

Mystery Shopping – Ziele, Prozess und Qualität eines Verfahrens zum Controlling der Dienstleistungsqualität

Dobbelstein, Thomas und Windbacher, Daniel

1 Einleitung und Zielsetzung

Um sich im Wettbewerb bewähren zu können wird für viele Unternehmen ein hohes Niveau an Dienstleistungs- und Servicequalität und damit eine hohe Kundenzufriedenheit zunehmend wichtiger. Hinter dieser Bemühung steht die Annahme, dass eine hohe Dienstleistungs- bzw. Servicequalität den Schlüssel zum wirtschaftlichen Erfolg darstellt (Stücken 2003, S. 45). Kunden, welche mit dem wahrgenommenen Service einer Unternehmung zufrieden sind, weisen eine höhere Kundentreue und darüber hinaus eine geringere Preissensibilität auf. Weiterhin ist der Prozess zur Gewinnung von Neukunden aufwändiger und zumeist auch kostenintensiver, als bestehende Kunden zufrieden zu stellen und somit langfristig an das Unternehmen zu binden (Schlich 2006, S. 274, Matzler, Kittinger-Rosanelli 2000, S. 221). Auf dieser Erkenntnis basierend sind Unternehmen bemüht, die vom Kunden wahrgenommene Dienstleistungs- bzw. Servicequalität gezielt zu steuern. Eine besondere Bedeutung wird daher auf die zuverlässige Messung dieser Konstrukte gelegt. Hierbei werden an die Messinstrumente bestimmte Anforderungen gestellt, welche die Besonderheiten von Dienstleistungen abbilden können. Sie müssen vor allem die Komplexität, Intangibilität, die Beteiligung des Kunden an der Leistungserstellung und den prozessualen Charakter einer Dienstleistung erfassen. In diesem Zusammenhang gewinnt in letzter Zeit das Instrument des Mystery Shoppings auch im deutschsprachigen Raum zunehmend an Bedeutung.

Ziel dieses Beitrags ist, das Marktforschungsinstrument Mystery Shopping als eine konkrete Form des Mystery Research zu definieren und gegenüber der klassischen Kundenzufriedenheitsbefragung abzugrenzen. Weiterhin wird der Prozess eines Mystery Shopping Projektes am Beispiel der Verkaufsgesprächsanalyse aufgezeigt. Dabei wird ein besonderes Augenmerk auf die Rahmenbedingungen gelegt, welche die Güte der Daten beeinflussen. Weiterhin werden die allgemeine Zuverlässigkeit sowie die Messgenauigkeit dieses Instrumentes näher betrachtet. Die Kernaussage bezieht sich auf die Gestaltung des Forschungsdesigns einer Mystery Analyse, welches die wesentlichen wissenschaftlich-theoretischen sowie praktischen Anforderungen berücksichtigt und die üblichen Qualitätsanforderungen einer Marktforschung erfüllt.

2 Definition Mystery Shopping und Abgrenzung zur klassischen Kundenzufriedenheitsbefragung

Zufriedenheit ist das Resultat eines Vergleichs zwischen der erwarteten und der Wahrnehmung der tatsächlich erhaltenen Leistung. Unzufriedenheit kann somit aus einer subjektiv aus Kundensicht nicht erfüllten Erwartung resultieren (Abbildung 1). Weiterhin kann die subjektiv wahrgenommene Leistung vom objektiven, tatsächlich vorhandenen Leistungsniveau abweichen (Gap 5 in Abbildung 1). Abbildung 1 verdeutlicht weiterhin die Ursachen, die trotz des Bestrebens eines Unternehmens, die Kundenerwartungen zu erfüllen, zu unzufriedenen Kunden führen können.

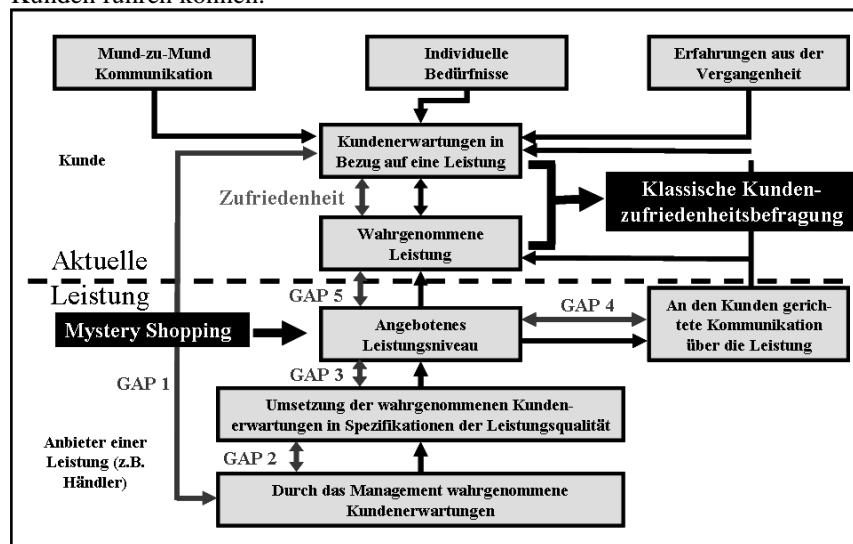


Abbildung 1: Gap- Modell als Basis einer Analyse des objektiven Leistungsniveaus und der subjektiven Kundenzufriedenheit (Zeithaml, Berry, Parasurman 1991, S.111)

Mystery Shopping ist eine Möglichkeit, die tatsächliche Qualität eines Dienstleistungsprozesses auf objektiver Ebene durch den Einsatz von Mystery Shoppern zu eruieren (Abbildung 1) (Bruhn 1997, S. 63f, Ecken 1998, S. 34, Siefke 1998, S. 106f). Mystery Shopper sind Testkunden, welche für die Mitarbeiter nicht erkennbare „reale“ Dienstleistungssituationen simulieren, um so den Dienstleistungsprozess bzw. das tatsächliche Leistungsniveau einer Unternehmung zu bewerten. Durch die verdeckte, aber teilnehmende Beobachtung werden die subjektiv erlebten Sachverhalte auf eine möglichst objektive Ebene gebracht (Bruhn 1996, S. 48). Die Ergebnisse einer Mystery Shopping Analyse, als ein Instrument des Qualitätscontrollings oder der Personalentwicklung, bilden die Grundlage einer objektiven Auswertung der Dienstleistungsqualität an

den Kundenschnittstellen (Stücken 2003, S. 45). Hierbei sollen die Testkäufer nur objektive oder objektivierbare Bewertungen vornehmen (Bauer, Urbahn, Markart 2003, S.16). Die Ergebnisse einer Mystery Analyse bilden somit objektiv den gegenwärtigen Zustand der Dienstleistungsqualität innerhalb der einzelnen Kundenschnittstellen ab. Klassische Kundenzufriedenheitsbefragungen hingegen liefern Informationen über Kundenerwartungen und deren Erfüllung in der Vergangenheit (Abbildung 1). In die subjektive Kundenzufriedenheit fließen nicht nur die aktuellen objektiven Leistungen eines Unternehmens mit ein, sondern ebenso die bisherigen Erfahrungen mit dem Unternehmen, das Unternehmensimage und die individuellen Erwartungen des jeweiligen Kunden. Somit können klassische Kundenzufriedenheitsbefragungen, welche auf Kundenwahrnehmungen basieren, nicht die objektive gegenwärtige Dienstleistungsqualität oder Einhaltung genauer Servicestandards abbilden. Aus diesem Grund empfiehlt es sich, beide Instrumente komplementär zueinander zu nutzen: Bei klassischen Kundenzufriedenheitsuntersuchungen stehen die subjektiven Meinungen des Kunden im Vordergrund, während bei der Mystery Analyse objektiv zu bewertende Kriterien aus Sicht der Zielgruppe, aber nicht aktueller tatsächlicher Kunden untersucht werden (Krüger, Stücken 2003, S. 32). Weiterhin stellt Mystery Shopping die bisher nur eingeschränkt vorhandene Verbindung zwischen quantitativer und qualitativer Marktforschung dar, da mit diesem Instrument gleichzeitig Informationen gebündelt sowie Ursachen tiefgründig (qualitativ) erforscht werden können (Lammers, Schubert 2003, S. 56). Mystery Shopping kann alle Kommunikationswege betrachten; die telefonische Beratung (Mystery Calls) kann ebenso getestet werden wie der E-Mail-Versand (Mystery Mailing) oder das Verhalten der Mitarbeiter und die Situation am POS (Mystery Shopping im engen Sinne).

3 Ziele des Mystery Shoppings

Mystery Shopping oder auch „Silent-Shopper-Verfahren“ genannt, ist ein Verfahren zur objektiven Erfassung der Dienstleistungsqualität durch Testkunden. Die Zielsetzung bzw. der Anwendungsbereich liegt daher in der:

1. Beurteilung der tatsächlichen Dienstleistungsqualität
2. Definition, Einführung und Überprüfung von Servicestandards
3. Benchmarking und Konkurrenzbeobachtung

(Miles 1993, S. 19 und Dawson, Hillier 1995, S. 417 und Wilson 2001, S. 721ff). Dienstleistungen weisen verschiedene Besonderheiten auf, welche im Mystery Shopping Prozess berücksichtigt und bewertet werden müssen, um die tatsächliche Dienstleistungsqualität beurteilen zu können. Intangibilität von Dienstleistungen bedeutet, dass die überwiegende Anzahl der Prüfkriterien,

welche zur Beurteilung der Qualität geeignet sind, nur während des Dienstleistungserstellungsprozesses wahrgenommen und bewertet werden können. Die Anzahl der Leistungsattribute, welche vor oder nach dem Kauf bzw. der Inanspruchnahme der Dienstleistung zur Verfügung stehen, sind dagegen eher gering (Matzler, Kittinger-Rosanelli 2000, S. 223). Eine Dienstleistung entsteht durch die direkte Interaktion des Unternehmens mit dem Kunden. Dieser ist an der Leistungserstellung direkt beteiligt, was die Erbringung ein und derselben Dienstleistung situationspezifisch differieren lässt. Die dritte charakterliche Eigenschaft einer Dienstleistung ist ihr prozessualer Charakter. Dies hat zur Folge, dass der Kunde die Qualität einer Dienstleistung nicht an einer vorab festgelegten Menge von einzelnen Leistungsmerkmalen beurteilt, sondern als zusammenhängenden Prozess bewertet (Stauss, Hentschel 1992, S.59ff).

Aus diesen Charakteristika resultiert, dass die teilnehmende Beobachtung als Messinstrument detailliertere Ergebnisse liefern kann als eine klassische Befragung. Der systematischen Ermittlung von Kundenerwartungen, deren Umsetzung in Servicestandards und der kontinuierlichen Messung der Einhaltung dieser Standards kommen im zweiten Anwendungsbereich eine große Bedeutung zu. Die im Zuge eines Mystery Shopping Projektes definierten Beurteilungskriterien können als Qualitätsstandards und Verhaltensgrundsätze für Mitarbeiter dienen (Berry, Zeithaml, Parasuraman 1990, S. 29ff). Somit können eigene Leistungen und die Leistungen der Konkurrenz einfach gemessen und überprüft werden. Während es oftmals schwierig ist, die Kunden des Mitbewerbers bezüglich dessen Dienstleistungsqualität zu befragen und mit der eigenen zu vergleichen, ermöglicht es Mystery Shopping einfach und ohne dass es der Konkurrent bemerkt, dessen Dienstleistungsqualität zu erfassen und auszuwerten (Schreier 2004, S. 30).

4 Prozess eines Mystery Shopping Projektes - dargestellt am Beispiel der Verkaufsgesprächsanalyse

Aus den dargestellten unterschiedlichen Zielen des Mystery Shopping wird im Folgenden die Überprüfung vorgegebener Service- bzw. Verhaltensstandards am Beispiel des Verkaufsgesprächs näher betrachtet. Der hierzu erforderliche

Prozess ist in Abbildung 2 in seinen drei Phasen aufgezeigt.

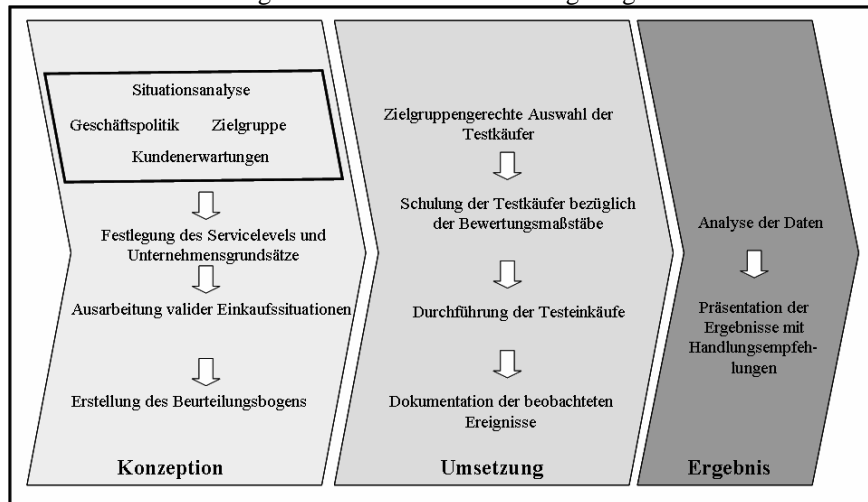


Abbildung 2: Mystery Shopping Prozess

Die folgenden Kapitel erläutern die wichtigsten Prozessschritte näher und zeigen die Vorgehensweise innerhalb dieser Arbeitsschritte auf.

4.1 Analyse der Ausgangssituation

Ausgangspunkt einer Mystery Shopping Studie kann z. B. das Erkennen einer Schwachstelle innerhalb des Dienstleistungsprozesses sein; sei es als Ergebnis einer Zufriedenheitsbefragung oder einer Analyse betriebswirtschaftlicher Zahlen (Schreier 2004, S. 30). Im Folgenden wird eine Schwachstelle innerhalb des Verkaufsgesprächs unterstellt, welches aus diesem Grund näher betrachtet und bezüglich der Gründe für die niedrige Kaufabschlussquote der Verkäufer detailliert untersucht werden soll. Bei der Konzeption eines Mystery Projektes ist darauf zu achten, möglichst objektive Ergebnisse zu erzielen. Aus diesem Grund muss bei der Festlegung des Forschungsdesigns eine besondere Bedeutung der Erstellung des Beurteilungsbogens, der Auswahl und Schulung der Testkäufer sowie der Erstellung der Testsituationen zugesprochen werden (Stücken 2003, S. 30).

Um dieser Bedeutung gerecht zu werden, sind zahlreiche qualifizierte Informationen über das zu beurteilende Unternehmen erforderlich. Grundlage der spezifischen Testkundenauswahl ist die Zielgruppe bzw. tatsächliche Kundenstruktur. Aus diesem Grund ist die Zielgruppe mittels geeigneter Kriterien, etwa relevanter soziodemographischer oder verhaltensorientierter bzw. psychographischer Merkmale zu erfassen. Zur Erstellung des späteren Beurteilungsbogens

und zur realitätsnahen Gestaltung der Einkaufssituationen sind weiterhin die geplante Geschäftspolitik und Positionierung als Grundlage vorgegebener Servicestandards zu erheben. Letztere dienen als grober Leitfaden zur Festlegung der zu prüfenden Kriterien.

4.2 *Operationalisierung des Konstruktes Fachkompetenz im Rahmen des Verkaufsgespräches*

Um einen aussagekräftigen Beurteilungsbogen erstellen zu können, welcher objektive und detaillierte Ergebnisse über die Gründe des Schwachpunktes liefert, ist es notwendig, das zu prüfende Konstrukt auf eine Ebene konkret messbarer Kriterien zu überführen. Im vorliegenden Beispiel wird der Leistungsparameter Verkaufsgespräch näher betrachtet. Vorab ist zu analysieren, welche Faktoren innerhalb des Verkaufsgesprächs aus Kundensicht als wichtig betrachtet werden. Wichtige Elemente eines qualifizierten Verkaufsgespräches sind:

1. das Kundenverständnis,
2. die Fachkompetenz,
3. die Kommunikationsfähigkeit und
4. die Verkaufskompetenz.

Um die Methodik der Operationalisierung zu verdeutlichen wird beispielhaft das theoretische Konstrukt Fachkompetenz herausgegriffen und weiter spezifiziert (Abbildung 3).

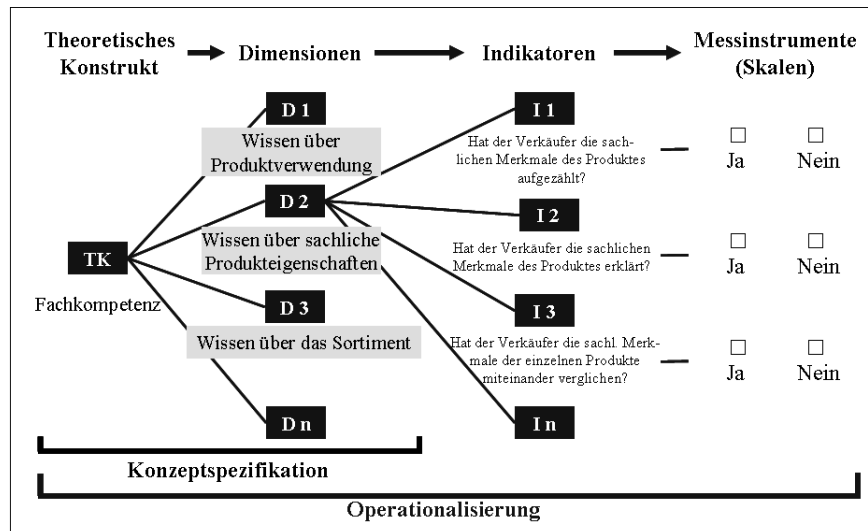


Abbildung 3: Überblick über die Stufen der Operationalisierung

In einem ersten Schritt sind die einzelnen Dimensionen der Fachkompetenz herauszuarbeiten und festzulegen:

1. Wissen über die Verwendung des Produktes
2. Wissen über sachliche Produkteigenschaften
3. Wissen über das geführte Sortiment
4. Wissen über die Produktparte im Allgemeinen
5. Wissen über aktuelle Geschehnisse in dieser Sparte.

Dimensionen

In einem zweiten Schritt der Operationalisierung werden diesen einzelnen Dimensionen Indikatoren zugewiesen. Im dritten Schritt sind die Indikatoren in Skalen zu überführen. Dabei sind dichotome Ja/Nein-Skalen zu bevorzugen, so dass beobachtbare Sachverhalte klar und eindeutig bewertet werden können. So z. B. für das Wissen über sachliche Produkteigenschaften:

1. hat der Verkäufer die sachlichen Merkmale des Produktes aufgezählt? (Ja/Nein)
2. hat der Verkäufer die einzelnen sachlichen Merkmale erklärt? (Ja/Nein)
3. hat der Verkäufer die verschiedenen sachlichen Merkmale unterschiedlicher Produkte miteinander verglichen? (Ja/Nein)

Indikatoren und Skalen

Diese nun konkret zu messenden Faktoren werden in einem Beurteilungsbogen zusammengefasst. Welche Besonderheiten hierbei, vor allem vor dem Hintergrund der Güte der Erhebung, zu beachten sind, wird im folgenden Kapitel beschrieben.

4.3 Erstellung des Beurteilungsbogens

Der Erstellung des Beurteilungsbogens kommt neben der Auswahl und Schulung der Testkäufer sehr hohe Bedeutung zu, da die Güte der Untersuchung, hier Objektivität, Reliabilität und Validität, stark von der Umsetzung dieser beiden Arbeitsschritte abhängt. Bedeutsam hierbei ist, dass die verwendeten Skalen keinen subjektiven Ermessungsspielraum für den Tester offen lassen, welche die objektiven Ergebnisse verfälschen könnten (Schlich 2006, S. 74). Aus diesem Grund empfiehlt sich auf erwähnte Ja-Nein-Skalen zurück zu greifen. Diese Antwort-Dichotomie verringert subjektiv interpretierbare Spielräume in hohem Maße und dient somit dem Ziel, alle Kriterien möglichst objektiv zu erfassen (Stücken 2003, S. 30). Werden Kriterien nicht dichotom sondern etwa auf 5er- Skalen gemessen, so ist jede einzelne Antwortkategorie klar und eindeutig zu definieren, so dass eine genaue Zuordnung der Werte zu den beobachteten Situationen erfolgen kann (Bauer, Urbahn, Markart 2003, S. 16). Diese definierten Antwortkategorien können als Handbuch ausgearbeitet werden, welches den Testeinkäufern als Nachschlagewerk zur Verfügung steht (Deckers 2003, S. 34). Neben der Zielsetzung der objektiven Datenerhebung können gezielt einige Fragen in den Beurteilungsbogen integriert werden, welche gewollt die subjektive Beurteilung durch den Testkäufer abbilden. Die Antwortinterpretation bei diesen Fragen ist jedoch in hohem Ausmaß von der Auswahl der Testeinkäufer abhängig (Stücken 2003, S. 30).

Bei der Entwicklung des Beurteilungsbogens ist weiterhin zu berücksichtigen, dass der Testkunde die Bewertung nicht während des Beobachtens dokumentieren kann, sondern diese erst unmittelbar nach Beendigung der Einkaufssituation vornimmt. Durch die geistige Kapazität der Testeinkäufer wird die Anzahl der zu beurteilenden Kriterien begrenzt (Matzler, Kittinger-Rosanelli 2000, S. 236). Durch den weitgehend standardisierten Aufbau des Beurteilungsbogens werden die Manipulationsmöglichkeiten der Testkäufer sowie das nachlässige Ausfüllen des Beurteilungsbogens reduziert (Schlich 2006, S. 74).

Insgesamt ist eine gewissenhafte und gründliche Auswahl der Testeinkäufer von hoher Bedeutung. Auf sie wird im Folgenden eingegangen.

4.4 Auswahl der Testeinkäufer

Mystery Einkäufe sind unter möglichst realen Bedingungen und unbemerkt durchzuführen, da bei Mystery Shopping die Gefahr der Einflussnahme des Testkäufers auf die Ergebnisse deutlich höher ist, als die eines Interviewers bei einer Befragung (Stücken 2003, S. 30). Aus diesem Grund ist es zwingend notwendig, die Testeinkäufer der Zielgruppe bzw. der tatsächlichen Kundenstruktur anzupassen, um zu vermeiden, dass diese durch das Verkaufspersonal erkannt und enttarnt werden. Grundlage dieser Auswahl stellt die zu Beginn ermittelte Zielgruppen- bzw. Kundenstrukturanalyse dar. Die Testkunden werden nach den ermittelten, z. B. soziodemographischen oder verhaltensorientierten bzw. psychographischen, Merkmalen der tatsächlichen Kunden ausgewählt (Matzler, Pechlaner, Kohl 2000, S. 172). Es ist auch darauf zu achten, dass der Testkunde im unmittelbaren geografischen Umfeld ansässig ist, um die Erhebung nicht durch untypische Verhaltensweisen oder Dialekt zu gefährden. Des Weiteren ist zu diskutieren, ob professionelle Testkäufer oder Laien, Kunden oder Nicht-Kunden als Testkunden eingesetzt werden. In einer Studie zur Verlässlichkeit von Testkaufdaten konnten Finn und Kayandé zeigen, dass Nicht-Kunden verlässlichere Ergebnisse liefern als reale Kunden und dass bei professionellen Testkäufern im Vergleich zu gut geschulten Laien-Testkäufern Ergebnisse von deutlich geringerer Verlässlichkeit generiert werden (Wieseke, Schmidt, Lingenfelder 2006, S. 42). In jedem Fall sind Testkäufer sorgfältig zu schulen. Eine häufige Ursache unzureichender Diagnosen der Dienstleistungsqualität liegt in mangelnder Kenntnis der Testkäufer. Ein Vorteil professioneller Testkäufer besteht in der zeitlichen Flexibilität gegenüber den Laien. Laien üben zumeist eine weitere hauptberufliche Tätigkeit aus, so dass die Verfügbarkeit nicht immer gewährleistet ist. Jedoch müssen professionelle Testkäufer immer in verschiedene Rollen schlüpfen, was nicht dem Kriterium entspricht, die Einkaufssituation unter realen Bedingungen durchzuführen (Krüger, Stücken 2003, S. 32). Bei Nicht-Kunden ist von Vorteil, dass sie Testeinkäufe unvorbelastet und ohne bisherige Erfahrungen, welche die objektive Beurteilung beeinflussen könnten, vornehmen. Ein Nachteil liegt darin, dass in manchen Fällen aufwendige Kundendaten in das EDV-System des Unternehmens eingespeist und später wieder gelöscht werden müssen, um reale Beratungssituationen simulieren zu können (Müller, Neuhaus, Löhnert 2006, S. 24). Im Unterschied dazu müssen reale Kunden aufwändig in verschiedenen Datenbanken ausgesucht und akquiriert werden, wobei eine Vielzahl an datenschutzrechtlichen Bestimmungen zu beachten ist (Schreier 2004, S. 30).

4.5 *Rechtliche Rahmenbedingungen*

Mystery Shopping ist kein rechtsfreier Raum, sondern unterliegt gesetzlichen Bestimmungen. Im Wesentlichen sind hier das allgemeine Persönlichkeitsrecht in seiner Gestalt als informationelles Selbstbestimmungsrecht (Grundgesetz Artikel 2, Abs. 1, 1 Abs. 1), das Bundesdatenschutzgesetz (insb. §4 Abs. 1, § 28, Abs. 3), das Betriebsverfassungsgesetz sowie arbeitsrechtliche Bestimmungen und Betriebsvereinbarungen relevant. Grundsätzlich ist es jedem Institut erlaubt, Mystery Shopping durchzuführen, jedoch sind bei der Dokumentation und Auswertung der Daten einige rechtliche Bestimmungen zu beachten (Bauer, Urbahn, Markart 2003, S. 16). Die Weiterleitung nicht anonymisierter Bewertungsergebnisse ist nur zulässig, wenn der betroffene Mitarbeiter zuvor eingewilligt hat oder eine vorab festgelegte Betriebsvereinbarung besteht. Seitens des Auftraggebers ist sicherzustellen, dass das Recht des Mitarbeiters auf informelle Selbstbestimmung gewahrt bleibt. Dies bedeutet insbesondere, dass der Mitarbeiter nicht stigmatisiert oder herabgesetzt werden darf. So muss eine neutrale Bewertung der Leistung des Mitarbeiters erfolgen, welche als Ausgangspunkt zur Verbesserung der Servicequalität genutzt wird und nicht, um arbeitsrechtliche Maßnahmen (Abmahnung oder Kündigung) gegen die Mitarbeiter einzuleiten. Weiterhin müssen alle technischen und organisatorischen Maßnahmen sichergestellt werden, welche die Richtlinien des Datenschutzes vorgeben. Neben den rechtlichen Bestimmungen hat die Europäische Gesellschaft für Meinungs- und Marktforschung einen Qualitäts- und Ehrenkodex veröffentlicht (www.esomar.org).

4.6 *Gestaltung der Testsituationen*

Eine Stärke des Mystery Shoppings liegt darin, Dienstleistungsqualität unter realen Bedingungen und im alltäglichen Umgang mit dem Kunden zu analysieren. Je höher die Realitätsnähe der Situationen, umso mehr wird jeder Einfluss, welcher das Ergebnis verzerren könnte, ausgeschaltet (Schlich 2006, S. 274). Dem Testkunden wird eine Einkaufssituation vorgegeben, welche er bei seinem Testeinkauf simuliert. Für die Rollendefinition erhalten die Testkäufer einen gewissen Freiraum, damit sie sich mit der ihnen zugeordneten Rolle identifizieren können (Müller, Neuhaus, Löhnert 2006, S. 24). Ebenfalls sollten die Situationen mit dem realen Vorkommen verschiedener Geschäftsvorfälle abgeglichen werden, um eine unnormale und damit auffällige Anhäufung bestimmter Kaufsituationen zu umgehen. Zahlreiche Szenarien sind nicht ohne Geschäftsabschluss durchführbar. Andererseits gibt es Einkaufssituationen, welche keinen unmittelbaren Kaufabschluss erfordern. Geschäftsabschlüsse bringen einen hohen administrativen Aufwand mit sich und binden unnötig Ressourcen. Der eigentliche Einkauf muss im Anschluss an den Testeinkauf rückgängig gemacht

werden, was erneut Ressourcen bindet und mit Kosten verbunden ist. Aus diesem Grund sollte er nur dort durchgeführt werden, wo er auch zwingend zur Beurteilung notwendig ist.

4.7 Analysemöglichkeiten und Handlungsempfehlungen

Für Mitarbeiter ist interessant zu wissen, welche Defizite in der eigenen Filiale festzustellen sind und welches - ggf. auch persönliches - Verbesserungspotenzial besteht. Je höher die Entscheidungsebene desto stärker werden Ergebnisse aggregiert, bis hin zur „One-Number“, welche es der obersten Ebene ermöglicht, einen Vergleich zwischen verschiedenen Geschäftseinheiten durchzuführen (Krüger, Stücken 2003, S. 32). Abbildung 4 gibt einen Überblick über ein entsprechendes Kennzahlensystem.

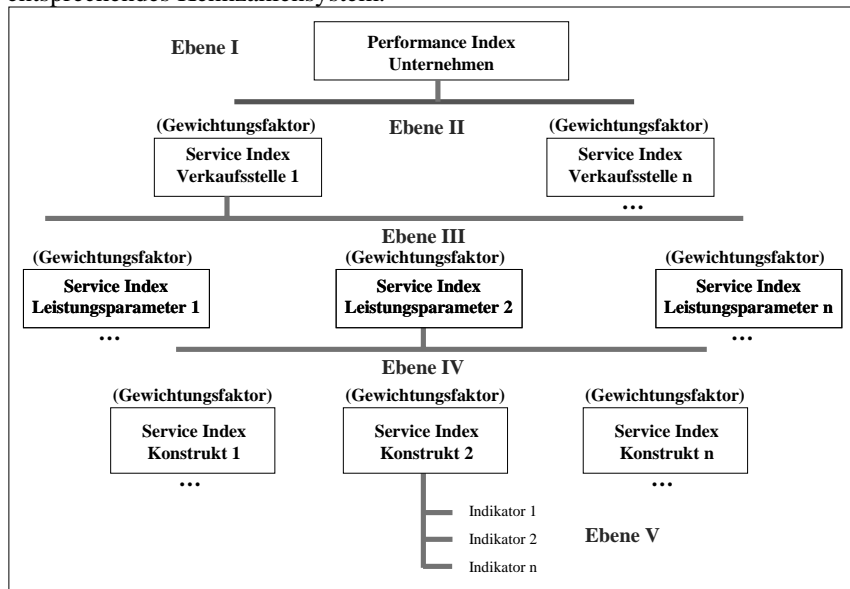


Abbildung 4: Unternehmensbezogener Performance Index des Mystery Shoppings

Die Bewertung der Einzelindikatoren (z. B. Erklärung der Merkmale eines Produktes, Vergleich der Merkmale verschiedener Produkte) ergibt den Indexwert für ein Konstrukt. Verschiedene Konstrukte (z. B. Fachwissen, Freundlichkeit) werden je nach ihrer aus der Unternehmensstrategie resultierenden Bedeutung gewichtet und zu Leistungsparametern (z. B. Beratungskompetenz, Kassierqualität) aggregiert. Die ebenfalls mit ihrer strategischen Bedeutung gewichteten Leistungsparameter ergeben den Service Index einer Verkaufsstelle. Über alle Verkaufsstellen betrachtet ergibt sich hieraus der Performance Index

des gesamten Unternehmens. Eine zusätzliche Analyse mit multivariaten Verfahren ermöglicht des Weiteren sehr spezifische Handlungsempfehlungen (Wieseke, Schmidt, Lingenfelder 2006, S. 46). So können etwa der Einfluss einzelner Leistungsparameter auf den Unternehmenserfolg berechnet und evtl. dringend erforderliche Verbesserungsmaßnahmen abgeleitet werden. Ergänzend zu den Beurteilungsbögen werden alle Beratungsbroschüren, -unterlagen und Beispielsrechnungen ausgewertet, welche der Testkunde bei seinem Testeinkauf erhält (Müller, Neuhaus, Löhner 2006, S. 24).

Liegen neben den objektiven Ergebnissen eines Mystery Shopping Projektes subjektive Zufriedenheitswerte einer Kundenbefragung vor, so können diese komplementär genutzt und in einer Handlungsrelevanzmatrix abgebildet werden (Abbildung 5).

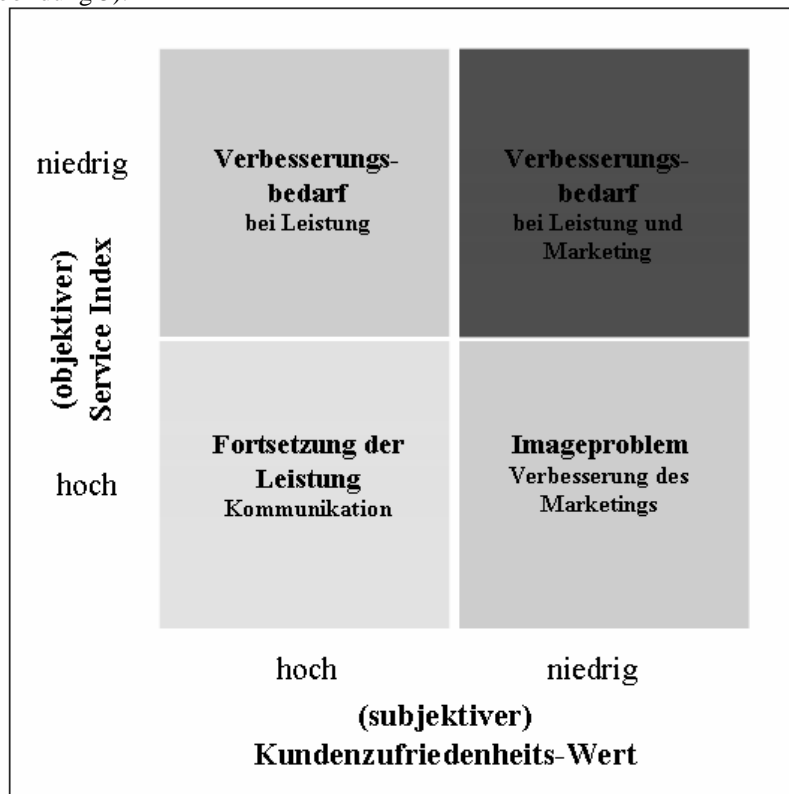


Abbildung 5: Handlungsrelevanzmatrix

Wird ein Konstrukt objektiv sowie subjektiv hoch bewertet, so können die Leistungserstellung fortgesetzt und die entsprechenden Leistungen als USP

kommuniziert werden. Wird ein Konstrukt objektiv hoch, aber subjektiv niedrig bewertet, so besteht ein Imageproblem. Die jeweilige Leistung wird vom Kunden schlechter eingeschätzt als diese in Wirklichkeit ist. Wird hingegen eine Leistung objektiv niedrig bewertet und subjektiv hoch, muss nachgebessert werden, um die gute subjektive Beurteilung der Kunden im Zeitverlauf dauerhaft aufrecht zu erhalten. Höchstes Verbesserungspotential bieten die Leistungsprozesse, welche objektiv und subjektiv niedrig bewertet werden. Werden diese Prozesse optimiert, so kann die Kundenzufriedenheit nicht nur objektiv kurzfristig, sondern auch subjektiv langfristig gesteigert werden (Stücken 2003, S. 30).

5 Beurteilung der Erhebungsqualität

Zur Beurteilung der Qualität des Mystery Shoppings werden die allgemein üblichen Kriterien Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) verwendet (Abbildung 6).

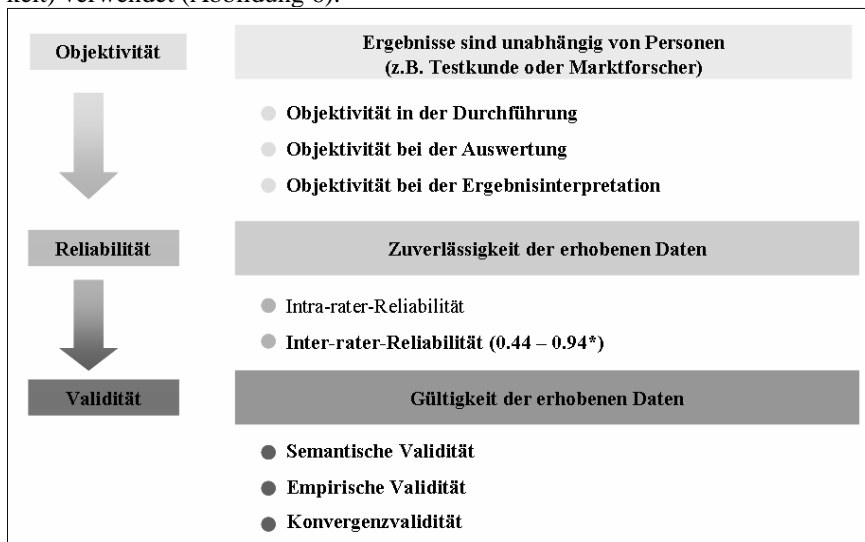


Abbildung 6: Überblick Güte der erhobenen Daten

5.1 Objektivität

Eine Datenerhebung ist objektiv, wenn sie unabhängig vom Beobachter oder Interviewer ist, somit unterschiedliche Personen zum selben Ergebnis kommen. Die Objektivität muss bei der Durchführung der Datenerhebung, bei der Auswertung der Daten und bei der Interpretation der Ergebnisse gegeben sein (Ab-

bildung 6). Abbildung 7 verdeutlicht die Anforderungen, welche zu berücksichtigen sind, um die Objektivität sicher zu stellen.

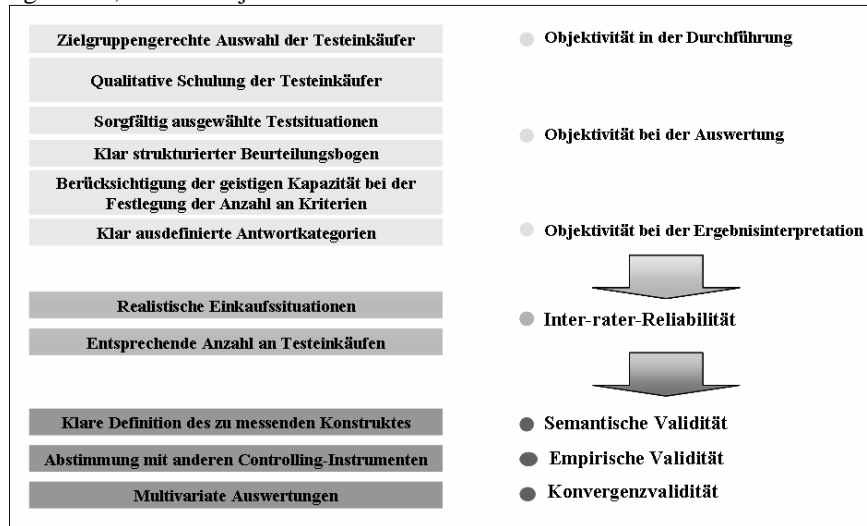


Abbildung 7: Maßnahmen zur Sicherstellung der Erhebungsqualität

Durch eine zielgruppengerechte Auswahl der Testeinkäufer wird der Grundstein für eine geheime und somit objektive Messung der Servicequalität gelegt. Hierdurch wird verhindert, dass ein Verkäufer einen Testkunden als solchen erkennt und sein Verhalten beeinflusst wird (Reaktivität). Durch eine ausführliche Schulung der Testeinkäufer bezüglich der Beurteilungsmaßstäbe werden gleiche Situationen von unterschiedlichen Personen gleich beurteilt. Ebenfalls wird bei der Festlegung der Anzahl der zu bewertenden Kriterien auf die geistige Leistung der Testeinkäufer eingegangen, um diese nicht durch eine zu hohe Anzahl an Kriterien zu überfordern und die Ergebnisse durch ungenaues Wiedergeben zu verfälschen. Die vorab klar definierten Situationen, welche den einzelnen Skalenwerten des Beurteilungsbogens zugeteilt wurden, unterbinden Diskrepanzen bei der Auswertung der erhobenen Daten sowie bei der Interpretation der Ergebnisse.

5.2 Reliabilität

Unter Reliabilität wird die Zuverlässigkeit der Datenerhebung verstanden. Ein ebenso fundiertes wie praktikables Verfahren, um die Zuverlässigkeit einer Marktforschung zu messen ist, zwei identische Untersuchungen durchzuführen und deren Ergebnisse miteinander zu vergleichen. Somit kann Reliabilität auf zweierlei Arten gemessen werden:

1. in wie weit verschiedene Beobachter den gleichen Sachverhalt gleich einordnen (Inter-Rater-Reliabilität) und
2. in wie weit ein Beobachter zu verschiedenen Zeitpunkten den gleichen Sachverhalt konsistent zuordnet (Intra-Rater-Reliabilität) (Grüner 1974, S. 56).

Die Intra-Rater-Reliabilität lässt sich beim Mystery Shopping nur schwer erheben, da ein mehrmaliges Auftreten ein und desselben Testkäufers in derselben Situation sehr auffällig wäre und der Testkauf nach einer Entdeckung des Testkäufers unter veränderten Bedingungen durchgeführt werden müsste (Matzler, Kittinger-Rosanelli 2000, S. 230).

Eine hohe Reliabilität lässt sich bei Mystery Shopping durch einen Beurteilungsbogen erreichen, welcher möglichst objektive Bewertungen sicherstellt. (Grove, Fisk 1992, S. 217- 224). Hierbei wird auf einfach zu beobachtende Kriterien zurückgegriffen, welche in möglichst einfache Skalen (Ja/Nein) einzuordnen sind. Werden detailliertere Skalen benutzt, so müssen die einzelnen zu beobachtenden Konstrukte möglichst klar operationalisiert und eindeutig definiert sein (Matzler, Pechlaner, Kohl 2000, S. 169). Eine Messung der Inter-Rater-Reliabilität bei einer Erhebung der Dienstleistungsqualität von Reisebürounternehmen ergab einen Reliabilitätskoeffizienten zwischen 0,44 und 0,94 (Lingenfelder, Wieseke, Schmidt 2003, S. 301). Bei fachgerechter Erhebung sind somit hinreichend zuverlässige Aussagen erzielbar.

5.3 Validität

Der dritte Qualitätsmaßstab, welcher auf Objektivität und Reliabilität als notwendige Voraussetzungen aufbaut, ist die Validität. Validität beschreibt die Gültigkeit der erhobenen Daten. Sie ist dann gegeben, wenn das Verfahren tatsächlich in der Lage ist, das zu messen, was es messen soll. In der Literatur wird eine Vielzahl verschiedener Arten der Validitätsmessung diskutiert. Für das vorliegende Mystery Shopping ist die Validität gegeben, wenn:

1. die Operationalisierung des zu beurteilenden Konstruktes zufrieden stellend gelingt (semantische Validität) (Kromrey 1998, S. 188),
2. das Messverfahren das zu messende Konstrukt unverfälscht abbildet (empirische Validität) (Kromrey 1998, S. 188) und
3. verschiedene Operationalisierungen desselben Konstruktes zu ähnlichen Ergebnissen führen (Konvergenzvalidität) (Deckers 2003, S. 34f).

Zur Sicherstellung der semantischen Validität empfiehlt sich, die Operationalisierungen der einzelnen Konstrukte mit Experten zu diskutieren sowie Pre-

Tests durchzuführen. Hiermit können Relevanz, Vollständigkeit sowie Eindeutigkeit der operationalisierten Kriterien überprüft werden (Deckers 2003, S. 34f). Die empirische Validität wird bei Mystery Shopping dadurch gewährleistet, dass die Daten in einem natürlichen Umfeld erhoben werden. Weiterhin kann durch eine entsprechende Schulung der Testeinkäufer und die Konstruktion realitätsnaher Einkaufssituationen sichergestellt werden, dass z.B. Interviewereinflüsse auf ein Minimum reduziert werden. So wird etwa durch die verdeckt teilnehmende Beobachtung die Reaktivität, d. h. das Verändern des Verhaltens seitens des Mitarbeiters aufgrund der Prüfungssituation, ausgeschlossen (Matzler, Kittinger-Rosanelli 2000, S. 231). Zur Überprüfung der Konvergenzvalidität kann ein bestimmtes Konstrukt einmal merkmalsorientiert über eine skalierte Globalbewertung und einmal ereignisorientiert über eine Critical-Incident Frage erhoben werden. Um eine Aussage über die Konvergenzvalidität machen zu können, wird eine Korrelation der Globalbewertung mit den positiven bzw. negativen Incidents berechnet. Die Validität der Ergebnisse ist gegeben, wenn mit der Anzahl der berichteten positiven bzw. negativen Incidents die Qualitätseinstufung ebenfalls steigt bzw. fällt (Deckers 2003, S. 34f).

6 Fazit und Ausblick

Am Beispiel des Verkaufsprozesses wurden Ziele, Prozess und Qualität des Mystery Shoppings dargestellt. Zusammenfassend kann Mystery Shopping als geeignetes Instrument zur Kontrolle der Dienstleistungsqualität und der Identifikation von Verbesserungspotenzialen angesehen werden. Auch wenn zu Reliabilität und Validität bisher erst wenige empirische Ergebnisse vorliegen, so deuten diese darauf hin, dass mit einer sorgfältigen Gestaltung des Mystery Shopping Prozesses die notwendigen Voraussetzungen für hinreichend zuverlässige und gültige Ergebnisse vorliegen. Hierzu gehört neben der Situationsanalyse, der zielgruppengerechten Auswahl der Testpersonen und der Planung der Testsituationen vor allem die fundierte Operationalisierung der zu messenden Dienstleistungen und ihrer Teilelemente. Werden diese Faktoren berücksichtigt, kann aus den Ergebnissen des Mystery Shoppings ein zuverlässiges Kennzahlensystem für die Dienstleistungsqualität vom einzelnen Mitarbeiter bis zum Gesamtunternehmen abgeleitet werden.

Das Instrument des Mystery Shoppings stammt aus den USA und ist dort seit ca. 25 Jahren als Controlling-Tool im Einsatz. Seit einiger Zeit lassen auch Unternehmen in Deutschland ihre Dienstleistungsqualität mittels Mystery Shopping überprüfen. Die Zuverlässigkeit des Instruments, das in diesem Beitrag nicht näher analysierte Kosten-Nutzen Verhältnis sowie die Diskussion in der Fachpresse lassen vermuten, dass Mystery Shopping auch in Deutschland weitere Verbreitung finden wird.

7 Literaturverzeichnis

- Bauer, F., Urbahn, J., Markart, V.: Die zehn häufigsten Missverständnisse zum Thema Mystery Analysen, in: Planung & Analyse, Nr. 05, 22.10.2003, S. 16
- Berry, L.J., Zeithaml, V.A., Parasuraman, A.: Five Imperatives for Improving Service Quality, in: Sloan Management Review, Vol. 29, Summer, 1990, P. 29-38
- Bruhn, M.: Qualitätsmanagement für Dienstleistungen: Grundlagen, Konzepte, Methoden, Berlin, 1996
- Bruhn, M.: Qualitätsmanagement für Dienstleistungen: Grundlagen, Konzepte, Methoden, 2. Auflage, Berlin et al., 1997
- Dawson, J., Hillier, J.: Competitor Mystery Shopping: Methodological Considerations and Implications for the MRS Code of Conduct, in: Journal of the Market Research Society, Vol. 37, No. 4, 1995, P. 417-422
- Deckers, R.: Taking the Mystery out of Mystery Shopping - Qualitätskriterien für das Testkauf-Verfahren, in: Planung & Analyse, Nr. 05, 22.10.2003, S. 34
- Ecken, Ch.: Mystery Shopping für eine neue Servicequalität: Marktforschungsinstitut Shop'n Check bringt Methode zur Qualitätskontrolle nach Deutschland, in: Horizont, Nr. 47, 19.11.1998, S. 34
- Grove, S.J., Fisk, R.P.: Observational Data Collection Methods for Services Marketing: An Overview, in: Journal of the Academy of Marketing Science, Vol. 20, 1992, P. 217-224
- Grüner, K.-W.: Beobachtung, Stuttgart, 1974
- Kromrey, H.: Empirische Sozialforschung, 8. Auflage, Opladen, 1998
- Krüger, S., Stücken, M.: Testberatung: Wo sind die Grenzen, in: Bank und Markt, Nr. 09, 01.09.2003, S. 32
- Lammers, M., Schubert, A.: Am Puls der Dienstleistungsqualität mit Mystery Shopping- Methode der Wahl auch bei komplexen Dienstleistungen, in: Planung & Analyse, Nr. 05, 22.10.2003, S. 56
- Lingenfelder, M., Wieseke, J., Schmidt, K.: Dienstleistungsqualität von Reisebüro-Unternehmen - Modellierung und Messung mittels Mystery Shopping - Am Fallbeispiel einer im Franchisesystem organisierten Reisebüroketten, in: Tourismus Journal, Nr. 03, 03/2003, S. 283-306
- Matzler, K., Kittinger-Rosanelli, C.: Mystery Shopping als Instrument zur Messung der wahrgenommenen Dienstleistungsqualität von Banken, in: Jahrbuch der Absatz- und Verbrauchswirtschaft, 03/2000, S. 220-241
- Matzler, K., Pechlaner, H., Kohl, M.: Formulierung von Servicestandards für touristische Dienstleistungen und Überprüfung durch den Einsatz von „Mystery Guests“, in: Tourismus Journal 4. Jg (2000), Nr. 02, S. 157-176
- Miles, L.: Rise of the Mystery Shopper: The Voice of the Customer can no longer be ignored and many Organisations are reacting accordingly, in: Marketing, 29.07.1993, P. 19-20

Müller, F., Neuhaus, M., Löhnert, R.: Prüfung nach den Vorschriften des Wertpapierhandelsgesetz, Mystery Shopping- ein neuer Weg in der Revisionsarbeit, in: Betriebswirtschaftliche Blätter, Nr. 01, 01/2006, S. 24

Schlich, A. W.: Ein Qualitätsinstrument kommt in Mode, Mystery Shopping bewertet Servicequalität von Banken, in: Betriebswirtschaftliche Blätter, Nr. 05, 05/2006, S. 274

Schreier, G.: Verdeckte Ermittler, in: Werben & Verkaufen, Nr. 15, 08.04.2004, S. 30

Siefke, A.: Zufriedenheit mit Dienstleistungen: Ein phasenorientierter Ansatz zur Operationalisierung und Erklärung der Kundenzufriedenheit im Verkehrsbereich auf empirischer Basis, Frankfurt a. M. et al., 1998

Stauss, B., Henschel, B.: Attribute-Based versus Incident-Based Measurement of Service Quality: Results of an Empirical Study in the German Car Service Industry, in: Kunst, P., Lemmink, J. (Eds.): Quality Management in Services, Assen, 1992, P. 59-78

Stücken, M.: Mystery Research oder Kundenzufriedenheitsbefragung- Eine Analyse der Methode, in: Planung & Analyse, Nr. 05, 22.10.2003, S. 45

Wieseke, J., Schmidt, K., Lingenfelder, M.: Was leistet Mystery Shopping, in: Absatzwirtschaft, 06/2006, S. 42-44

Wilson, A.M., Gutmann, J.: Public Transport: The Role of Mystery Shopping in Invest Decisions, in: Journal of the Market Research Society, Vol. 40, No. 4, 1998, P. 5-17
www.esomar.de

Zeithaml, V., Berry, L., Parasuraman, A.: Kommunikations- und Kontrollprozesse bei der Erstellung von Dienstleistungsqualität, in: Bruhn, M, Stauss, B. (Hrsg): Dienstleistungsqualität, Wiesbaden 1991, S. 107-136.

Autoren:

Thomas Dobbelstein, Prof. Dr.
Wissenschaftlicher Leiter Customer Research 42 e. K. und
Stv. Direktor der Berufsakademie Ravensburg
Kontakt: dobbelstein@cr42.de

Daniel Windbacher, Dipl. Betriebsw. (BA)
Wissenschaftlicher Mitarbeiter Customer Research 42 e. K.
Kontakt: Windbacher@cr42.de