive Wahrnehmung (von Kunden), sondern eine möglichst objektive und standardisierte Beurteilung von Qualitätsaspekten das besondere Merkmal des Verfahrens. Mystery Shopping weist insbesondere deshalb folgende Vorteile auf (Bruhn 2008, S. 132ff; Finn 2001, S. 318; Finn/Kayande 1999, S. 195 ff.; Haas 2002, S. 279; Haas 2003, S. 24 ff.; Schmidt 2007, S. 174; van der Wiele/Hesselink/van Iwaarden 2005):

- Mystery-Shopping ist insbesondere bei der Messung von Dienstleistungen auch schon bei relativ kleinen Kontaktzahlen ein sehr reliables Messinstrument, da sich die ausgebildeten bzw. geschulten Mystery Shopper durch eine standardisierte Vorgehensweise auf die einzelnen Kriterien konzentrieren können und so objektive Fakten gut erfassen können.
- Mystery Shopping eignet sich sowohl zur Erhebung qualitativer (z. B. Freundlichkeit) als auch quantitativer (z. B. Wartezeiten) Merkmale und führt aufgrund des hohen Sachverstands der geschulten Tester zu sehr differenzierten Ergebnissen mit einer vergleichsweise hohen Anzahl an zu testenden Merkmalen.
- Mystery-Shopping ist gut mit anderen Marktforschungsinstrumenten (z. B. SERVQUAL/SERVPERF) kombinierbar; Insofern ist gerade die Kombination einer wissenschaftlich fundierten und anerkannten Operationalisierung des Konstrukts der Servicequalität einerseits und der objektiven Erhebungsmethode andererseits die Grundlage für eine hohe Validität und Reliabilität der Studien.

⁶ So stellt die Bewertung einer einfachen Dienstleistung wie die Bedienung an einer Tankstelle oder einem Coffee-Shop andere Anforderungen an die Tester als die Bewertung einer komplexen Dienstleistung wie die Beratung beim Autokauf oder bei der Auswahl einer Finanzdienstleistung. Ebenso spielt es eine Rolle, ob die Servicequalität einer Kette, bei der zentrale Vorgaben zur Erbringung der Dienstleistung gemacht werden, bewertet werden sollen, oder ob einzelne Outlets mit sehr unterschiedlichen Gestaltungen der Dienstleistung vergleichend evaluiert werden.

3. Testmethodik Deutsches Institut für Service-Qualität

3.1 Testdesign: vielschichtiger kanal- und branchenspezifischer Analyseansatz

Kompetente Beratung, übersichtliche Präsentation von Waren, gute telefonische Erreichbarkeit, schnelle Beantwortung von E-Mails, informative Internetseiten: Service ist wie oben dargelegt in vielen Branchen mehr als freundliches Auftreten. Die Messung von Dienstleistungsqualität erfordert daher einen vielschichtigen, branchen- und kanalspezifischen Analyseansatz. Das Deutsche Institut für Service-Qualität verwendet ein umfangreiches Testdesign und prüft Kriterien auf bis zu fünf verschiedenen Kontaktkanälen (Filiale, Telefon, E-Mail, Internet, App, Video-Call, Versand) und anhand zahlreicher Dimensionen.

Im Vorfeld der Untersuchung werden zunächst die Untersuchungszielsetzung und die zu testenden Kriterien sowie deren Gewichtungen in Absprache mit dem beauftragenden Medienpartner festgelegt. Bei der Prüfung der Servicequalität von Unternehmen kommen dann stets mehrdimensionale Analysen und praxisrelevante Verfahren aus der empirischen Sozialforschung zur Anwendung. Beim Testdesign wird grundsätzlich ein wirkungsanalytischer Ansatz verfolgt, d. h. primär wird die Service-Wirkung auf Kunden und Interessenten untersucht.

Die SERVQUAL/SERVPERF-Dimensionen werden dazu je nach Branche und Untersuchungsinhalt sowie durch die oben beschriebenen Erkenntnisse der Verhaltensökonomie modifiziert. Beispielsweise wird beim Kriterium „Beratungserlebnis“ der Erinnerungswert sowie der Wohlfühlfaktor erhoben. Hierdurch wird das Kriterium Einfühlungsvermögen aus dem SERVQUAL/SERVPERF-Modell durch eine Bewertung besonders positiver oder negativer Momente sowie der emotionalen Wirkung der Dienstleistung nach den Erkenntnissen der Verhaltensökonomie angereichert. Ein weiteres Beispiel ist der Aufwand bei der Inanspruchnahme einer Dienstleistung. Dieser Forderung aus der Verhaltensökonomie wird die Testmethodik insofern gerecht, indem zum einen das Kriterium Zuverlässigkeit aus dem SERVQUAL/SERVPERF-Modell in das Kriterium „Lösungsqualität“ umbenannt wird und dort zum anderen die Erhebung der Relation von Gesprächsdauer und -ergebnis (bzw. von Antwortumfang und Informationsqualität) bzw. die Einfachheit der Lösung des Anliegens erfolgt.

Die Kriteriengestaltung und die Gewichtung werden regelmäßig durch Fokusgruppen- und Expertengespräche auf Relevanz überprüft. Diese Überprüfung bildet die Basis für die Gestaltung der standardisierten Fragebögen sowie die Konzeption der Rollenspiele im Rahmen der Datenerhebung.

| Filiale | Telefon | E-Mail | Chat | Video-Call | Internet | App | Versand |
|--|---------|--------|------|--|----------|--|---------|
| Umfeld | | | | Virtuelles Umfeld | | | |
| Angebot | | | | | | | |
| Beratungskompetenz Richtigkeit von Aussagen, vollständige Beantwortung etc. | | | | Internet-analyse Kontaktarten, Umfang und Zugriff auf Informationen etc. | | Bedienungsfreundlichkeit Einstieg, Menüführung, Design, Sprache etc. | |
| Lösungsqualität gezielte Bedarfsanalyse, Aufzeigen des Produktnutzens etc. | | | | Nutzerbetrachtung Optik, Navigation, Verständlichkeit etc. | | Versandqualität Versanddauer, Kosten, Zustellungsqualität etc. | |
| Kommunikationsqualität Freundlichkeit, Namensansprache, aktive Gesprächsführung etc. | | | | Nutzwert Funktionen und Inhalte | | Bestellbedingungen Rückgaberecht, Zahlungsarten etc. | |
| Wartezeiten und Erreichbarkeit Wartezeiten, Öffnungszeiten, Termineinhaltung etc. | | | | | | | |
| Beratungserlebnis Mehrwert der Beratung, Erinnerungswert etc. | | | | | | | |

3.2 Datengewinnung und -Analyse: Durchführung, Rollenspiele und Fragebogen

Das Beurteilungsverfahren basiert auf Mystery-Tests (Beratung/Kauf vor Ort, Mystery-Mails, Mystery-Calls, Mystery-Chat-Gespräche) und Internetanalysen, die von unternehmensunabhängigen, objektiven Testern aus externer Sicht durchgeführt werden. Die Erhebung erfolgt anhand eines standardisierten Fragebogens, der in großen Teilen für alle Branchen einheitlich ist – ergänzt um branchenspezifische Kriterien. Nicht die subjektive Wahrnehmung, sondern eine möglichst objektive Beurteilung von Qualitätsaspekten ist zentraler Gegenstand des Verfahrens.

Die Testdurchführung und Bewertung erfolgt nach einer standardisierten Vorgehensweise. Hierzu werden den Mystery-Testern Rollenspiele mit den zu stellenden Fragen vorgegeben. Die Themen beziehen sich auf die jeweilige Branche und entsprechen typischen Verbraucheranliegen. Zur Ermittlung dieser Anliegen führen wir regelmäßig Fokusgruppengespräche durch. Mit Hilfe von Pretests wird die Durchführbarkeit der Rollenspiele untersucht und durch Schulungen sichergestellt, dass unterschiedliche Tester dasselbe Analysemerkmal (Interrater-Reliabilität) und dass derselbe Tester zeitlich versetzt dasselbe Analysemerkmal (Intrarater-Reliabilität) gleich bewertet.

Für jedes Rollenspiel eines Mystery-Tests wird ein Testleitfaden erstellt. Der Testleitfaden enthält alle für die Durchführung relevanten Informationen und gibt Auskunft über weitere zu ermittelnde Bewertungskriterien (z.B. Geschäftszeiten oder behindertengerechter Eingangsbereich). Außerdem beinhaltet der Testleitfaden die nötigen Hintergrundinformationen, um in der Interaktionsphase die Fragen des Beraters zu beantworten. Hierfür werden im Vorwege konkrete Profile definiert. Im Falle einer Baufinanzierungsberatung wären dies beispielsweise Angaben zur Immobilie sowie zur finanziellen, beruflichen und familiären Situation des zu Beratenden. Falls soziodemografische Aspekte (z.B. Alter oder Geschlecht) eine Rolle spielen, wie dies bei Beratungen zur Altersvorsorge der Fall ist, werden die Tester so ausgewählt, dass sie diesem Profil entsprechen. Zum Beginn der Testphase findet zudem ein Briefing für alle teilnehmenden Tester statt.

Direkt im Anschluss an die Testdurchführung halten die Tester ihre Erkenntnisse und Eindrücke in einem Fragebogen fest. Die Fragen sind so formuliert, dass sie eindeutig und unmissverständlich beantwortet werden können. Dies ist die Grundlage für reliable und valide Messergebnisse. Zusätzlich sind Plausibilitätskontrollen integriert. Die Art der Erfassung der Merkmalsausprägungen ist auf das jeweilige Kriterium abgestimmt. Daher kommen verschiedene Fragetypen zum Einsatz:

- Die Erfahrungen der Mystery Tester bezüglich der Beratung und des Beratungsumfelds werden mit einer Likert-Skala gemessen. Hierzu verwenden wir eine bipolare 5er-Skalierung.
- Objektiv zu erfassende Kriterien werden mit dichotomen Fragen (bei zwei Antwortkategorien, z.B. Ja und Nein) sowie mit Alternativfragen (mehr als zwei Antwortkategorien) erhoben. Ohne die Vorgabe von Antwortmöglichkeiten werden Kriterien erfasst wie Wartezeit bis zum Beratungsbeginn, Zinssatz entsprechend des Testprofils oder Öffnungszeiten.
- Zusätzlich werden offene Fragen gestellt, die der qualitativen Leistungsanalyse und der Plausibilitätskontrolle dienen. Hierzu zählen die sinngemäße Wiedergabe der Mitarbeiterantworten, die Beschreibung des Mitarbeiterverhaltens und des Kontaktumfelds sowie sonstige Auffälligkeiten und Originalzitate.

3.3 Umfassende Qualitätssicherung

Unser Testdesign sowie unsere Erhebungsverfahren werden regelmäßig wissenschaftlich überprüft. Hierzu kommen sowohl klassische Verfahren wie Faktoren- oder Regressionsanalysen zum Einsatz, als auch State of the Art-Methoden wie PLS-basierte Strukturgleichungsmodelle. Beispielsweise haben wir in Kooperation mit der Universität Hamburg zwei Projekte zur Untersuchung unseres Testdesigns durchgeführt. Die Analysen bestätigen die hohe Reliabilität (Messgenauigkeit) und Validität (Gültigkeit der Messung) unserer Methodik.

Damit weiterhin sichergestellt ist, dass alle Tester das gleiche Verständnis von der Kriterienbedeutung und den Bewertungsmöglichkeiten haben, analysieren wir fortlaufend die Ergebnisse unserer Tester anhand von quantitativen und qualitativen Plausibilitätsprüfungen und führen regelmäßig gezielte Coachingmaßnahmen durch. Bevor neue Tester für unsere Servicestudien eingesetzt werden und ihre Bewertungen in die Ergebnisanalysen einfließen, durchlaufen sie verschiedene Schulungsmaßnahmen. Das einheitliche Bewertungsverständnis führt zu einer hohen Interrater- und Intrarater-Reliabilität, die gleichzeitig für die Objektivität der Methodik sprechen.

Insgesamt wird größter Wert auf umfangreiche und gezielte Maßnahmen zur Qualitätssicherung bei der Vorbereitung, Durchführung und Auswertung der Tests gelegt:

- Orientierung an den ESOMAR-Richtlinien (veröffentlicht von der European Society for Opinion and Marketing Research); in diesem

international anerkannten Verhaltenskodex verpflichten sich die Marktforscher zur Selbstregulierung, um das Vertrauen in die Öffentlichkeit zu stärken und die Objektivität und Redlichkeit ihrer Studien zu garantieren

- Kontinuierliche Überprüfung der Relevanz von Testinhalten und der Validität des Testdesigns durch regelmäßige explorative Kundenbefragungen und Expertendiskussionen sowie eine Begutachtung durch Wissenschaftler.
- Einsatz geschulter Tester nach Auswahlprozessen inkl. Probe-tests, regelmäßiges Coaching, stichprobenweise Kontrolle vor Ort (Test-Ort und -Uhrzeit werden dem Tester vorgegeben).
- Gezielte Zusammensetzung des Testerpanels, um Verzerrungen zu vermeiden: Der prozentuale Anteil des Testvolumens eines Testers pro Testdurchgang liegt unter 5 Prozent.
- Sicherung der Objektivität durch standardisierte Befragungsmethodik, inhaltliche und statistische Plausibilitätskontrollen und Fragebogen-Pre-Tests.
- Fortlaufende Dokumentation der Erhebung, stichprobenartige Erhebungskontrollen, projektbezogene Testerleitfäden für die Erhebung, Beschränkung der Testanrufe und Testbesuch pro Tester und Tag.
- Vier-Augen-Prinzip bei der Dateneingabe, Plausibilitätsprüfung der Antworten, 100-prozentige Kontrolle der Testbögen vor Dateneingabe.

4. Detaillierte Testkriterien

Die einzelnen Testkriterien sind wie beschrieben weitestgehend standardisiert. Einzelne Kriterien und Gewichtungen werden branchen- und untersuchungsabhängig in Absprache mit dem beauftragenden Medienpartner angepasst. Die genaue Untersuchungsmethodik ist den jeweiligen Dokumentationen zu entnehmen, die für jede Studie erstellt werden. Im Folgenden werden exemplarische Untersuchungsdesigns für verschiedene Kanäle dargestellt.

4.1 Filial-Test

4.1.1 Beratungstest am Beispiel Banken und Versicherungen

Im Rahmen von Filial-Tests werden Beratungssituationen analysiert. Die Gespräche werden auf Basis von definierten Rollenspielen und vorgegebenen Persönlichkeitsprofilen geführt.

Die Bewertung der Beratungsqualität kann anhand folgender Kategorien und Gewichtungen erfolgen.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, Glaubwürdigkeit des Beraters)
- Lösungsqualität (z.B. gezielte Analyse der finanziellen Situation, Aufzeigen des Kunden- oder Produktnutzens)
- Kommunikationsqualität (z.B. Freundlichkeit, Motivation der Mitarbeiter, aktive Gesprächsführung)
- Wartezeiten und Erreichbarkeit (z.B. Anteil der Beratungen, die innerhalb von 5 Minuten begannen, Wartezeiten, Öffnungszeiten)
- Qualität des Umfelds (z.B. Gestaltung und Sauberkeit der Räumlichkeiten, Raumatmosphäre, diskrete Beratungsatmosphäre, Erscheinungsbild der Mitarbeiter)
- Beratungserlebnis (z.B. Mehrwert der Beratung, Erinnerungswert)

4.1.2 Beratungstest in Filialen mit Einkaufsmöglichkeit

Die Bewertung der Beratungsqualität erfolgt anhand folgender Kategorien und Gewichtungen.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, Kenntnis der Kostenstruktur, Glaubwürdigkeit)
- Lösungsqualität (z.B. gezielte Bedarfsanalyse, Aufzeigen des Kunden- oder Produktnutzens, Relation Gesprächsdauer und -ergebnis)
- Kommunikationsqualität (z.B. Freundlichkeit, Motivation der Mitarbeiter, Reaktion auf Beschwerden)
- Wartezeiten und Erreichbarkeit (z.B. aktives Hilfsangebot, Wartezeit bis zur Beratung und an der Kasse, Öffnungszeiten)
- Qualität des Umfelds (z.B. Sauberkeit, Raumatmosphäre, Orientierungsmöglichkeiten, übersichtliche Präsentation der Waren, barrierefreies Umfeld)
- Zusatzservices (z.B. familiengerechte Ausstattung, Lieferservice, Kundentoilette)
- Kulanz beim Umtausch
- Beratungserlebnis (z.B. Authentischer Berater, Mehrwert der Beratung, Erinnerungswert)
- Angebot (z.B. Angebotsvielfalt, Aktionsangebote, Befüllung der Regale, Auszeichnung der Ware mit Preisen)

4.2 Telefonische Kontaktqualität

Mystery-Calling stellt die Servicequalität der telefonischen Beratung anhand der folgenden Kriterien auf den Prüfstand.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, inhaltliche Verständlichkeit, Glaubwürdigkeit)
- Lösungsqualität (z.B. gezielte Bedarfsanalyse, Aufzeigen des Kunden- oder Produktnutzens, Relation Gesprächsdauer und -ergebnis)
- Kommunikationsqualität (z.B. Namensansprache, Freundlichkeit, Interaktionsqualität, Motivation)
- Wartezeiten und Erreichbarkeit (z.B. zustande gekommene Gespräche, durchschnittliche Wartezeit, Erreichbarkeit)
- Beratungserlebnis (z.B. Authentischer Berater, Mehrwert der Beratung, Erinnerungswert)

4.3 E-Mail-Beantwortung

Mystery-E-Mailing misst die Servicequalität der Beantwortung von E-Mails. Alle zu testenden Unternehmen bekommen dieselben E-Mails an gleichen Tagen mit der Bitte um Beantwortung zugeschickt.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, inhaltliche Verständlichkeit, orthografisch und grammatikalisch fehlerfreie Antwort)
- Lösungsqualität (z.B. bedarfsgerechte Antworten, einfache Lösung des Anliegens, Relation Antwortumfang und -ergebnis)
- Kommunikationsqualität (z.B. Anrede mit Namen, Freundlichkeit, aktives Gesprächsangebot, Nennung wichtiger Kontaktdaten)
- Wartezeiten und Erreichbarkeit (z.B. eingegangene Antworten, Dauer bis zum Erhalt der Antwort)
- Beratungserlebnis (z.B. Mehrwert der Antwort, Erinnerungswert)

4.4 Online-Chat-Gespräche

Mystery-Chat-Gespräche prüfen die Servicequalität bei der Beratung in Online-Chats.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, inhaltliche Verständlichkeit, orthografisch und grammatikalisch fehlerfreie Antwort)
- Lösungsqualität (z.B. bedarfsgerechte Antworten, einfache Lösung des Anliegens, Relation Chatdauer und -ergebnis)
- Kommunikationsqualität (z.B. Mitarbeiter mit Name und Bild vorgestellt, Freundlichkeit, Interaktionsqualität)
- Wartezeiten und Erreichbarkeit (z.B. Sichtbarkeit/Zugänglichkeit des Chats, Durchschnittliche Wartezeit bis Chat-Beginn)
- Beratungserlebnis (z.B. Mehrwert der Beratung, Erinnerungswert)

4.5 Video-Call

Die Bewertung der Beratungsqualität erfolgt anhand folgender Kategorien.

- Beratungskompetenz (z.B. Richtigkeit der Aussagen, Vollständigkeit der Aussagen, Kenntnis der Kostenstruktur, Glaubwürdigkeit)
- Lösungsqualität (z.B. gezielte Bedarfsanalyse, Aufzeigen des Kunden- oder Produktnutzens, Relation Gesprächsdauer und -ergebnis)
- Kommunikationsqualität (z.B. Freundlichkeit, Motivation der Mitarbeiter, Reaktion auf Beschwerden)
- Wartezeiten und Erreichbarkeit (z.B. Wartezeit bis zur Video-Beratung, Gesprächsunterbrechungen)
- Qualität des virtuellen Umfelds (z.B. diskrete Beratungsatmosphäre, Erscheinungsbild des Arbeitsplatzes oder des Beraters)
- Beratungserlebnis (z.B. Authentischer Berater, Mehrwert der Beratung, Erinnerungswert)

4.6 Internet-Test

Der Online-Service wird im Rahmen einer Internetanalyse und aus Nutzersicht untersucht.

Internetanalyse

- Kontaktarten (z.B. kostenpflichtige Telefonnummer, Call-Back, Kontaktformular, Beschwerdemöglichkeit)
- Allgemeine Informationen (z.B. AGB, FAQ, Suchfunktion)
- Spezielle Informationen je nach Untersuchungsthema

Nutzerbetrachtung

- Bedienungsfreundlichkeit (z.B. Navigation, Übersichtlichkeit, Optik)
- Inhalte (z.B. Strukturierung, Verständlichkeit, Umfang)

4.7 App-Test

Die Apps von Anbietern werden anhand von Nutzerbetrachtungen auf Bedienungsfreundlichkeit und Nutzwert analysiert.

- Bedienungsfreundlichkeit (z.B. Einstieg, Menüführung, Design, Sprache)
- Nutzwert (z.B. Funktionen und Inhalte)

4.8 Versandqualität

Versandtests ermitteln die Servicequalität bei der Bestellung von Informationen oder Produkten. Alle zu testenden Unternehmen bekommen die gleichen Anfragen.

- Reaktionszeit in Tagen
- Fehlerquote (Unterlagen nach fünf Werktagen nicht erhalten)
- Konfektionierung (angefordertes Informationsmaterial erhalten)
- Korrekte Datenerfassung

5. Literatur

Arambewela, Rodney; Hall, John (2006): A comparative analysis of international education satisfaction using SERVQUAL, in: Journal of Services Research 6, S. 141-163

Baumgartner, Hans; Sujan, Mita; Padgett, Dan (1997): Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgement, in: Journal of Marketing Research, May 1997, S. 219-232

Beck, Hanno (2014): Behavioural Economics – Eine Einführung, SpringerGabler, Wiesbaden 2014

Berekoven, Ludwig; Eckert, Werner; Ellenrieder, Peter (2006): Marktforschung – Methodische Grundlagen und praktische Anwendung, 11. Auflage, Wiesbaden: Gabler 2006

Brodie, Roderick J.; Hollebeek, Linda D.; Juric, Biljana; Ilic, Ana (2011): Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research, in: Journal of Service Research 14 (3), S. 252-271

Brown, T.J./Churchill Jr., G.A./Peter, J.P. (1993): Research note: Improving the measurement of service quality, in: Journal of Retailing, Vol. 69, S. 127-140

Bruhn, Manfred (2010): Marketing – Grundlagen für Studium und Praxis, 10. Auflage, Wiesbaden: Gabler

Bruhn, Manfred (2008): Qualitätsmanagement für Dienstleistungen – Grundlagen, Konzepte, Methoden, 7. Auflage, Berlin: Springer

Cronin, J.J., Taylor, St.A. (1994): SERVPERF versus SERVQUAL, in Journal of Marketing, Vol. 58, S. 125-132.

Dagger, Tracey S.; Sweeney, Jillian C.; Johnson, Lester W. (2007): A hierarchical model of health service quality: Scale development and investigation of an integrated model, in: Journal of Service Research 10 (2), S. 123-142

Dawson, Janet; Hillier, Jill (1995): Competitor Mystery Shopping: Methodological Considerations for MRS Code of Conduct, in: Journal of the Market Research Society 37 (4), S. 417-427

- Ding, David Xin; Hu, Paul Jen-Hwa; Sheng, Olivia R. Liu (2011): e-SELFQUAL: A scale for measuring online self-service quality, in: Journal of Business Research 64 (5), S. 508-515
- Dixon, Matthew; Freeman, Karen; Toman, Nicholas (2010): Stop Trying to Delight Your Customers, in: Harvard Business Review, July/August, 2010, S. 116-122
- Finanztest (2010): Die Blamage geht weiter – Banken im Test, in: Finanztest 8/2010, S. 24-29
- Finkelman, Daniel P.; Golland, Anthony R. (1990): How not to satisfy your customers, in McKinsey Quarterly, Winter 1990, S. 2-12
- Finn, A. (2001): Mystery Shopper Benchmarking of Durable-Goods Chains and Stores, in: Journal of Service Research, Vol. 3, S. 310-320
- Finn, A.; Kayande U. (1999): Unmasking a phantom: A psychometric assessment of mystery shopping, in: Journal of Retailing 75 (2) Jg. 1999, S. 195-217
- Getty, J. M.; Getty, R. L. (2003): Lodging quality index (LQI) – Assessing customers' perceptions of quality delivery, in: International Journal of Contemporary Hospitality Management 15 (2), S. 94-104
- Gierl, Heribert (1993): Zufriedene Kunden als Markenwechsler, in: Absatzwirtschaft 36 (2), S. 90-92.
- Haas, A. (2002): Analyse von Verkaufssituationen mit Mystery Shopping. In: Jahrbuch der Absatz- und Verbrauchsforschung, 48. Jg., 3/2002, S. 277-294
- Haas, A. (2003): Wie valide sind Urteile von Mystery Shoppern?, in: Planung & Analyse – Zeitschrift für Marketing, 30. Jg., S. 24-29
- Homburg, Christian (2012): Marketingmanagement: Strategie – Instrumente – Umsetzung – Unternehmensführung, 4. Auflage, Wiesbaden: Springer Gabler
- Ingram, H.; Daskalakis, G. (1999): Measuring quality gaps in hotels: the case of Crete, in: International Journal of Contemporary Hospitality Management 11 (1), S. 24-31
- Just, David R.; Sigirici, Ozge; Wansink, Brian (2015): Peak-end pizza: prices delay evaluations of quality, in: Journal of Product and Brand Management 24 (7), S. 770-778
- Kahnemann, Daniel; Fredrickson, Barbara L.; Schreiber, Charles A.; Redelmeier, Donald A. (1993): When More Pain is Preferred to Less, in: Psychological Science 4 (6), S. 401-405.
- Koch, Jörg (2012): Marktforschung – Grundlagen und praktische Anwendungen, 6. Auflage, München: Oldenbourg

- Lau, P. M.; Akbar, A. K.; Yong Gun Fie, D. (2005): Service Quality: A Study of the Luxury Hotels in Malaysia, in: The Journal of American Academy of Business 7 (2), S. 46-56
- McAlexander, J.H.; Kaldenburg, D.O.; Koenig, H.F. (1994): Service Quality Measurement, in: Journal of Health Care Management, Vol. 14, S. 34-39.
- Meng, Juan; Summey, John H.; Herndon, Neil C.; Kwong, Kenneth K. (2009): Some retail service quality expectations of Chinese shoppers, in: International Journal of Market Research 51 (6), S. 773-796.
- Nambisan, Priya; James H. Watt (2011): Managing customer experiences in online product communities, in: Journal of Business Research 64 (8), S. 889-895
- Parasuraman, A.; Zeithaml, Valarie A.; Berry, Leonard L. (1988): SERVQUAL - A multiple item scale for measuring consumer perceptions of service quality, in: Journal of Retailing 64 (1) Jg. 1988, S. 12-40
- Saleh, F.; Ryan, C. (1991): Analysing Service Quality in the Hospitality Industry Using the SERVQUAL Model, in: Service Industries Journal 11 (3), S. 324-346
- Schmidt, Karsten (2007): Mystery Shopping Leistungsfähigkeit eines Instruments zur Messung der Dienstleistungsqualität, Marburg: Deutscher Universitäts-Verlag
- Shaw, Colin; Hamilton, Ryan: The Intuitive Customer, Palgrave Macmillan, London 2016
- Seth, Nitin; Deshmukh, S.G.; Vrat, Prem (2005): Service quality models: a review, in: International Journal of Quality & Reliability Management 22 (9), Jg. 2005, S. 913-949
- Stum, David L.; Thiry, Alain (1991): Building Customer Loyalty, in: Training & Development Journal, April 1991, S. 34-36
- Ting-Kwong Luk, S.; Layton, R. (2004): Managing both Outcome and Porcess Quality is Critical to Quality of Hotel Service, in: Total Quality Management & Business Excellence 15 (3), S. 259-279
- van der Wiele, T./Hesselink, M./van Iwaarden, J. (2005): Mystery Shopping: A Tool to Develop Insight into Customer Service Provision, in: Total Quality Management & Business Excellence, Vol. 16, No. 4, S. 529-541
- van Dyke, T.P.; Kappelman, L.A.; Prybutok, V.R. (1997): Measuring information systems service quality: Concerns on the use of the SERVQUAL questionnaire, in: MIS Quarterly (21), S. 195-209

Vivek, Shiri D.; Beatty, Sharon E.; Morgan, Robert M. (2012): Customer Engagement: Exploring Customer Relationships beyond Purchase, in: Journal of Marketing Theory and Practice 20 (2), S. 127-125

Zhou, L. (2004): A dimension-specific analysis of performance-only measurement of service quality and satisfaction in China's retail banking, in: The Journal of Services Marketing 18, S. 534-546