

Repräsentative Umfragen: Auswahlverfahren und Ausschöpfung



Dr. Georges Ulrich
Geschäftsführer
gfs-befragungsdienst AG,
Zürich

Im folgenden Artikel wird der Begriff Repräsentativität bei quantitativen Umfragen reflektiert, oft gehörte Argumente zur Erreichung von Repräsentativität werden thematisiert und die Komplexität des Themas wird aufgezeigt. Es besteht nicht der Anspruch, «die Lösung» zu präsentieren, doch soll der Artikel zum Darüber-Nachdenken anregen.

Nach Tiede/Voß (2000: S. 84) haftet einem Ergebnis, das nicht repräsentativ ist, ein Makel an. «Repräsentativität ist als Gütesiegel für ein Forschungsergebnis begehrt, weil das repräsentative Ergebnis den Eindruck der Verlässlichkeit, Glaubwürdigkeit und Verallgemeinerungsfähigkeit erweckt.» Dabei muss man sich allerdings zuerst fragen, ob die Auseinandersetzung mit der für das Erkenntnisinteresse relevanten Fragestellung überhaupt Repräsentativität verlangt. Repräsentativität steht zum Beispiel bei vielen qualitativen Forschungsdesigns gar nicht im Zentrum des Interesses. Erhebungsmethoden wie Gruppendiskussionen, Tiefeninterviews, Expertenbefragungen usw. sind per se nicht repräsentativ, weil der Fokus der Forscher und Forscherinnen nicht darin besteht, die Erkenntnisse daraus auf eine ganze Grundgesamtheit zu übertragen, sondern Informationen zu Motivationen und Prozessen zu erhalten, die bei einer ganz spezifischen Fragestellung von Bedeutung sind. Bei Bevölkerungsbefragungen jedoch impliziert schon die definierte Zielgruppe, dass eine zufällig gezogene Stichprobe Grundbedingung ist, um repräsentative und reliable Aussagen machen zu können.

Aussage 1: Repräsentativität impliziert ein Abbild der Grundgesamtheit

Einigkeit besteht bei den Praktikern lediglich darin, dass Repräsentativität auf ein möglichst exaktes Abbild der Grundgesamtheit abzielt. Worin sich die Geister scheiden ist, wie dies erreicht werden soll. Bei einer bevölkerungsrepräsentati-

ven Stichprobe entscheiden auf theoretischer Ebene nebst der Stichprobengrösse (Lippe/Kladroba 2002, S. 227–238) zusätzliche Faktoren über die Güte der später gewonnenen Aussagen. Während die Qualität der formulierten Aussagen auf der Angemessenheit von Forschungs- und Methodendesign beruht, auf guten Fragestellungen und vor allem auf den Kompetenzen analyseerfahrener Forscher/-innen, definiert sich die Datenqualität über eine qualitativ hochwertige Gewährleistung von Repräsentativität. Repräsentativität bedeutet nichts anderes, als dass jede Person aus der definierten und bekannten Grundgesamtheit die gleiche Chance haben muss, an der Befragung teilnehmen zu können (Berekoven et al. 1999; Diekmann 2004; Neumann/Kreuger 2003; Malhorta 2000 usw.). Die Annahme beruht auf der Gauß'schen Normalverteilung bzw. auf der Logik, dass es sich anteilmässig in die Befragung überträgt, wenn jedes Individuum die gleiche Teilnahmechance hat, wodurch ein Abbild der Grundgesamtheit quasi systemisch produziert wird. Werden bei der Stichprobenziehung und dem darauf folgenden Auswahlverfahren systematisch Gruppen der Grundgesamtheit ausgeschlossen, ist eine Befragung nicht repräsentativ (Zentes 1996, S. 383).

In diesem Artikel soll auf zwei Elemente der Repräsentativität speziell eingegangen werden, weil bei ihrer Handhabung in der Praxis sehr grosse Unterschiede bestehen (Beeler 2007): Einerseits betrifft dies das Auswahlverfahren und andererseits die Ausschöpfung. Vereinfacht

gesagt beschreibt das Auswahlverfahren die Vorgehensweise, wie man versucht, systematisch die richtige Person aus einer Grundgesamtheit bzw. einer Stichprobe auszuwählen. Die Ausschöpfung beschreibt, wie gut man es geschafft hat, mit der gewählten Methode die Zielperson zur Teilnahme zu mobilisieren. In der Praxis wird unter Ausschöpfung der Anteil (Prozentsatz) der gezogenen Stichprobe verstanden, der an der Befragung tatsächlich teilgenommen hat.

Aussage 2: Das Auswahlverfahren ist entscheidend für die Repräsentativität der Resultate

Man unterscheidet grob gesagt zwei Arten von Auswahlverfahren: Bevölkerungsrepräsentativität durch Zufallsziehung auf der einen Seite und auf der anderen Seite die sogenannte Quotenrepräsentativität, bei der die Eckwerte auf der Basis von bekannten soziodemografischen Verteilungen zu sichern versucht werden (Bausch 1995). Bei der Quotenrepräsentativität ist die meist getroffene darunterliegende Annahme, dass die Grundgesamtheit durch möglichst genaue Wiedergabe der in der Grundgesamtheit vorliegenden Soziodemografie – also Verteilung von Elementen wie Geschlecht, Wohnort, Alter, Einkommen usw. – abgebildet werden kann. Psychografische Typologien wie Lifestyle, Einstellungen und spezifische Verhaltensweisen müssen vorgängig bekannt sein, wenn man danach suchen will, was der Natur der Sache nach meist nicht möglich ist – oft sind sie ja gerade Gegenstand der Untersuchung. Da die

Die vier Stufen von Repräsentativität ohne Berücksichtigung der Stichprobengrösse

	1. Zielgruppe identifizieren	2. Auswahlverfahren/ Suchstrategien	3. Erreichbarkeit	4. Mobilisierung
Ideale Vorstellung	Definition Grundgesamtheit. Es besteht eine vollständige Datenbank (jede Person ist 1-mal darin vertreten)	Jede Person (\neq Haushalt und \neq Telefonnummer) muss dieselbe Chance haben, befragt zu werden	Die Person muss erreichbar sein	Die Person muss am Interview teilnehmen
Herausforderung	Eindeutigkeit ist oft schwierig	Chancengleichheit ist in der Realität schwierig, da eine vollständige und aktuelle Datenbank oft nicht vorhanden ist	Ungleiches Verhalten der Zielgruppen	Einstellung gegenüber Umfragen
Massnahmen	Klare Abgrenzung	Multiple Suchstrategien und Methoden	Multiple Kontaktierungsstrategien	Schulung Interviewer, Anreize (Incentives, Wettbewerb, Zustellen Auswertungsergebnisse usw.)
Weitere Probleme	Grundgesamtheiten sind oft gar nicht bekannt (Online-User, Handynutzer usw.), Soziodemografie als Auswahlkriterium ist oft ungenügend	Intervenierende Variablen wegen Redundanzen (z. B. auf unterschiedlichen Datenbanken) und Methodeneffekten		Interviewereffekte, intervenierende Variablen wegen unterschiedliche Reaktion auf Incentive

Abbildung 1

Quelle: G. Ulrich, gfs-befragungsdienst AG

retrospektive Beweisführung als wissenschaftlich unhaltbar gilt, fällt – da in der Regel unbekannt – die Vorgehensweise der psychografischen Quotenvorgabe von vornherein meistens weg (Berekoven et al. 1999, S. 50).

In der Praxis wird die Selektion der Befragungssubjekte oft durch die Interviewer und Interviewerinnen vorgenommen, in der Regel auf der Basis von soziodemografischen Merkmalen wie Region, Alter und Geschlecht. Diese Vorgehensweise gilt jedoch – ableitend aus bereits geführter Argumentation – für Studien mit sehr hohem Anspruch an Repräsentativität als ungenügend (Höpflinger 2007), weil die Auswahl der Zielperson nicht einer mathematisch-systematischen Zufälligkeit entspricht, sondern einem beliebigen Zufall. So suchen sich Interviewer und Interviewerinnen beispielsweise tendenziell Personen, die einfach erreichbar sind (das heisst bei Face-to-Face-Befragungen z. B. vor Einkaufszentren, an Bahnhöfen usw.) und verzerren dadurch die Stichprobe dahingehend, dass Personen, die sich an solchen Plätzen aufhalten, unabhängig von ihrer soziodemografischen Verteilung überrepräsentiert werden. Dies heisst nicht per se, dass die so erlangten Interviews eine reduzierte

Repräsentativität aufweisen. Es bedeutet aber, dass das Verfahren so ausgestaltet sein muss, dass die Selektion der Zielpersonen den oben definierten Anforderungen der Zufallsauswahl entspricht und dass es Gründe der Praktikabilität und Kosten gibt, warum es einen Kompromiss zwischen dem Anspruch an hohe Repräsentativität und Umsetzbarkeit abzuwägen gilt (Schnell et al. 1999; Diekmann 2004; Werani 2000; Steffens-Duch 2000; Keller et al. 2000 usw.).

Ausgangslage für die Zufallsziehung bei bevölkerungsrepräsentativen Telefonumfragen bildet in der Regel das öffentlich zugängliche elektronische Telefonbuch, das mit zusätzlichen Adressen angereichert wird, um einen möglichst hohen Anteil der Haushaltsanschlüsse zu erreichen. Auch wenn oft das Gegenteil behauptet wird: In Ermangelung einer aktuellen und zugänglichen Datenbank, die die Bevölkerung eindeutig abbildet, wird mit dem elektronischen Telefonverzeichnis lediglich ein Annäherungswert an Repräsentativität erreicht. Jedoch zeigen zum Beispiel Wahlumfragen trotz der theoretisch reduzierten Repräsentativität sehr genaue Ergebnisse. Dies wird jedoch in der Regel nur von sehr erfahrenen Forschern erreicht.

Aus dem elektronischen Telefonverzeichnis wird nach einem Zufallsprinzip eine Ausgangsstichprobe gezogen. In der Praxis gibt es unterschiedliche Vorgehensweisen, es empfiehlt sich aber, die ausgewählten Haushalte wegen der Erreichbarkeit zu verschiedenen Tageszeiten maximal 10-mal anzuwählen; mehr als 10 Anrufe ergeben kaum ein besseres Ergebnis (Beeler 2007). Erfüllt eine Person im Haushalt die Voraussetzung, zur Grundgesamtheit zu gehören, wird die Adresse verwendet. Erfüllen mehrere Personen im Haushalt die Voraussetzung, wird zum Beispiel jene Person befragt, die als letzte im Kalenderjahr Geburtstag hat. Durch diese zweite Stufe der Zufallsauswahl wird verhindert, dass systematische Gewohnheitseffekte im Telefonverhalten zu einer Stichprobenverzerrung führen. Um Verzerrungen wegen der unterschiedlichen Erreichbarkeit verschiedener Gruppen vermeiden zu können, werden in einer dritten Stufe Maximalquoten für Geschlecht und Alter vorgegeben, die nicht überschritten werden können. Letzteres wird in der Praxis nicht zuletzt aus Zeit- und Kostengründen gemacht, da man sonst rund 80% des Aufwandes für die letzten 20% der Interviews hätte. Geht man also von vorgängigen Argumenten aus, kommt man zum Schluss, dass ins-

Kreislauf der Verallgemeinerung bei repräsentativen Umfragen

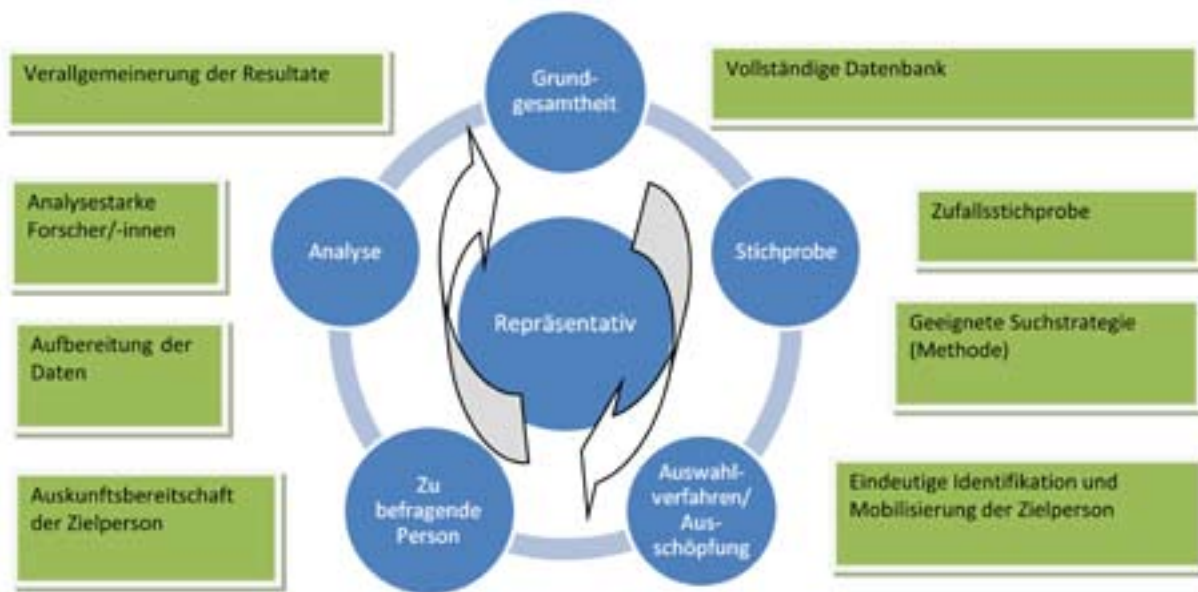


Abbildung 2

Quelle: G. Ulrich, gfs-befragungsdienst AG

besondere die Zufallsauswahl ideal für eine hohe Repräsentativität ist. Wie die Abbildung 1 zeigt, sind dabei vier Stufen mit jeweils eigenen Herausforderungen zu unterscheiden.

Grundvoraussetzung Nummer eins ist eine vollständige und aktuelle Adressdatenbank der Grundgesamtheit. Die zweite Grundvoraussetzung ist, dass genau die per Zufall ausgewählte Person erreichbar ist und beim Interview auch tatsächlich mitmacht – hier ist die Rolle der Interviewer entscheidend. Drittens muss der Stichprobenumfang genügend gross sein, um mit einem möglichst geringen Stichprobenfehler die der Grundgesamtheit unterliegende Struktur wiederzugeben (Höpflinger 2007). Letzteres ist vor allem dann ein Kriterium, wenn man zum Ziel hat, eine möglichst hohe strukturelle Gleichheit zwischen Grundgesamtheit und Stichprobe zu erreichen.

Allerdings sind die in Abbildung 2 aufgezeigten Grundvoraussetzungen in der Praxis nie wirklich gegeben (Lippe/Kladroba 2002) und es werden auch vermehrt Defizite in der Datenbasis der Grundgesamtheit zum Thema. Zum Beispiel weiss man relativ wenig über die Struktur der Gruppe sogenannter Handy-only-User,

also Personen, die nicht über einen Festanschluss erreichbar sind.

Was Onlineumfragen betrifft, so ist – wenn die User nicht vorgängig in einer repräsentativen Telefonbefragung rekrutiert worden sind – die Grundgesamtheit der Onlineuser weder bekannt noch hat man eine umfassende Datenbank. Oft wird nicht einmal eine Stichprobe gezogen und man weiss kaum etwas über diejenigen Personen, die nicht an der Umfrage teilnehmen. So reduziert sich auch die Aussagekraft von Online-Analyse-Tools, wenn es um die Frage von Repräsentativität geht. Denn es handelt sich in erster Linie um ein retrospektives Abbild von Internetusern, die das Internet nach einem Suchbegriff absuchen. Die Liste liess sich weiterführen, doch das sind, zusammen mit der in Studentenumfragen oft beobachteten Nachgewichtung von Quoten, diejenigen Argumente, die eine bevölkerungsrepräsentative Umfrage via Internet schon im Hinblick auf die ersten beiden Grundvoraussetzungen erschweren. Das bedeutet nicht, dass Onlineumfragen per se schlecht sind, doch die Fallstricke sind in punkto Repräsentanz doch erheblich gravierender als bei CATI-Umfragen, die eine bessere Kontrolle über das Auswahlverfahren ermöglichen

(Ulrich 2009). In der Praxis wird bei repräsentativen Umfragen der Fokus in der Regel auf die Defizite der Erreichbarkeit der Zielpersonen gelegt und darauf, wie man Probanden zur Teilnahme an Interviews mobilisieren kann (Höpflinger 2007; Koch 1991; Ulrich 2008).

Nachdem das Auswahlverfahren am Beispiel von Quoten und Zufallsziehung ausgeführt wurde, soll in der Folge das Thema Ausschöpfung behandelt werden.

Aussage 3: Eine hohe Ausschöpfung ist kein Garant für hohe Repräsentanz

Die Ausschöpfung, also das Verhältnis zwischen kontaktierten (\neq erreichte!) Zielpersonen und tatsächlich durchgeführten Interviews, wird in Prozenten über alle soziodemografischen Quoten gemessen (Schenk 1992, S. 84).

Den obigen Ausführungen folgend, müsste idealerweise die Ausschöpfung immer 100% sein, denn jede per Zufall ausgewählte Zielperson sollte, um den Anspruch der Zufälligkeit nach Gauß'scher Normalverteilung nicht zu verzerren, auch tatsächlich teilnehmen. Dies ist jedoch aus folgenden, nicht abschliessenden Gründen eigentlich nie der Fall:

1. Die Zielperson ist nicht erreichbar. Vor allem junge, gut ausgebildete Männer sind in der Regel schwieriger erreichbar, da sie einerseits mehrheitlich berufstätig sind und andererseits durch eine aktivere Freizeitgestaltung auch am Abend oft nicht zu Hause sind.
2. Die Zielperson verweigert die Teilnahme. Gründe dafür sind beispielsweise fehlendes Interesse, keine Zeit, grundsätzliche Abneigung gegen Umfragen und vieles mehr.
3. Überrepräsentierte Singlehaushalte verzerren die Stichprobe. Einerseits ist die Wahrscheinlichkeit, dass eine Person in einem Mehrpersonenhaushalt den Telefonhörer abnimmt, grösser, andererseits sind jedoch Singlehaushalte vor allem in Stadtgebieten überrepräsentiert. Kommt hinzu, dass in Mehrpersonenhaushalten tendenziell eher eine Frau den Hörer abnimmt.
4. Interviewereffekte wirken unterschiedlich, motivierend oder demotivierend. Die Interviewer/-innen haben einen wesentlichen Einfluss darauf, ob jemand teilnimmt oder nicht. Abgesehen davon zieht jeder Interviewer bzw. jede Interviewerin durch ihre ganz spezifische Art gewisse Leute an und andere nicht. Deshalb sollten einzelne Personen nicht zu viele Interviews durchführen.
5. Der allgemein anerkannten Definition von Ausschöpfung und dem Umgang mit diesem Thema wird ungenügend Rechnung getragen. Zum Beispiel braucht es für die Gruppe der jungen Männer mit guter Ausbildung ein X-Faches an Kontaktversuchen im Vergleich zu älteren Frauen, um eine entsprechende Anzahl an Interviews zu erreichen (Diekmann 2004; Schnell et al. 1999; Koschnik 1998 usw.).

Eine Untersuchung an der Hochschule für Wirtschaft Zürich HWZ (Beeler 2007) zeigt auf der Basis von 5 repräsentativen Befragungswellen (Zufallsziehung bis zu dem Zeitpunkt, wenn die erste Quote geschlossen wird), dass Frauen rund 2- bis 3-mal häufiger bereit sind, Auskunft zu geben, und somit bei diesen Befragungen mit einer reinen Zufallsauswahl kein Abbild der Bevölkerung zu erreichen war.

Die Studie untersucht in extenso das Verhältnis von Erreichbarkeit und Ausschöpfung und kommt zu sehr interessanten Ergebnissen:

- So bestätigt die Studie ganz klar, dass verschiedene Personengruppen zu verschiedenen Tageszeiten sehr unterschiedlich gut erreichbar sind und widerlegt die oft gehörte These, dass der Abend die beste Befragungszeit sei.
- Der Aufwand, um Männer unter 35 Jahren zu erreichen, ist mindestens 3-mal höher als bei über 35-Jährigen.
- Interessant ist auch die Erkenntnis, dass zwar junge Männer am Abend oder am Samstag eher erreichbar sind, dann aber bei Befragungen eher nicht mitmachen. Männer dieser Zielgruppe sind am Morgen zwar schwieriger erreichbar, aber sie nehmen dann eher am Interview teil. Dies gilt in verstärktem Masse für Männer mit höherer Bildung. Interessanterweise ist die Ausschöpfungsquote morgens zwischen 8 und 10 Uhr am höchsten.
- Besonders spannend ist auch, dass ein deutlich verbessertes Abbild der Bevölkerung gemäss soziodemografischen Merkmalen erreicht werden konnte, wenn der Ausschöpfung weniger Beachtung geschenkt wurde und, überspitzt formuliert, wenn verstärkt Personen befragt wurden, die einfach das Kriterium einer erhöhten Affinität zum Thema erfüllten und die Ausschöpfung somit sogar unter 30% fiel. Somit ist die Legitimation der Forderung nach Ausschöpfungsmaximierung aufgrund einer breiten Datenbasis erschwert.

Beeler schliesst seine Arbeit zur Ausschöpfung somit mit einem Zitat Levenssteins: «Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital.» Es stellt sich tatsächlich die Frage, ob die bisherige Sichtweise, einfach die Ausschöpfung maximieren zu müssen, vielleicht zu wenig weit greift.

Fazit und Ausblick

Allzu oft wird mit dem Begriff Repräsentativität salopp umgegangen. Repräsentativität per Definition ist und bleibt aber

eine ideale Vorstellung. Es gibt nicht eine einzige Lösung für das Vorgehen in der Praxis, wie Repräsentativität annäherungsweise erreicht werden kann, auch wenn dies manchmal behauptet wird. Erfahrene Forscher und Forscherinnen tragen wesentlich dazu bei, dass die Ergebnisse der Markt- und Meinungsforschung dennoch sehr präzise und von hoher Relevanz sind. Sollte dieser Artikel dazu angeregt haben, bestehende Wahrheiten zu reflektieren und sich in einem Interaktionsraum wie dem Berufsverband vsms Verband Schweizer Markt- und Sozialforscher auszutauschen, ist das Ziel erreicht.

Referenzliteratur

- Bausch, T.: Auswahlverfahren in der Marktforschung, in: Tietz, B./Köhler R./Zentes J. (Hrsg.): Handwörterbuch des Marketing, Stuttgart 1995, Sp. 155-166.
- Beeler, M.: Was heisst schon gute Ausschöpfung, HWZ-Hochschule für Wirtschaft, Diplomarbeit, Zürich 2007.
- Berekoven, L./Eckert, W./Ellenrieder, P.: Marktforschung: Methodische Grundlagen und praktische Anwendung, 8. Aufl., Wiesbaden 1999.
- Diekmann, A.: Empirische Sozialforschung (clone), Rowohlt-Taschenbuch-Verlag, Reinbek bei Hamburg 2004.
- Höpflinger, F.: Ausfälle und Verweigerungen bei Befragungen, www.mypage.bluewin.ch/hoepf/fhmethod1C.html, 2007.
- Keller, B./Lerch S./Matzke, S.: Umfrage, Kundenbindung und Wechselbereitschaft, in: Die Bank 6/2000, S. 376-381.
- Koschnik, W.: Standard-Lexikon für Mediaplanung und Mediaforschung, München u.a. 1998.
- Koch, A.: Zum Zusammenhang von Interviewermerkmalen und Ausschöpfungsquoten, in: ZUMA Nachrichten 28, Mannheim 1991.
- Lippe, P./Kladroba, A.: Repräsentativität von Stichproben, in: Marketing 24/2002, S. 227-238.
- Malhorta, N.: Marketing Research, Halow u.a. 2000.
- Neumann, W./Kreuger, L.: Social work, research methods, qualitative and quantitative applications, Pearson Education 2003.
- Schenk, M.: Soziale Netzwerke und Massenmedien, Moor Siebeck 1992.
- Schnell, R./Hill, P./Esser, E.: Methoden der empirischen Sozialforschung. München u.a. 1999.
- Steffens-Duch, S.: Commitment – die Bank im Urteil der Mitarbeiter, in: Die Bank 3/2000, S. 183-185.
- Tiede, M./Voß, M.: Schliessen mit Statistik – Verstehen, München/Wien 2000.
- Ulrich, G.: Umfrage, nein danke! Von sein und schein, in: Alpha – der Kadermarkt, Beilage zu Tages-Anzeiger und Sonntagszeitung, 14./15. März 2009, S. 1.
- Ulrich, G.: Die Interviewer und Interviewerinnen – was bewirken sie wirklich?, in: vsms-Jahrbuch 2008 «Markt- und Sozialforschung», S. 24-27.
- Werani, T.: Der Wert von kooperativen Geschäftsbeziehungen in industriellen Märkten, in: Die Unternehmung 2/2000, S. 123-143.
- Zentes J.: Grundbegriffe des Marketings, 4. Auflage, Stuttgart 1996.