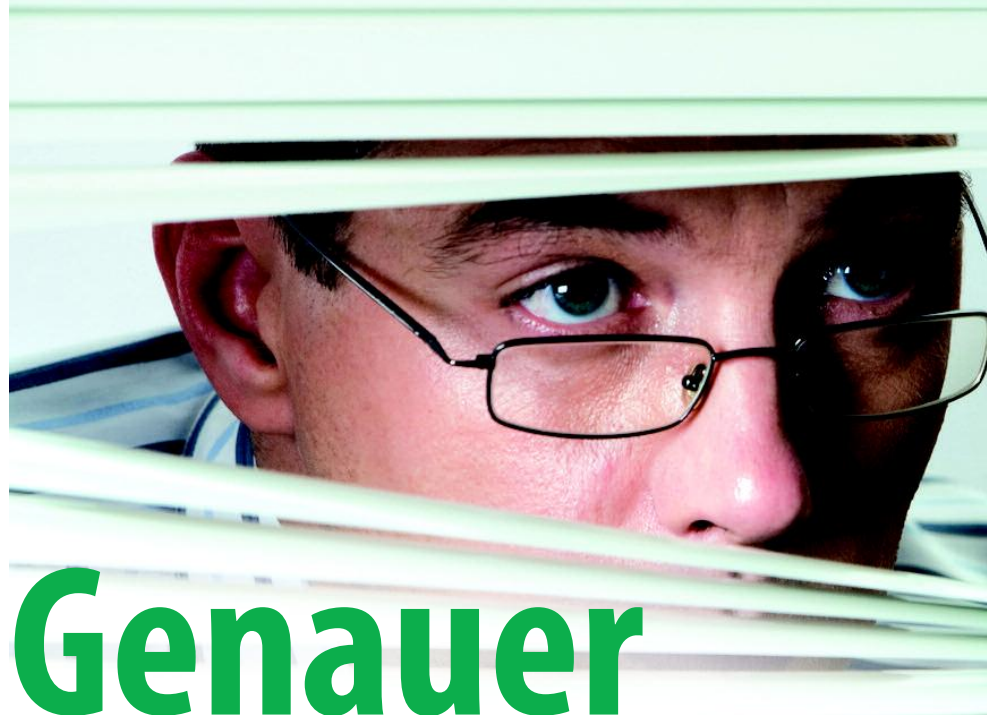


Häufig werden bei Mystery Shopping Testkunden eingesetzt, die über das Internet rekrutiert wurden und über die nur begrenzte Informationen vorliegen. **Gunnar Grieger** und **Ralf Deckers** zeigen anhand einer Studie aus dem Fast Food-Kontext, ob bei Testkunden aus verschiedenen soziodemografischen Gruppen Unterschiede in der Ergebnisqualität bestehen.

Mystery Shopping erfreut sich als Verfahren zur Erhebung der Dienstleistungsqualität in der Praxis zunehmender Beliebtheit. Es wird branchenübergreifend überall dort eingesetzt, wo Unternehmen ihren Kunden begegnen (B2C), wie in Einzelhandelsgeschäften, Gaststätten, Hotels, Call-Centern, Krankenhäusern oder Werkstätten. Betrachtet man die Anzahl der Mitglieder, die beim Berufsverband Deutscher Markt- und Sozialforscher als Anbieter von Mystery Shopping-Dienstleistungen registriert sind, so ist zwischen 1997 und 2008 ein Anstieg von drei auf 112 Unternehmen zu verzeichnen. Das deutsche Marktvolumen wird für das Jahr 2010 auf 75 Millionen Euro prognostiziert.

Trotz der Präsenz in der Praxis wurde das Thema Mystery Shopping in der wissenschaftlichen Literatur bisher nur wenig untersucht. Vergangene Forschungsbemühungen widmen sich eingehend dem Nachweis der Gütekriterien Reliabilität und Validität des Instruments. Eine Betrachtung der als Myste-



Genauer hingeschaut

Studie zur Ergebnisqualität von Mystery Shoppern

ry Shopper eingesetzten Personen blieb bisher aus.

Unternehmen, die Mystery Shopping einsetzen, bedienen sich häufig Agenturen, die ihr Testkundenfeld über das Internet rekrutieren. In der Regel sind diesen Agenturen lediglich soziodemografische Merkmale der registrierten Testkunden bekannt, und die Qualität der Arbeit sowie die Güte der Ergebnisse können erst im Nachhinein beurteilt werden. Basierend auf diesen Überlegungen stellt sich daher die Frage, wie sich die Mystery Shopper aus unterschiedlichen soziodemografischen Gruppen in der Qualität der abgelieferten Ergebnisse unterscheiden.

Mehrkomponenten-Modell zur Ermittlung der Ergebnisqualität

Zur Beantwortung dieser Fragestellung wurde ein Mehrkomponenten-Modell zur Ermittlung der Ergebnisqualität von Mystery Shoppern konzipiert. Dieses Modell basiert auf unterschiedlichen Ansätzen der Organisations- und Sozialpsychologie und wurde um pragmatische Überlegungen ergänzt (Abb. 1). Anhand der einzelnen Komponenten lässt sich die Er-

gebnisqualität von Mystery Shoppern abbilden und vergleichen.

Mit der Mittelwertstendenz lassen sich Urteilsgewohnheiten bezüglich milder oder strenger Bewertung spezifizieren. Die Streuungstendenz gibt Aufschluss über die Einheitlichkeit der Bewertung. Mit der Korrelationstendenz kann der Einfluss einzelner Dienstleistungsepisoden auf das Globalurteil gezeigt werden. Während diese als Urteilstendenzen bezeichneten Komponenten Informationen über das Beurteilerverhalten liefern, lässt sich mit der Komponente Beurteilerzuverlässigkeit ein konkreter Wert bezüglich der Übereinstimmung von zwei Mystery Shoppern, die die gleiche Dienstleistungssituation beobachtet haben, ermitteln. Dafür wird die Interrater-Reliabilität als Messgröße herangezogen. Die verbleibenden, aus pragmatischen Überlegungen resultierenden Komponenten geben Hinweise auf die Vollständigkeit des ausgefüllten Beobachtungsbogens und die Gewissenhaftigkeit (Übereinstimmung identischer Items an unterschiedlichen Stellen im Beobachtungsbogen) der Mystery Shopper.

Untersuchung im Fast Food-Kontext

Die für die Auswertung erforderlichen Daten wurden in einer bundesweiten Mystery Shopping-Studie für eine Fast Food-Kette erhoben. Um die Interrater-Reliabilität ermitteln zu können, wurden die Testkunden paarweise so zu-





sammengesetzt, dass jeweils spezifische soziodemografische Merkmale identisch waren (wie etwa gleiches Geschlecht, gleiche Altersklasse, gleiche Affinität zum Testobjekt, gleicher Mystery Shopping-Erfahrungsstand).

Die Analyse der gewonnenen Daten ergab, dass sich für den gewählten Untersuchungskontext alle Merk-

malsträger in ihrer Ergebnisqualität unterscheiden. Während mit den Urteilstendenzen, die ein Indikator für erforderliche Schulungsmaßnahmen sind, die Urteilsgeohnheiten der Merkmalsträger beschrieben werden, lassen die Komponenten Beurteilerzuverlässigkeit und Sorgfaltstendenz Aussagen über die Arbeitsqualität der eingesetzten

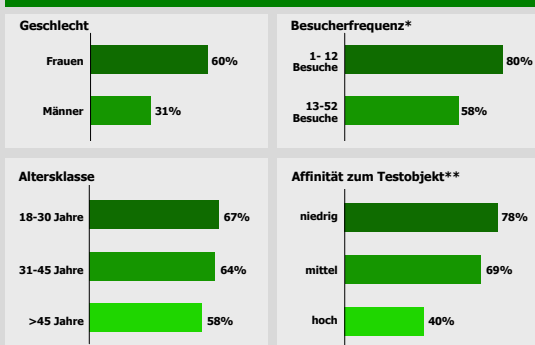
Mystery Shopper zu. Herauszustellen sind ausgewählte Ergebnisse der Interrater-Reliabilität der unterschiedlichen Merkmalsträger (Abb. 2).

Demnach beurteilen Männer, die als Testpaar die gleiche Situation beobachten, nur in rund einem Drittel der Fälle gleichartig. Frauen sind mit 60

Prozent gleichartig beurteilter Items zwar deutlich zuverlässiger, gleichwohl ist das Ergebnis nur wenig befriedigend. Jüngere Mystery Shopper wären älteren vorzuziehen. Testkunden, die häufig mit einem Testobjekt (in diesem Fall Fast Food-Restaurant) in Kontakt stehen und eine hohe Affinität zum spezifischen Testobjekt aufweisen, sind weniger geeignet als solche, die über niedrige Merkmalsausprägungen verfügen.

Die Erkenntnisse hinsichtlich der Affinität erscheinen schlüssig, da angenommen werden kann, dass mit zunehmender Involvierung eine Trübung der Objektivität einhergeht. Bezüglich der Altersklasse müsste eine rückläufige Objektivität bei steigendem Alter unterstellt werden. Bezogen auf das Geschlecht möchten die Autoren den Lesern die Interpretation überlassen. Allerdings soll angemerkt werden, dass Frauen sich im Rahmen dieser Untersuchung als erheblich zuverlässiger erwiesen haben. Abschließend sei betont, dass die Ergebnisse sich auf den Untersuchungskontext in nur einer Kette aus der Fast Food-Branche beziehen und daher nur mit gebotener Skepsis zu verallgemeinern sind. ■

Abb. 2 Anteil zuverlässig beurteilter Items



Quelle: MYSTERYPANEL und ABH Marketingservice
* Besuchshäufigkeit in irgendeinem Fast Food-Restaurant binnen 12 Monaten, ** Affinität zur getesteten Kette



Dr. Gunnar Grieger ist Inhaber der Testkundenagentur MYSTERYPANEL und schrieb seine Dissertation 2008 nebenberuflich zum Thema Mystery Shopping. www.mysterypanel.de



Dr. Ralf Deckers ist Leiter Market Intelligence bei der ABH Marketingservice GmbH in Köln und schrieb seine Dissertation 1998 über methodische Aspekte des Testkundenverfahrens. www.abh.de