

Online-Shopping
Einflussfaktoren auf Auswahl- und Nutzungsentscheidungen
von Electronic-Commerce-Anwendungen im
Dienstleistungsbereich

Dissertation
zur
Erlangung des akademischen Grades
doctor rerum politicarum (Dr. rer. pol.)
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Rostock

vorgelegt von

Gunther Held, geboren am 05.01.1964 in Esslingen/Neckar

aus Hamburg

Rostock, 2009

Erstgutachter: Prof. Dr. Friedemann Nerdinger
Lehrstuhl für Wirtschafts- und Organisationspsychologie
Wirtschafts- und Sozialwissenschaftliche Fakultät
Universität Rostock

Zweitgutachter: Prof. Dr. Udo Konradt
Lehrstuhl für Arbeits- und Organisationspsychologie
Philosophische Fakultät
Christian-Albrechts-Universität zu Kiel

Tag der mündlichen Prüfung: 26.06.2009

Inhaltsverzeichnis

Abbildungsverzeichnis	VIII
Tabellenverzeichnis.....	X
Abkürzungsverzeichnis	XII
1. Einleitung	1
2. Theoretischer und empirischer Hintergrund	5
2.1 Electronic Commerce	6
2.1.1 Daten zur Internetnutzung	6
2.1.2 Klassifikation im Electronic Commerce	7
2.1.3 Vor- und Nachteile des Electronic Commerce aus Kundensicht	11
2.2 Dienstleistungen im Electronic Commerce.....	12
2.2.1 Konstitutiva und Abgrenzungsmerkmale in der Dienstleistungsforschung	13
2.2.1.1 Klassische Dienstleistungsforschung	13
2.2.1.2 Dienstleistungsforschung im Internet.....	15
2.2.2 Definition von Dienstleistungen im Internet.....	16
2.2.3 Typologisierung von Dienstleistungsarten im Internet	17
2.2.3.1 Vorüberlegungen und inhaltliche Fokussierung.....	18
2.2.3.2 Das 4C-Net-Modell nach Wirtz im Dienstleistungskontext.....	18
2.2.3.3 Dienstleistungsarten im Internet.....	23
2.2.4 Electronic Commerce im Gesundheitswesen.....	26
2.2.5 Electronic Commerce und Krankenversicherungen.....	27
2.2.6 Zwischenfazit	27
2.3 Konsumentenverhalten.....	28
2.3.1 Grundlagen des Konsumentenverhaltens	28
2.3.2 Involvement.....	31
2.3.2.1 Faktoren des Involvement	32
2.3.2.2 Involvement im Internetkontext	33
2.3.3 Auswirkungen des Involvement auf Informations- und Entscheidungsverhalten ..	34
2.3.3.1 Involvement und Informationsverhalten	34
2.3.3.2 Involvement und Entscheidungsverhalten.....	35
2.3.3.3 High- und Low-Involvement bei Informations-und Entscheidungsprozessen.	37
2.3.4 Konsumentenverhalten im Electronic Commerce.....	38
2.3.4.1 Aktivierende Prozesse im Electronic Commerce	39
2.3.4.2 Extensives vs. limitiertes Kaufverhalten	40

2.3.4.3 Habitualisiertes Kaufverhalten vs. Impulskäufe.....	42
2.4 Kundenorientierung im Electronic Commerce	43
2.5 Erfolgsfaktorenforschung im Untersuchungsfeld Electronic Commerce	45
2.5.1 Begriffsklärung.....	45
2.5.2 Methoden der Erfolgsfaktorenforschung	47
2.5.3 Kritik an der Erfolgsfaktorenforschung	49
2.5.4 Erfolgsfaktoren kommerzieller Websites.....	51
2.5.4.1 Nutzungszufriedenheit.....	51
2.5.4.2 Nutzungsintention.....	52
2.5.4.3 Nutzung	53
2.5.4.4 Auswahl.....	53
2.5.5 Zwischenfazit	54
2.6 Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle	54
2.6.1 Charakteristika reflektiver Messmodelle	55
2.6.2 Charakteristika formativer Messmodelle	57
2.6.3 Fehlspezifikation des Messmodells.....	58
2.6.4 Indexentwicklung zur Operationalisierung formativer Konstrukte	61
2.6.4.1 Konstruktdefinition.....	61
2.6.4.2 Bestimmung der Indikatoren	62
2.6.4.3 Multikollinearität.....	64
2.6.4.4 Schätzung des formativen Messmodells.....	64
2.6.4.5 Indexberechnung	65
2.7 Usability	66
2.7.1 Begriffliche und konzeptionelle Basis	66
2.7.2 Europäische NORM EN ISO 9241	67
2.7.3 Heuristiken nach Nielsen	70
2.7.4 Ergänzende Heuristiken	75
2.7.5 Usability als Erfolgsfaktor von Webangeboten	76
2.8 Methoden der Usabilityevaluation	78
2.8.1 Mikroevaluation und summarische Evaluation.....	78
2.8.2 Deskriptive vs. prädiktive Usability-Evaluationsmethoden.....	79
2.8.2.1 Deskriptive Evaluationsmethoden.....	80
2.8.2.2 Prädiktive Evaluationsmethoden	81
2.9 Fragebogenbasierte Usabilityevaluation	81
2.9.1 Questionnaire for User Interface Satisfaction (QUIS).....	82
2.9.2 Software Usability Measurement Inventory (SUMI).....	83
2.9.3 Web Analysis and MeasureMent Inventory (WAMMI).....	85
2.9.4 IBM After-Scenario Questionnaire (ASQ)	87
2.9.5 IBM Post-Study System Usability Questionnaire (PSSUQ) & Computer System Usability Questionnaire (CSUQ).....	88

2.9.6 AttrakDiff 2	89
2.9.7 Wahrgenommene Orientierung in Online-Shops (WOOS)	91
2.9.8 Usability Fragebogen für Online-Shops (ufosV2)	92
2.10 Nomologisches Modell	95
2.10.1 Vertrauen.....	98
2.10.2 Reputation	99
2.10.3 Ästhetik	100
2.10.4 Nutzungsvergnügen.....	101
2.11 Zusammenhänge innerhalb des nomologischen Modells.....	102
2.11.1 Usability und Vertrauen	102
2.11.2 Reputation und Vertrauen	103
2.11.3 Usability und Nutzungsvergnügen	104
2.11.4 Ästhetik und Usability.....	104
2.11.5 Ästhetik und Nutzungsvergnügen	105
2.11.6 Moderierende Einflüsse	106
2.11.6.1 Begriffsklärung.....	106
2.11.6.2 Analyse latenter Moderatormodelle	107
2.11.6.2.1 Regressionsanalytischer Ansatz.....	107
2.11.6.2.2 Analysemethoden für Strukturgleichungsmodelle.....	108
3. Forschungsfragen	111
3.1 Ziele der Untersuchung	111
3.2 Inhaltliche Hypothesen.....	112
4. Untersuchungsplanung	117
4.1 Unabhängige Variable Website.....	117
4.1.1 Websites gesetzlicher Krankenkassen als Untersuchungsobjekt	117
4.1.2 Festlegung der Websites.....	118
4.2 Anpassung des Itempools der ufosV2a-Skalen.....	120
4.2.1 Vorüberlegungen.....	120
4.2.2 Überprüfung der Eignung und Modifikation der Items des ufosV2	121
4.2.2.1 Analyse der Anteile fehlender Werte	121
4.2.2.2 Analyse der inhaltlichen Verwendbarkeit	122
4.2.2.3 Umformulierung von Items	122
4.2.2.4 Analyse inhaltlich redundanter Indikatoren	123
4.2.3 Ergänzung des Itempools	123
4.2.4 Berücksichtigung von Störeffekten aufgrund der Polung bei der Itemformulierung.....	124
4.2.5 Antwortformat der Usability-Items.....	124
4.2.6 Voruntersuchung zur Verwendbarkeit der Items	125

4.2.7 Darstellung der verwendeten formativen und reflektiven Items.....	126
4.3 Operationalisierung der Variablen auf Prädiktorenebene	127
4.3.1 Erfassung der Ästhetik	128
4.3.2 Erfassung des Vertrauens	128
4.3.3 Erfassung des Involvement	129
4.3.4 Erfassung des Nutzungsvergnügens.....	130
4.3.5 Erfassung der Reputation	131
4.4 Operationalisierung der abhängigen Variablen.....	131
4.4.1 Reflektive Erfassung der Nutzerzufriedenheit	132
4.4.2 Reflektive Erfassung der Nutzungsintention.....	132
4.4.3 Operationalisierung der Auswahlentscheidung.....	133
4.5 Erhebung weiterer Variablen	134
4.5.1 Vorerfahrung im Bereich PC-/Internetnutzung und Online-Shopping	134
4.5.2 Bekanntheit des Webanbieters	135
4.5.3 Aktueller Versicherungsgeber.....	136
4.5.4 Wahrgenommener Preis	136
4.5.5 Demographische Merkmale der Teilnehmer.....	137
4.5.6 Kontrollvariablen zur Erhebungssituation	137
4.6 Versuchsszenario.....	137
4.7 Berücksichtigung möglicher Reihenfolgeeffekte.....	140
4.7.1 Reihenfolge der Websitedarbietung.....	141
4.7.2 Reihenfolge der Usabilityabfrage reflektiv versus formativ	141
4.7.3 Reihenfolge der Konstruktabfrage	142
4.8 Versuchsplan	144
5. Untersuchungsdurchführung	147
5.1. Teilnehmerstichprobe.....	147
5.1.1. Rekrutierung der Teilnehmer	147
5.1.1.1 Wege der Rekrutierung.....	147
5.1.1.2 Ablauf der Rekrutierung.....	148
5.1.1.3 Probleme bei der Teilnehmerrekrutierung.....	148
5.1.2 Beschreibung der Stichprobe	149
5.2 Computergestützte Durchführung des Versuchs.....	149
5.3 Versuchsablauf.....	152
6. Statistische Auswertung der Daten	157
6.1 Analyse und Behandlung fehlender Werte.....	158
6.1.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte	159
6.1.2 Fehlendmechanismen und ihre Konsequenzen	159
6.1.3 Imputation der fehlenden Daten.....	160

6.2 Vorgehen zur Identifikation problematischer Fälle	161
6.2.1 Fallweise Analyse der Varianz im Antwortverhalten	161
6.2.2 Konsistenz des Antwortverhaltens	161
6.3 Verteilungsprüfung.....	162
6.4 Prüfung auf Vorliegen eines Common Method Bias	163
6.5 Analyse der Güte reflektiver Maße auf Prädiktorenebene	164
6.5.1 Explorative Faktorenanalyse zur Analyse der Skalenstruktur	164
6.5.1.1 Voraussetzungen.....	164
6.5.1.2 Bestimmung der Faktorenanzahl und Faktorenqualität.....	165
6.5.2 Güteprüfung entsprechend der klassischen Testtheorie	166
6.5.3 Prüfung der Skalenstruktur mit Hilfe konfirmatorischer Faktorenanalyse	167
6.5.3.1 Voraussetzungen zur Berechnung einer CFA	167
6.5.3.2 Gütekriterien der CFA.....	168
6.6 Kausalanalyse mit Hilfe von Partial Least Squares (PLS).....	170
6.6.1 Voraussetzungen zur Anwendung des Partial Least Squares Algorithmus	170
6.6.2 Gütebeurteilung von PLS-Modellen	171
6.6.2.1 Beurteilung des Strukturmodells	172
6.6.2.2 Beurteilung reflektiver Messmodelle	174
6.6.2.3 Beurteilung formativer Messmodelle	174
6.7 Logistische Regression.....	175
6.7.1 Voraussetzungen der LR	176
6.7.2 Gütekriterien der LR	176
6.8 Weitere Testverfahren	178
6.8.1 t-Test.....	178
6.8.2 Einfaktorielle Varianzanalyse	178
6.8.3 Levene-Test.....	179
7. Ergebnisse	181
7.1 Vorbereitende Datenanalysen	181
7.1.1 Zusammenfassung der Bewertungen	181
7.1.2 Analyse und Behandlung fehlender Werte.....	181
7.1.2.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte	181
7.1.2.2 Analyse des Fehlendmechanismus	182
7.1.2.3 Ersetzen fehlender Werte	182
7.1.3 Verteilungen der erhobenen Indikatoren.....	183
7.1.4 Prüfung auf Vorliegen eines Common Method Bias	183
7.1.5 Identifikation problematischer Fälle	184
7.2 Reihenfolgeeffekte	184
7.3 Erfassung der Variablen auf Prädiktorenebene.....	185
7.3.1 Prüfung der Skalenstruktur mit Hilfe explorativer Faktorenanalyse	186

7.3.2	Prüfung der Skalengüte auf Einzelskalenebene	189
7.3.3	Prüfung der Skalenstruktur mit Hilfe konfirmatorischer Faktorenanalyse	190
7.3.4	Diskriminationsfähigkeit der reflektiven Skalen.....	194
7.3.5	Erfassung der wahrgenommenen Höhe des Beitragssatzes	196
7.4	Erfassung der abhängigen Variablen.....	196
7.4.1	Reflektive Erfassung der Nutzerzufriedenheit	196
7.4.2	Reflektive Erfassung der Nutzungsintention.....	196
7.4.3	Erfassung der Auswahlentscheidung	197
7.5	Festlegung der formativen Skala ufosV2fa.....	198
7.5.1	Endgültige Bestimmung der ufosV2fa-Indikatoren	198
7.5.1.1	Elimination von Indikatoren aufgrund eines zu hohen Anteils fehlender Werte	198
7.5.1.2	Eliminierung inhaltlich redundanter Indikatoren	199
7.5.2	Behandlung von Multikollinearität zwischen den ufosV2fa-Items.....	200
7.5.3	PLS-Zwei-Konstrukt-Modell zur Validierung.....	202
7.6	Untersuchung des nomologischen Modells mit Hilfe von PLS	204
7.6.1	Beurteilung der Messmodelle.....	204
7.6.2	Beurteilung der Strukturmodelle.....	207
7.7	Logistische Regression zur Vorhersage der Auswahlentscheidung.....	211
7.7.1	Prüfung der Voraussetzungen und Ausreißerdiagnostik.....	213
7.7.2	Sequentielle Berechnung der LR.....	213
7.8	Übersicht der Befunde.....	216
8.	Diskussion	219
8.1	Diskussion der Befunde	219
8.1.1	Befunde zur Güte der ufosV2a-Skalen.....	219
8.1.1.1	Reliabilität der reflektiven Usability-Skala ufosV2ra	219
8.1.1.2	Faktorielle Validität der reflektiven ufosV2ra-Skala	220
8.1.1.3	Diskriminationsfähigkeit der reflektiven ufosV2ra-Skala.....	220
8.1.1.4	Überprüfung der formativen Skala ufosV2fa	221
8.1.1.5	Kriteriumsvalidität der ufosV2a-Skalen.....	222
8.1.1.6	Nomologische Validität der ufosV2a-Skalen.....	223
8.1.2	Befunde zu den weiteren Hypothesen im nomologischen Modell.....	225
8.1.3	Befunde zu möglichen Effekten aufgrund der Itemreihenfolge.....	227
8.2	Bewertung des methodischen Vorgehens	227
8.2.1	Interne Validität des Quasi-Experiments	227
8.2.2	Externe Validität des Quasi-Experiments	229
8.2.2.1	Realitätsbezug der quasi-experimentellen Nutzungssituation.....	230
8.2.2.2	Operationalisierung der Auswahlentscheidung	231
8.2.2.3	Repräsentativität der Stichprobe.....	231

8.2.2.4 Abhängigkeit der Ergebnisse von der Auswahl der Websites.....	232
8.2.3 Vorgehen der statistischen Auswertung.....	233
8.3 Zusammenfassende Bewertung der Ergebnisse zu den ufosV2a-Skalen.....	235
8.4 Erfolgswirksamkeit der Usability bei Websites	237
8.5 Implikationen für Forschung und Praxis	237
8.6 Weiterführende Fragestellungen	239
Zusammenfassung der Untersuchung	241
Summary	243
Literaturverzeichnis.....	245
Anhang A: Ergänzende Darstellungen zum Hintergrund	289
Anhang B: Ergänzende Darstellungen zur ufosV2-Entwicklung	299
Anhang C: Ergänzende Darstellungen zur Untersuchungsplanung	307
Anhang D: Ergänzende Darstellungen zur Durchführung	313
Anhang E: Ergänzende Ergebnisdarstellungen	331

Abbildungsverzeichnis

Abb. 2-1:	Häufigkeit (%-Angaben) der im Internet gekauften Waren und Dienstleistungen in Deutschland (AGOF, 2008)	9
Abb. 2-2:	Typologisierung von Dienstleistungen im Internet nach Digitalisierungsgrad und echter Interaktion (nach Breithaupt 2005, S. 125)	23
Abb. 2-3:	Relevante individuelle Zustände und Prozesse des Konsumentenverhaltens (angelehnt an Trommsdorff, 1998, S. 32).	29
Abb. 2-4:	Ausmaß an kognitiver Steuerung und Involvement bei unterschiedlichen Typen von Kaufentscheidungen (Angelehnt an Kuß & Tomczak, 2000, S. 97).....	36
Abb. 2-5:	Gegenüberstellung von High- und Low-Involvement bei Informations- und Entscheidungsprozessen (Eigene Darstellung in Anlehnung an Trommsdorff, 1998).....	38
Abb. 2-6:	Methoden zur Identifikation von Erfolgsfaktoren (Haenecke, 2002)	47
Abb. 2-7:	Latentes Konstrukt mit reflektiven Indikatoren	55
Abb. 2-8:	Latentes Konstrukt mit formativen Indikatoren	57
Abb. 2-9:	Zwei-Konstrukt-Modell zur Bestimmung der Validität einer formativen LV (Diamantopoulos & Winklhofer, 2001)	65
Abb. 2-10:	Klassifikation von Methoden zur Usabilityevaluation (in Anlehnung an Hamborg & Gediga, 2006).....	79
Abb. 2-11:	Zwei Items aus dem QUIS-Fragebogen 7.0 (Harper et al., 1997).....	83
Abb. 2-12:	Zwei Items aus dem SUMI-Fragebogen 3.9 (Kirakowski, 1997)	84
Abb. 2-13:	Zwei Items aus dem WAMMI-Fragebogen 3.9 (Kirakowski & Claridge, 2001)	86
Abb. 2-14:	Auszug aus dem After-Scenario Questionnaire (Lewis, 1995).....	88
Abb. 2-15:	Zwei Items aus dem AttrackDiff 2-Fragebogen (www.attrakdiff.de).....	90
Abb. 2-16:	Beispielitems aus dem ufosV2-Fragebogen (Christophersen, 2007)	94
Abb. 2-17:	Untersuchungsmodell	96
Abb. 3-1:	Zwei-Konstrukt-Modell zur Validierung der formativen Skala ufosV2fa anhand der reflektiven Usability-Skala ufosV2ra	113
Abb. 3-2:	Zuordnung der Hypothesen im nomologischen Modell.....	115
Abb. 4-1:	Schematischer Aufbau des Fragebogens zur Bewertung der Websites	143
Abb. 4-2:	Zusammensetzung des Teilnehmercodes.....	145
Abb. 5-1:	Screenshot einer Seite der Hypertext -Versuchsumgebung.....	150
Abb. 5-2:	Ausschnitt aus dem Fragebogen	151
Abb. 5-3:	Schematische Darstellung des Versuchsablaufs	152
Abb. 7-1:	CFA-Modell zur Überprüfung der Skalenstruktur.....	191
Abb. 7-2:	Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala ufosV2fa mit Hilfe der reflektiven Usability-Skala ufosV2ra	203

Abb. 7-3:	Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen	207
Abb. 7-4:	Schritte der sequentiellen logistischen Regressionsanalyse unter Angabe der jeweils eingesetzten Prädiktoren	212

Tabellenverzeichnis

Tab. 2-1:	Teilbereiche des Electronic Commerce (nach Kuß & Tomczak, 2000, S. 156). .. 8	8
Tab. 2-2:	Absicht, Waren und Dienstleistungen im Internet zu kaufen (AGOF, 2008) 10	10
Tab. 2-3:	Entscheidungshilfen zur Frage, ob ein Konstrukt reflektiv oder formativ ist 62	62
Tab. 2-4:	Übersicht der Heuristischen Regeln nach Nielsen (1994)..... 70	70
Tab. 2-5:	Evaluation des Messmodells der CFA (Christophersen, 2007, S. 182) 95	95
Tab. 3-1:	Inhaltliche Hypothesen im nomologischen Modell zur Güte der ufosV2a-Skalen (H5 bis H10) 114	114
Tab. 3-2:	Inhaltliche Hypothesen im nomologischen Modell zu den Zusammenhängen der weiteren Konstrukte (H11 bis H19) 116	116
Tab. 4-1:	Im Rahmen der Untersuchung eingesetzte Websites 120	120
Tab. 4-2:	Reflektive Indikatoren zur Erfassung der Usability 126	126
Tab. 4-3:	Formative Indikatoren zur Erfassung der Usability 127	127
Tab. 5-1:	Geschlecht und Alter der Versuchsteilnehmer 149	149
Tab. 6-1:	Übersicht über die korrespondierenden Abschnitte zu den Hypothesen..... 158	158
Tab. 6-2:	Anspruchsniveaus der herangezogenen Gütemaße zur Beurteilung des CFA-Modells..... 170	170
Tab. 6-3:	Gütemaße der logistischen Regression..... 178	178
Tab. 7-1:	Missing Data im Datensatz - Aufteilung der Teilnehmer 182	182
Tab. 7-2:	Abfolge der reflektiven und formativen Usability-Skalen 185	185
Tab. 7-3:	Übersicht über alle berechneten explorativen Faktorenanalysen zur Überprüfung der angenommenen Skalenstruktur..... 187	187
Tab. 7-4:	Mustermatrix der fünften berechneten Hauptachsenanalyse..... 188	188
Tab. 7-5:	Skalen- und Itemkennwerte der reflektiven Skalen auf Prädiktorenebene 189	189
Tab. 7-6:	Standardisierte Faktorladungen und Messfehlervarianzen der CFA zur Analyse der Skalenstruktur 192	192
Tab. 7-7:	Evaluation des Messmodells der CFA 193	193
Tab. 7-8:	Konstruktinterkorrelationen auf Basis der CFA..... 193	193
Tab. 7-9:	Prüfung der Normalverteilungsannahme für die reflektiven Skalenwerte 194	194
Tab. 7-10:	Ergebnisse der Levene-Tests auf Varianzhomogenität 195	195
Tab. 7-11:	Ergebnisse der mehrdimensionalen einfaktoriellen Varianzanalyse..... 195	195
Tab. 7-12:	Skalen- und Itemkennwerte der reflektiven Skala zur Nutzerzufriedenheit 196	196
Tab. 7-13:	Skalen- und Itemkennwerte der reflektiven Skala zur Nutzungsintention 197	197
Tab. 7-14:	Wahlhäufigkeit der Websites 198	198
Tab. 7-15:	Formative Usability Indikatoren mit mehr als 10 fehlenden Werten 199	199
Tab. 7-16:	Indikatorpaarungen mit hoher bivariater Korrelation und inhaltlicher Redundanz 200	200

Tab. 7-17:	Kennwerte des PLS-Messmodells zum Zwei-Konstrukt-Modell.....	202
Tab. 7-18:	Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte.....	205
Tab. 7-19:	Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability	206
Tab. 7-20:	Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: aufgeklärte Varianz (R^2).....	208
Tab. 7-21:	Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Pfadkoeffizienten.....	208
Tab. 7-22:	Übersicht über die mit Hilfe der PLS-Modelle überprüften Hypothesen.....	211
Tab. 7-23:	Kennwerte der sequentiellen LR Schritte 1 bis 2c	214
Tab. 7-24:	Items der endgültigen reflektiven Skala ufosV2ra	216
Tab. 7-25:	Items der endgültigen formativen Skala ufosV2fa	216
Tab. 7-26:	Hypothesen zu Güte und Wirkungszusammenhängen der ufosV2a-Skalen	217
Tab. 7-27:	Weitere Hypothesen zu Zusammenhängen im nomologischen Modell	218

Abkürzungsverzeichnis

α	Cronbachs Alpha
β	Pfadkoeffizient
δ	Delta-Wert in der explorativen Faktorenanalyse
ζ	Messfehler auf Ebene der latenten Variable
ε	Messfehler auf Indikatorebene
λ	Ladung
γ	Gewicht
η	latente Variable, Konstrukt
χ^2	Chi-Quadrat-Wert
λ_j	Eigenwert in der Faktorenanalyse
ADF	Asymptotically distribution-free
AGFI	Adjusted-Goodness-of-Fit-Index
AGOF	Arbeitsgemeinschaft Online-Forschung e.V.
a_{ij}	Ladungsgewichte in der Faktorenanalyse
ANOVA	Einfaktorielle Varianzanalyse
ASQ	After Scenario Questionnaire
Aufl.	Auflage
AV	Abhängige Variable
AVE	Average Variance Extracted
B	Regressionskoeffizient
B2C	Business-to-Consumer
bzgl.	bezüglich
bzw.	beziehungsweise
ca.	circa
CFA	konfirmatorische Faktorenanalyse
CFI	Comparative-Fit-Index
CSUQ	Computer System Usability Questionnaire
CUSI	Computer User Satisfaction Inventory
$DEV(\xi_1)$	durchschnittlich erfasste Varianz eines Faktors
df	Freiheitsgrad
e	Messfehler des Indikators
e.V.	eingetragener Verein
E-Commerce	Electronic Commerce
ed.	Edition (englisch: Auflage)
Ed.	Editor (englisch: ein Herausgeber)
EFA	Explorative Faktorenanalyse

et al.	et alii (lateinisch für: und andere)
etc.	et cetera (und so weiter)
Ex	Exzess (Kurtosis)
Exp(B)	Exponentialfunktionswert des Regressionskoeffizienten
F	F-Wert
f.	und folgende Seite
ff.	und folgende Seiten
GFI	Goodness-of-Fit-Index
GLS	Generalized Least Square
GUI	Graphical User Interface (englisch: graphische Benutzerschnittstelle)
H	Hypothese
h^2	Kommunalität
HAA	Hauptachsenanalyse
HCI	Human-Computer Interaction
HKA	Hauptkomponentenanalyse
HQ	Hedonische Qualität
HQ-I	Hedonische Qualität - Identität
HQ-S	Hedonische Qualität - Stimulation
Hrsg.	Herausgeber
KMO	Kaiser-Meyer-Olkin-Koeffizient
LISREL	Linear Structural Relations
LV	Latente Variable
LR	Logistische Regression
M	Mittelwert
MAR	Missing At Random
MCAR	Missing Completely At Random
ML	Maximum Likelihood
MLR	Multinomiale Logistische Regression
MNAR	Missing Not At Random
MSA	Measure of Sampling Adequacy
n	Anzahl
NA	Not Applicable (englisch: nicht beantwortbar)
NFI	Normed-Fit-Index
p	Signifikanzniveau mit folgenden Klassifikationen: p<0.10, * p<0.05, ** p<0.01, *** p<0.001
PEUO	Perceived Ease of Use
PLS	Partial Least Squares
p_m	Itemschwierigkeit
PQ	Pragmatische Qualität

PU	Perceived Usefulness
PUTQ	Purdue Usability Testing Questionnaire
QUIS	Questionnaire for User Interface Satisfaction
r	Korrelationskoeffizient
R ²	Determinationskoeffizient (aufgeklärte Varianz)
rel(x _i)	Indikatorreliabilität in der konfirmatorischen Faktorenanalyse
r _{it}	korrigierte Trennschärfe
RMSEA	Root-Mean-Square-Error-of-Approximation
S.	Seite
Sch	Schiefe
SD	Standardabweichung
SE	Standardfehler
SRMR	Standardized-Root-Mean-Square-Residual
SUMI	Software Usability Measurement Inventory
t	t-Wert
Tab.	Tabelle
TAM	Technology Acceptance Model
uf	Präfix eines formativen Items der Usability-Skala ufosV2f
uf_in	Präfix eines formativen Unterindizes der Skala ufosV2f zur Behandlung von Multikollinearität
ufos	Usability Fragebogen für Online-Shops
ufosV1	erste Version des Usability Fragebogen für Online-Shops
ufosV2	zweite Version des Usability Fragebogen für Online-Shops
ufosV2a	Weiterentwickelte Version des Usability Fragebogen ufosV2 für Dienstleistungs-Websites
ufosV2f	formative Skala des ufosV2
ufosV2r	reflektive Skala des ufosV2
ufosV2fa	formative Skala des ufosV2a
ufosV2ra	reflektive Skala des ufosV2a
ur	Präfix eines reflektiven Items der Usability-Skala ufosV2r
URL	Uniform Resource Locator (Internetadresse)
UV	unabhängige Variable
vgl.	vergleiche
Vpn	Versuchsperson(en)
vs.	versus
Wald	Wert der Wald-Statistik
WAMMI	Website Analysis and Measurement Inventory
WOOS	Wahrgenommene Orientierung in Online-Shops
x	reflektiver Indikator
y	formativer Indikator

zResid

standardisierte Residuen

1. Einleitung

Das Internet gewann in den letzten Jahren immer stärker an Bedeutung. Laut der AGOF-Studie Internet Facts 2008 I, an der 110.947 Internetnutzer ab 14 Jahren teilnahmen, liegt die durchschnittliche Reichweite des Internet in den Altersgruppen ab 14 Jahren bei annähernd 63% (AGOF, 2008). Für die junge Altersgruppe der 14-29-Jährigen ist die Online-Nutzung inzwischen etwas Selbstverständliches. Diese Gruppe ist mit nahezu 95% fast vollständig im Internet präsent. Dienstleistungsangebote und kommerzielle Angebote profitieren in zunehmendem Maße von den interaktiven Möglichkeiten des Internets. Immer mehr Internetnutzer führen bislang offline erledigte Einkäufe und Geschäftskontakte auch im Internet aus. So gaben 62 Prozent der Befragten an, in den zurückliegenden 3 Monaten online eingekauft zu haben und rund 55 Prozent haben Online-Banking genutzt.

Allerdings nimmt ein großer Teil der Webnutzer häufig oder gelegentlich Probleme während der Internetsitzung wahr, so z.B. aufgrund von unübersichtlichen Homepages, aufwändiger Informationssuchen oder aufgrund eines langsamen Seitenaufbaus (vgl. Dzida & Wandke, 2006, S. 483; Van Eimeren, Gerhard & Frees, 2002). In verschiedenen Studien, die sich mit der Identifikation von Erfolgsfaktoren für die Akzeptanz von Electronic-Commerce-Angeboten beschäftigten, wurde die Erfolgswirksamkeit des Faktors Usability empirisch untermauert. Die Nutzerfreundlichkeit bzw. Usability ist als einer von mehreren kundennahen Erfolgsfaktoren im Segment Business-to-Consumer anzusehen (vgl. Chen, Gillenson & Sherrell, 2004; Petre, Minocha & Roberts, 2006). Die Optimierung der Usability von Electronic-Commerce-Angeboten ist damit eine wichtige Fragestellung in der Electronic-Commerce-Forschung.

Als Definition der Usability wird vielfach auf die europäische Norm EN ISO 9241 „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ verwiesen. Diese zielt auf die Tauglichkeit eines Arbeitsmittels für den Gebrauch im Nutzungskontext ab und umfasst die Effektivität, Effizienz und Zufriedenheit als Komponenten der Usability (DIN EN ISO 9241-11, 1999). Die Evaluation der Usability kann mit Hilfe unterschiedlicher Verfahren erfolgen. Diese unterscheiden sich in vielerlei Hinsicht etwa dahingehend, inwieweit, konkrete Usability-Probleme in der Interaktion identifiziert und Vorschläge zur Verbesserung abgeleitet werden sollen oder hinsichtlich der Expertise des eingebundenen Personenkreises.

Für die Evaluation der Usability von Software existieren unterschiedliche standardisierte *Fragebogenverfahren*, die eine Bewertung der Gebrauchstauglichkeit aus Sicht der Nutzer ermöglichen. Standardisierte Fragebogenverfahren ermöglichen den Zugang zu nicht oder nur eingeschränkt beobachtbaren Daten wie z.B. Einstellungen der Testperson (Nielsen, 1993). Sie zeichnen sich dadurch aus, dass ein Nutzerurteil zur Gebrauchstauglichkeit einer Software quantitativ erfasst wird und die Ergebnisse anhand statistischer Gütekriterien überprüfbar sind.

Im Rahmen der vorliegenden Arbeit wird die Bedeutung der Usability von Electronic-Commerce-Anwendungen im Dienstleistungsbereich untersucht. Hierbei wird Bezug genommen auf die Ergebnisse der Arbeit von Christophersen (2007). Der Autor entwickelte mit dem *Usability Fragebogen für Online-Shops2*, kurz *ufosV2*, ein Verfahren, das die Erfassung der kundenseitigen Bewertung der Usability von Online-Shops ermöglicht. Mit der vorliegenden Arbeit sollte die Anwendbarkeit der Usability-Skala *ufosV2* auf den Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich untersucht werden. Ein Ziel der durchgeführten Untersuchung war es, die Skala hinsichtlich der Anforderungen von Dienstleistungen zum *ufosV2a*¹ weiterzuentwickeln. In der Folge sollten die Einflussfaktoren auf die Auswahl- und Nutzungsentscheidungen von E-Commerce-Anwendungen im Dienstleistungsbereich, am Beispiel der gesetzlichen Krankenversicherung, untersucht werden. Dabei wurde dem Einfluss des Kundeninvolvement auf die betrachteten Entscheidungsprozesse eine besondere Berücksichtigung zugemessen.

Mit der Skala *ufosV2a* sollte ein valides Instrument zur Evaluation der Usability von Electronic-Commerce-Anwendungen im Dienstleistungsbereich aus Kundensicht bereitgestellt werden. Auf der Grundlage des verfolgten Validierungsansatzes kann die Relevanz der Usability im Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich untermauert werden. Zusammenhänge zwischen der Usability und anderen kundennahen Erfolgsfaktoren, die aufgrund vorangegangener Studien als empirisch abgesichert gelten, erfahren auch bei Verwendung der weiterentwickelten *ufosV2a*-Skalen erneut Bestätigung. Die vorliegende Untersuchung untermauert diese Befunde für den Untersuchungsbereich der Online-Dienstleistungen am Beispiel der gesetzlichen Krankenversicherung und ergänzt sie, indem sie den Einfluss des Involvement auf die Usability mit einbezieht.

Im Folgenden wird der Aufbau der vorliegenden Arbeit skizziert. Nach diesem einleitenden, zum Thema hinführenden Kapitel wird der theoretische und empirische Hintergrund der Untersuchung vorgestellt (Kapitel 2). Der Schwerpunkt liegt auf einer Einführung in das Kundenverhalten im Electronic-Commerce sowie auf der Usability und ihrer fragebogenbasierten Erfassung. Der Erfolgsfaktor Usability wird in ein konzeptuelles Erfolgsmodell zur Vorhersage der Nutzungsintention integriert, mit dessen Hilfe die weiterentwickelten Usability-Skalen *ufosV2a* validiert werden sollen. Schließlich werden die Zusammenhänge innerhalb des konzeptuellen Erfolgsmodells erläutert.

Im dritten Kapitel werden die Ziele der Untersuchung dargestellt und die inhaltlichen Hypothesen abgeleitet. Kapitel vier dient der Vorstellung der Untersuchungsplanung. Die *UfosV2a*-Skalen werden an die Erfordernisse von Electronic-Commerce-Anwendungen im Dienstleistungsbereich angepasst. Desweiteren werden die Operationalisierungen der Erfolgs-

¹ Die hinsichtlich der Anforderungen von Electronic-Commerce-Anwendungen im Dienstleistungsbereich weiterentwickelten *ufosV2*-Skalen werden im weiteren Verlauf durch ein angehängtes „a“ gekennzeichnet.

faktoren, die neben der Usability in das konzeptuelle Erfolgsmodell einfließen, sowie der Aufbau des Versuchsplans dargestellt.

Die Durchführung der Untersuchung wird in Kapitel fünf dokumentiert. Hierbei wird die Teilnehmerstichprobe hinsichtlich ihrer Rekrutierung und Zusammensetzung beschrieben sowie die computergestützte Durchführung des Versuchs und der Versuchsablauf dargestellt.

In Kapitel sechs werden die bei der Analyse der Daten eingesetzten statistischen Verfahren sowie ihre Voraussetzungen und Gütekriterien dargestellt. Von zentraler Bedeutung sind hierbei der Partial Least Squares Algorithmus (PLS) zur Schätzung von Strukturgleichungsmodellen sowie das Verfahren der Logistischen Regression (LR). Darüber hinaus werden im Rahmen der Analyse weitere Verfahren angewandt, die ebenfalls beschrieben werden.

Die Untersuchungsergebnisse werden in Kapitel sieben vorgestellt. Es kommen die in Kapitel sechs benannten statistischen Verfahren zur Anwendung, um die Hypothesen zur Güte der weiter entwickelten Usability-Skalen zu überprüfen. In Kapitel acht wird eine Überprüfung der Ergebnisse vorgenommen. Es erfolgt eine Diskussion der Befunde, inwieweit die angestrebten Untersuchungsziele erreicht wurden, sowie eine kritische Auseinandersetzung mit dem methodischen Vorgehen. Desweiteren werden Implikationen für die weitere Forschung und die Praxis dargestellt. Abschließend wird in Kapitel neun die gesamte Arbeit zusammengefasst.

2. Theoretischer und empirischer Hintergrund

Mit der vorliegenden Arbeit wird die Bedeutung der Usability im Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich untersucht. Hierbei wird Bezug genommen auf die Ergebnisse der Arbeit von Christophersen (2007). Der Autor entwickelte mit der Usability-Skala uvoSv2 ein Instrument zur fragebogenbasierten Erfassung des Konstruktes Usability aus Nutzersicht. Ein zentrales Anliegen der vorliegenden Arbeit besteht darin, die Anwendbarkeit der Usability-Skala uvoSv2 auf den Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich, hier am Beispiel der gesetzlichen Krankenkassen, zu überprüfen und bei Bedarf an die Anforderungen von Dienstleistungen anzupassen. In Abschnitt 2.1 wird zunächst ein allgemeiner Überblick über das Forschungsfeld Electronic-Commerce gegeben, bevor in Abschnitt 2.2 das Untersuchungsfeld der Dienstleistungen im Bereich Electronic-Commerce dargestellt wird. Dabei erfahren die Anforderungen und Potenziale des Electronic-Commerce im Gesundheitswesen und im Bereich der Krankenversicherungen eine besondere Betrachtung. In Abschnitt 2.3 erfolgt, verankert in der allgemeinen Konsumentenverhaltenstheorie, eine Betrachtung empirischer und theoretischer Erkenntnisse des Konsumentenverhaltens im Electronic-Commerce. Dabei erfährt der Einfluss des Involvement auf das Informations- und Entscheidungsverhalten der Konsumenten eine besondere Würdigung. Die Grundlagen einer kundenorientierten Gestaltung der Electronic-Commerce Prozesse als Voraussetzung für eine erfolgreiche Online-Geschäftstätigkeit werden in Abschnitt 2.4 erörtert. In Abschnitt 2.5 erfolgt ausgehend von einer allgemeinen Auseinandersetzung mit der Erfolgsfaktorenforschung eine Betrachtung kritischer Erfolgsfaktoren kommerzieller Websites. Abschnitt 2.6 stellt dar, dass es bei der Erfassung latenter Konstrukte wie der betrachteten kundennahen Erfolgsfaktoren kommerzieller Websites zwischen zwei Formen von Messmodellen zu differenzieren gilt. Eine Nichtberücksichtigung dieser Unterscheidung kann zu negativen Auswirkungen auf die Messgüte einer Skala führen. In Abschnitt 2.7 wird das Kernkonstrukt Usability begrifflich bestimmt und seine Relevanz im Kontext des Electronic-Commerce beleuchtet. Um die Usability von Softwareanwendungen zu evaluieren, stehen unterschiedliche Methoden zur Verfügung. Ein Überblick über das Spektrum der unterschiedlichen Evaluationsansätze wird in Abschnitt 2.8 gegeben. In Abschnitt 2.9 wird daran anschließend ein besonderer Focus auf die fragebogenbasierte Usabilityevaluation gerichtet. In der vorliegenden Arbeit wird ein Ansatz zur Validierung der zu entwickelnden Skalen verfolgt, bei dem die zu messenden Konstrukte in einen Zusammenhang gestellt werden. In Abschnitt 2.10 wird deshalb ein nomologisches Modell dargestellt, in dem die Usability zu weiteren kundennahen Erfolgsfaktoren kommerzieller Websites in Beziehung gesetzt wird. Die in dem nomologischen Modell angenommenen Zusammenhänge zwischen einzelnen Erfolgsfaktoren, sowie moderierende Einflüsse des Involvements auf ausgewählte Zusammenhänge werden in Abschnitt 2.11 empirisch untermauert.

2.1 Electronic Commerce

Electronic Commerce (E-Commerce) wird nach Hermanns und Gampgenrieder (2002, S. 70f.) als der absatzgerichtete Teil des E-Business verstanden. Der Begriff Electronic Business (E-Business) wird in der Literatur allerdings uneinheitlich verwandt. Eine Gegenüberstellung von Definitionsansätzen findet sich beispielsweise bei Wirtz (2001, S. 33). Unter E-Business wird im Folgenden die elektronische Anbahnung, die Verhandlung sowie Durchführung jeglicher kommerzieller Aktivitäten unter Nutzung des Internet verstanden. E-Commerce konzentriert sich folglich auf die Unterstützung und Erweiterung von Unternehmensaktivitäten im Absatzbereich eines Unternehmens mit Hilfe des Internet (Wiedemann, Frenzel & Buxel, 2001, S. 420).

In Anlehnung an Meffert (1992, S. 24) umfasst der Absatzprozess eines Unternehmens marktverbindende Aktivitäten, die im weitesten Sinne darauf abzielen, Markttransaktionen zu schaffen, zu erhalten und zu intensivieren. Die Ausgestaltung dieses Prozesses ist nicht nur aus Sicht der Kunden, sondern auch aus Unternehmenssicht bedeutsam, da hier Erfolg bzw. Misserfolg des Unternehmens in hohem Maße beeinflusst werden. Überträgt man diese Überlegungen auf den E-Commerce-Prozess, so ergibt sich folgendes Verständnis:

Der E-Commerce Prozess umfasst sämtliche Aktivitäten eines Unternehmens, die dazu bestimmt sind, die von ihm geschaffenen Produkte unter Nutzung des Internet am Markt zu verwerten. Dabei ist unter einem Produkt im E-Commerce all das zu verstehen, was sich einem potentiellen Abnehmer anbieten lässt, um ein Bedürfnis oder einen Wunsch zu befriedigen (Schmid, 2001). Im weiteren Verlauf der vorliegenden Arbeit bezieht sich die Darstellung auf die Möglichkeiten des Internet zur Unterstützung, Verbesserung und Erweiterung existierender Geschäftsmodelle. Im nachfolgenden Abschnitt werden zunächst ausgewählte Daten zur Nutzung des Internet in Deutschland dargestellt.

2.1.1 Daten zur Internetnutzung

Mit zunehmendem Interesse der Forschung am Konsumentenverhalten im Internet entwickelt sich die Verfügbarkeit diesbezüglicher kunden- bzw. nutzerbezogener Daten. Da es keine einheitliche Erhebungsweise und Grundlage gibt und die Daten schnell veralten, variieren diese teilweise. Dennoch sollen nachfolgend einige ausgewählte Daten zur Internetnutzung in Deutschland aufgezeigt werden, um die Entwicklungen in der Internetnutzung deutlich zu machen.

Laut der AGOF-Studie Internet Facts 2008 I, an der 110.947 Internetnutzer ab 14 Jahren teilnahmen, liegt die durchschnittliche Reichweite des Internet in den Altersgruppen ab 14 Jahren im Durchschnitt bei annähernd 63% (AGOF, 2008). Für die jungen Altersklassen ist die Online-Nutzung inzwischen etwas Selbstverständliches. Die Gruppe der 14-29-Jährigen ist mit nahezu 95% fast vollständig im Internet präsent. Aber nicht nur die Anzahl der Internetnutzer hat über die Jahre kontinuierlich zugenommen, auch der Anteil an Personen

mit großer Nutzungserfahrung ist gewachsen. Mittlerweile sind bereits 69 Prozent der Internetnutzer seit mehr als drei Jahren im Netz.

Nach wie vor zeichnen sich Internetnutzer durch ein höheres Bildungsniveau aus: 29,5 Prozent von ihnen haben Abitur bzw. einen Universitätsabschluss. Darüber hinaus weisen sie gehobene Einkommensstrukturen auf: 28,9 Prozent der Internetnutzer verfügen über ein Haushaltsnettoeinkommen von 3.000 Euro und mehr. Gegenüber den Anfängen des Internet nutzen zunehmend mehr Frauen und ältere Menschen das Internet. Sowohl in Bezug auf das Geschlecht als auch auf das Alter der Internetnutzer erfolgt eine kontinuierliche Anpassung an das tatsächliche Profil der deutschen Bevölkerung. Nach AGOF waren im 1. Quartal 2008 bereits 45% der Internetnutzer weiblich, bei einem Anteil von 51% in der Gesamtbevölkerung.

Sinkende Kosten für die Infrastruktur und Nutzung des Internets, sowie verbesserte Übertragungstechnologien 2007 nutzten 69% der Internetnutzer DSL sowie 20% ISDN und der damit verbundene leichte und kostengünstige Zugang zum Internet, haben die Anzahl der Internetnutzer in den letzten Jahren kontinuierlich anwachsen lassen.

An erster Stelle steht für 88,9 Prozent der Internetnutzer das Senden und Empfangen von privaten E-Mails, gefolgt von der Informationsrecherche bei 88 Prozent der User. Auch Dienstleistungsangebote und kommerzielle Angebote profitieren zunehmend von den interaktiven Möglichkeiten des Internets. Immer mehr Internetnutzer führen bislang offline erledigte Geschäfte jetzt auch im Internet aus: 62 Prozent haben im 1. Quartal 2008 online eingekauft und 54,8 Prozent haben Online-Banking genutzt.

Ein Vergleich der Produktpräferenzen bei Internetnutzern und bei Nicht-Internetnutzern zeigt, dass bei beiden Gruppen nahezu die gleichen Artikel, mehrheitlich Produkte des täglichen Bedarfs, im Fokus des alltäglichen Interesses stehen. Bei Internetnutzern finden sich außerdem Musik-CDs, Urlaubs- und Last-Minute-Reisen. Sie zeigen darüber hinaus überdurchschnittliches Interesse an Computer-Software, gebührenpflichtigen Musik- und Filmdownloads, Computer-Hardware, Computer- und Videogames. Diese Interessenschwerpunkte symbolisieren die unterschiedlichen Lebenswelten der beiden Gruppen: während Nicht-Internetnutzer ihren Fokus auf alltägliche Produkte der Offline-Welt setzen, sind Internetnutzer Produkten aus der technischen und digitalen Welt eng verbunden (AGOF, 2008).

2.1.2 Klassifikation im Electronic Commerce

Kuß und Tomczak (2000, S. 156) unterscheiden die Akteure eines E-Commerceprozesses in die Gruppen Endverbraucher, Unternehmen und öffentliche Verwaltung. Je nach Klassifikation der beteiligten Anbieter und Nachfrager einer Leistung lassen sich somit verschiedene Teilbereiche des E-Commerce differenzieren. Tabelle 2-1 gibt diesbezüglich einen Überblick.

Tab. 2-1: Teilbereiche des Electronic Commerce (nach Kuß & Tomczak, 2000, S. 156).

		Nachfrager der Leistung		
		Consumer	Business	Administration
Anbieter der Leistung	Consumer	Consumer-to-Consumer (z.B. Internet-Kleinanzeigenmarkt)	Consumer-to-Business (z.B. Jobbörsen mit Anzeigen von Arbeitssuchenden)	Consumer-to-Administration (z.B. Steuerabwicklung von Privatpersonen)
	Business	Business-to-Consumer (z.B. Verkauf von Krankenversicherungsschutz)	Business-to-Business (z.B. Beschaffung von Bürobedarf durch Unternehmen)	Business-to-Administration (z.B. Steuerabwicklung von Unternehmen)
	Administration	Administration-to-Consumer (z.B. Abwicklung von Unterstützungsleistungen)	Administration-to-Business (z.B. Beschaffung von Bürobedarf durch öffentliche Institutionen)	Administration-to-Administration (z.B. Transaktionen zwischen öffentlichen Institutionen)

Im Rahmen des E-Commerce kann beispielsweise ein Unternehmen mit anderen Unternehmen, mit Bereichen der öffentlichen Verwaltung, aber auch mit Endverbrauchern in Beziehung stehen. Man spricht von Business-to-Business (B-to-B), wenn es sich beim Nachfrager der Leistung um ein Unternehmen handelt und von Business-to-Administration (B-to-A), wenn es sich beim Nachfrager der Leistung um den öffentlichen Sektor handelt. Geht es jedoch um einen Endverbraucher, liegt Business-to-Consumer (B-to-C) vor (Weber & Weber, 2002, S. 10f.). Während im B-to-B vor allem Transaktionen mit großem Volumen und wenigen Nachfragern stattfinden, die meist über einen Rahmenvertrag langfristig festgelegt sind, ist der Bereich des B-to-C hingegen durch häufige Transaktionen von Gütern oder Dienstleistungen mit einer großen Anzahl von Kunden und eher kleinen Volumina gekennzeichnet (Merz, 2002, S. 24f.). Die Betrachtungen in dieser Arbeit beschränken sich auf den B-to-C-Bereich. Sie erfolgt am Beispiel der Nutzung des Internet für die Interaktion mit Endkunden in der gesetzlichen Krankenversicherung. Unternehmen der gesetzlichen Krankenversicherung sind Körperschaften öffentlichen Rechts und damit Unternehmen, die selbstverwaltet unter Aufsicht einer öffentlichen Landes- oder Bundesbehörde stehen (SBG 5, §4). Die Wertschöpfungsprozesse im E-Commerce von Krankenkassen sind nach Einschätzung des Verfassers im Kontext einer zunehmenden Wettbewerbsorientierung im Gesundheitswesen (GKV-WSG, 2007), insbesondere im Hinblick auf digitale Vertriebsformen, vergleichbar mit den E-Commerce-Anwendungen im klassischen Business-to-Consumer-Bereich, etwa bei Unternehmen der Privaten Krankenversicherung (ebenso Schulenburg & Greiner, 2000).

Der Business-to-Consumer Bereich hat zwischenzeitlich eine beachtliche Größe erreicht. Es ist dabei nicht nur die Anzahl der Internetnutzer gestiegen, sondern immer mehr Nutzer verwenden das Internet neben der Durchführung von Recherchen auch zum Einkaufen. So gaben laut einer Umfrage der AGOF (2008) aus dem Jahre 2008 62,0% der Befragten an, in den letzten 3 Monaten gelegentlich oder häufig online eingekauft zu haben, während 88,0% angaben, das Internet für Recherchen genutzt zu haben. Bei den regelmäßigen Internetnutzern gaben 73,6% der Befragten an, in den vergangenen 12 Monaten Produkte über das Internet gekauft zu haben. In Abbildung 2-1 werden die im Internet in Deutschland verkauften Waren und Dienstleistungen entsprechend ihrer Häufigkeit dargestellt.

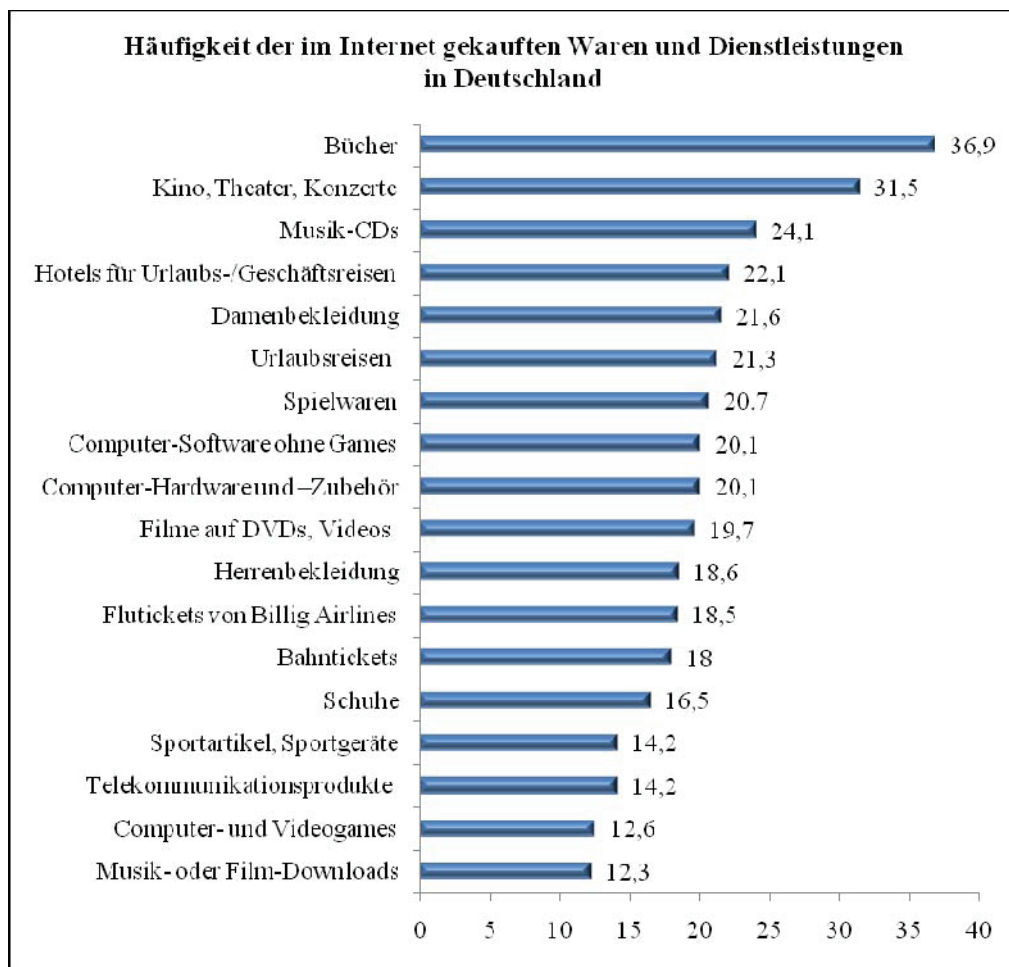


Abb. 2-1: Häufigkeit (%-Angaben) der im Internet gekauften Waren und Dienstleistungen in Deutschland (AGOF, 2008)

Anmerkung: Lesebeispiel „36,9% der deutschen Internetnutzer haben in den letzten 12 Monaten Bücher online gekauft“; Erhebungszeitraum der Internet Facts 2008-I-Studie: 01.01. bis 31.03.2008.

Betrachtet man die in der Umfrage ebenfalls erhobene Absicht, in den folgenden 12 Monaten über das Internet Produkte oder Dienstleistungen zu kaufen, so wird deutlich, dass das Internet produktübergreifend hohe Käuferpotenziale bietet (siehe Tabelle 2-2).

Tab. 2-2: Absicht, Waren und Dienstleistungen im Internet zu kaufen (AGOF, 2008)

Top-Produkte Kaufplanung der Internetnutzer			Top Produkte Kaufplanung der Nicht-Internetnutzer		
	%	Index		%	Index
1 Musik-CDs	62,3	116	1 Musik-CDs	38,4	72
2 Urlaubsreisen und auch Last-Minute-Reisen	54,9	115	2 Urlaubsreisen und auch Last-Minute-Reisen	35,1	74
3 Möbel, Gegenstände z. Wohnungseinrichtung	54,1	125	3 Hotels für Urlaubs- oder Geschäftsreisen	26,8	70
4 Filme auf DVDs, Videos	53,9	130	4 Spielwaren	24,6	69
5 Hotels für Urlaubs- oder Geschäftsreisen	45,1	117	5 Möbel, Gegenstände z. Wohnungseinrichtung	24,6	57
6 Heimwerkerbedarf, Heimwerkergeräte	44,6	121	6 Heimwerkerbedarf, Heimwerkergeräte	23,7	64
7 Spielwaren	42,0	118	7 Filme auf DVDs, Videos	20,0	48
8 Sportartikel, Sportgeräte	38,6	133	8 Fernseher mit Flachbildschirm (LCD, Plasma)	19,7	83
9 Telekommunikationsprodukte	32,2	132	9 Haushaltsgroßgeräte, wie z.B. Kühlschrank	18,2	80
10 Handytarife, Handyverträge	29,3	135	10 Sportartikel, Sportgeräte	12,5	43
11 Fernseher mit Flachbildschirm (LCD, Plasma)	25,9	110	11 Digitale Fotoapparate	11,8	59
12 Haushaltsgroßgeräte, wie z.B. Kühlschrank	25,2	111	12 Telekommunikationsprodukte	11,2	46
13 Digitale Fotoapparate	24,6	124	13 Geldanlagen, Aktien, Wertpapiere, Fonds	9,9	63
14 Gebrauchtwagen	22,9	127	14 Gebrauchtwagen	9,8	54
15 DVD-Player/-Recorder, Festplattenrecorder	20,2	127	15 Navigationssysteme	9,5	64
16 Geldanlagen, Aktien, Wertpapiere, Fonds	19,1	122	16 DVD-Player/-Recorder, Festplattenrecorder	8,5	53
17 DSL- oder and. Breitband-Internetanschluss	18,6	128	17 Handytarife, Handyverträge	8,5	39
18 Navigationssysteme	17,7	120	18 Neuwagen	8,4	77
19 Andere Versicherungen	16,8	137	19 DSL- oder and. Breitband-Internetanschluss	7,1	49
20 Mietwagen	14,7	140	20 Triple Play (Fernsehen, Telefon, Internet)	6,9	59

Quelle: AGOF 2008, S. 10.

Es wird deutlich, dass Internetnutzer sich generell durch eine breite Kaufplanung über unterschiedliche Gruppen von Waren und Dienstleistungen auszeichnen. So plant mehr als die Hälfte der befragten Personen den Kauf von Musik-CDs, Urlaubsreisen zu buchen oder Möbel bzw. Gegenstände zur Wohnungseinrichtung zu erwerben. Die breite Produktpalette bei der Kaufplanung und das hohe Kaufinteresse machen Internetnutzer zu einer attraktiven Zielgruppe für Online-Anbieter.

Andererseits gibt es auch Produkte und Dienstleistungen, zu denen sich zwar ein beachtlicher Teil der Internetnutzer online informiert, diese aber nur begrenzt online erwirbt. Zu diesen Produktgruppen gehören beispielsweise Kredite, Lebens- und Rentenversicherungen zur privaten Altersvorsorge sowie Krankenversicherungen. Hierbei handelt es sich um Produkte, die in der Regel langfristige Investitionen bedingen und den Internetnutzern in der Folge vor einem Kauf häufig ein persönliches Beratungsgespräch erforderlich scheint. So gaben in der zitierten Untersuchung (AGOF, 2008) 19,7% der Internetnutzer an, sich im Internet für Krankenversicherungen zu interessieren und 9% planten, in den nächsten 12 Monaten einen Krankenversicherungstarif online zu kaufen. Rückblickend gaben allerdings nur knapp 3% an, in den letzten 12 Monaten tatsächlich einen Krankenversicherungstarif im Internet abgeschlossen zu haben.

Für Werbetreibende dieser Produktkategorien kann dies zur Folge haben, dass das Internet nur begrenzt einen direkten Vertriebskanal darstellt. Vielmehr kann das Internet als Plattform bei der Kaufvorbereitung genutzt werden, indem durch eine Online-Präsenz Grundlagen für spätere Offline-Käufe geschaffen werden.

2.1.3 Vor- und Nachteile des Electronic Commerce aus Kundensicht

Die Vorteile des E-Commerce gegenüber dem klassischen Offline-Kauf liegen aus Kundensicht nach Wirtz und Vogt (2001, S. 117) in geringeren Informations- und Transaktionskosten sowie einer gesteigerten Transparenz, z.B. in der Möglichkeit, Preisvergleiche vornehmen zu können (ebns. Su, 2008). So bewirken beim Online-Shopping deutlich geringere Informationskosten eine höhere Preistransparenz auf Seiten der Konsumenten. Dies legt den Schluss nahe, dass vor allem preisorientierte Konsumenten das Internet nutzen, um ihr Bedürfnis nach preisgünstigen Einkäufen zu befriedigen (Lingenfelder, 2001, S. 377). Swoboda (1999) bezeichnet das Streben der Kunden nach einer möglichst bequemen Art des Einkaufens und Gebrauchens von Produkten mit dem Begriff Convenienceorientierung. Die Absicht, im Internet einzukaufen, müsste entsprechend umso größer sein, je stärker die Convenienceorientierung eines Kunden ist (vgl. Lingenfelder, 2001, S. 377). Dies belegt auch eine Studie von Li, Kuo und Russell (1999). Internet-Vielkäufer zeigten eine signifikant höhere Convenienceorientierung als Nicht- bzw. Wenig-Käufer.

Im Rahmen einer repräsentativen Erhebung der Postbank wurden im Zeitraum von Juli bis August 2004 die Konsumenteneinstellungen zum Online-Handel in Deutschland analysiert. Basis der repräsentativen telefonischen Befragung waren 1.020 Privatpersonen in Deutschland, die Online-Shopping regelmäßig nutzen (Deutsche Postbank AG, 2004; vgl. auch Emnid, 2000). Die Befragten beurteilten die Vorteilhaftigkeit verschiedener Merkmale des Online-Einkaufs. Die Vorteilserwartungen lagen im Bereich von transparenten und günstigeren Preisen, Zeitersparnissen und Angebotsverbesserungen (Umfang, Verfügbarkeit und Bequemlichkeit). Geordnet nach dem Anteil der Zustimmung begrüßten die Befragten insbesondere die Zeitersparnis (66,9%), die größere Auswahl (66,5%), fehlende Ladenschlusszeiten (65,7%), den Zugriff auf spezielle Produkte (65,4%), höhere Preistransparenz (64,6%) sowie geringere Kosten des Einkaufsprozesses (64,4%). Wichtig für die Einordnung dieser Angaben ist, dass es sich hierbei nicht um erlebte Vorteile, sondern um erwartete Vorteile handelt.

Den erwarteten Vorteilen aus Kundensicht stehen Nachteile gegenüber, die sich weitgehend auf erwartete Risikosituationen beziehen. Wie jede andere Form des Distanzhandels auch, beinhaltet der Online-Handel größere Risiken als der Kauf im stationären Einzelhandel. Neben dem funktionalen Risiko, dass das Produkt nicht die gewünschten Eigenschaften aufweist oder nur bedingt funktionstüchtig ist, spielen insbesondere das Übertragungs- und das Datenrisiko beim E-Commerce aus Kundensicht eine besonders große Rolle (Lingenfelder, 2001, S. 379 f.) Unter Übertragungsrisiko wird das Risiko verstanden, dass bei der Bestell- und Zahlungsabwicklung persönliche Daten aufgrund von Sicherheitsmängeln an unberechtigte Dritte gelangen können, unter Datenrisiko, dass Online-Anbieter sensible Kundendaten erfassen, speichern und ggf. an andere Unternehmen veräußern. Aufgrund der hohen Bedeutung für den Einkauf im Internet sind deshalb Sicherheitsbedenken von Seiten

der Internet-Nutzer an erster Stelle zu nennen. So stimmen 39,7 % der Internet-Nutzer der Aussage „Der Zahlungsverkehr über das Internet ist noch zu unsicher“ voll und ganz zu (Postbank, 2004). Drei Viertel aller Online-Shopper haben nach wie vor Bedenken beim Einkauf im Internet. Dies gilt insbesondere bei unbekanntem Online-Shops sowie bei unbekanntem Bezahlverfahren. Die Nutzer orientieren sich aufgrund der eigenen Unsicherheit eher an bekannten und vertrauten Marken.

Internet-Anbieter müssen also weiter daran arbeiten, die Nutzer von der Sicherheit ihrer Website zu überzeugen. Nach wie vor bestehen diesbezüglich Probleme, die es insbesondere seitens der Anbieter zu lösen gilt (vgl. Bliemel & Fassott, 2000, S. 19). Dies wird auch durch die Ergebnisse der zitierten Untersuchung der Postbank (2004) bestätigt. So wurden von den Befragten neben dem fehlenden Einkaufserlebnis insbesondere die nachfolgenden Aspekte, geordnet nach dem Anteil der Nennungen, als erwartete Nachteile angeführt: Gefahr des Datenmissbrauchs (71%), geringes Vertrauen in unbekannte Online-Anbieter (67%), Unsicherheit des Zahlungsverkehrs (63%) sowie mangelndes Vertrauen in die Zuverlässigkeit des Online-Anbieters in Bezug auf die Einhaltung der versprochenen Leistungen (48%). Angesichts dieser Ergebnisse ist davon auszugehen, dass Konsumenten um so eher dazu neigen, im Internet einzukaufen, je geringer ihrerseits das wahrgenommene Risiko bezüglich des Online-Kaufes ist. Dholakia und Dholakia (2001, S. 429) konnten diesen Zusammenhang in einer Studie bestätigen. So wiesen erfahrene Online-Käufer systematisch geringere Bedenken gegenüber dem Internet im Hinblick auf Sicherheit, Kriminalität und Missbrauch der Privatsphäre auf als Nicht-Käufer. Die Autoren schließen daraus, dass zunehmende Erfahrung im Online-Shopping das wahrgenommene Risiko abbaut. Es ist jedoch ebenso denkbar, dass Personen, die weniger Bedenken haben, eher zu Online-Käufern werden als Personen, die starke Bedenken gegenüber dem Einkauf in diesem Medium haben.

2.2 Dienstleistungen im Electronic Commerce

Der Dienstleistungssektor hat in den vergangenen Jahrzehnten im Vergleich zu den anderen Wirtschaftssektoren eine dominante Entwicklung bezüglich der Bruttowertschöpfung und Beschäftigung erfahren (u.a. Meffert & Bruhn, 2006, S. 20ff.) Von zentraler Bedeutung für diese Entwicklung ist der technologische Fortschritt. Analog zum industriellen und landwirtschaftlichen Sektor hat dieser auch den Dienstleistungssektor in den letzten Jahrzehnten in hohem Maße beeinflusst (Lovelock, 2001, S. 26f.; Zeithaml & Bitner, 2003, S. 14ff.). So stellt der Technologieeinsatz in vielen klassischen Dienstleistungsbranchen einen zentralen Wettbewerbsvorteil dar. Insbesondere am Beispiel der Bankdienstleistungen zeigen sich die prägenden Einflüsse von Informations- und Kommunikationstechnologien (Batiz-Lazo & Wood, 2003; Grofmann, Schäfers & Viktorin, 1999). Der Finanzdienstleistungssektor konnte durch diese Technologien sowie die damit verbundenen Effektivitäts- und Effizienzvorteile bestehende Leistungen verbessern und neue entwickeln. Das Internet, selbst ein Resultat der

rasanten Entwicklung von Informations- und Kommunikationstechnologien, trägt zu dieser Entwicklung bei.

Im nachfolgenden Teil der Arbeit wird auf Dienstleistungen im E-Commerce eingegangen. Um das Forschungsfeld einordnen zu können, werden zunächst Abgrenzungsmerkmale in der klassischen Dienstleistungsforschung allgemein charakterisiert und hinsichtlich ihrer Übertragbarkeit auf Dienstleistungen im Internet diskutiert (Abschnitt 2.2.1). Anschließend werden Ansätze zur Definition von Dienstleistungen im Internet dargestellt (Abschnitt 2.2.2). Dem folgt in Abschnitt 2.2.3 eine Typologisierung von Dienstleistungsarten im Internet. In den Abschnitten 2.2.4 und 2.2.5 werden abschließend Entwicklungen des E-Commerce im Gesundheitswesen im Allgemeinen sowie im Bereich der Krankenversicherungen im Besonderen dargestellt.

2.2.1 Konstitutiva und Abgrenzungsmerkmale in der Dienstleistungsforschung

Zunächst werden in Abschnitt 2.2.1.1 Ansätze aus der Dienstleistungsforschung zur Abgrenzung zwischen klassischen Dienstleistungen und Sachgütern dargestellt. Daran anschließend wird in Abschnitt 2.2.1.2 die Übertragbarkeit der Ansätze zur Abgrenzung aus der klassischen Dienstleistungsforschung auf den Internetkontext erörtert.

2.2.1.1 Klassische Dienstleistungsforschung

Dienstleistungen waren in den letzten Jahrzehnten verstärkt Gegenstand von Forschungsarbeiten (u.a. Hilke, 1989; Lovelock, 2001; Meffert & Bruhn, 2003; Zeithaml & Bitner, 2003). Dennoch gibt es kein einheitliches Verständnis des Dienstleistungsbegriffs, vielmehr herrschen unterschiedliche Definitionen vor (eine Übersicht gibt Rück, 2000, S. 4ff.). Die Problematik, den Dienstleistungsbegriff zu fassen, ist unter anderem in der Heterogenität des Forschungsfeldes begründet. So sind beispielsweise die Produktion und das Angebot von Sachgütern und Dienstleistungen oftmals miteinander verbunden, so dass eine klare Abgrenzung teilweise nicht möglich ist (Maleri, 1997, S. 43). Dieser Sachverhalt wird von Hilke (1989, S. 7f.) anhand des Marketing-Verbundkastens aufgezeigt, der das Kontinuum zwischen Gütern mit überwiegend hohem Sachleistungsanteil bei zugleich geringem Dienstleistungsanteil und solchen mit ausschließlichem Dienstleistungsanteil darstellt.

In der Literatur wird die Abgrenzung von Dienstleistungen zu Sachgütern häufig über das Herausarbeiten von das Wesen der Dienstleistung bestimmenden Konstitutiva versucht (u.a. Meffert & Bruhn, 2000, S. 27ff.). Dabei werden unterschiedlich viele Kriterien zur Abgrenzung heran gezogen. Neben den beiden Kriterien der Intangibilität und der Integrativität des externen Faktors, auf die oft zurück gegriffen wird und die auch in der Definition von Hentschel enthalten sind, werden auch andere Besonderheiten von Dienstleistungen diskutiert (u.a. Grönroos, 1982; Rück, 2000, S. 228ff.; Zeithaml, 1981). Nachfolgend werden die beiden ersten Kernkonstitutiva, Immaterialität und Integrativität, vorgestellt, während auf die anderen Merkmale nur kurz eingegangen werden soll. Immaterialität und Integrativität von

Dienstleistungen orientieren sich an der Dichotomie der Prozess- und Ergebnisorientierung von Dienstleistungen (Rück, 2000, S. 179ff.). Die *Immaterialität von Dienstleistungen* (auch Intangibilität) wird häufig im Hinblick auf das Ergebnis des Leistungserstellungsprozesses als dienstleistungsimmanentes Konstitutivum in Abgrenzung zu materiellen Sachgütern angeführt (Hilke, 1989, S. 26ff.). Unter Immaterialität wird dabei etwas nicht physisch Präsenes verstanden. Der Konsument erlangt kein Eigentum an tangiblen Ergebnisbestandteilen, zumal sich der Nutzen der Dienstleistung vielmehr am Ergebnis eines Prozesses manifestiert und nicht an tangiblen Teilen (Lovelock, 2001, S. 9f.).

Hentschel (1992, S. 28f.) weist auf zwei Ebenen der Immaterialität hin. Neben einer physischen Immaterialität benennt er zum anderen eine intellektuelle Immaterialität. Diese bezieht sich auf vom Kunden nur schwer zu erfassende oder nachzuvollziehende Elemente der Leistung. Aufgrund der Intangibilität kann die Dienstleistung nicht sensual erfasst werden, sondern es kommt vielmehr zur Wahrnehmung des Prozessergebnisses, ihrer Performance (Gabbot & Hogg, 1998, S. 279).

Durch die mit der Intangibilität einhergehende Schwierigkeit, die Qualität der Leistung vor ihrer Erstellung zu beurteilen, verstärkt sich das kundenseitig wahrgenommene Risiko und entsprechend die Unsicherheit des Kunden. Des Weiteren hat die Art der Leistungserstellung bei Dienstleistungen unter Einbezug des Kunden in den Erstellungsprozess eine Variabilität des Ergebnisses, vor allem in qualitativer Hinsicht, zur Folge. Dies birgt einerseits die Möglichkeit zur Individualisierung, andererseits jedoch auch die Problematik, einen einheitlichen Qualitätsstandard zu erbringen. Als weiteres dienstleistungsimmanentes Konstitutivum wird die *Integrativität eines externen Faktors* in den Dienstleistungsprozess herangezogen, die sich als prozessbezogen darstellt. Die Erstellung von Dienstleistungen wird bestimmt durch temporale, tätigkeitsbezogene Faktoren, welche der Befriedigung von Konsumentenbedürfnissen dienen (Bruhn, 2000, S. 24; Meffert & Bruhn, 1997, S. 82). Dienstleistungen sind also Prozesse, in die der Konsument als externer Faktor einbezogen wird (u.a. Engelhardt & Freiling, 1995; Maleri, 2001, S. 132ff.). Er bringt dabei Informationen, Objekte, Rechte und ggfs. auch sich selbst als Produktionsfaktor in den Leistungserstellungsprozess mit ein. Diese sogenannten kundenseitigen Faktoren, die als Verrichtungsobjekt oder auch subjekt nicht vollständig im Verfügungsbereich des Anbieters stehen, dienen der Konkretisierung der individuellen Bedarfslage des Konsumenten (Engelhardt & Freiling, 1995; Stauss, 1996). Über die Spezifizierung des Leistungsziels nehmen sie auf den Leistungserstellungsprozess Einfluss. Entscheidend ist, dass der externe Faktor als nicht vom Anbieter disponierbarer Faktor in den Erstellungsprozess mit eingeht und zu einem Produktionsfaktor wird, der im Verlauf des Prozesses eine Veränderung erfährt (Hilke, 1989, S. 12). Eine einmal erstellte Dienstleistung kann somit nicht mehr an Dritte weitergegeben werden, es entfällt dadurch in der Regel die *Drittverwendungsfähigkeit* (Maleri, 2001, S. 140).

Oftmals werden neben den beiden vorgestellten Konstitutiva Immaterialität und Integrativität weitere Besonderheiten von Dienstleistungen benannt, die teilweise in engem Zusammenhang mit den beiden Kernkonstitutiva stehen (u.a. Grönroos, 1982; Lovelock, 2001, S. 8ff.; Zeithaml & Bitner, 2003, S. 20ff.). Eine weitere den Dienstleistungen inhärente Besonderheit ist die Gleichzeitigkeit von Produktion und Konsum, das *Uno-actu-Prinzip*, das sich aus der Integrativität des Kunden in den Leistungserstellungsprozess und der Immaterialität der Leistung ergibt (Grönroos, 1982). Damit eng verbunden sind die Nicht-Transportfähigkeit und Nicht-Lagerfähigkeit von Dienstleistungen.

2.2.1.2 Dienstleistungsforschung im Internet

Die *Immaterialität*, wie in Abschnitt 2.2.1.1 dargestellt, stellt in Bezug auf digitale Dienstleistungen im Internet aufgrund dessen Eigenschaften ein schwaches Abgrenzungsmerkmal gegenüber digitalen Gütern dar. Vielmehr ist Immaterialität sowohl als Konstitutivum digitaler Sachgüter als auch digitaler Dienstleistungen anzusehen (Luxem, 2000, S. 16). Brännback und Puhakainen (1998) hingegen gehen bei elektronischen Dienstleistungen von Tangibilität aus, da diese „can be displayed, stored and communicated.“ Außerdem, so führen die Autoren an, kann der Kunde „get a feel for the service by navigating through it“. Hierbei setzen die Autoren jedoch Website und Leistung gleich. Der Aussage, dass Websites heruntergeladen und abgespeichert werden können, ist zwar zuzustimmen, allerdings stellt diese nicht den Schwerpunkt der Leistung dar. Dieser liegt vielmehr in den hinter der Website liegenden Prozessen und damit verknüpften Funktionalitäten und lässt sich nicht speichern.

Die *Integrativität* stellt, wie bereits dargestellt, ein zentrales prozessbezogenes Konstitutivum von Dienstleistungen dar. Bei Dienstleistungen im Internet lässt sich diese Eigenschaft ebenfalls aufzeigen, wobei sich die Art der Kundenbeteiligung verändert. Wegmann (2002, S. 248) charakterisiert mit Integrativität von E-Services das Ausmaß der Leistungsbestandteile, die in Form eines elektronischen Informationsaustausches abgebildet werden und bezeichnet Integrativität entsprechend als „Informationalisierungsgrad“.

Im Gegensatz zu vielen Dienstleistungen in der offline-Welt kommt es im Internet nicht zu einer direkten *räumlichen Integration* des externen Faktors im Dienstleistungserstellungsprozess. Wenngleich das Internet auch ubiquitär ist und die Dienstleistungserstellung damit von physischen Orten unabhängig, so kann dennoch nicht auf einen Wegfall der räumlichen Ebene geschlossen werden (Spence, 1999). Vielmehr verlagert sich diese in den virtuellen Raum des Internet und der Konsument muss diesen Ort bewusst aufsuchen, um die Dienstleistung in Anspruch nehmen zu können (Venkatesh, 1998). Vergleichbar schreibt Flemming (1998, S. 1): „knowingly or unknowingly, we perceive the web as a space“ und zeigt dies anhand von sprachlichen Indikatoren wie Site Maps oder „online gehen“ auf.

Auch die *zeitliche Integration* erfährt im Internetkontext eine besondere Ausprägung. Aufgrund der Möglichkeiten im Internet, vorproduzierte Prozesse und Inhalte bei der Dienstleistungserstellung zu nutzen ist das uno-actu-Prinzip, die Gleichzeitigkeit von Produktion und Konsum, nicht durchgängig aufrechtzuerhalten. Andererseits erfolgt die Erstellung der eigentlichen Leistung aufgrund der kundenseitigen Endkombination digital bereitgestellter Nutzenpotenziale nach wie vor simultan. So ergibt sich zwar aus Anbietersicht eine asynchrone Leistungserstellung, aus Sicht des Kunden allerdings findet eine Echtzeitinteraktion mit anbieterseitig zur Verfügung gestellten Potenzialen statt. Analog zum Konsum traditioneller Dienstleistungen bedingt also auch der Konsum einer Dienstleistung im Internet das temporäre Zusammentreffen von Kunde und technischen Ressourcen des Internetanbieters (Isoniemi & Snellman, 2000, S. 2).

Die *Nicht-Transportfähigkeit* und die *Nicht-Lagerfähigkeit* von Dienstleistungen im Internet stellen sich aufgrund des oben angeführten Umstandes, dass der Ort der Leistungserstellung eine Neuinterpretation erfahren muss und der Anteil vorproduzierter Leistungsinhalte groß und leicht zu vervielfältigen ist, anders dar als in der offline-Welt.

Zusammenfassend kann festgestellt werden, dass die Kernkonstitutiva Immaterialität und Integration des externen Faktors nicht ausreichend zur Charakterisierung von Dienstleistungen im Internet und ihrer Abgrenzung zu digitalen Sachgütern geeignet sind. Dienstleistungen im Internet zeichnen sich zwar durch Digitalität und damit Immaterialität aus, dies ergibt jedoch keine ausreichende Abgrenzung gegenüber digitalen Sachgütern.

2.2.2 Definition von Dienstleistungen im Internet

Auf Basis der dargestellten Konstitutiva von Dienstleistungen im Internet soll nun eine Definition der Dienstleistung im Internet erfolgen. An Stelle der im klassischen Dienstleistungskontext dominierenden Kernkonstitutiva Immaterialität und Integrativität stehen nun die Online-Darbietung und die im Zuge der externen Faktorintegration bedingte Interaktivität, die sich als Spezialfall der ursprünglichen Merkmale fassen lassen.

In der Literatur finden sich unterschiedliche Definitionsansätze für Dienstleistungen im Internet bzw. Electronic-Services (Bruhn, 2002, S. 6ff.). So bezeichnen manche Autoren alle digitalen Angebote als Dienstleistungen. Bei Grönroos, Heinonen, Isoniemi und Lindholm (1999, S. 4) heißt es beispielsweise: "the offering of any physical good or service over the internet is a service". Auch Mittal und Sawhney (2001) untersuchen digitale Sachgüter und Dienstleistungen gemeinsam als so genannte Electronic Information Products or Services (EIPS), die als "a product or service accessed through an electronic user interface and contains significant information content" gelten. Ähnlich beschreibt Alpar (1999, S. 273) das Angebot einer website als "a website to be a service that provides a user with information and communication".

Definitionsansätze zu Electronic-Services finden sich auch bei Anbietern von IT-Lösungen. Die Firma Hewlett-Packard stellte auf ihren Web-Seiten eine umfassende Betrachtung von E-Services bereit (Hewlett-Packard, 2000 nach Breithaupt, 2005, S. 109). E-Service werden als "Services, Business processes, Applications and Information technology resources available via the net" gefasst. "An electronic service available via the net that completes tasks, solves problems, or conducts transactions. E-Services can be used by people, business, and other e-services and can be accessed via a wide range of information appliances". Analog definieren Picinelli und Stammers (2000, S. 1) E-Services als „any asset that is made available via the Internet to drive new revenue streams or create new efficiencies“, wobei diese sich modular darstellen und kombiniert werden können. Die Autoren grenzen E-Services von Web-Services ("self-contained, modular applications that can be described, published, located, and invoked over a network, generally, the Web") und E-Processes ("any business process sustaining the operational aspects of an Internet centred business model") ab.

Manche Autoren beziehen explizit verschiedene Technologien mit in die Definition ein. So versteht Hennig-Thurau (1998, S. 381f.) unter Internet Services "sämtliche Formen der Informationsübermittlung an Konsumenten (...), die unter Einsatz eines Internet-Browsers erfolgen". Boyer, Halowell und Roth (2002) beziehen die Interaktivität der Dienstleistung mit ein; sie definieren E-Services als "all interactive services that are delivered on the Internet using advanced telecommunications, information, and multimedia technologies."

Andere Autoren beziehen explizit die Integration des externen Faktors in die Definition mit ein; so bezeichnet Wegmann (2002, S. 247) E-Services als Dienstleistungen, "bei denen die Leistungserstellung elektronisch, d.h., computergestützt, erfolgt. Dies umfasst auch, dass die notwendige Integration des Nachfragers in den Leistungserstellungsprozess über elektronische Medien erfolgt".

Für diese Arbeit soll die von Breithaupt (2002, S.185) entwickelte Definition gelten, die bewusst aus Kundensicht formuliert ist. Er führt den Begriff "Dienstleistungen im Internet" ein als "an einem bestimmten virtuellen Ort im WWW von einem Anbieter bereitgestellte digitalisierte Leistungspotenziale, welche der Kunde in teilweise oder vollkommen selbständig zu durchlaufenden Prozessen temporär bewusst zur Bedürfnisbefriedigung nutzt. Im Rahmen des Prozesses bringt der Kunde interaktiv Informationen in den Prozess mit ein, um als Ergebnis ihm zugute kommende Entscheidungen, Speicherleistungen, Selektionen oder Ähnliches zu erlangen".

2.2.3 Typologisierung von Dienstleistungsarten im Internet

Im folgenden Kapitel sollen, aufbauend auf den bisherigen allgemeinen Ausführungen zu Dienstleistungen, Typologisierungen der Arten von Dienstleistungen im Internet dargestellt werden. Nach einer inhaltlichen Abgrenzung (Abschnitt 2.2.3.1) wird in Abschnitt 2.2.3.2. als Basis für die Typologisierung das 4 C-Net-Modell nach Wirtz dargestellt, das einen ersten

Überblick über verschiedene Typen von Dienstleistungen im Internet gibt. In Abschnitt 2.2.3.3 werden daran anschließend Möglichkeiten zur Typologisierung von Dienstleistungen im Internet aufgezeigt und eine Fokussierung für die vorliegende Arbeit vorgenommen.

2.2.3.1 Vorüberlegungen und inhaltliche Fokussierung

Häufig begründen sich Dienstleistungen im Internet auf Anpassung und Übertragung bestehender Dienstleistungen aus der offline-Welt (u.a. Fitzsimmons & Fitzsimmons, 2000, S. 241f.). Teilweise lassen sich Dienstleistungen dabei komplett online erstellen, häufig werden jedoch nur Ausschnitte des Gesamtprozesses über das Internet abgewickelt (u.a. Meyer & Pfeiffer, 1998, S. 301). Ein ähnlicher Sachverhalt liegt der Unterscheidung in Kernleistungen und produkt- bzw. leistungsbegleitende Dienstleistungen zugrunde. Wirtz und Olderog (2002, S. 516ff.) unterscheiden diesbezüglich im Internet Value-added-Services, als begleitende Dienstleistungen des eigentlichen Kernproduktes bzw. der Kernleistung, und Standalone-Services, die unabhängig von anderen Sachgütern und Dienstleistungen angeboten werden. Darüber hinaus wurden auch völlig neue internetbasierte Dienstleistungsgeschäftsmodelle entwickelt. Wirtz (2001, S. 217) bezeichnet die Anbieter solcher Leistungen als "pure player". Im Rahmen dieser Arbeit sollen in erster Linie Kernleistungen, auf das Internet übertragene Dienstleistungen der offline-Welt, und begleitende Dienstleistungen im Mittelpunkt stehen.

Wie bereits in Abschnitt 2.1.3 dargestellt beschränken sich die Betrachtungen in dieser Arbeit auf den B-to-C-Bereich. Wirtz (2001, S. 39ff.) unterscheidet in seiner Typologie der Aktivitäten des Electronic Business für den B-to-C-Bereich vier Formen, die im nachfolgenden Abschnitt erörtert werden.

2.2.3.2 Das 4C-Net-Modell nach Wirtz im Dienstleistungskontext

Wirtz (2001, S. 39ff.) unterscheidet in seiner Typologie der Aktivitäten des Electronic Business für den B-to-C-Bereich folgende vier Formen:

Electronic Commerce ("die elektronische Unterstützung von Aktivitäten, die in direktem Zusammenhang mit dem Kauf und Verkauf von Gütern und Dienstleistungen via elektronischer Netze in Verbindung stehen"),

Electronic Communication ("die entgeltliche und unentgeltliche Bereitstellung und Nutzung netzwerkbasierter und elektronischer Kommunikationsplattformen),

Electronic Education ("die Transferierung von Aus- und Weiterbildungsleistungen an Dritte mittels elektronischer Netzwerke"), sowie

Electronic Information/Entertainment ("die Bereitstellung von informierenden und/oder unterhaltenden Inhalten und Konzepten für Dritte mittels elektronischer Netze").

Auf diesen vier grundlegend verschiedenen Aktivitäten aufbauend entwickelt Wirtz das 4C-Net-Modell für den B-to-C-Bereich (Wirtz, 2001, S. 218ff.). Das Modell unterscheidet für das

Internet die vier Basisgeschäftsmodelle Content (1), Commerce (2), Context (3) und Connection (4), die nachfolgend dargestellt werden.

(1) Content

Technische Entwicklungen und damit einhergehend immer schnellere Telekommunikationsverbindungen sowie ein hoher Standard in der Software ermöglichen bei einem wachsenden Umfang digitalisiert zur Verfügung stehender Informationen zunehmend mehr Funktionalitäten in Bezug auf den Content. So nehmen die Möglichkeiten der Darstellung, Selektion und Verarbeitung von häufig datenbankbasierten Informationen im Internet kontinuierlich zu. Waren zunächst überwiegend Anbieter online dargebotener Sachgüter im Internet (beispielsweise Zeitungen, online-Publishing, online-Datenbanken) vertreten, nahm der Anteil von Dienstleistungsanbietern mit echt interaktiven Angeboten kontinuierlich zu. Nach Wirtz (2001, S. 219ff.) findet das Geschäftsmodell, das sich daran orientiert, Kundenbedürfnisse besser und umfassender zu befriedigen, zunehmende Anwendung. Beispielhaft seien die Bereiche Electronic Education (virtuelle Universität) und Electronic Information (beispielsweise online-Zeitschriften) sowie Electronic Entertainment (online-Spiele, Filme, Musik) angeführt. Trotz aller Entwicklungen hin zu einer verstärkten Interaktivität der Angebote sind viele contentbezogene Geschäftsmodelle aufgrund einer Dominanz informationsbasierter Inhalte durch einen begrenzten Grad der Interaktivität gekennzeichnet.

Ein Bereich, der derzeit noch überwiegend im Content-Bereich steht, ist der Gesundheitsbereich. Ärzte, Krankenhäuser, Krankenkassen und auch Gesundheitsportale stellen vorwiegend Informationen bei begrenztem Interaktivitätsgrad zur Verfügung. Eine positive Entwicklung konnte in einer Untersuchung der TNS Emnid / Infratest aus dem Jahre 2005 für den Bereich der Krankenkassen festgestellt werden (TNS Infratest, 2005). So gaben 40% der Online-Nutzer von Krankenkassenwebsites als Grund für den Besuch der Website die Nutzung des geschützten persönlichen Bereiches an. Als Hauptmotive für die Nutzung wurden neben dem Abruf von online-Formularen, die Inanspruchnahme persönlicher Services sowie die Suche nach aktuellen individualisierten Informationen benannt.

Aufgrund des großen Nutzerpotenzials, einer förderlichen Gesetzgebung und der weiteren zu erwartenden technischen Entwicklungen, insbesondere im Bereich der Sicherheitsmechanismen, so beispielsweise bei der virtuellen Patientenakte, ist allerdings zu erwarten, dass der Gesundheitsbereich zunehmend Commerce-Charakter entwickeln wird.

(2) Commerce

Ursprünglich haben Handelsunternehmen das Internet schwerpunktmäßig als neuen Vertriebskanal genutzt und aufgebaut. So waren Versandhäuser, die über ähnliche Erfahrungen aus Katalog-Lösungen und die erforderlichen Infrastrukturen verfügten, bereits früh als Händler im Internet vertreten. Ebenso entwickelten sich "real player", bspw. der OnlineBuchhändler Amazon, die sich teilweise ohne auf vergleichbare Erfahrungen aus dem offline-Handel

zurückgreifen zu können, mit innovativen elektronischen Vertriebsprozessen dauerhaft haben etablieren können. Andererseits konnten Vertriebsformen der klassischen offline-Welt, die sich mit Hilfe von Internettechnologien effizienter und umfassender anbieten lassen, wie beispielsweise Auktionen, im Kontext des E-Commerce an Bedeutung gewinnen (Bapna, Goes & Gupta, 2001; Klein, 2000; Lührig & Dholakia, 2002; Roth & Pastowski, 2002). So wird das Potenzial handelsbezogener Dienstleistungen im Internet am Erfolg von online-Auktionshäusern und am Beispiel von Börsen besonders deutlich.

Dienstleistungen eignen sich unterschiedlich gut für den Vertrieb über das Internet. Nach Meffert und Bruhn (2000, S. 406) kommen insbesondere Dienstleistungen für den Vertrieb über das Internet in Frage, bei denen Anrechte wie beispielsweise Fahrkarten oder Eintrittskarten gehandelt werden. Scansaroli und Eng (1997) zeigen auf, dass sich insbesondere standardisierte durable Objekte mit wenig Anforderungen an Kundenservice für den Online-Vertrieb eignen. Digitale Informationsangebote können ebenfalls erfolgreich online vermarktet werden, weil sie unerschöpflich reproduzierbar sind, grenzenlos individualisierbar und keine Distributionskosten haben (Gogan, 1997).

Ein weiterer Bereich, dessen Internetdienstleistungen vorwiegend im Commerce-Feld angesiedelt sind, sind Anbieter von Finanzdienstleistungen (Lamberti, 2000; Linn, Krottsch & Stein, 2001). Gerade Banken, deren Dienstleistungen auf Grundlage der Bereitstellung und dem Austausch von Informationen erbracht werden, profitieren besonders von standardisierbaren Prozessen, die im Internet automatisiert und digitalisiert genutzt werden können.

Vergleichbares gilt für Versicherungsdienstleistungen im Internet. Prozesse der Information, Beratung über den Vertragsabschluss bis zur Schadensfallabwicklungen werden über das Internet angeboten und erstellt. So besteht für Krankenversicherer die Möglichkeit, das Internet im Rahmen der Kundenbetreuung und Leistungsabwicklung einzusetzen und damit schnellere und effizientere Geschäftsabläufe zu erzielen, wodurch eine höhere Kundenbindung erreicht werden kann (vgl. Lürzer, 2000). Der Verfasser der vorliegenden Arbeit teilt diese Meinung jedoch nur eingeschränkt, da gerade der Ersatz einer persönlichen Beratung durch eine technische Schnittstelle den Vertrauensaufbau erschwert und Kundenbindung damit eher beeinträchtigt.

Auch beim Vertrieb ihrer Produkte über das Internet sind den Versicherungsunternehmen Grenzen gesetzt. Insbesondere die vertriebliche Passivität des Internet bei häufig hoher Erklärungsbedürftigkeit der Versicherungsprodukte sowie der begrenzte Bedarf des einzelnen Konsumenten an Versicherungen führen dazu, dass für den Internetvertrieb von Versicherungsleistungen neue Absatzgestaltungen vorgenommen werden müssen (Lürzer, 2000). Bei ausschließlich internetgestütztem Versicherungsvertrieb ist die fehlende Möglichkeit des Nutzers zur individuellen Nachfrage ein Manko. Ebenso kann dem bei komplexeren Versicherungsprodukten häufig gegebenen Beratungsbedarf im standardisierten Internet-

vertrieb nur schwer Rechnung getragen werden, weshalb auf Call-Center oder ähnliche serviceunterstützende Funktionen nicht verzichtet werden kann (Hartwig, 2001).

Hiervon geht ein erheblicher Veränderungsdruck auf die Versicherer aus, ihre Prozesse von der Produktentwicklung bis hin zur Vertriebsstrategie auf die Erfordernisse des Internet anzupassen. Inzwischen hat sich die Anbieterlandschaft aufgrund dieser Anforderungen bereits verändert. So haben sich Direktversicherungsunternehmen wie etwa die Cosmos Direkt AG etabliert, die bei konsequenter und ausschließlicher Nutzung des Internet als Vertriebskanal sowie gleichzeitiger Auslagerung von nicht zu den Kernkompetenzen von Versicherern gehörenden Funktionen vergleichsweise kostengünstige Alternativen der Geschäftsaufnahme und -ausweitung bieten (vgl. Knospe, 2000). Im Marktsegment der Krankenzusatzversicherungen konnten sich Unternehmen wie die Quelle-Karstadt Versicherung AG oder die Envivas Krankenversicherung AG erfolgreich als Direktversicherer im Internet etablieren. Beide Unternehmen konzentrieren ihre Vertriebswege auf Kooperationen mit Unternehmen der gesetzlichen Krankenkassen indem sie deren Sortiment gegenüber dem bestehenden Versichertenpotenzial durch Krankenzusatzversicherungsangebote ergänzen. Diese Vertriebsweise ermöglicht niedrigere Prämien und zugleich verschiebt sich die Achse der Informationsasymmetrie aufgrund der Fülle an Information, insbesondere unter Nutzung spezieller Vergleichsdienste, zu Gunsten der Konsumenten.

Bezüglich der Entwicklung des Internetanteils am Vertrieb von Versicherungsleistungen gibt es unterschiedliche Prognosen. Langfrist-Prognosen kommen meist zu der Einschätzung, dass die über das Internet abgeschlossenen oder induzierten Verträge einen Marktanteil von 3-10% erreichen können. So prognostiziert Forrester research inc. (2006) für den deutschen Online-Versicherungsmarkt einen Anstieg der über das Internet abgeschlossenen Versicherungspolicen von 2,2 Mio. im Jahre 2005 um 68% auf 3,7 Mio. im Jahre 2011.

(3) Context

Hoffman, Novak und Chatterjee (1995) differenzieren zwischen destination sites (z.B. online-Shops oder Internetinformationsangebote) und web traffic control sites (bspw. Suchagenten), die den Nutzer auf destination sites lenken sollen. Anbieter von web traffic control sites nutzen das Geschäftsmodell "Context" als Basis und sind dadurch gekennzeichnet, dass sie nicht primär eigene Inhalte anbieten, sondern ihre Leistung darin besteht, die Verbindung zu anderen Web-Anbietern herzustellen. Nach Wirtz (2001, S. 242ff.) können in diesem Zusammenhang Suchmaschinen und Web-Kataloge unterschieden werden, wobei er die beiden Formen nach der Art ihrer Aufbereitung differenziert. Während Suchmaschinen allein auf der Eingabe von kundenseitigen Anfragen arbeiten, sind Web-Kataloge redaktionell aufbereitet.

Desweiteren ist aufgrund der Masse an datenbasierten Informationen, die über das Internet verfügbar sind, ein Bedarf an Strukturierung und Transparenz und damit ein neues Feld für

Dienstleister im Internet entstanden (Wirtz, 2001, S. 16). Der Großteil der Anbieter solcher Leistungen ist der Gruppe der Intermediäre zuzuordnen (Picot & Neuburger, 2002, S. 96; Weber & Zhiqiang, 2007). Diese agieren im Internet mit dem Ziel, durch Aggregation von Informationen über die zu vermittelnden Leistungen einen Match zwischen Anbietern und Nachfragenden im Internet herzustellen, aus dem Transaktionen oder Interaktionen zwischen den Gruppen resultieren. Sogenannte Portale aggregieren die Internetangebote verschiedener Anbieter und fungieren damit als Eingangsseite für potenzielle Kunden, indem sie Suchaufwand reduzieren (Schubert, 2000, S. 135ff.).

Eine weitere Anbieterform, die in diesem Zusammenhang von Relevanz ist, ist die der sogenannten Trusted Third Party (TTP) nach Ba, Whinston und Zhang (2000, S. 193ff.). Diese Anbieter können ebenfalls zur Gruppe der Intermediäre gezählt werden. Sie erbringen ihre Leistung für Anbieter im Internet, indem sie diese im Hinblick auf sicherheits- und vertrauensrelevante Aspekte bewerten und häufig auch beraten. Das Bewertungsergebnis in Form eines von der TTP vergebenen Zertifikates ist wiederum für die Kunden des zertifizierten Anbieters relevant, denen es auf Grundlage der Fremdbewertung möglich ist, Unsicherheit und wahrgenommenes Risiko zu reduzieren (Kim, 2008).

(4) Connection

Neben der Bereitstellung und Lokalisierung von Informationen ist auch die Möglichkeit des netzwerkbasierten Austauschs von Informationen im Internet von erhöhter Bedeutung. "Dabei können die herzustellenden Verbindungen sowohl technologischer, kommerzieller als auch rein kommunikativer Art sein" (Wirtz, 2001, S. 252). Wirtz unterscheidet diesbezüglich Intertion und Intra-Connection als Subgeschäftsmodelle. Unter Intertion versteht er die infrastrukturelle Bereitstellungsleistung, die Zugangsleistungen von Providern umfasst, während letzteres für Community-Angebote steht, wie Opinionportals, Customer exchange oder Customer chat sowie Mailing Services innerhalb des Internet. Communities können sowohl von Unternehmen als auch von Kunden betrieben werden und werden auch als virtuelle Gemeinschaften bezeichnet: "Virtuelle Gemeinschaften beschreiben den Zusammenschluss von Individuen oder Organisationen, die gemeinsame Werte und Interessen miteinander teilen und die über längere Zeit mittels elektronischer Medien, ort- und (teilweise auch) zeitungebunden in einem gemeinsamen semantischen Raum (gemeinsame Begriffswelt) kommunizieren" (Schubert, 2000, S. 30). In der Praxis existieren Community-Angebote in unterschiedlicher thematischer Breite. Ebenso ist der Kommunikationsprozess in unterschiedlicher Weise strukturiert und wird teilweise von einer Person oder Organisation moderiert. Die Mitglieder der Community kommunizieren auf der Basis von E-Mail oder Chats.

2.2.3.3 Dienstleistungsarten im Internet

In den vorangegangenen Abschnitten wurde deutlich, dass im Internet verschiedene Dienstleistungsarten identifiziert werden können. Die an früherer Stelle dargestellten Merkmale des Digitalisierungsgrades - das Ausmaß, in dem zentrale nutzenstiftende Funktionen der Dienstleistung digitalisierbar, d.h. rein informationstechnisch abwickelbar sind - und der echten Interaktivität können zur Typologisierung von Dienstleistungen im Internet dienen. Breithaupt (2005, S. 124ff.) wählt entsprechend eine Gegenüberstellung der beiden Dimensionen „Digitalisierungsgrad“ und „Grad der echten Interaktion“ mit den jeweils möglichen Ausprägungen gering versus hoch.

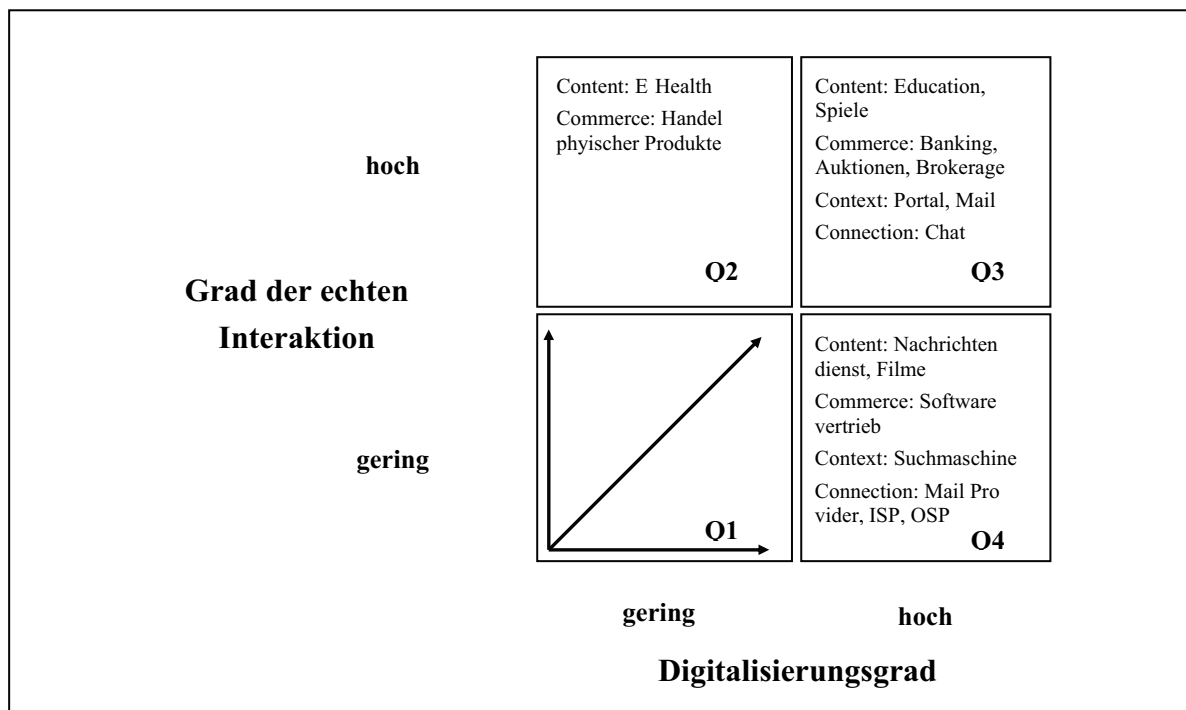


Abb. 2-2: Typologisierung von Dienstleistungen im Internet nach Digitalisierungsgrad und echter Interaktion (nach Breithaupt 2005, S. 125)

Nach Breithaupt (2005, S. 70) liegt „echte Interaktivität im Internet [liegt] vor, wenn der Kunde die Möglichkeit hat, im Hinblick auf die Leistungserstellung synchron oder asynchron Programmabläufe zu empfangen, individuell zu steuern und innerhalb einer wechselseitigen kausalen Kommunikation mit dem Anbieter Informationen reflektiv senden zu können.“

Aus Abbildung 2-2 sind die vier resultierenden Quadranten ersichtlich. Der Quadrant Q1 enthält Angebote geringen Digitalisierungsgrades und niedriger echter Interaktivität. Damit sind beide Konstitutiva von Dienstleistungen im Internet nur gering ausgeprägt. Wie bereits im vorangegangenen Abschnitt aufgezeigt, gibt es in den Bereichen Content und Commerce Dienstleistungen, die nicht komplett im Internet erstellt werden können, von denen jedoch weite Teile unter interaktiver Beteiligung des Kunden über das Internet erbracht werden (Q2). Gleichzeitig gibt es Dienstleistungen, die über einen hohen Grad an Digitalisierung verfügen,

jedoch nur über einen geringen Grad an echter Interaktivität (Q4). So sollte bspw. die Bereitstellung einer Provider-Dienstleistung nach ihrer Einrichtung nur noch eine geringe Interaktivität zwischen Kunde und Dienstleister erforderlich machen. Das Feld Q3 verbindet hohe echte Interaktivität mit einem hohen Grad der Digitalisierung. Breithaupt (2005, S.124) bezeichnet die dort angesiedelten Dienstleistungen als „ideale Dienstleistungen“ im Internet. Der vorliegende Untersuchungsgegenstand der Websites von Krankenkassen positioniert sich je nach Anwendung über alle vier Quadranten. So ist die interaktive Bearbeitung individueller krankenversicherungsbezogener Vorgänge bspw. die Verarbeitung persönlicher Daten oder die Nutzung persönlicher Services - vergleichbar mit den Nutzungsprozessen des Online-Banking in Q3 anzusiedeln. Die Nutzung als Content-Anbieter ist dagegen eher in den Bereichen Q1 und Q2 einzuordnen. Die Teilnahme an Bonusprogrammen wäre bei einem hohen Grad an Interaktion und einem eingeschränkten Grad an Digitalisierung in Q2 anzusiedeln.

Auf die Unterscheidung zwischen Such-, Erfahrungs- und Vertrauensgütern, die auf Nelson (1974) sowie Darby und Karni (1973) zurückgeht, wird auch im Kontext digitaler Güter zurückgegriffen (Choi, Stahl & Whinston, 1997, S. 138; Luxem, 2000, S. 30). Ausgangspunkt dieser Betrachtung ist eine Unterscheidung von Arten der Produktqualität nach ihren informationsökonomischen Eigenschaften in den Kategorien Such-, Erfahrungs- und Vertrauenseigenschaften (Kuß & Tomczak, 2000, S. 103ff.; Weiber & Adler, 1995, S. 61f.). Eigenschaften, die der Konsument bereits vor dem Konsum evaluieren kann, werden Sucheigenschaften genannt, während Erfahrungseigenschaften hingegen solche Merkmale bezeichnen, die vom Konsumenten erst im Laufe des Konsums beurteilt werden können. Vertrauenseigenschaften entziehen sich gänzlich der sicheren Beurteilung durch den Konsumenten. Nach Choi, Stahl und Whinston (1997, S. 138) sind digitale Produkte zumeist Erfahrungsgüter, da ihre Qualität aufgrund der Immaterialität und der Erfordernis der Integration erst nach dem Konsum beurteilt werden kann, zumal eine Nutzung häufig nur einfach erfolgt. Im Rahmen der vorliegenden Arbeit liegt der Schwerpunkt der Betrachtung auf digitalen Erfahrungsgütern. So ist die materielle Krankenversicherungsleistung zwar als Vertrauensgut einzuschätzen, die der Konsumentenbetrachtung unmittelbar zugrunde liegende eigentliche Dienstleistung einer digitalisierten und interaktiven Leistungsbereitstellung ist jedoch eher von Erfahrungseigenschaften geprägt (vgl. Kortzfleisch & Winand, 2000).

Breithaupt (2005, S. 128ff.) erweitert die oben genannten Differenzierungsmerkmale, indem er das von Aleff (2002, S.103ff.) angeführte Typologisierungskriterium des zeitlichen Charakters der Dienstleistungserstellung auf den Internetkontext überträgt und diesbezüglich eine Verbindung zur konsumentenseitigen Wahrnehmung der Dienstleistungsqualität herstellt. Als übergreifende Systematisierung dient die Dreiteilung der Dienstleistungsgenerierung nach Hilke (1989, S. 10ff.) in die Phasen Potenzial, Prozess und Ergebnis. Daraus resultieren vier entsprechende Dienstleistungstypen im Internet, wobei die Ergebnisphase zwei Dienstleistungstypen unterschiedlicher Prägung umfasst.

Als ersten Dienstleistungstyp im Internet benennt der Autor die *potenzialorientierte Dienstleistung im Internet*. Diese ist durch ihren Infrastrukturcharakter bestimmt. Im Rahmen der potenzialorientierten Dienstleistung werden schwerpunktmäßig technologiebasierte Leistungspotenziale vorgehalten, die es dem Kunden ermöglichen, das Internet und seine Dienstleistungen in Anspruch zu nehmen. Potenzialorientierte Leistungen im Internet sind häufig durch einen hohen Grad der Digitalisierung und einen niedrigen Grad echter Interaktivität gekennzeichnet (Quadrant Q4 aus Abbildung 2-2).

Prozessorientierte Dienstleistungen im Internet sind durch ihren Erlebnischarakter bestimmt. Für den Kunden steht weniger das Ergebnis der Leistungserstellung im Mittelpunkt, als vielmehr der Erstellungsprozess der Dienstleistung an sich. Dieser Dienstleistungstyp ist dadurch gekennzeichnet, dass Konsumenten Zeit "nicht als Verlust, sondern als Erlebnisraum, der mit emotional oder intellektuell befriedigenden Eindrücken erfüllt sein sollte" (Stauss, 1991) begreifen. Sie stehen damit im Gegensatz zu auf Zeitersparnis ausgerichteten Dienstleistungen und sind vorwiegend durch Erfahrungseigenschaften bestimmt.

Analog zu den Ausführungen von Aleff (2002, S. 117ff.) unterscheidet Breithaupt (2005, S. 128f.) zwei Formen *ergebnisorientierter Dienstleistungen im Internet*. *Ergebnisorientierte Dienstleistungen im Internet mit Zeitsparcharakter*, beispielsweise Suchmaschinen oder Wissensportale, sind durch einen hohen Grad der Digitalisierung und eine eher geringe Interaktivität gekennzeichnet (Quadrant Q4 aus Abbildung 2-2). Dabei steht neben dem funktionalen Ergebnis auch die Dauer der Transaktion im Fokus. Der Leistungstyp ist durch Such- und Erfahrungseigenschaften bestimmt und findet in der Regel in informaler Beziehung zwischen Kunde und Anbieter statt. Dabei kommt es zu diskreten Leistungserstellungen, die häufig einfach im Zuge der Navigation durch das WWW vom Kunden genutzt werden.

Ergebnisorientierte Dienstleistungen im Internet mit Problemlösungscharakter sind hingegen durch eine hohe Interaktivität gekennzeichnet, wobei der Grad der Digitalisierung uneinheitlich ausgeprägt ist (Quadrant Q2 bzw. Quadrant Q3 aus Abbildung 2-2). Als Beispiele benennt der Autor Anwendungen aus dem Gesundheitsbereich oder online-Schulungen. Da der Konsument im Zuge der Leistungsanspruchnahme vielfach keine Transparenz bezüglich Anzahl und Qualität alternativer Leistungsanbieter hat und auch das Leistungsergebnis vielfach intellektuell intangibel ist, sind ergebnisorientierte Dienstleistungen im Internet mit Problemlösungscharakter hauptsächlich durch Vertrauenseigenschaften bestimmt.

Im Rahmen der vorliegenden Arbeit liegt der Schwerpunkt der Betrachtung auf *ergebnisorientierten Dienstleistungen* im Internet. Der Untersuchungsgegenstand der Websites von Krankenkassen verfügt sowohl über Elemente von ergebnisorientierten Dienstleistungen mit Zeitsparcharakter als auch mit Problemlösungscharakter. So ist der allgemein zugängliche Bereich einer Krankenkassenwebsite eher von niedrigen Interaktionsmöglichkeiten und standardisierten Angeboten gekennzeichnet und damit eher als ergebnisorientierte

Dienstleistung im Internet mit Zeitsparcharakter geprägt. Der versichertenindividuelle zugangsgeschützte Bereich hingegen ist, vergleichbar mit den Anwendungen des online-Banking, durch einen hohen Grad an Interaktivität und einen hohen Individualisierungsgrad gekennzeichnet und eher von einem Problemlösungscharakter geprägt.

2.2.4 Electronic Commerce im Gesundheitswesen

Die Nutzung von Online-Medien bietet den Nachfragern von Gesundheitsleistungen vielfältige Möglichkeiten. Insbesondere erhalten sie Zugang zu einem nahezu unbegrenzten Angebot an gesundheitsspezifischen Informationen und Serviceleistungen. Auf einer Vielzahl von Websites können Informationen zu gesundheitsspezifischen sowie gesundheitsnahen Themen eingesehen werden. Im Mittelpunkt der Nutzung gesundheitsbezogener Online-Medien stehen die Suche nach Informationen über Krankheitsbilder und Behandlungsoptionen, sowie nach Eigenschaften von Medikamenten und anderen Medizinprodukten, gefolgt vom Themenkreis Prävention, insbesondere gesunder Lebenswandel durch entsprechende Ernährung und Fitness (Poensgen & Larsson, 2001). Außerdem besteht ein Bedarf an unterstützenden Informationen bei der Auswahl eines geeigneten Leistungserbringers, insbesondere bei der Auswahl eines Krankenhauses oder Arztes (Burchert, 1998). Dies wird durch die Ergebnisse der bereits zitierten ACTA-Studie aus dem Jahr 2004 bestätigt. So gaben 40% der Befragten an, regelmäßig Gesundheitsinformationen über das Internet zu beziehen (ACTA, 2004).

Bei der Nutzung gesundheitsbezogener Online-Medien gelten die Webseiten von Non-Profit-Organisationen bei den Konsumenten in der Regel als die glaubwürdigsten Informationsquellen im Gesundheitswesen. Insbesondere die Webseiten von Universitätskliniken, medizinischer Fachgesellschaften und medizinischer Fachverlage sind bei den Konsumenten anerkannt, während kommerziellen Anbietern wie Online-Portalen oder Gesundheitsinformations-Providern sowie Pharmaunternehmen am wenigsten Vertrauen geschenkt wird (Poensgen & Larsson, 2001; TNS Infratest, 2005; Zahedi & Song, 2008).

Folgt man der Typologisierung von Wirtz (2000, S. 87) für Aktivitäten des Electronic Business im B-to-C-Bereich, so sind die Online-Leistungsangebote im Gesundheitswesen vornehmlich dem Bereich der Content-Angebote zuzuordnen. Neben den alteingesessenen Akteuren, wie Krankenhäuser, Pharmaunternehmen und Versicherungen haben sich hier auch neue Akteure, wie die Portalbetreiber Gesundheitsscout24 oder Onmeda am Markt etabliert. So hat bspw. das Gesundheitsportal Onmeda als größtes deutsches Gesundheitsportal mit 1,8% Reichweite im Jahre 2006 in Deutschland Rang 56 unter den Content-Anbietern belegt (AGOF, 2007).

Context-Anbieter im Gesundheitswesen unterstützen den Nutzer bei der Suche nach spezifischen Informationen und Leistungen im Internet. Aufgrund der zunehmenden Komplexität im Gesundheitsmarkt und der Vielzahl von Angeboten im Internet ist das Ziel

eine Verbesserung der Markttransparenz sowie die Unterstützung der Konsumenten bei der Suche und Strukturierung von Gesundheitsinformationen (Evans & Wurster, 1999; Sillence, Briggs, Harris & Fishwick, 2007). Während Connectivity-Anwendungen im Gesundheitsmarkt die Netzwerkarchitekturen der digitalen Ökonomie nutzen, um Akteure im Gesundheitswesen - Versicherungsunternehmen, Leistungserbringer und Versicherte beispielsweise im Rahmen der digitalen Patientenakte zu verbinden, sind Commerce-Anwendungen digitale Vertriebsformen, die sowohl Anbietern als auch Konsumenten „ganzheitliche Bequemlichkeit“ und Kosteneinsparungen versprechen (Chain, Sarasohn-Kahn & Wayne, 2000, S. 50; Parente, 2000, S. 91).

2.2.5 Electronic Commerce und Krankenversicherungen

Die Krankenversicherungen erhalten durch das Internet und die damit verbundenen Möglichkeiten der digitalen Informations- und Kommunikationstechnologie eine Chance, sich auf Grundlage ihrer Stärken zu einer gestaltenden Größe im Gesundheitsmarkt zu entwickeln. Im Kontext einer zunehmenden Wettbewerbsorientierung im Gesundheitswesen bietet sich hier die Möglichkeit, die eigene Wettbewerbsfähigkeit und Wettbewerbsposition zu stärken. Bei der Ausgestaltung eigener Internetauftritte kommen den Krankenversicherungen ihre Kenntnisse des Gesundheitsmarktes und der Konsumentenbedürfnisse zu Gute. In der Rolle eines Navigators können sie die Konsumenten durch das Überangebot an Gesundheitsinformationen führen und in ihren Entscheidungsprozessen qualitativ unterstützen (Poensgen & Larsson, 2001, S. 7f.).

Vornehmlich agieren diese als Anbieter von gesundheitspezifischen Contents, insbesondere aus den Themenfeldern Prävention und Versorgung. Im Context-Bereich unterstützen sie ihre Versicherten bspw. beim Zugang zu Qualitätsinformationen im Themenbereich der stationären Versorgung, während digitale Vertriebsformen im Rahmen von E-Commerce-Anwendungen insbesondere in der Versichertengewinnung sowie im Vertrieb von Ergänzungs- und Wahltarifen genutzt werden (TNS Infratest, 2005).

Dabei ist zu berücksichtigen, dass die Qualität der von Krankenversicherungen angebotenen Informationsinhalte aufgrund ihrer Interessen in der Leistungs- und Ausgabensteuerung im Allgemeinen nur als mittelmäßig wahrgenommen wird. Für die Krankenversicherungen eröffnen sich Potenziale der Einflussnahme sowohl auf die Konsumenten als auch auf die Leistungserbringer, die zu Kostensenkungen führen können. So kann die Steuerung von Patienten in Disease Management Programme oder Maßnahmen der Integrierten Versorgung effizienzsteigernd wirken, indem diese auf ein gesundheitsbewusstes und therapiekonformes Verhalten abstellen (vgl. Preuß, 1997, S. 276).

2.2.6 Zwischenfazit

Die vorangegangenen Abschnitte haben gezeigt, dass Dienstleistungen im Internet trotz vieler Parallelen zur offline-Welt eigenständig charakterisiert werden können. Ausgehend von der

Beschreibung der Merkmale von Dienstleistungen der offline-Welt und digitaler Güter wurden in Abschnitt 2.2.1 Konstitutiva von Dienstleistungen im Internet dargestellt. In der Abgrenzung zu Dienstleistungen der offline-Welt wurde unter Einbeziehung der Internetspezifika Digitalität und Virtualität eine eigene Perspektive auf Dienstleistungen in der online-Welt eingenommen. Diese diente, neben der Darstellung von allgemeinen Besonderheiten des Interneteinsatzes im Dienstleistungskontext, der inhaltlichen Abgrenzung des Untersuchungsbereiches. In Abschnitt 2.2.2 wurde die im Kontext dieser Arbeit verwendete Definition von Dienstleistungen im Internet vorgestellt. Aufbauend auf den bisherigen Ausführungen wurde in Abschnitt 2.2.3 ein Modell zur Typologisierung von Dienstleistungen im Internet dargestellt und eine Einordnung des Untersuchungsgegenstandes vorgenommen, bevor in den Abschnitten 2.2.4 und 2.2.5 die Potenziale des E-Commerce im Gesundheitswesen im Allgemeinen und für den Anwendungsbereich der Krankenkassen im Besonderen beschrieben wurden.

Im nachfolgenden Abschnitt 2.3 erfolgt, verankert in der allgemeinen Konsumentenverhaltenstheorie, eine Betrachtung empirischer und theoretischer Erkenntnisse des Konsumentenverhaltens im Electronic-Commerce. Dabei erfährt der Einfluss des Involvement auf das Informations- und Entscheidungsverhalten der Konsumenten eine besondere Würdigung.

2.3 Konsumentenverhalten

Zunächst soll eine einführende Betrachtung des Konsumentenverhaltens erfolgen. Nach einem Überblick zur Thematik (siehe Abschnitt 2.3.1) werden die Faktoren des Involvement sowie Spezifika im Internetkontext betrachtet (siehe Abschnitt 2.3.2). In Abschnitt 2.3.3 wird ein Überblick über die Bestimmungsfaktoren des Informations- und Entscheidungsverhaltens von Konsumenten im Kaufprozess gegeben. Hierbei wird der Einfluss des Involvement auf das Informations- und Entscheidungsverhalten der Konsumenten näher dargestellt. Dieses steht im Zentrum der Arbeit und erfährt bei der empirischen Untersuchung eine besondere Betrachtung. In Abschnitt 2.3.4 wird der Frage nachgegangen, inwieweit die dargestellten Erkenntnisse zum Konsumentenverhalten auf Situationen im E-Commerce übertragbar sind. Dies erfolgt anhand der Darstellung empirischer und theoretischer Erkenntnisse des Konsumentenverhaltens im E-Commerce.

2.3.1 Grundlagen des Konsumentenverhaltens

Konsumentenverhalten umfasst nach der Definition von Bagozzi (1975) "Handlungen, Prozesse und soziale Beziehungen, die von Individuen, Gruppen und Organisationen beim Erwerb, Gebrauch und der Sammlung von Erfahrungen mit Produkten, Dienstleistungen und anderen Ressourcen gezeigt werden". Allgemeiner gehalten bezeichnen Kroeber-Riehl und Weinberg (1999, S. 3) das Konsumentenverhalten als „das Verhalten der Menschen beim Kauf und Konsum von wirtschaftlichen Gütern“.

Nachfolgend werden die konsumentenindividuellen Ausgangsbedingungen des Kaufverhaltens, sowie die Einflussfaktoren auf das Informations- und Entscheidungsverhalten der Konsumenten im E-Commerce beschrieben. Hierbei erfährt der Einfluss des Involvement auf den Kaufentscheidungsprozess eine besondere Betrachtung. Ein Kaufentscheidungsprozess stellt einen Spezialfall eines Informationsverarbeitungsprozess dar, wobei davon auszugehen ist, dass er für den Konsumenten selten ohne Vorwissen beginnt (Trommsdorff, 1998, S. 25). Vielmehr spielen Informationen vor, während und nach einer Kaufentscheidung für den Konsumenten eine wichtige Rolle. Häufig verfügen sie über eine Fülle an Informationen und entwickeln unter Berücksichtigung der Erfahrungen aus früheren Kaufentscheidungsprozessen und auf Grundlage tangierter Werte und Einstellungen bestimmte Strategien zur Informationsaufnahme, Wahrnehmung und zur Entscheidungsfindung. Diese individuellen Ausgangsbedingungen nehmen maßgeblich Einfluss auf das Kaufverhalten des Konsumenten.

Einen Überblick über die wesentlichen Erklärungsgrößen gibt Abbildung 2-3. Trommsdorff (1998, S. 32) unterscheidet *Zustände* als statische Erklärungsgrößen und *Prozesse* im Sinne dynamischer Erklärungsgrößen, die sich gegenseitig beeinflussen.

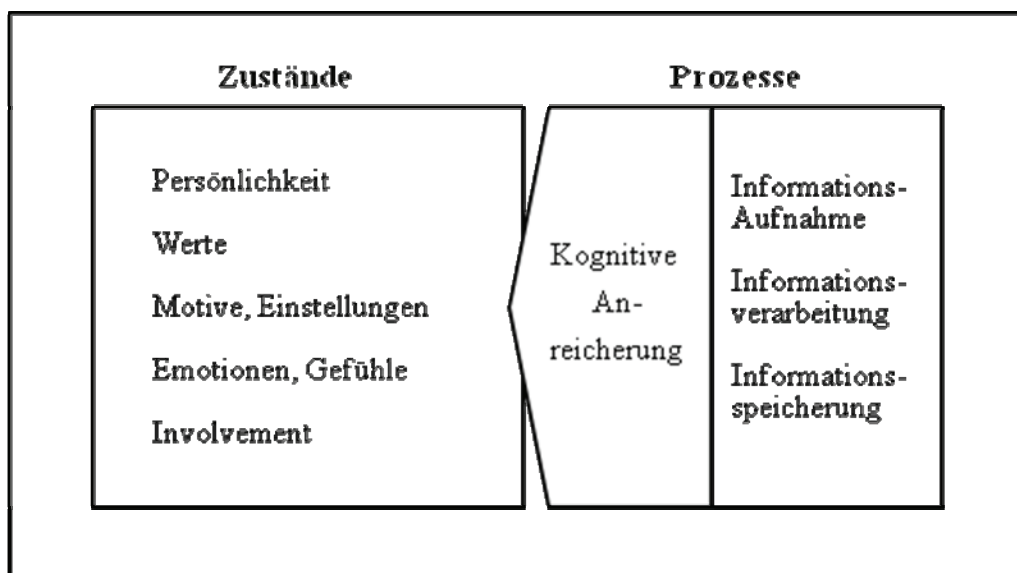


Abb. 2-3: Relevante individuelle Zustände und Prozesse des Konsumentenverhaltens (angelehnt an Trommsdorff, 1998, S. 32 und Raffee und Silberer, 1981, S. 43 ff.).

Nachfolgend werden die einzelnen Zustände kurz definiert und die Zusammenhänge zwischen den einzelnen Erklärungsgrößen aufgezeigt. In Abschnitt 2.3.3 werden die Prozesse der Informationsaufnahme, -verarbeitung und -speicherung als dynamische Erklärungsgrößen des Konsumentenverhaltens beschrieben.

- *Emotionen und Gefühle*: Emotionen bzw. Gefühle wie Freude, Lust, Geborgenheit, Liebe, Trauer, Ekel, Furcht etc. "sind innere Erregungen, die angenehm oder unangenehm empfunden und mehr oder weniger bewusst erlebt werden." (Kroeber-Riel, 1999, S. 106). Erkenntnisse aus der Emotionspsychologie werden bei der Vermittlung emotionaler Konsumerlebnisse und bei der emotionalen Produktdifferenzierung eingesetzt (Trommsdorff,

1998, S. 60-78). So wird etwa in der Werbung regelmäßig genutzt, dass Emotionen die Leistungsfähigkeit der Konsumenten erhöhen können und deren Aufnahmebereitschaft für Informationen dadurch stimulieren können. Darüber hinaus werden Emotionen in allen absatzpolitischen Instrumenten gezielt eingesetzt; angefangen bei Produktpräsentationen und beschreibungen über emotional geprägte Ladengestaltungen bis hin zur Attribuierung von Produkten mit Emotionen mit Hilfe des Design von Produkt und Verpackung sowie ihrer materiellen Ausgestaltung. Bezogen auf Webanwendungen wird der Einfluss von Emotionen und Gefühlen als Determinanten des Nutzungsverhaltens in Abschnitt 2.10.4 weitergehend dargestellt.

- *Einstellungen und Motive:* Einstellungen lassen sich definieren "als innere Haltung eines Individuums, gegenüber bestimmten Reizen relativ fest gefügte, stabile positive oder negative Reaktionen zu zeigen" (Bänsch, 1996, S. 39). Motive hingegen stellen wirksame Antriebskräfte des Konsumentenverhaltens dar, indem sie Emotionen oder Triebe einer Person mit zielorientiertem Handeln verbinden.

Da Einstellungen und Motive Verhalten direkt beeinflussen, ist deren Erfassung und Beeinflussung aus Marketingsicht von großem Interesse (Kroeber-Riel & Werner 1999, S. 211ff.). Durch die Clusterung homogener Einstellungen von Konsumenten lassen sich Marktsegmente abgrenzen, die in der Markenkommunikation gezielt angesprochen werden können. Aufgabe des Marketings ist es, entsprechende Marktsegmente zu identifizieren und das eigene Produkt bei der Produktpositionierung mit korrespondierenden Eigenschaften zu belegen (Trommsdorff, 1998, S. 159f.).

- *Werte:* "Ein Wert ist ein konsistentes System von Einstellungen mit normativer Verbindlichkeit ... ein Wert ist der Zustand der Bereitschaft, sich Einstellungsobjekten gegenüber konstant positiv oder negativ zu verhalten." (Trommsdorff, 1998, S. 175).

Werte bündeln also ein System von Einstellungen und sind in Abgrenzung zu einzelnen Einstellungen verbindlicher. Bestehen in einer Gesellschaft werthomogene Gruppen, ist deren Identifizierung und Segmentierung anhand von Werten möglich. So fassen beispielsweise soziale Milieus nach dem Sinus-Institut Gruppen von Menschen zusammen, die sich in ihrer Lebensweise und Lebensauffassung ähneln (Diezinger & Mayr-Kleffel, 1999, S. 89f.). Im Ergebnis wird die Bevölkerung nach unterschiedlichen Wertorientierungen und Lebensstilen segmentiert. Ebenso wichtig ist es für das Marketing, bereits frühzeitig Anzeichen eines Wertewandels zu erkennen und diese richtig zu interpretieren (Silberer, 1985, S. 119ff.).

- *Persönlichkeit:* Die Begriffsbestimmung der Persönlichkeit als Einflussgröße des Konsumentenverhaltens wird in der Literatur stark von der wissenschaftlichen Grundposition der Forscher beeinflusst. Trommsdorff (1998, S. 202) selbst beschreibt die Persönlichkeit in seiner Arbeitsdefinition als "die Gesamtheit der für eine Person als typisch angesehenen, fest eingepägten und normalerweise nicht zu ändernden Verhaltensmuster (insbesondere Reaktions- und Kommunikationsmuster)." Neben den persönlichen Anlagen gehören des

Weiteren auch die Einstellungen, Normen, Werte und der Lebensstil zur Persönlichkeit. Dabei sollten auch die Umweltbedingungen (Kultur, soziale Schicht, Lebenszyklus etc.), innerhalb derer sich eine Persönlichkeit entwickelt, berücksichtigt werden.

Die Nutzung diesbezüglicher Informationen zur Persönlichkeit von Konsumenten für Zwecke des Marketings kann durch Segmentierung von Persönlichkeitstypen erfolgen. Anhand von soziodemographischen Kriterien und Verhaltensmustern abgegrenzte Zielgruppen können wichtige Beiträge zur Beschreibung von Konsumenten und zur Erklärung ihres Verhaltens liefern.

- *Aktiviertheit und Involvement*: Aktiviertheit als physiologische Erregung wirkt auf die Prozesse des Erwerbs und der Verarbeitung von Informationen ein und ist geeignet, die oben angeführten Zustände zu beeinflussen. Aktiviertheit alleine, beispielsweise aufgrund eines Marketinganreizes, führt nur selten direkt zu einem Kauf. Dies wäre der Fall bei einem Impulskauf (siehe auch Abschnitt 2.3.4.3 zu Impulskäufen im E-Commerce). Vielmehr ist sie geeignet, andere Zustände, wie Motive oder Einstellungen zu verändern (Trommsdorff, 1998, S. 43). Eine hohe Aktiviertheit der Konsumenten in einem Prozess wird auch als hohes Involvement bezeichnet. Im folgenden Abschnitt werden zunächst die Einflussfaktoren des Involvement dargestellt. Daran anschließend werden die Auswirkungen des Involvement auf das Informations- und Entscheidungsverhalten im Allgemeinen sowie seine Bedeutung bei Dienstleistungen im Internet im Besonderen beschrieben.

2.3.2 Involvement

In der Literatur besteht weitestgehend Übereinstimmung darüber, dass Involvement ein Konstrukt in der Kaufverhaltensforschung darstellt, das auf die verschiedensten Bereiche des Konsumentenverhaltens nachhaltig einwirkt und diese beeinflusst (Engel, Blackwell & Miniard, 1993, S. 275; Homburg & Kebbe, 2001; Kroeber-Riel, 1992, S. 168). Der Einfluss des Involvement erstreckt sich dabei sowohl auf die Gewinnung, Wahrnehmung, und Verarbeitung von Informationen als auch auf das Verhalten der Konsumenten.

Kroeber-Riel und Weinberg (1999, S. 248) bezeichnen Involvement als eine komplexe, mehrdimensionale Größe und verstehen darunter „die Ich-Beteiligung oder das Engagement, das mit einem Verhalten verbunden ist, zum Beispiel die innere Beteiligung, mit der jemand eine Kaufentscheidung fällt.“ Entsprechend verstehen Meffert und Bruhn (1997, S. 81) Involvement als ein Bündel von aktivierenden, kognitiven und persönlichkeitsbezogenen Komponenten (Meffert & Bruhn, 1997, S. 81).

Involvement bezieht sich auf das Ausmaß der persönlichen Bedeutung eines Objektes oder eines Verhaltens für ein Individuum in Bezug auf seine Ziele, Werte und sein Selbstkonzept (Blackwell, Miniard & Engel, 2001, S. 91). Es steht damit auf einer anderen Ebene als Einstellungen, die die Art der Beziehung zu dem Objekt (positiv versus negativ) wiedergeben. Allgemein wird das Involvement anhand der Dimensionen der subjektiven Wichtigkeit für

den Konsumenten, des Risikos möglicher negativer Konsequenzen, des hedonistischen Werts und des symbolischen Werts, so beispielsweise bei Markenprodukten, beschrieben (McColl-Kennedy & Fetter, 2001). Auf eine Kaufentscheidung bezogen zielt Involvement auf die persönliche Bedeutung des Objekterwerbs für den individuellen Konsumenten. Entsprechend kann Involvement von Kunde zu Kunde variieren, durch die jeweilige Situation beeinflusst werden und hat als motivationale Variable direkten Einfluss auf das Kaufverhalten (Zaichkowski, 1985).

2.3.2.1 Faktoren des Involvement

Als mehrdimensionale Größe wird Involvement bestimmt durch personenspezifische, produkt-/leistungsspezifische, situationsspezifische und medienspezifische Einflussfaktoren (Neibecker, 1990). Kroeber-Riel (1996) unterscheidet persönliche, reizabhängige und situative Komponenten wobei er Produkt-/Leistungsinvolverment als reizabhängiges Involvement einordnet. Während persönliches Involvement und Produkt-/Leistungsinvolverment sich transaktionsunabhängig darstellen, wird situatives Involvement von transaktionsspezifischen Konstellationen beeinflusst (McColl-Kennedy & Fetter, 2001).

Persönliches Involvement: Personen unterscheiden sich bezüglich ihrer individuellen Informationsneigung. Konsumenten mit einer höheren Informationsneigung gehen im Kaufentscheidungsprozess informationsbewusster vor und suchen eher mehr Informationen als der durchschnittliche Konsument (Thorelli & Thorelli, 1977). Das hinter der Informationsneigung stehende Engagement bezeichnen Kroeber-Riel und Weinberg (2003, S. 250) als persönliches Involvement. Das persönliche Involvement ist eine Neigung der individuellen Prädisposition und stellt wie die Neigung, Risiken einzugehen oder die Neigung zu Kritik ein Persönlichkeitsmerkmal dar.

Produkt-/Leistungsinvolverment: Je stärker die Eigenschaften eines Produktes oder einer Dienstleistung die zentralen Werte von Konsumenten ansprechen, umso mehr interessieren und engagieren sie sich für diese. Es wird daher teilweise auch von Ego-Involvement gesprochen (Neibecker, 1990). In Abgrenzung zu Purchase-Involvement, das sich auf das Interesse und die Auseinandersetzung des Konsumenten mit dem Kaufprozess bezieht, wird unter Ego-Involvement die Bedeutung des Objektes für das Individuum, sein Selbstbild und seine Werte verstanden (Beatty, Kahle & Homer, 1988). Nach Trommsdorff (1998, S. 53f.) lässt sich das Produkt-/Leistungsinvolverment anhand von fünf Dimensionen beschreiben:

1. Interesse am Produkt bzw. an der Leistung
2. Verstärkung, Spaß und Belohnung bei der Entscheidung bzw. beim Konsum
3. Identifikation mit dem Produkt bzw. der Leistung und Unterstützung der persönlichen Ausdrucksmöglichkeit
4. Wahrscheinlichkeit, mit dem Erwerb bzw. der Nutzung ein Risiko einzugehen
5. Risikokosten im Risikofall.

Das Aktivierungspotenzial, das ein Produkt oder eine Dienstleistung für einen Konsumenten besitzt, kann für das Involvement von größerer Bedeutung sein als der Anschaffungspreis (Kroeber-Riel & Weinberg, 2003, S. 409ff.). Andererseits wird das Involvement wesentlich durch das vom Konsumenten wahrgenommene Risiko der Kauf- oder Nutzungsentscheidung geprägt, in welches auch das Preisrisiko einfließt. Je größer die Unsicherheit aufgrund eines potentiellen Risikos eingeschätzt wird, umso größer sind auch das Interesse und der Antrieb, im Vorfeld der Kaufentscheidung zusätzliche Informationen einzuholen und diese in den Beurteilungsprozess einzubeziehen (Burnkrant & Sawyer, 1983; Neibecker, 1990).

Das *situative Involvement* bezieht sich auf die temporäre Relevanz eines Objektes für eine Person. Es hängt von den Gegebenheiten der aktuellen Beurteilungssituation ab und ist somit zeitabhängigen Schwankungen unterworfen (Enders, 1997; Liechtenstein, Netemeyer & Burton, 1990; Nerdinger, 2001, S. 26). Situatives Involvement ist in Konflikten begründet, die bei der Informationssuche entstehen können, wenn beim Vergleich der verfügbaren Informationen mit den Erwartungswerten eines subjektiven kognitiven Bezugssystems eine Diskrepanz besteht und diese mit zusätzlichen Informationen zu schließen versucht wird (Kroeber-Riel & Weinberg, 2003, S. 250f.). Während Konsumenten sich, solange kein Bedarf besteht, für viele Produkte und Dienstleistungen im Normalfall kaum interessieren, kann sich situationsbedingt ein verstärktes Interesse an diesen entwickeln. So steigt beispielsweise das Interesse für Autos an, wenn der Kauf eines neuen Autos ansteht. Andererseits kann ein situationsbedingter Zeitdruck oder eine Überlastung mit Informationen zu einem temporären Rückgang des Involvement führen (Kroeber-Riel & Weinberg, 2003, S. 380; Nerdinger, 2001, S. 26).

2.3.2.2 Involvement im Internetkontext

Bei der Reflexion des Konstruktes Involvement im Kontext des Internet stellt sich die Frage, ob neben den benannten Formen des persönlichen, produkt-/leistungsbezogenen und situativen Involvements weitere spezifische Formen des Involvement durch das Internet hinzukommen.

So stellen Bauer, Fischer & Sauer (2000, S. 1140) im Kontext des Kaufverhaltens im Online-Shopping das Konstrukt des *Technologieinvolvements* vor. Bei Technologieinvolvement handelt es sich um das Ausmaß, in dem sich ein Konsument der Technologie des Internet zuwendet und sich der Auseinandersetzung mit dem Internet widmet. Mit zunehmendem Technologieinvolvement der Konsumenten nimmt die Wahrscheinlichkeit zu, Transaktionen online auszuführen. Griffith, Krampf und Palmer (2001) beziehen sich mit dem Konstrukt *Interface Involvement* stärker auf die technische Ausstattung, wie beispielsweise die eingesetzte Hardware und die den Content repräsentierende Software und deren unterstützende Wirkung bei der Auseinandersetzung mit dem Content. Sie definieren das Interface Involvement als „the ability of a user interface to facilitate users' involvement with the information content presented to them“.

Neben der technisch fokussierten Betrachtung des Involvement im Internetkontext ist auch eine medienbezogene Dimension zu berücksichtigen. So findet sich in der Literatur die Annahme, dass aufgrund der Neuartigkeit des Internet und einer damit verbundenen Unsicherheit oder eines gesteigerten Interesses bei den Konsumenten das Produktinvolvement bei Internetangeboten durch das Medieninvolvement überlagert werden kann (Diehl, 2002, S. 116f.). Dholakia und Fortin (2002, S. 162ff.) entwickeln in einem Modell zur interaktiven Werbung ein Verständnis von Involvement, das aus den medienbezogenen interaktiven Möglichkeiten des Internet resultiert. Den Autoren folgend entwickelt ein Konsument aufgrund seiner interaktiven Möglichkeiten, wie beispielsweise den Content zu kontrollieren oder mit dem Anbieter in Kontakt zu treten, ein situatives und medienbezogenes Involvement.

Breithaupt (2005, S. 205ff.) vertritt allerdings die Ansicht, dass sich durch den Internetkontext keine weitere Form des Involvement ergibt, sondern das Internet vielmehr als Determinante auf das persönliche, produkt-/leistungsbezogene und situative Involvement wirkt. Nach Einschätzung des Autors wird das persönliche Involvement dahingehend beeinflusst, dass sich die Frage der Offenheit des Konsumenten gegenüber der Nutzung einer Technologie, in diesem Falle der Technologie Internet, stellt. Desweiteren können einzelne, die Internetnutzung betreffende Aspekte, Teil der Leistung sein und damit in das Produkt-/Leistungsinvolverment eingehen, welches wesentlich durch das Interesse bestimmt wird, das ein Konsument einem Produkt oder einer Dienstleistung entgegenbringt. Im Rahmen der vorliegenden Arbeit liegt der Schwerpunkt der Betrachtung auf dem Produkt-/Leistungsinvolverment, das Konsumenten den Angeboten von Krankenkassen entgegenbringen.

Im nachfolgenden Abschnitt 2.3.3 wird ein Überblick über die Bestimmungsfaktoren des Informations- und Entscheidungsverhaltens von Konsumenten in Kaufprozessen gegeben, sowie der Einfluss des Involvement diesbezüglich erörtert.

2.3.3 Auswirkungen des Involvement auf Informations- und Entscheidungsverhalten

Zunächst wird in Abschnitt 2.3.3.1 der Einfluss des Involvement auf Umfang und Intensität des Informationsverhaltens dargestellt. In Abschnitt 2.3.3.2 wird der Einfluss des Involvement auf Kaufentscheidungen beschrieben und in der Folge eine Typologisierung von Kaufentscheidungen abgeleitet. In Abschnitt 2.3.3.3 werden die Einflüsse, unterschieden nach der Intensität in High- und Low-Involvement, zusammengefasst und einander gegenübergestellt.

2.3.3.1 Involvement und Informationsverhalten

Der Grad des Involvement eines Konsumenten wirkt sich auf dessen Informationssuche, -aufnahme und -verarbeitung aus. In Abhängigkeit des Involvement können die Aufmerksamkeit und die Intensität der Bemühungen zur Informationsgewinnung, -verarbeitung und -speicherung im Kaufentscheidungsprozess variieren (Gabott & Hogg, 1999; Roßmanith, 2001). So definiert Trommsdorff (1998, S. 50) Involvement als „Aktivierungsgrad bzw. die

Motivstärke zur objektgerichteten Informationssuche, -aufnahme, -verarbeitung und -speicherung“. Der Ablauf der Informationsaufnahme hängt davon ab, welche aktivierenden Kräfte dabei wirksam werden und nach welchen kognitiven Programmen die Informationsaufnahme gesteuert wird. So bestimmt die Stärke der aktivierenden Kräfte, die Intensität und den Umfang der Informationssuche und -aufnahme, während die kognitiven Programme die Auswahl der Informationsquellen bestimmen. Dabei nimmt das wahrgenommene Kaufrisiko wesentlichen Einfluss auf die Intensität der Informationssuche; "je größer das wahrgenommene Kaufrisiko ist, um so stärker ist der Antrieb, zusätzliche Informationen zu suchen." (Kroeber-Riel & Weinberg, 2003, S. 251 f.).

Hoch involvierte Konsumenten sind demnach zu höherem kognitivem Aufwand bei der Informationsgewinnung und -verarbeitung bereit. Konsumenten mit geringem Involvement hingegen bevorzugen eher oberflächliche, periphere Informationen, da ihr Informationsbedarf geringer ausgeprägt ist als bei Konsumenten mit hohem Involvement (Enders, 1997; Park, Lee & Han, 2007). Eine Person mit hohem Involvement setzt sich eher aktiv und kritisch mit Informationen auseinander, während eine Person mit geringem Involvement diesen eher passiv und unkritisch begegnet (Trommsdorff, 1995). Während bei geringem Involvement Bilder wichtige Marketingfunktionen unterstützen können, da sie für den Konsumenten Informationen leichter vermitteln als Texte und weniger Aufwand bedeuten als das persönliche Gespräch, spielt bei hohem Involvement die persönliche Kommunikation eine größere Rolle (Kroeber-Riel & Weinberg, 1999, S. 252).

2.3.3.2 Involvement und Entscheidungsverhalten

Eines der bekanntesten Modelle zur Erklärung von Kaufentscheidungen wurde von Blackwell und Miniard (1995) entwickelt. Es beschreibt die psychischen Vorgänge bei Konsumenten im Verlauf eines Kaufentscheidungsprozesses und basiert auf den drei Teilprozessen der Informationsverarbeitung, -bewertung und Entscheidung. Der Entscheidungsprozess erfolgt aus der Problemerkennung, wenn der Konsument eine Abweichung zwischen angestrebtem Ideal- und tatsächlichem Istzustand feststellt. Hat er keine angemessene Problemlösung, setzt die Informationssuche ein. Deren Intensität hängt u.a. vom Grad des Involvement sowie von den mit der Informationssuche verbundenen Informationskosten und dem erwarteten Nutzen ab. Die beschafften Informationen werden im Rahmen der Informationsaufnahme laufend selektiert, mit den eigenen Überzeugungen und Einstellungen abgeglichen und bewertet. Dabei kommt es zu Informationsverzerrungen und -verlusten. Abhängig von Aktivierungsgrad und Involvement gehen die neu aufgenommenen Informationen in das Gedächtnis ein. Sobald die gewonnenen Informationen eine Alternativenbewertung erlauben geht die Informationssuche in den Prozess der Bewertung über. Anhand der Bedeutung für die Ziele, Motive, Einstellungen und Absichten des Konsumenten, beeinflusst durch externe Faktoren des aktuellen Lebensumfeldes, erfolgt die Bewertung der Alternativen.

Kaufentscheidungen können sehr unterschiedlich ablaufen: Während einige Entscheidungen nur wenige Sekunden benötigen, erstrecken sich andere über mehrere Tage oder Wochen. Teilweise werden diese Entscheidungen spontan und ohne vorherige Informationsauswertung getroffen. Zum Teil erfolgen sie aber auch auf Grundlage einer umfangreichen und sorgfältigen Analyse der Informationsquellen. Während manche Entscheidungen eher unwichtig sind, sind andere von großer Bedeutung. Die Komplexität und Vielfalt an Kaufentscheidungsprozessen in der Realität machte eine Kategorisierung erforderlich.

Eine erste Kategorisierung erfolgte durch Katona (1960, S. 57f.). Der Autor differenziert zwischen "echten Entscheidungen" anlässlich der Wahrnehmung einer neuen Situation und der Lösung eines daraus resultierenden Problems einerseits und andererseits einem "habituellen Verhalten" in ähnlich bereits erlebten Situationen. Diese Sichtweise spiegelt sich wider bei extensiven, umfassend durchdachten Kaufentscheidungsprozessen, wie sie beispielsweise bei Neuproduktkäufen ablaufen bzw. in einem habitualisierten Kaufverhalten bei Wiederholungskäufen. Der Ansatz wurde später um limitierte Kaufentscheidungen und Impulskäufe erweitert (vgl. Kroeber-Riel & Weinberg, 2003, S. 384 ff.; Kuß & Tomczak, 2000, S. 94 ff.; Nerdinger, 2001, S. 54 ff.). Die unterschiedlichen Typen von Kaufentscheidungen lassen sich nach dem Ausmaß der kognitiven Steuerung, bzw. des persönlichen Involvements wie folgt anordnen:

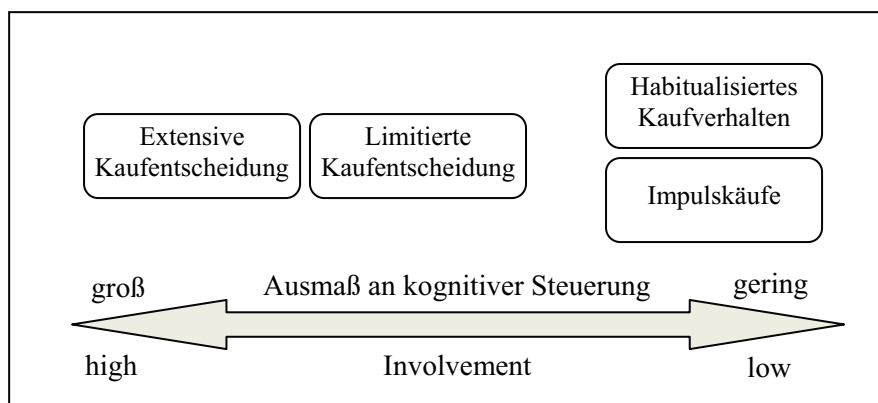


Abb. 2-4: Ausmaß an kognitiver Steuerung und Involvement bei unterschiedlichen Typen von Kaufentscheidungen (Angelehnt an Kuß & Tomczak, 2000, S. 97)

Während *extensive Kaufentscheidungen* willentlich und auf Basis umfassender und durchdachter Überlegungen erfolgen, liegen *limitierte Kaufentscheidungen* dann vor, wenn der Konsument auf eine begrenzte Auswahl von Entscheidungskriterien zurückgreifen kann. Eine Limitierung kann darin begründet sein, dass der Konsument auf Erfahrungen aus ähnlichen Produktkäufen in der Vergangenheit zurückgreifen kann, oder nur eine bestimmte Auswahl an Produktalternativen in Betracht gezogen wird. Während bei limitierten Kaufentscheidungen eine eingeschränkte kognitive Produktbewertung erfolgt, reicht bei rein *habitualisierten Käufen* nach Kroeber-Riel eine "automatische Reaktion nach Wiedererkennung" z.B. von Markenattributen aus (Kroeber-Riel & Weinberg, 2003, S. 401 f.).

Habitualisierte Kaufgewohnheiten entwickeln sich aufgrund der Sammlung eigener positiver Erfahrungen mit bestimmten Produkten aus früheren Kaufentscheidungsprozessen oder durch eine Übernahme beobachteter Verhaltensmuster anderer Personen (Kroeber-Riel & Weinberg, 2003, S. 403). Starke Produkt- oder Markentreue spiegelt sich in der Regel in einem habitualisierten Kaufverhalten wider, da bei gewohnheitsmäßigen Käufen häufig bereits bekannte Marken ausgewählt werden, ohne dass ein bewusster Entscheidungsprozess vorausgeht (Dietrich, 1986).

Impulskäufe unterliegen ebenfalls kognitiv wenig gesteuerten Entscheidungen. In der Regel erfolgen sie in Reizsituationen ungeplant unter emotionaler Aufladung und Aktivierung sowie weitgehend reaktiv, also durch automatisches Reagieren in der Kaufsituation (Weinberg, 1981, S. 14). Nach Trommsdorff neigen Konsumenten zum Impulskauf, wenn ein geringes Involvement vorliegt, kein Anlass zur Informationsverarbeitung besteht, die Kosten der Informationsbeschaffung und -verarbeitung hoch sind, Zeitdruck besteht, nur ein geringes persönliches oder soziales Risiko mit dem Kauf verbunden ist, die Produktalternativen sich nur unwesentlich unterscheiden oder "der betreffende Konsument allgemein dazu neigt, beim Kaufen nicht nachzudenken." (Trommsdorff, 1998, S. 308).

2.3.3.3 High- und Low-Involvement bei Informations- und Entscheidungsprozessen

Blackwell, Miniard und Engel (2001, S. 247) unterscheiden geringes und hohes Kundeninvolvement nach der Intensität der Bemühungen im Vorfeld des Kaufs. Eine intensive Beteiligung der Konsumenten in einem Kaufprozess wird als *hohes Involvement* bezeichnet. In diesem Fall werden im Entscheidungsprozess in der Regel mehr Informationen herangezogen und gleichzeitig verarbeitet. Das Ausmaß der kognitiven, bewussten Informationsverarbeitung steigt an, was sich in einer verlängerten Entscheidungszeit auswirkt (Deimel, 1989). Hohes Involvement ist gekennzeichnet durch ein größeres Interesse und eine höhere Intensität, relevante Informationen über das angestrebte Objekt zu erhalten und zu verarbeiten. Vor einer Kaufentscheidung erfolgt eine aktive Informationssuche und Auseinandersetzung mit den Informationen. Die aufgenommenen Informationen werden kritisch analysiert und bewertet bevor eine Entscheidung getroffen wird. So stellen Konsumenten bei hohem Involvement einen bewussteren Vergleich konkurrierender alternativer Angebote an, bevor sie eine Auswahlentscheidung treffen und nehmen dabei größere Unterschiede in den Markenattributen wahr, als Konsumenten mit niedrigem Involvement (Robertson, 1976).

Demgegenüber bewertet ein Konsument eine Kaufentscheidung bei *geringem Involvement* als unbedeutend für seine Person. Die Situation ist durch eine geringere Aktivierung gekennzeichnet, was zu einer eingeschränkten, kognitiv weniger bewussten Informationsaufnahme führt (Kroeber-Riel & Weinberg, 2003, S. 92). Niedriges Involvement wird deutlich in einem geringen Bemühen, relevante Informationen über das betreffende Objekt zu erlangen sowie in einem eingeschränkten Bewusstsein über die Unterschiede in den Objekteigenschaften und den verschiedenen Marken (Zaichkowsky, 1985). Low-Involvement-Produkte weisen

demnach meistens einen entwickelten Lebenszyklus, wenig psychische Produktdifferenzierung, wenige kaufentscheidende Merkmale, wenig intensiv ausgeprägte Einstellungen beim Konsumenten sowie ein gering empfundenes Kaufrisiko auf. In Abbildung 2-5 werden die Erkenntnisse zu Informations- und Entscheidungsprozessen bei High- und Low-Involvement zusammengefasst.

High-Involvement	Low-Involvement
<ul style="list-style-type: none"> ▪ Aktive Informationssuche ▪ Aktive Auseinandersetzung ▪ Hohe Verarbeitungstiefe ▪ Markenbewertung vor dem Kauf ▪ Viele Merkmale beachtet ▪ Bewusster Vergleich konkurrierender Alternativen ▪ Optimierungsziel ▪ Stark verankerte Einstellungen ▪ Hohe persönliche Bedeutung ▪ Hohe Gedächtnisleistung ▪ Längere Entscheidungszeit 	<ul style="list-style-type: none"> ▪ Passives Informationsverhalten ▪ Geringe Auseinandersetzung ▪ Geringe Verarbeitungstiefe ▪ Keine Markenbewertung vor dem Kauf ▪ Wenige Merkmale beachtet ▪ Kein bewusster Vergleich konkurrierender Alternativen ▪ Anspruchsniveaueziel ▪ Gering verankerte Einstellungen ▪ Geringe persönliche Bedeutung ▪ Geringe Gedächtnisleistung ▪ Kürzere Entscheidungszeit

Abb. 2-5: Gegenüberstellung von High- und Low-Involvement bei Informations- und Entscheidungsprozessen (Eigene Darstellung in Anlehnung an Trommsdorff, 1998)

Inwieweit Involvement Voraussetzung oder Folge eines Kauf- oder Nutzungsprozesses ist, hängt von den Eigenschaften des angestrebten Produktes ab. In verschiedenen Untersuchungen konnte übereinstimmend aufgezeigt werden, dass im Vorfeld von Kaufentscheidungen eine Suche nach Informationen häufig nicht oder nur in geringem Umfang erfolgt (Raffee & Silberer, 1981, S. 283 ff.). Für unterschiedliche Produktarten wurde festgestellt, dass mehr als die Hälfte der Konsumenten sich vor einem Kauf nicht aktiv um Informationen bemühten. Das Informationsverhalten im Vorfeld eines Kaufes wird in hohem Maße durch die persönliche Beteiligung beeinflusst, die mit der Kaufentscheidung für den Konsumenten verbunden ist.

2.3.4 Konsumentenverhalten im Electronic Commerce

Es stellt sich die Frage, inwieweit die angeführten allgemeinen Erkenntnisse zum Konsumentenverhalten auf Situationen des E-Commerce übertragbar sind. Zahlreiche Veröffentlichungen in der Konsumentenverhaltensforschung widmen sich den Spezifika, die mit der Nutzung von Angeboten im Internet einhergehen (u.a. Aimichai-Hamburger, 2002; Foscht, 1999; Keeney, 1999; Mathwick, 2002; Pereira 1999; Raffee & Jacobs, 1986). Dabei

findet sich häufig die Ansicht, dass Online Consumer Behavior trotz der Besonderheiten des E-Commerce über viele Parallelen zur klassischen Konsumentenverhaltensforschung verfügt. Über die Bereiche des Konsumentenverhaltens, die durch das Internet Veränderungen unterworfen sind, besteht jedoch keine Einigkeit.

So geht Gräf (1999, S. 82ff.) von einer umfassenden internetbedingten Veränderung der Konstrukte der Konsumentenverhaltensforschung aus. Er argumentiert, dass bei Internetkunden die Einstellungen gegenüber den einzelnen Angebotsaspekten eher in den Hintergrund treten, während die Einstellungen gegenüber dem Medium in den Vordergrund kommen. Breithaupt (2005, S. 167) hingegen führt aus, dass der Konsument einerseits eine Nutzerrolle einnimmt, verbunden mit einer gegenüber der Nutzung des Internet verankerten Einstellung, während andererseits in seiner Rolle als Konsument im Internet produkt- und leistungsspezifische Einstellungen zum tragen kommen. Fritz (2001, S. 88ff.) hingegen stellt in Bezug auf das Konsumentenverhalten im Internet drei medienspezifische Einflussfaktoren heraus: Erfahrung im Umgang mit dem Medium Internet, Involvement sowohl gegenüber dem Medium Internet als auch dem angestrebten Objekt sowie ein Flowerleben in der Handhabung des Internet.

Detaillierter betrachtet stellt sich die Frage, ob der verhaltenswissenschaftliche Bezugsrahmen zum Konsumentenverhalten dem interaktiven Medium Internet und insbesondere den spezifischen Anforderungen des E-Commerce gerecht wird. Und wenn ja, ist zu beantworten, wie aktivierende und kognitive Prozesse im Bereich des E-Commerce ablaufen. Entsprechend soll das Konsumentenverhalten im Online-Handel nachfolgend analog zum traditionellen Konsumentenverhalten, getrennt nach den Aspekten aktivierender Prozesse (Abschnitt 2.3.4.1), kognitiver Prozesse bei extensivem bzw. limitiertem Kaufverhalten (Abschnitt 2.3.4.2) sowie bei habitualisiertem Kaufverhalten bzw. Impulskäufen (Abschnitt 2.3.4.3), dargestellt werden.

2.3.4.1 Aktivierende Prozesse im Electronic Commerce

Um Angebote des E-Commerce zu nutzen, ist grundsätzlich eine Aktivierung erforderlich. Nachfolgend ist die Aktivierung von Interesse, die erforderlich ist, um zu einer Webseite eines E-Commerce-Anbieters zu gelangen. Dabei spielt neben der klassischen Massenkommunikation die Internetwerbung über Werbebanner als „Aktivierung zum Clickthrough“, der Einsatz von Suchmaschinen sowie der persönliche Austausch unter den Internetnutzern eine Rolle, um den E-Commerce-Anbieter als Markenname bekannt zu machen. Allerdings ist der Besuch der Website eines Anbieters in Folge einer Suchmaschinenabfrage wahrscheinlicher, wenn dem Konsumenten der Anbieter, bzw. die Website bereits bekannt ist (Esch, Langner & Jungen, 1998).

Nachdem der Nutzer die Homepage eines E-Commerce Anbieters aufgerufen hat, kommt es darauf an, ihn durch weitere Aktivierung involviert zu halten, so dass er weitere Seiten aufruft und Angebote nutzt (vgl. Hühnerberg, 1997, S. 115). Entscheidend hierfür ist neben der

inhaltlichen Qualität die ansprechende Gestaltung der Darstellung, eine hohe Benutzerfreundlichkeit sowie die Vernetzung mit anderen Internetseiten oder anders ausgedrückt eine "haushaltsgerechte Oberfläche des Electronic-Commerce-Systems" (Ring & Winand, 2000). Eine entsprechende Involvierung des Nutzers ist eher zu erwarten, wenn

- die Benutzeroberfläche optisch ansprechend gestaltet ist, das Design dabei einfach bleibt und sowohl Einsteigern als auch Experten in ihren Anforderungen möglichst gerecht wird (Schwarz & Müller, 2007),
- unterschiedliche Szenarien angelehnt an die Nutzungsabsichten des Nutzers entsprechend seiner Rollenmerkmale, Motive und Einstellungen gewählt werden können,
- die Angebote realitätsnah und direkt erfahrbar dargestellt werden und gegebenenfalls von visuellen Reizen unterstützt werden, ohne dabei vom Wesentlichen abzulenken (Winand & Pohl, 1998),
- eine schnelle Informationssuche möglich ist und adäquate Zusatzinformationen, z.B. mit Hilfe von Hyperlink, eingeholt werden können (Stewart, 2006).

Wie bereits in Abschnitt 2.3.3 für den Offline-Handel dargestellt, werden auch im E-Commerce Bewertungen von Informationen im Rahmen von Kaufentscheidungsprozessen je nach Ausprägung des Involvements auf einer unterschiedlichen Basis getroffen.

In der Folge werden Bewertungen im Rahmen eines Kaufvorgangs je nach Ausprägung des Involvements auf unterschiedlicher Basis getroffen. So stellen Ranaweera, McDougall & Bansal (2004; 2005) für Kaufprozesse im E-Commerce bei Kunden mit hohem Involvement eine enge Verknüpfung des Online-Kaufverhaltens mit der Kundenzufriedenheit bzw. zwischen der Usability einer Website und der Nutzerzufriedenheit fest. Die Autoren schreiben in Folge dessen dem Involvement eine moderierende Rolle auf die Beziehungen zwischen Kundenzufriedenheit und Online-Kaufverhalten bzw. Usability und Nutzerzufriedenheit zu.

2.3.4.2 Extensives vs. limitiertes Kaufverhalten

Extensive und teilweise auch limitierte Kaufprozesse sind auf Konsumentenseite durch hohes Involvement, häufig verbunden mit einem hohen wahrgenommenen Risiko, gekennzeichnet, was einen dementsprechend ausführlichen kognitiven Produktbewertungsprozess zur Folge hat (Bänsch, 1996, S. 76f.; Kuß & Tomczak, 2000, S. 97).

Wie bereits beschrieben, ist eine naheliegende Strategie zur Reduzierung des Risikos die Beschaffung von Informationen. Während traditionell bestimmte Informationen teilweise erheblichen Aufwand und Kosten verursachen, z.B. aufgrund von Fahrten zu Kaufhäusern, Anfragen bei Experten etc., liegen die relevanten Informationen häufig im Internet vor. Darin liegt jedoch zugleich auch ein neues Problem. "Begrenzungen in Menge und Verfügbarkeit von Informationen wurden durch begrenzte Ressourcen zu deren Aufnahme und Bewertung abgelöst" (Friedewald & Kolo, 2000). Es sind Ressourcen erforderlich, um die hohe

quantitative Anzahl an Informationen zu verarbeiten und qualitativ zu bewerten. Neben einer persönlichen Aufnahme- und Selektionsfähigkeit sind desweiteren gute Surffähigkeiten erforderlich, um schnell und zuverlässig relevante Informationen zu finden (Esch, Langner & Jungen, 1998). Neben den Beschränkungen, die sich im Rahmen des Produktbewertungsprozesses aufgrund zur Verfügung stehender Ressourcen ergeben, ist die Frage der Produkteignung für einen Bewertungsprozess im E-Commerce zu betrachten.

Eine Bewertung der Online-Eignung verschiedener Produkte und Dienstleistungen findet sich bei Ahlert et al. (2000, S. 24) und Loos (1998, S. 52f.). Eine Analyse der von Ahlert dargestellten Einflussfaktoren des Kaufprozesses bzw. der Produkte oder Dienstleistungen ist nicht möglich, da eine detaillierte Aufstellung der Teilkriterien fehlt. Loos ordnet Konsumgütersegmente nach zwei Dimensionen, der Emotionalität und der Komplexität der zugehörigen Produkteigenschaften, in eine Matrix ein. Während die emotionale Komponente differenziert, inwiefern ein Produkt ausschließlich rational beschrieben und bewertet werden kann oder emotional erfasst wird und individuell aktivierende Prozesse auslöst, beschreibt die Dimension Komplexität den Erklärungs- und Beratungsbedarf eines Produktes (Loos, 1998, S. 53).

Vor dem Hintergrund des der vorliegenden Arbeit zugrunde gelegten nomologischen Modells (siehe Abschnitt 2.10) ist es jedoch als kritisch zu bewerten, dass wichtige Einflusskriterien des Kaufprozesses im E-Commerce, wie Involvement, wahrgenommenes Risiko und informationsökonomische Aspekte nicht berücksichtigt werden. Die Schlussfolgerung, dass komplexe und rationale Produkte eine hohe Online-Eignung aufweisen, ist für große Bereiche der Soft- und Hardware nachvollziehbar. Komplexe Beratungsleistungen hingegen lassen sich jedoch nur schwerlich adäquat internetbasiert substituieren. Andererseits werden auch emotionale Produkte erfolgreich über das Internet vertrieben, wenn sie als Markenprodukte das Vertrauen der Konsumenten genießen. Die Einschätzung der Online-Eignung kann nicht nur über die Produkteigenschaften erfolgen, sondern ergibt sich vielmehr aus einer relativen Vorteilhaftigkeit gegenüber dem traditionellen Handel (Licharz, 2002, S. 132).

Hier bietet eine Übertragung der informationsökonomischen Kriterien aus Abschnitt 2.3.2 auf den Anwendungsbereich des E-Commerce einen Rahmen für die Beurteilung produktbezogener Risiken bei der Einordnung der Online-Produkteignung (Licharz, 2002, S. 132; Wilke, 2000). Dabei ist es von Interesse, inwiefern sich die informationsökonomischen Eigenschaften eines Produktes beim Übergang vom traditionellen Handel zum Online-Handel verändern und inwieweit der Online-Handel gegenüber dem traditionellen Handel geeignet ist, Unsicherheiten beim Konsumenten zu reduzieren.

Bei der Online-Präsentation eines Produktes ist häufig eine Verschlechterung der informationsökonomischen Qualität gegenüber dem traditionellen Handel zu erwarten (Kortzfleisch & Winand, 2000). Sucheigenschaften, wie die Farbe, die Haptik, der Geruch lassen sich oft nicht adäquat virtualisieren und werden damit zu Erfahrungseigenschaften.

Andererseits muss aus einem Übergang von einer Sucheigenschaft zu einer Erfahrungseigenschaft im Online-Handel nicht unbedingt ein Nachteil für die Konsumenten entstehen, da entsprechende Informationssubstitute wie beispielsweise Testurteile oder Kundenbeurteilungen leicht zur Verfügung gestellt werden können (Clemons, 2007). Auch sind Fälle denkbar, in denen bestimmte Vertrauenseigenschaften eines Produktes oder einer Dienstleistung durch Virtualisierung erfahrbar gemacht werden. So können beispielsweise durch Simulation von Leistungsfällen spezifische Leistungsdimensionen einer Versicherung und das dabei abgedeckte Risiko erfahrbar gemacht werden (Kortzfleisch & Winand, 2000).

Bleibt die informationsökonomische Qualität einer Eigenschaft gleich, kommt es im Falle von Sucheigenschaften darauf an, diese dem Kunden schnell, vollständig und vor allem einfach zugänglich zu machen. Bei Erfahrungs- und Vertrauenseigenschaften sind, wie bereits angeführt, die Online-Möglichkeiten an leistungsbezogenen und leistungsübergreifenden Informationssubstituten gegenüber dem Offline-Handel zu nutzen.

2.3.4.3 Habitualisiertes Kaufverhalten vs. Impulskäufe

Wie bereits in Abschnitt 2.3.3 dargestellt wurde, sind Impulskäufe und habitualisiertes Kaufverhalten im Vergleich zu extensivem oder auch limitiertem Kaufverhalten durch ein eher geringes Involvement des Konsumenten gekennzeichnet. Beim habitualisierten Kaufverhalten resultiert dies aus der Gewohnheit aufgrund in der Vergangenheit gemachter Erfahrungen mit bestimmten Produkten. Beim Impulskauf hingegen aus dem mit der Kaufentscheidung verbundenen geringen wahrgenommenen Risiko.

Allerdings können höhere Risiken des Online-Handels besonders ins Gewicht fallen, da der Konsument in diesen Fällen von habitualisierten oder Impulskäufen im Internet absieht und diese stattdessen traditionell durchführt oder wie sonst eher bei extensiven oder auch limitierten Kaufprozessen einen entsprechenden ausführlichen kognitiven Informationsverarbeitungsprozess vollzieht (Licharz 2002, S. 133).

Anbieter im E-Commerce können für den Betrachtungsfall des habitualisierten Kaufes Vorteile erzielen, indem sie die Möglichkeiten detaillierter Datenanalysen nutzen. Dies können Online-Anbieter aufgrund des digitalen Umfeldes grundsätzlich schneller, einfacher und vor allem günstiger realisieren als traditionelle Anbieter. "Marketing in the Web era is different because of the incredible amount of real-time information that is potentially available about the behavior and interests of customers. E-CRM for example, helps companies to gather insightful information on customers, to segment them, and base interactions with them on their preferences and needs. This is a marketer's dream." (Varianini & Vaturi, 2000, S. 93). Allerdings zeigen die Autoren in einer Analyse verschiedener Online-Anbieter auch, dass dieser Traum in der Realität nur sehr komplex umzusetzen ist.

Konsumenten entscheiden sich bei vergleichbarem Preis vor allem dann für Online-Angebote, wenn diese ihren Bedürfnissen entsprechen und sie die erforderlichen Transaktionen schneller

und mit weniger Aufwand durchführen können als traditionell (Keeney, 1999). Die Zielgruppen des Impulskaufes sind im E-Commerce Nutzergruppen, die zum "Browsing" tendieren. "Sie durchwandern eine Hypermedia-Basis ohne ein klar umrissenes Suchziel." (Esch, Langner & Jungen, 1998). Einem Online-Anbieter bietet sich die Chance, diese durch eine aktivierende Gestaltung seines Angebots zu einem Kaufinteresse zu bewegen. Die Herausforderung liegt jedoch in der "impulskaufgerechten" Abwicklung der mit dem Kauf verbundenen Transaktionen. Analog zum traditionellen Handel akzeptiert der Impulskäufer keinen größeren Aufwand an Zeit, Geld oder Risiko (Trommsdorff, 1998, S. 308f.). Aufwändige und zeitraubende Eingabemasken oder komplizierte Abrechnungspraktiken werden daher kaum angenommen.

Die zentralen Aspekte einer kundenorientierten Gestaltung des E-Commerce Prozesses und die Zusammenhänge mit der Zufriedenheit des Nutzers werden im folgenden Kapitel 2.4 dargestellt.

2.4 Kundenorientierung im Electronic Commerce

Das der Kundenorientierung zugrunde liegende Prinzip besteht darin, die Aktivitäten des Unternehmens an den spezifischen Erwartungen und damit Anforderungen der Kunden auszurichten (Bruhn, 2002, S. 38; Kühn, 1991). Unter Kundenorientierung ist somit die grundsätzliche Ausrichtung der Unternehmensaktivitäten an den Kundenerwartungen bei der Gestaltung der Interaktion und dem Aufbau einer langfristigen Beziehung zu verstehen (Bruhn, 2003, S. 12). Bezogen auf die Gestaltung der Interaktionen im E-Commerce-Prozess steht damit die Fähigkeit eines Unternehmens im Vordergrund, die Anforderungen an die Gestaltung aus Kundensicht kontinuierlich zu erfassen, in Unternehmensleistungen umzusetzen und eine langfristige Kundenbeziehung aufzubauen.

Der zentrale Aspekt eines kundenorientierten E-Commerce-Prozesses sind die Kundenerwartungen, wobei der Begriff Kundenerwartung sich unterschiedlich interpretieren lässt (vgl. Bruhn, 2001, S. 64). Hier sei als Kundenerwartung die Leistungsart und das Leistungsniveau bezeichnet, die bzw. das sich der Kunde von seinem Anbieter wünscht (Parasuraman, Zeithaml & Berry, 2001, S. 467f.). In diese Erwartungen ist auch der E-Commerce-Prozess eingebunden, weil sich Kunden auch vom E-Commerce-Prozess einen bestimmten Nutzen versprechen. Ob der E-Commerce-Prozess kundenorientiert gestaltet ist, hängt letztlich davon ab, ob die Erwartungen durch den wahrgenommenen Wert der Leistung erfüllt oder sogar übertroffen werden. Erfüllt oder übertrifft die wahrgenommene Leistung die Kundenerwartungen, stellt sich Zufriedenheit ein und der Kaufprozess wird fortgeführt. Werden die Kundenerwartungen nicht erfüllt, ist der Kunde unzufrieden und der Kaufprozess wird abgebrochen (vgl. Bruhn, 2001, S. 63). Kundenzufriedenheit stellt somit immer das Resultat eines komplexen Vergleichsprozesses während des gesamten Kaufprozesses dar (vgl. Homburg, Becker & Hentschel, 2003; Meffert & Bruhn, 2003, S. 195; Nerdinger, 2001, S.

59). Ist der Kunde zufrieden, wirkt sich das in seinem Verhalten dahingehend aus, dass er die Beziehung zu dem Unternehmen aufrechterhält und damit an das Unternehmen gebunden werden kann (Homburg & Bruhn, 2000, S. 5ff.). Kundenbindung resultiert somit aus Kundenzufriedenheit, die wiederum ein kundenorientiertes Verhalten des Unternehmens voraussetzt.

Die positiven Zusammenhänge zwischen dem Grad der Kundenorientierung der Unternehmensaktivitäten und der langfristigen Kundenbindung sowie dem Erfolg eines Unternehmens konnten in verschiedenen empirischen Studien gezeigt werden (u.a. Gupta & Kim, 2007; Jaworski & Kohli, 1993; Massad, Heckman & Crowston, 2006; Reichheld & Sasser, 1990). Übertragen auf den Anwendungsbereich des E-Commerce bedeutet dies, dass von entscheidender Bedeutung für den Erfolg eines online-Anbieters ist, den E-Commerce-Prozess kundenorientiert auszurichten, um den Kunden zufrieden zu stellen und damit an das Unternehmen zu binden.

Eine Studie der Boston Consulting Group, deren Erkenntnis auf einer Befragung von 2.876 US-Internetkunden basiert, beschreibt es als für eine erfolgreiche Positionierung im E-Commerce von Bedeutung, dem Kunden mit bestimmten Leistungsaspekten ein positives Einkaufserlebnis zu vermitteln (Silverstein, Stanger & Abdelmessih, 2001, S. 5). Dies ist dann möglich, wenn das Unternehmen nicht nur durch seine Produkte, sondern auch durch die Gestaltung des E-Commerce-Prozesses selbst einen Leistungsvorteil anstrebt. So kann eine langfristig ausgerichtete Beziehung zum Kunden entstehen, die eine (mehr oder minder starke) Interaktion ermöglicht (Diller, 1997, S. 573). Beide Aspekte - Interaktion und Langfristigkeit haben für den Aufbau eines Leistungsvorteils im E-Commerce eine zentrale Bedeutung.

Der E-Commerce-Prozess muss dafür grundsätzlich so gestaltet werden, dass der Kunde in der Interaktion mit dem Unternehmen einen Leistungsvorteil erfährt. Ein großer Teil der Besucher von Online-Shops beginnt den Kaufvorgang zwar, bricht diesen jedoch wieder ab, da ihnen kein Vorteil aus der Interaktion entsteht (Christophersen & Konradt, 2004). So wurde in dem Internetshopping Report 2001 ermittelt, dass benutzerergonomische Mängel besonders häufig die Ursache für den Abbruch des E-Commerce-Prozesses sind (Kliemann, 2001, S. 27). Dabei spielen neben Mängeln in der Darstellung, Aspekten der Einfachheit des Kaufprozesses auch Sicherheitsbedenken eine Rolle. Mehr als die Hälfte der befragten Internetkunden gaben an, den Kaufprozess in der Vergangenheit abgebrochen zu haben, weil sie das Produkt nicht schnell genug gefunden haben. Fast die Hälfte bemängelte die zur Verfügung gestellten Informationen, da Produkte nicht genau genug beschrieben wurden, Abbildungen fehlten oder das Produktangebot zu unübersichtlich war. Defizite in der Einfachheit beziehen sich sowohl auf technische Aspekte wie Ladezeiten, als auch auf Aspekte der Verständlichkeit und Benutzerfreundlichkeit.

Durch eine aus Kundensicht unbefriedigende Gestaltung der Interaktionen im E-Commerce-Prozess geht den Unternehmen bei vorzeitigem Abbruch des Kaufprozesses ein erhebliches Potenzial verloren, denn die meisten Besucher einer Webseite betreten diese bereits mit einer festen Kaufabsicht. Eine Online-Umfrage mit 1.013 Teilnehmern zeigte, dass fast 71 Prozent der Internetnutzer ihren Online-Kauf im Voraus geplant hatten und die Seite eines Unternehmens bereits mit einer festen Kaufabsicht besuchten. Nur 29 Prozent surfen ohne ein mehr oder minder festes Ziel durch das Web und kamen so zu einem Kauf (Wolfenbarger & Gilly, 2001).

Eine kundenorientierte Gestaltung des E-Commerce-Prozesses ist somit sowohl für die erfolgreiche Interaktion als auch für den Aufbau einer langfristigen Kundenbeziehung notwendig. Folglich kann eine Orientierung an den Anforderungen der Konsumenten im E-Commerce-Prozess als eine zentrale Voraussetzung für eine erfolgreiche online-Geschäftstätigkeit gesehen werden.

2.5 Erfolgsfaktorenforschung im Untersuchungsfeld Electronic Commerce

Mit der vorliegenden Arbeit wird die Bedeutung der Usability im Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich untersucht. Ein zentrales Anliegen besteht darin, die Anwendbarkeit der Usability-Skala ufosV2 (Christophersen, 2007) auf den Kontext von Electronic-Commerce-Anwendungen im Dienstleistungsbereich zu überprüfen, Befunde aus der vorangegangenen Untersuchung zu replizieren und die Usability-Skala ufosV2 bei Bedarf an die Anforderungen von Dienstleistungen anzupassen.

Hierbei stellt die Usability einen Erfolgsfaktor dar, der zum Zwecke der Skalvalidierung in Beziehung zu weiteren anwendernahen Erfolgsfaktoren kommerzieller Websites gesetzt wird. Entsprechend ist die Arbeit der internetbezogenen Erfolgsfaktorenforschung zuzuordnen. Nachfolgend wird in Abschnitt 2.5.1 ein allgemeines Verständnis des Begriffs in der Literatur skizziert. Daran anschließend wird dargestellt, welche allgemeinen methodischen Ansätze existieren (Abschnitt 2.5.2) und in welcher Hinsicht Kritik an der Erfolgsfaktorenforschung geäußert wird (Abschnitt 2.5.3). In Abschnitt 2.5.4 werden ausgewählte Erfolgsfaktoren kommerzieller Websites dargestellt, bevor abschließend in Abschnitt 2.5.5 ein Zwischenfazit gezogen wird.

2.5.1 Begriffsklärung

Die Identifikation kritischer Erfolgsfaktoren (auch: »strategischer Erfolgsfaktoren« oder »Schlüsselerfolgsfaktoren«) ist eine Methode, die Unternehmen in der Praxis in die Lage versetzen soll, sich auf eine begrenzte Anzahl von Handlungsfeldern zu fokussieren, die einen Wettbewerbsvorteil gegenüber der Konkurrenz begründen (Digman, 1990). Grundlegendes Ziel ist es, die Determinanten zu ermitteln, die den Erfolg eines Unternehmens langfristig wirksam beeinflussen. Grundannahme der Erfolgsfaktorenforschung ist dabei, dass sich der

Erfolg oder Misserfolg eines Unternehmens auf einige wenige Einflussfaktoren zurückführen lässt.

Um den Erfolg eines Unternehmens bewerten zu können und die Einflussfaktoren bestimmen zu können, werden in einer Erfolgsfaktorenstudie zunächst Indikatoren bestimmt, die den Erfolg messbar und damit auch vergleichbar machen. Darauf aufbauend wird untersucht, durch welche Variablen ein Erfolgsindikator nachhaltig determiniert wird. Dabei werden interne und externe Variablen unterschieden. Interne Variablen sind unternehmensbezogene Größen, die durch das Unternehmen beeinflussbar sind. Externe Variablen hingegen sind umweltbezogene Größen, die das Unternehmen nicht oder nur in geringem Maße beeinflussen kann. Diejenigen internen und externen Variablen, die den Unternehmenserfolg wesentlich bestimmen, werden als Erfolgsfaktoren bezeichnet (Haennecke, 2002).

Die Idee, Unternehmenserfolg auf bestimmte Einflussfaktoren zurückzuführen und daraus Empfehlungen an das Management abzuleiten, geht zurück auf die Beratungspraxis von McKinsey in den 60er Jahren und wurde später von Rockart (1979) mit seinem Konzept der kritischen Erfolgsfaktoren aufgegriffen. Im Rahmen des PIMS-Programms („Profit Impact of Marketing Strategies“) wurde in den 60er Jahren begonnen, in inzwischen etwa 300 Unternehmen mit rund 3000 strategischen Geschäftseinheiten, systematisch Unternehmensdaten sowie unternehmensexterne Variablen zu erfassen. Mit Hilfe der multiplen linearen Regression werden die Einflüsse dieser Größen auf den ROI („Return On Investment“) ermittelt (Buzzell & Gale, 1989).

In den folgenden Jahren wurde die Suche nach kritischen Erfolgsfaktoren mit den verschiedensten Methoden betrieben (vgl. für einen Überblick Trommsdorff, 1990). Gleichzeitig nahm der Anspruch an die Reichweite der Empfehlungen zum Einsatz kritischer Erfolgsfaktoren zu. Stand anfangs noch die Suche nach Erfolgsfaktoren für einzelne Unternehmen im Vordergrund, so erlangte die Erfolgsfaktorenforschung in den 80er Jahren eine verstärkte Aufmerksamkeit und es wurde zunehmend der Versuch unternommen, dieses Vorhaben auf ganze Branchen auszuweiten, oder noch allgemeiner, für jede Art von Unternehmen. In der Literatur finden sich verschiedene Zusammenstellungen zu branchenübergreifenden und branchenspezifischen Untersuchungen (vgl. Fritz, 1990; Göttgens, 1996, S. 475ff.; Schröder, 1994). Dabei wird deutlich, dass sich die unterschiedlichen Studien nicht nur im Hinblick auf die Untersuchungsansätze sondern teilweise auch bezüglich der Ergebnisse erheblich unterscheiden. In der Folge wurde die Erfolgsfaktorenforschung seit Anfang der 90er Jahre vielfach heftig kritisiert (Fritz, 1990; Haennecke, 2002; Kube, 1991, S. 4ff.; Schröder, 1994).

Als weiterer Kritikpunkt wurde ihr entgegengehalten, dass der betriebswirtschaftliche Erfolg nicht auf einige wenige isolierte Erfolgsfaktoren zurückgeführt werden könne. Vielmehr wird angeführt, dass dieser durch eine Vielzahl interdependenter interner und externer Variablen bestimmt werde, ohne dass dabei die Erfolgswirksamkeit einzelner Variablen isoliert werden könne. Angesichts dieser multiplen Kausalität einer Vielzahl von Faktoren auf den Unter-

nehmenserfolg wird der Erklärungsgehalt der Erfolgsfaktorenforschung grundsätzlich eingeschränkt. Diese Einschränkung wird in den Augen vieler Autoren noch dadurch verstärkt, dass Erfolgsfaktorenstudien teilweise nur ungenügend theoretisch fundiert sind. Die Studien erfolgten häufig nicht hypothesengeleitet auf Basis vorliegender gesicherter Ergebnisse. Statt dessen hätten sie oftmals nur einen erkundenden und datenorientierten Charakter und die zugrunde liegenden Wirkungsbeziehungen zwischen der Vielzahl an Faktoren würde häufig nur unzureichend aufgedeckt (Fritz, 1990, S. 103; Grabner-Kräuter, 1993; Kube, 1991, S. 55f.; Schröder, 1994). Angesichts der bestehenden Mängel erfolgt im nachfolgenden Abschnitt eine Betrachtung der Methoden der Erfolgsfaktorenforschung und der Aussagekraft ihrer Ergebnisse.

2.5.2 Methoden der Erfolgsfaktorenforschung

Bis in die Mitte der 90er Jahre fehlte ein einheitliches Verständnis darüber, welcher Methoden sich die Erfolgsfaktorenforschung bedienen sollte. Die Vielzahl heterogener Untersuchungsmethoden war ein wesentlicher Kritikpunkt an der Erfolgsfaktorenforschung. Schlüssige Systematisierungen der Untersuchungsmethoden zur Identifikation von Erfolgsfaktoren in Forschungssituationen lassen sich in der Literatur Mitte der 90er Jahre etwa bei Grünig, Heckner und Zeus (1996) sowie Heckner (1998, S. 69ff.) finden. Darauf aufbauend unterscheidet Haenecke (2002) in seiner Systematisierung fünf Methoden zur empirischen Identifikation strategischer Erfolgsfaktoren (vgl. Abbildung 2-6).

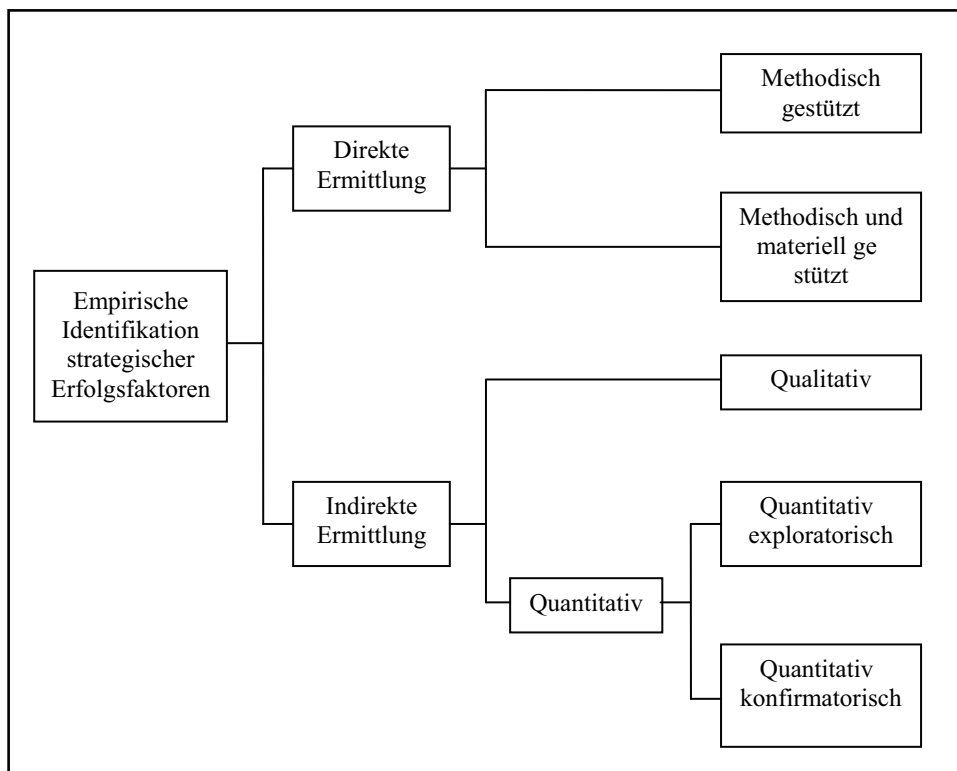


Abb. 2-6: Methoden zur Identifikation von Erfolgsfaktoren (Haenecke, 2002)

Die Methoden werden zunächst nach der Art der Ermittlung der Erfolgsfaktoren differenziert, wobei die Erfolgsfaktoren direkt oder indirekt ermittelt werden. Bei den *direkten Ermittlungsmethoden* werden die den Erfolg beeinflussenden Variablen direkt in Expertenbefragungen erhoben. Die direkte Ermittlung kann methodisch gestützt, etwa durch Kreativitäts- oder Problemlösungstechniken, oder methodisch und materiell gestützt, beispielsweise anhand von Checklisten, erfolgen.

Im Rahmen der *indirekten Ermittlungsmethoden* wird mit Hilfe gedanklicher Analyse und statistischer Verfahren untersucht, welche Faktoren den Erfolg wirksam beeinflussen. Im Gegensatz zu den direkten Ermittlungsmethoden wird also nicht direkt nach den erfolgsbeeinflussenden Variablen gefragt. Die Methoden zur indirekten Ermittlung strategischer Erfolgsfaktoren können je nach Art der Erhebung weiter in qualitative und quantitative Erhebungsmethoden klassifiziert werden.

Qualitative Studien untersuchen keine quantitativen Messgrößen, sondern stellen qualitative Aussagen in den Mittelpunkt der Betrachtung. Als zentrale Arbeit im Bereich der qualitativen Studien kann die Arbeit von Peters und Waterman (2000) angesehen werden. Die Autoren wählten zunächst Unternehmen aus, die sie anhand einer Anzahl definierter Erfolgskriterien als besonders erfolgreich bewerteten. Anschließend untersuchten sie die Unternehmen auf gemeinsame Eigenschaften, die sie als Erfolgsfaktoren interpretierten (vgl. Frese, 1985; Fritz, 1990).

In *quantitativen Studien* hingegen werden quantifizierte Unternehmensdaten erhoben, deren Anteil am Unternehmenserfolg mit Hilfe mathematischer Analysemethoden bestimmt wird. Quantitative Untersuchungsmethoden werden entsprechend der Art des Untersuchungsansatzes unterschieden in explorative, kausalstrukturentdeckende Verfahren und konfirmatorische, kausalstrukturüberprüfende Verfahren.

Quantitativ-explorative Studien verfolgen das Ziel, unter den möglicherweise erfolgswirksamen Variablen diejenigen zu identifizieren, die den Erfolg tatsächlich beeinflussen. *Quantitativ-konfirmatorische Studien* hingegen zielen darauf ab, empirisch und theoretisch bereits gut untersuchte Kausalzusammenhänge mit Hilfe kausalanalytischer Verfahren zu überprüfen. Da aufgrund von Voruntersuchungen auf ein tiefergehendes Verständnis der Kausalstrukturen zurückgegriffen werden kann, reichen im Unterschied zu explorativen Studien weniger Variablen in der Betrachtung aus. Während in einer konfirmatorischen Untersuchung die vorliegenden Ergebnisse und Theorien genutzt werden, um Kausalhypothesen aufzustellen, sollten diese im Rahmen einer explorativen Studie hingegen genutzt werden, um einen möglichst vollständigen Katalog möglicher Einflussgrößen aufzustellen und diesen zu überprüfen (Kube, 1991, S. 55). Nach Haenecke (2002) wird die höchste Aussagekraft bei einem quantitativ-konfirmatorischen Vorgehen erzielt. Sollte ein solches Vorgehen aufgrund einer unzureichenden theoretischen Fundierung nicht möglich sein, empfiehlt er ein quantitativ-exploratives Vorgehen.

Ob ein konfirmatorisches Vorgehen möglich ist, hängt entscheidend davon ab, ob der in der Literatur dokumentierte Entwicklungsstand der Theorie ausreichend ist, um aus früheren Studien abgeleitete Hypothesen zu bestätigen oder zu falsifizieren. Nur wenn auf Grundlage der vorliegenden Theorie zufriedenstellend Kausalhypothesen begründet werden können, ist ein konfirmatorisches Vorgehen möglich (Backhaus, Erichson, Plinke & Weiber, 2000, S. 491). Anhand des in der Literatur dokumentierten Forschungsstands ist somit die grundlegende Entscheidung zu treffen, inwieweit zunächst Hypothesen erkundet werden müssen oder ob auf Basis gesicherten Wissens und theoretischer Grundlagen zufriedenstellend Hypothesen gebildet werden können. Fehlen diese Grundlagen, sollten die Möglichkeiten des explorativen Untersuchungsansatzes für die Erkenntnisgewinnung genutzt werden (Schröder, 1994).

Nach den Prinzipien des kritischen Rationalismus ist Kausalität nicht messbar. Es kann lediglich versucht werden, Hypothesen über Ursache-Wirkung-Beziehungen zu überprüfen und zu falsifizieren. Werden die Hypothesen in einer Reihe von Überprüfungen nicht falsifiziert, können sie als bestätigt angesehen werden (Albert, 1987; Popper, 1935).

Übertragen auf die Erfolgsfaktorenforschung bedeutet dies, dass die Ursachen des Erfolgs nicht direkt nachgewiesen werden können. Vielmehr kann eine Erfolgsfaktorenstudie nur „mögliche“ Erfolgsursachen aufzeigen. Ob es sich dabei um „tatsächliche“ Erfolgsursachen handelt, kann erst durch mehrfaches Überprüfen der Hypothesen bestätigt werden (Hildebrandt, 1999, S. 48f.; Kube, 1991, S. 46). An das Vorgehen in einer Erfolgsfaktorenstudie können daher die Anforderungen abgeleitet werden, theoriegeleitet und hypothesentestend unter Verwendung eines Bezugsrahmens vorzugehen (Haenecke, 2002).

Im Rahmen der vorliegenden Arbeit wird auf Grundlage des in der Literatur vorliegenden Theoriestandes ein quantitativ-konfirmatorisches hypothesengeleitetes Vorgehen gewählt.

2.5.3 Kritik an der Erfolgsfaktorenforschung

Wesentliche Kritikpunkte an der Erfolgsfaktorenforschung waren lange Zeit die Vielzahl heterogener Forschungsmethoden und die mangelnde Aussagekraft aufgrund der multiplen Kausalität einer Vielzahl von Faktoren auf den Unternehmenserfolg sowie aufgrund einer ungenügenden theoretischen Fundierung der Studien (Fritz, 1990; Kube, 1991, S. 4ff.; Schröder, 1994).

Die Diskussion zur Bewertung der Methoden zur Identifikation kritischer Erfolgsfaktoren fokussiert daher vornehmlich auf den Nutzen bzw. die Aussagekraft der Methoden sowie auf den mit der Untersuchung einhergehenden Aufwand (Grünig, Heckner & Zeus, 1996; Haenecke, 2002; Heckner, 1998, S. 81ff.). Haenecke (2002) formuliert aufbauend auf der Arbeit von Grünig, Heckner, und Zeus (1996) sechs Kriterien, die er zur Bewertung von Methoden zur Identifikation von Erfolgsfaktoren heranzieht.

- ⇒ 1. *Aufdecken der Kausalstruktur*: Um mögliche Ursachen des Erfolges aufdecken zu können, sollte im Rahmen einer Erfolgsfaktorenstudie theoriegestützt und hypothesentestend vorgegangen werden. Nach den Prinzipien des kritischen Rationalismus ist Kausalität nicht direkt nachweisbar (Albert, 1987). Vielmehr sind Hypothesen über Ursache-Wirkung-Beziehungen zu prüfen. Werden diese Hypothesen bei wiederholter Überprüfung nicht falsifiziert, können sie als bewährt akzeptiert werden. Die Methoden der indirekten Ermittlung von Erfolgsfaktoren bieten diese Möglichkeit, Kausalität theoretisch zu begründen. Aufbauend auf vorhandenen Kenntnissen kann die Analyse, durch einen Bezugsrahmen geleitet, hypothesentestend erfolgen (Haenecke, 2002).
- ⇒ 2. *Berücksichtigung aller Perspektiven*. Vielen Erfolgsfaktorenstudien ist gemein, dass der Blick auf das betrachtete Untersuchungsobjekt unvollständig ist. Häufig werden nur ausgewählte Mitarbeiter des untersuchten Unternehmens befragt. Stattdessen sollten die Perspektiven aller an der Entstehung des Erfolges beteiligten Parteien, insbesondere die Sicht der Kunden und Anwender, einbezogen werden (Haenecke, 2002).
- ⇒ 3. *Berücksichtigung qualitativer und quantitativer Erfolgsfaktoren*. In quantitativen Erfolgsfaktorenstudien werden qualitative Aspekte des Erfolgs oft nur unzureichend berücksichtigt (Kube, 1991, S. 53). Der Autor leitet die Erfordernis ab, dass neben den leicht operationalisierbaren „harten“ Erfolgsfaktoren auch qualitative „weiche“ Größen in die Untersuchung einbezogen werden. Abstrakte Charakteristika eines Untersuchungsobjektes, die zur Erklärung der Beobachtung erforderlich sind, einer unmittelbaren Messung aber nicht zugänglich sind, werden mit Hilfe theoretischer Konstrukte beschrieben (Homburg & Pflesser, 1999, S. 636f.). Um im Verlauf von explorativen Studien in Regressionsanalysen mit theoretischen Konstrukten arbeiten zu können, können der Regressionsanalyse Faktorenanalysen vorgeschaltet werden.
- ⇒ 4. *Überprüfung der zeitlichen Stabilität*. Aufgrund des stetigen Wandels von Unternehmens- und Umweltsituationen sind Erfolgsfaktoren regelmäßig bzgl. ihrer Wirkungsintensität und kausalen Wirkungsinteraktion zu überprüfen. Dabei ist zu berücksichtigen, dass die Wirkungsintensität einer durchgeführten Maßnahme nicht unabhängig von der Situation ist, in der sie erfolgt (Beck, Brüderl & Woywode, 2002). Nicolai (2002) spricht in diesem Zusammenhang von einer gegenseitig beeinflussenden Simultanität von Umfeldsituation und Erfolg beeinflussender Maßnahme.
- ⇒ 5. *Objektivität*. Der allgemeinen Anforderung in der empirischen Forschung entsprechend sollte eine Erfolgsfaktorenstudie objektiv sein, das Ergebnis also nicht von der Durchführung beeinflusst werden (Berekhoven, Eckert & Ellenrieder, 1996, S. 86; Bortz & Döring, 2006, S. 194). Dem entgegen steht die vielfach in der Erfolgsfaktorenforschung angewandte Praxis der monopersonalen Datenerhebung, die vielfältige Verzerrungen begünstigt (Ernst, 2001; Kube, 1991, S. 58). Woywode (2002) spricht in diesem Zusammenhang von einem „key informant bias“.

⇒ 6. *Reliabilität*. Um sicherzustellen, dass bei einer Wiederholung einer Erfolgsfaktorenstudie die gleichen Ergebnisse geliefert werden, ist es erforderlich, Zufallsfehler so weit als möglich auszuschließen (Berekhoven, Eckert & Ellenrieder, 1996, S. 87).

Die Identifikation von Erfolgsfaktoren ist nur dann zukünftig erfolgsrelevant, wenn der Wettbewerb dessen Wert nicht bereits antizipiert hat. Die Erfolgsfaktorenforschung reproduziert jedoch häufig nur die Regeln, die ohnehin bereits in einer Branche verbreitet sind (Trommsdorff, 1990, S. 2). Dabei hat außergewöhnliche Performance häufig sehr viel mit Einzigartigkeit zu tun (Porter, 1996). Ein wirklicher Wettbewerbsvorteil besteht häufig deshalb, weil ihn andere Unternehmen nicht ohne weiteres imitieren können. Sei es weil sie aufgrund von Mobilitätsbarrieren daran gehindert werden, diese Positionierung einzunehmen (Dess & Davis, 1984) oder weil schwer imitierbare Ressourcen erforderlich sind, die auf den Faktormärkten nicht erhältlich sind (Barney, 1991). Immer wieder zeigt sich, dass außergewöhnlich erfolgreiche Unternehmen mit den vermeintlich generell gültigen Regeln ihrer Branche brechen und diese neu definieren. Führt man sich vor Augen, dass es das Ziel der Erfolgsfaktorenforschung ist, gesetzesmäßige Zusammenhänge aufzudecken, so werden hier zugleich ihre Grenzen deutlich.

Für die vorliegende Arbeit kann festgehalten werden, dass eine Studie zur Erforschung von Erfolgsfaktoren verschiedenen methodischen Anforderungen gerecht werden muss. Desweiteren ist es erforderlich, dass die Untersuchungsergebnisse Relevanz für die Praxis besitzen und eine angemessene Definition der Erfolgskriterien vorliegt.

2.5.4 Erfolgsfaktoren kommerzieller Websites

In der vorliegenden Arbeit wird untersucht, welche Faktoren von maßgeblicher Bedeutung für den Erfolg kommerzieller Websites sind. Im nachfolgenden Abschnitt soll deshalb eine Auseinandersetzung mit möglichen Erfolgsgrößen erfolgen. In der Literatur findet sich kein einheitliches Verständnis bezüglich der Indikatoren zur Erfassung des Erfolgs kommerzieller Websites. Vielmehr wird unterschieden in Kriterien, die auf die Zufriedenheit eines Anwenders mit der Nutzung einer Website, auf die Intention, diese zu nutzen bzw. auf eine tatsächliche Nutzung abzielen.

2.5.4.1 Nutzungszufriedenheit

Als Erfolgsgröße kommerzieller Websites wird häufig die *Zufriedenheit des Anwenders* mit der Website und den Interaktionen angesehen (Ba, 2001; Francis & White, 2002; Massad, Heckman & Crowston, 2006; Shankar, Smith & Rangaswamy, 2003; Wolfenbarger & Gilly, 2003). Homburg und Stock (2001) verstehen Kundenzufriedenheit „als eine Einstellung, die sich aus dem abwägenden Vergleich zwischen der erwarteten Leistung (Soll) und der tatsächlich wahrgenommenen Leistung (Ist) ergibt“. Der Gebrauch eines Produktes oder die Nutzung einer Dienstleistung führt zu Folgen, die sich entweder in Kundenzufriedenheit oder Kundenunzufriedenheit äußern. Das Ausmaß der Zufriedenheit bzw. Unzufriedenheit hat

Rückwirkungen auf die Überzeugungen des Konsumenten hinsichtlich der Leistung und seine Alternativenbewertung bei der nächsten Auswahlentscheidung. Zufriedenheit begründet oder verstärkt eine positive Einstellung gegenüber der Leistung und erhöht die Wahrscheinlichkeit für eine Wiederwahl bzw. einen Wiederkauf (Engel, Blackwell & Miniard, 1990, S. 481).

Stauss (1999) schreibt der Kundenzufriedenheit Verhaltensrelevanz in Kaufentscheidungsprozessen zu. Diese zeigt sich in einer erhöhten Wahrscheinlichkeit, Wiederkäufe zu tätigen, in Weiterempfehlungen und Cross-Buying-Verhalten aufgrund einer Generalisierung der positiven Erfahrungen auf andere Produkte oder Dienstleistungen des Anbieters (Kuß & Tomczak, 2000, S. 146; Mittal & Kamakura, 2001; Oldenburger, Lehto, Feinberg, Lehto & Salvendy, 2008; Simon & Homburg, 1998, S. 19f.). Zufriedene Kunden sind außerdem weniger anfällig gegenüber Marketingmaßnahmen der Konkurrenz und weniger preisempfindlich (Reichheld, 1997).

Diese positiven Konsequenzen von Kundenzufriedenheit werden in der Literatur meist in dem Begriff *Kundenloyalität* zusammengefasst (vgl. Braunstein, 2001, S. 33; Homburg & Rudolf, 1998, S. 51; Rapp, 1995, S. 13). Kundenzufriedenheit ist somit ein wichtiges Konstrukt bei der Vorhersage von Kaufverhalten und ein Schlüssel zum Aufbau von Kundenloyalität im Onlinebereich (Kassim & Abdullah, 2008; Shankar, Smith & Rangaswamy, 2003).

2.5.4.2 Nutzungsintention

In vielen Studien zum Online-Shopping wird als Erfolgsgröße anstelle tatsächlich getätigter Käufe die *Kaufintention*, bzw. übertragen auf Websites die *Nutzungsintention* erhoben (z.B. Gefen & Straub, 2000; Jarvenpaa & Todd, 1997a; Jarvenpaa, Tractinsky, Saarinen & Vitale, 1999; Jarvenpaa, Tractinsky & Vitale, 2000; Konradt, Wandke, Balazs & Christophersen, 2003; Koufaris & Hampton-Sosa, 2004; Li, Kuo & Russell, 1999; Lohse & Spiller, 1999). Insbesondere wenn Daten bezüglich einer tatsächlichen Nutzung nicht verfügbar sind, wird in Studien häufig die Nutzungsintention in der Annahme verwendet, dass eine Verhaltensintention unmittelbaren Einfluss auf das tatsächliche Verhalten nimmt (Chau, 1996; Hu, Chau, Sheng & Tam, 1999).

Nach Ajzen und Fishbein (1980) wird die Intention einer Person, ein bestimmtes Verhalten zu realisieren als unmittelbarer Prädiktor für ihr tatsächliches Verhalten angesehen. Die Theorie des überlegten Handelns (Fishbein & Ajzen, 1975) sowie die Theorie des geplanten Verhaltens (Ajzen, 1985) sind wichtige psychologische Theorien zur Vorhersage konkreten Verhaltens. Beide Theorien beziehen sich auf Verhalten, dessen Ausführung vollständig willentlich beeinflusst wird. Externe oder interne Faktoren, die die Ausführung eines Verhaltens beeinflussen können, werden nicht berücksichtigt. Im Falle willentlich kontrollierten Handelns ist die Ausführung eines Verhaltens eine direkte Folge der Intention, dieses Verhalten zu vollziehen. Die Vorhersagevalidität der Verhaltensintention für das konkrete Verhalten ist somit sehr hoch (Ajzen, 1985, S. 12). Die Genauigkeit der Verhaltensvorhersage

ist dabei umso größer, je geringer der zeitliche Abstand zwischen der Messung einer Verhaltensintention und der Beobachtung des zugehörigen Verhaltens ausfällt.

2.5.4.3 Nutzung

Zur Beurteilung des Effektes von Informationstechnologien auf das organisationale und individuelle Leistungsvermögen wird häufig die *Nutzung* der Technologie als geeigneter Indikator herangezogen (Devaraj & Kohli, 2003; Straub, Limayem, & Karahanna-Evaristo, 1995). Dabei ist zu unterscheiden zwischen der tatsächlich gemessenen Systemnutzung und einer anwenderseitig eingeschätzten Nutzung, die teilweise erheblich voneinander abweichen können. Wenngleich die Erfassung des Kriteriums häufig anhand von Nutzereinschätzungen erfolgt, wird eine Messung der tatsächlichen Nutzung hinsichtlich Kriterien wie Dauer, Häufigkeit und Intensität als geeigneten Indikatoren für die Beurteilung von Informationstechnologien empfohlen (Devaraj & Kohli, 2003; Straub et al., 1995; Szajna, 1996; Venkatesh, Brown, Maruping & Bala, 2008).

Bezogen auf den Kontext des E-Commerce finden sich ebenfalls Studien, die die Nutzung der Anwendung als Erfolgsgröße heranziehen (Christophersen, 2007; Kim & Prabhakar, 2004). Vielfach wird als Erfolgsgröße bei kommerziellen Websites die Wandlungsquote, also die Anzahl der Bestellungen je Nutzung der Website, sowie daraus abgeleitet ökonomische Erfolgsgrößen wie der Umsatz je Besucher oder die Gewinnung des Erstbesuchers für den gezielten Wiederholungskontakt bis hin zum Online-Stammkunden, verwendet (Gräf, 1999, S. 145ff.). Diese Messgrößen tragen dem Umstand Rechnung, dass eine Website im Falle kommerzieller Nutzung trotz hoher Nutzungsraten ökonomisch nicht erfolgreich sein muss, solange nicht entsprechende Rentabilitätsziele erreicht werden.

2.5.4.4 Auswahl

Die Mehrzahl der empirischen Studien zum Konsumentenverhalten im E-Commerce fokussiert auf die Identifikation und Untersuchung der Faktoren und Wirkungszusammenhänge, die die generelle Akzeptanz und Nutzung von Webangeboten begründen (vgl. u.a. Gefen, 2000; Gefen & Straub, 2000; Grabener-Kräuter & Kaluscha, 2003; Jarvenpaa et al., 2000; Koufaris & Hampton-Sosa, 2002; De Ruyter, Wetzels & Kleijnen, 2001). Seltener hingegen werden die Prozesse untersucht, die zur Auswahl des spezifischen Webangebotes oder Online-Shops führen, mit Hilfe dessen die gewünschte Transaktion getätigt wird (Lohse & Spiller, 1999; Lowengart & Tractinsky, 2001). Vor dem Hintergrund des wettbewerblichen Umfeldes von E-Commerce-Anwendungen ist die Kenntnis der Faktoren, die einen potentiellen Konsumenten zur Auswahl eines Webangebotes aus einem konkurrierenden Set an Angeboten veranlassen, von Bedeutung für die Ausgestaltung eines Webangebotes (Lowengart & Tractinsky, 2001).

In der vorliegenden Untersuchung soll deshalb neben den Faktoren Nutzungszufriedenheit und Nutzungsintention, die auf die spezifische Akzeptanz eines Webangebots zielen, auch die

Auswahl eines Webangebots aus einem konkurrierenden Set als Erfolgsgröße im Rahmen der Untersuchung der Erfolgsfaktoren und Wirkungszusammenhänge herangezogen werden.

2.5.5 Zwischenfazit

Immer mehr Internet-Nutzer verwenden das Internet zur Durchführung von Recherchen sowie zum online-Einkauf. Dabei besteht ein Unterschied zwischen dem Anteil an Internet-Nutzern, die das Internet für Recherchen nutzen (88,0%) und denen, die online einkaufen (62,0%). Online erworben werden sehr unterschiedliche Produkte und Dienstleistungen. Es gibt aber auch Produkte und Dienstleistungen, zu denen sich ein beachtlicher Teil der Internetnutzer online informiert, diese aber nur begrenzt online erwirbt. Zu diesen gehören beispielsweise Lebens- und Rentenversicherungen zur privaten Altersvorsorge, aber auch Krankenversicherungen. Es handelt sich hierbei um Leistungen, die in der Regel langfristige Investitionen bedingen und den Internetnutzern in der Folge vor einem Kauf häufig ein persönliches Beratungsgespräch erforderlich scheint. Hier übernimmt das Internet neben seiner Funktion als Vertriebskanal, die Funktion einer Plattform bei der Kaufvorbereitung, indem durch eine Online-Präsenz Grundlagen für spätere Offline-Käufe geschaffen werden.

Die Nutzung von Online-Medien bietet den Nachfragern von Gesundheitsleistungen Zugang zu einem nahezu unbegrenzten Angebot an gesundheitsspezifischen Informationen und Serviceleistungen. Hier agieren Krankenversicherungsunternehmen als Anbieter von gesundheitsspezifischen Contents und unterstützen ihre Versicherten bspw. beim Zugang zu Qualitätsinformationen in der medizinischen Versorgung. Digitale Vertriebsformen werden im Rahmen von E-Commerce-Anwendungen hauptsächlich in der Versichertengewinnung sowie im Vertrieb von Ergänzungs- und Wahlтарifen genutzt. Die Erfolgsfaktorenforschung im E-Commerce geht insbesondere der Frage nach, welche Erfolgsfaktoren maßgeblich beeinflussen, ob Anwender einer kommerziellen Website mit deren Nutzung zufrieden sind, sowie eine Intention ausprägen, diese wiederholt zu nutzen bzw. Käufe zu tätigen. Es ist naheliegend, dass diejenigen Anbieter kommerzieller Websites erfolgreich sein dürften, die in der Lage sind, die Vor- und Nachteile des E-Commerce angemessen zu adressieren. Ein wesentlicher Einfluss kommt dabei der Usability zu, die im Abschnitt 2.7 dargestellt wird.

Zuvor wird im nachfolgenden Abschnitt 2.6 dargestellt, dass es bei der Erfassung latenter Konstrukte wie der betrachteten kundennahen Erfolgsfaktoren kommerzieller Websites zwischen zwei Formen von Messmodellen zu differenzieren gilt. Eine Nichtberücksichtigung dieser Unterscheidung kann zu negativen Auswirkungen auf die Messgüte einer Skala führen.

2.6 Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle

Unter einer latenten Variablen wird ein hypothetisches Konstrukt verstanden, das nicht direkt beobachtbar und damit auch nicht direkt messbar ist (Homburg & Dobratz, 1998). Um ein

latentes Konstrukt zu messen bedarf es eines Messmodells, mit Hilfe dessen das Konstrukt anhand von beobachtbaren Indikatoren erfasst wird. Da sich latente Konstrukte in der Beziehung zu ihren Indikatoren unterscheiden können, kann die Erfassung einerseits mittels formativer und andererseits mittels reflektiver Messungen erfolgen (Diamantopoulos & Winkelhofer, 2001).

Bei einem reflektiven Messmodell wird davon ausgegangen, dass die Indikatoren aus dem latenten Konstrukt resultieren (Nunally & Bernstein, 1994). Wohingegen einem formativen Messmodell die Prämisse zugrundeliegt, dass die Indikatoren die latente Variable verursachen (Bollen & Lennox, 1991). Die beiden Modellansätze unterscheiden sich in der Annahme der Kausalität zwischen latenter Variable und Indikatoren und damit in der Richtung der Wirkungsbeziehungen.

2.6.1 Charakteristika reflektiver Messmodelle

Bei einem reflektiven Messmodell folgen die Indikatoren aus der latenten Variable. Dem Messmodell liegt die Prämisse zugrunde, dass die latente Variable die Ausprägungen der beobachtbaren Indikatoren kausal verursacht (Edwards & Bagozzi, 2000). In Abbildung 2-7 zielen die Pfeilspitzen deshalb von der latenten Variable ausgehend auf die Indikatoren.

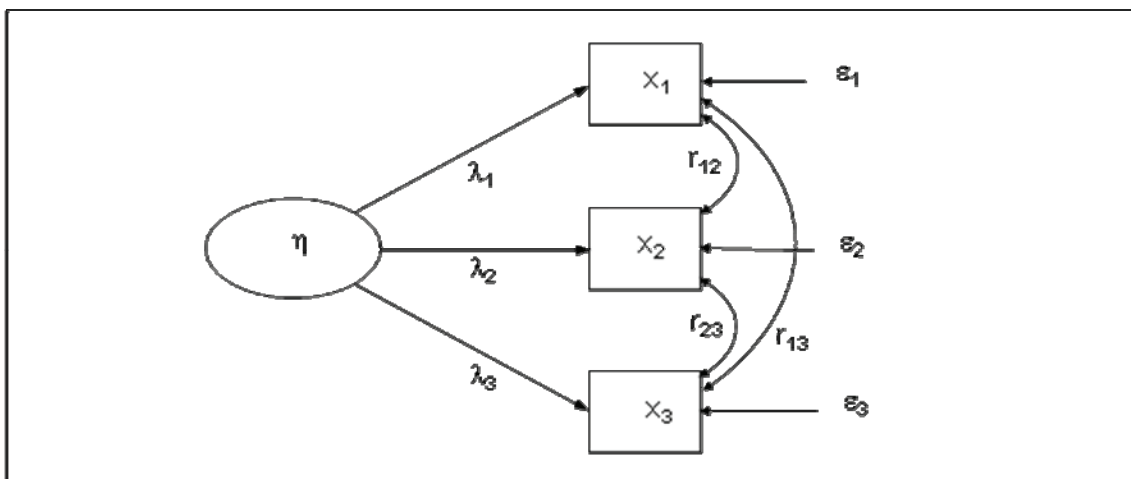


Abb. 2-7: Latentes Konstrukt mit reflektiven Indikatoren

Anmerkung: Latente Variable (η), Faktorladung (λ), reflektiver Indikator (x), Messfehler auf Indikatorebene (ε), Korrelation zwischen den Indikatoren (r).

Jeder Indikator stellt also eine fehlerbehaftete Messung der zugeordneten latenten Variablen dar. Aufgrund dieser Kausalitätsannahme führt eine Veränderung der latenten Variablen zu einer Veränderung aller Indikatorausprägungen, da jeder einzelne Indikator den Wert der zugehörigen latenten Variable reflektiert. Entsprechend sind die Indikatoren austauschbar und hohe Korrelationen zwischen den Indikatoren sind zu erwarten (Bollen & Lennox, 1991). Im Falle fehlerfreier Messungen der latenten Variable durch die reflektiven Indikatoren wären die Korrelationskoeffizienten unter den Indikatoren eins. Je höher hingegen der Messfehler ε_i

eines reflektiven Indikators x_i ausfällt, umso geringer ist ceteris paribus die Korrelation mit den übrigen Indikatoren (Eggert & Fassot, 2003).

Die Bestimmung der reflektiven Indikatoren erfolgt nach der Formel:

$$(1) \quad x_i = \lambda_i \eta + \varepsilon_i$$

wobei:

x_i	reflektiver Indikator i ,
λ_i	die i -te Faktorladung,
η	die latente Variable,
ε_i	der i -te Messfehler auf Indikatorebene

Gemäß der klassischen Testtheorie werden für die Überprüfung der Angemessenheit einer reflektiven Skala das Maß der internen Konsistenz, die Forderung der Eindimensionalität und der Reliabilität herangezogen (Homburg & Giering, 1996). Entsprechend sollen zur Optimierung der Skalenreliabilität einzelne, besonders messfehlerbehaftete Indikatoren mit geringer Faktorladung aus der Skala eliminiert werden (Churchill, 1979).

Diese Prozedur kann zu fehlerhaften Ergebnissen führen, wenn, um eine hohe interne Konsistenz gemäß *Cronbachs α* zu erreichen, aufgrund von Umformulierungen ein und desselben Indikators semantisch kaum mehr zu unterscheidende Items herangezogen werden (Albers & Hildebrandt, 2006). Der tatsächliche Nutzen durch einen Informationszugewinn solcher Items ist gering. Die Entwicklung von Multi-Item-Skalen unter Verwendung des Gütekriteriums der internen Konsistenz führt in diesem Falle nicht zur gewünschten validen Messung der latenten Variable (Bagozzi, 1994).

Typische Beispiele reflektiver Variablen sind psychologische Konstrukte wie Einstellungen oder Persönlichkeitseigenschaften. Beispielsweise stellt in der Self-Awareness-Forschung die Kompetenz „Lernsituationen aktiv zu suchen“ ein latentes Konstrukt dar, das sich anhand der Indikatoren Kreativität, Erfolgsorientierung und Veränderungsbereitschaft beschreiben lässt (Gray & Bromwich, 2001). So zeigen sich Personen, die Lernsituationen aktiv aufsuchen, in der Regel als kreativ, erfolgsorientiert und veränderungsbereit.

Ein Beispiel aus der Marketingwissenschaft für ein mehrdimensionales, reflektives Messmodell stellt MARKOR, ein Konzept zur Implementierung des Marketinggedankens in Unternehmen, dar (Kohli, Jaworski & Kumar, 1993). Die Autoren präsentieren in einem verhaltensorientierten Ansatz drei reflektive Dimensionen der Marktorientierung, die wiederum mit Hilfe reflektiver Items erhoben werden: (a) die Generierung marktrelevanter Informationen, insbesondere zu Kunden und Wettbewerbern, (b) die unternehmensinterne Kommunikation der Informationen sowie (c) die Reaktionsfähigkeit des Unternehmens auf die Informationen und deren Nutzung im Rahmen der Marktbearbeitung.

2.6.2 Charakteristika formativer Messmodelle

In einem formativen Messmodell verursachen die beobachtbaren Indikatoren in einer gewichteten Zusammensetzung das latente Konstrukt (Bollen, 1989; siehe Abbildung 2-8). Die Indikatoren formen dementsprechend als Bausteine das Konstrukt (Rossiter, 2002). Das Messmodell unterscheidet sich von einem reflektiven Messmodell durch die Richtung der Wirkungsbeziehungen zwischen Konstrukt und Indikatoren (Bollen & Lennox, 1991). Aus dieser Kausalitätsannahme resultiert, dass eine Veränderung eines einzelnen Indikators eine Veränderung des Konstrukts verursacht, wobei die Ausprägungen der übrigen Indikatoren hiervon unbeeinflusst bleiben können.

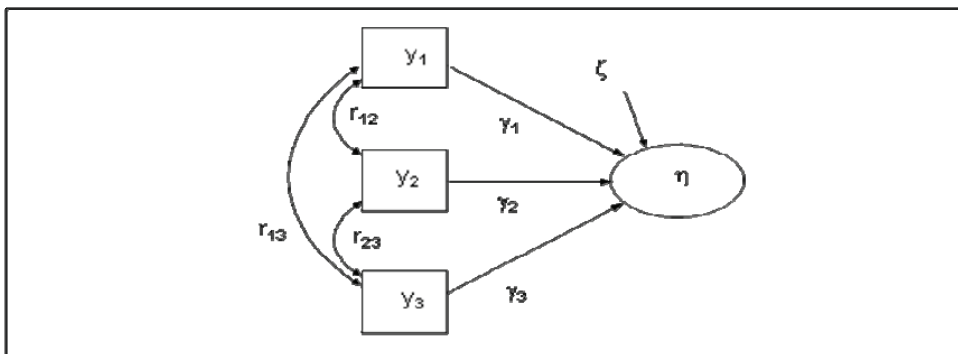


Abb. 2-8: Latentes Konstrukt mit formativen Indikatoren

Anmerkung: Latente Variable (η), Gewicht (γ), formativer Indikator (y), Messfehler auf Ebene der latenten Variable (ζ), Korrelation zwischen den Indikatoren (r).

Mathematisch ist die latente Variable (η) als Linearkombination der formativen Indikatoren und einem Fehlerterm definiert (Bollen & Lennox, 1991). Die Spezifikation eines formativen Messmodells erfolgt nach der Formel:

$$(2) \quad \eta = \gamma_1 y_1 + \gamma_2 y_2 + \dots + \gamma_n y_n + \zeta$$

wobei:

- η die latente Variable,
- y_n der n-te formative Indikator
- γ_n das n-te Gewicht
- ζ der Messfehler auf Ebene der latenten Variable

Während reflektiv gemessene Konstrukte aufgrund der Forderung nach interner Konsistenz der Skalen immer nur eine Facette messen, sollen mit Hilfe von formativen Indikatoren unterschiedliche Facetten eines Konstrukts erfasst werden, die gemeinsam wiederum das Konstrukt definieren. Eine hohe Korrelation der Indikatoren wie im Falle des reflektiven Messmodells ist dabei, wenn auch möglich, nicht als notwendig zu erwarten (Diamantopoulos, 1999). Vielmehr können die Korrelationen r zwischen den Indikatoren alle Werte im Bereich -1 bis +1 annehmen (Nunnally & Bernstein, 1994, S. 489).

Da formative Indikatoren unterschiedliche Facetten eines Konstrukts erfassen, geht die Entfernung eines Indikators immer auch mit dem Verlust einer inhaltlichen Facette der unter-

suchten Variable einher (Rossiter, 2002). Die Nichtberücksichtigung eines formativen Indikators führt zu einer Änderung des konzeptionellen Inhalts des zu operationalisierenden Konstrukts und zu einer Verringerung der Messgüte (Eggert & Fassot, 2003). Entsprechend ist eine Skalenbereinigung nach den Kriterien der klassischen Testtheorie, insbesondere die Eliminierung von intern nicht konsistenten Indikatoren, nicht möglich. Eine Bereinigungsverfahren wie sie bei reflektiven Indikatoren angemessen ist, ist bei formativen Indikatoren also nicht akzeptabel, da sie möglicherweise wichtige Facetten eines Konstrukts ausschließt. Vielmehr empfiehlt Rossiter (2005) zur Beurteilung der Messgüte eine Einschätzung zur Güte des zugrunde liegenden Skalenentwicklungsprozesses vorzunehmen.

Im Falle eines formativen Messmodells ist bereits zu Beginn eines Untersuchungsprozesses eine klare und inhaltlich erschöpfende Definition des zu untersuchenden Konstrukts mit all seinen Facetten vorzunehmen (Rossiter, 2002). Die Einbindung von Experten aus dem Forschungsgebiet sowie die Durchführung von Voruntersuchungen wie beispielsweise Fokus-Gruppen, schriftlichen Befragungen oder Interviews können hierbei hilfreich sein. Dabei ist zu beachten, dass die Facetten eines Konstrukts sowohl den inhaltlichen Umfang als auch die Reichweite der Aussagen bestimmen, die im Rahmen einer Untersuchung gewonnen werden können (Hildebrandt, 1986). Dies gilt insbesondere, wenn im Rahmen der Erfolgsfaktorenforschung konkrete Handlungsempfehlungen für das Management abgegeben werden sollen. Sollen die Treiber identifiziert werden, die den Erfolg beeinflussen, so ist es erforderlich, dass die exogenen Indikatoren formativer Natur sind und danach ausgesucht werden, dass sie inhaltlich beeinflussbar und potentiell erfolgsrelevant sind (Albers, 2007; Albers & Hildebrandt, 2006).

Ein Beispiel für formative Messmodelle ist der *sozioökonomische Status* einer Person. Dieser setzt sich unter anderem zusammen aus Messungen zu den Größen Bildung, Einkommen und Prestige des Berufs (Hauser, 1973, S. 268). Wenn eine der Größen sich in ihrer Ausprägung verändert, verändert sich auch der Wert des sozioökonomischen Status. Entsprechend bedingt eine Änderung des sozioökonomischen Status einer Person eine Veränderung der Ausprägung mindestens einer der Ausgangsgrößen, jedoch nicht zwingend aller Messwerte.

Eine weitere etablierte Anwendung formativer Messmodelle ist der *American Customer Satisfaction Index (ACSI)*. Er ermittelt für den Wirtschaftsraum USA über 43 Branchen den Grad der Kundenzufriedenheit und führt hierzu Messungen zu verschiedenen Größen wie Kundenerwartungen, Qualitäts-, Zufriedenheits- und Nutzeneinschätzungen zu einem Wert zusammen (Fornell, 2007).

2.6.3 Fehlspezifikation des Messmodells

Bei lediglich eindimensionalen Messmodellen lassen sich zwei Arten von fehlspezifizierten Modellen beschreiben (Eberl, 2004). Einerseits kann irrtümlicherweise ein formatives Messmodell angenommen werden, obwohl tatsächlich die latente Variable die Ursache der

Indikatoren darstellt und entsprechend ein reflektives Messmodell angemessen wäre. Andererseits kann ein in der Realität formativer Zusammenhang zwischen latenter Variable und Indikatoren fälschlicherweise im Modell als reflektiv spezifiziert werden.

Untersuchungen zeigen, dass es bei Fehlspezifikationen von Messmodellen zu fehlerbehafteten Parameterschätzungen der Beziehungen zwischen den Konstrukten und zwangsläufig zu falschen Interpretationen und Ableitungen bezüglich der postulierten Modellbeziehungen kommen kann (Jarvis, Mackenzie & Podsakoff, 2003; Law & Wong, 1999). So zeigen Law und Wong (1999) in einer Untersuchung zweier kovarianzbasierter Strukturgleichungsmodelle, die sich dahingehend unterscheiden, dass bei identischen Indikatoren für die LV ein reflektives bzw. ein formatives Messmodell angenommen werden, dass die Parameterschätzungen für die Strukturbeziehungen signifikant voneinander abweichen. In einer ähnlichen Untersuchung von Jarvis et al. (2003) kam es im Falle von Fehlspezifikationen ebenfalls zu einem substantiellen Bias bei den Strukturparametern. Die Ergebnisse weisen darauf hin, dass die Koeffizientenwerte von Pfaden, die auf eine reflektive fehlspezifizierte LV hinführen, im Falle von Fehlspezifikation unterschätzt werden, wohingegen sie im Falle von wegführenden Pfaden eher überschätzt werden.

Albers und Hildebrandt (2006) argumentieren, dass letztendlich nicht geprüft werden könne, ob eine Konstrukt-Operationalisierung als reflektiv oder formativ falsch oder richtig sei. Vielmehr könne die Operationalisierung einer fehlspezifizierten LV zu einem unpassenden Modell führen, das die Qualität der Untersuchungsergebnisse einschränke, in seiner inhaltlichen Bedeutung deshalb aber nicht unbedingt falsch sein müsse.

Eine irrtümlich formative Spezifikation eines Konstrukts wäre insofern problematisch, als dass die für den Fall eines reflektiven Messmodells anzuwendenden Empfehlungen zur Skalenbildung nach Churchill (1979), insbesondere zur Skalenbereinigung über das Maß der internen Konsistenz, nicht beachtet würden. Dies kann zu einer Beibehaltung irrelevanter Indikatoren sowie zu einer Verzerrung der Parameterschätzer führen (Eberl, 2004).

Im Gegensatz dazu führt eine irrtümlich reflektive Spezifikation eines tatsächlich formativen Messmodells zu einer Anwendung der klassischen Skalenbereinigungsprozedur nach dem Prinzip der internen Konsistenz. Wird diese auf Indikatoren formativer Natur angewandt, kann dieses Vorgehen zu einer unangemessenen Indikatorelimination führen (Eggert & Fassot, 2003). Durch die Fehlspezifikation eines formativen Messmodells werden möglicherweise wichtige Facetten des Konstrukts vernachlässigt, sodass sich die Ergebnisse inhaltlich verändern. Die Annahme reflektiver statt richtigerweise formativer Indikatoren schränkt das Modell in seiner Aussagekraft ein (Albers, 2007; Albers & Hildebrandt, 2006). So begrenzt eine unangemessene Elimination von Indikatoren den Bedeutungsinhalt des Konstrukts und man verzichtet auf den Informationsgewinn bezüglich des Erfolgsbeitrags einzelner Indikatoren.

Verschiedene Metastudien setzen sich mit der Häufigkeit der Anwendung fehlspezifizierter Messmodelle in wissenschaftlichen Veröffentlichungen auseinander (Eberl, 2004; Eggert & Fassot, 2003; Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Unabhängig von der Frage der Angemessenheit der verwendeten Modellspezifikation ist eine deutliche Dominanz des reflektiven Messmodells bei über 95% der untersuchten Konstrukte festzustellen. Entsprechend gering fällt der Anteil der Publikationen aus, in denen Messmodelle fälschlicherweise formativ spezifiziert werden. Der Anteil an Studien, in denen die verwendeten Konstrukte fälschlicherweise reflektiv spezifiziert werden, obwohl die Indikatoren eher formativer Natur sind, fällt hingegen teilweise sehr hoch aus. So stellen beispielsweise Eggert und Fassot (2003) auf Grundlage der Durchsicht von 30 Artikeln in der Zeitschrift *Marketing ZfP* in der überwiegenden Mehrheit von 79,6% fest, dass die untersuchten Messmodelle fälschlicherweise reflektiv spezifiziert sind. Dabei ist zu vermuten, dass die in der Betriebswirtschaftslehre verwendeten Konstrukte aufgrund des Bemühens um Identifikation erfolgsrelevanter Indikatoren zur Ableitung gestalterischer Maßnahmen tendenziell eher formativer Natur sind (Albers & Hildebrandt, 2006).

Diese aus messtheoretischer Sicht nicht gerechtfertigte Dominanz und Fehlspezifikation führen Eggert und Fassot (2003) neben einer Unkenntnis der Unterschiede der Messmodellvarianten unter anderem darauf zurück, dass die gängigen Softwarepakete zur Analyse von Strukturgleichungsmodellen wie LISREL oder AMOS (Arbuckle & Wothke, 1999; Jöreskog & Sörbom, 1996) standardmäßig reflektive Messmodelle anbieten. Diese standardisierten Softwarepakete basieren auf dem Verfahren der Kovarianzstrukturanalyse und sind in der Anwendung auf formative Indikatoren problematisch, da unter anderem ein Identifizierbarkeitsproblem impliziert wird (Eberl, 2004; MacCallum & Browne, 1993). Eine Berücksichtigung formativer Konstrukte ist in LISREL zwar grundsätzlich möglich, allerdings sehr aufwändig (Albers & Hildebrandt, 2006). Die Einbindung der formativen Indikatoren hat dabei als eigenständige LVn zu erfolgen (MacCallum & Browne, 1993)

Neben den kovarianzbasierten Verfahren zur Analyse von Strukturgleichungsmodellen besteht in Partial Least Squares (PLS) ein regressionsbasiertes Verfahren, das dem Anbieter standardmäßig beide Messmodelle anbietet (Wold, 1966). Somit können sowohl reflektive als auch formative Messmodelle problemlos bearbeitet werden. Allerdings kann es auch bei PLS-Modellen zu stark fehlerbehafteten Parameterschätzungen der Strukturbeziehungen aufgrund einer fälschlichen Bereinigung der Indikator-Menge kommen, wenn ein formatives Konstrukt irrtümlich als reflektiv fehlspezifiziert wird (Albers & Hildebrandt, 2006).

PLS zeichnet sich gegenüber den kovarianzbasierten Verfahren durch weitere Vorteile aus, wie beispielsweise geringere Anforderungen an die Datenverteilung und den Stichprobenumfang sowie eine bessere Eignung zur Prognose abhängiger LVn. Nachteilig ist zu erwähnen, dass die Parameterschätzer im statistischen Sinne nicht konsistent sind und die Korrelationen zwischen den LV eher unterschätzt werden (Dijkstra, 1983; Wold, 1982).

Aufgrund der besonderen Eignung des Verfahrens Partial Least Squares zur Modellierung sowohl reflektiver als auch formativer Messmodelle wird in der aktuellen Untersuchung das Verfahren *SmartPLS* (Hansmann & Ringle, 2004) eingesetzt.

2.6.4 Indexentwicklung zur Operationalisierung formativer Konstrukte

Angesichts der dargestellten Häufigkeit und Einschränkungen fälschlicherweise als reflektiv spezifizierter Messmodelle soll nachfolgend ein Verfahren zur Entwicklung von Indizes zur Erfassung formativer LVn dargestellt werden. Das Verfahren orientiert sich an Diamantopoulos und Winkelhofer (2001) und erfolgt in fünf Prozessschritten. Auf der Grundlage einer sorgfältigen Definition des Konstrukts erfolgt im zweiten Schritt eine Definition der relevanten Indikatoren. Diese sind auf Multikollinearität zu überprüfen. Nach einer Schätzung und Prüfung der Güte des Messmodells erfolgt abschließend die Indexberechnung.

Die Entwicklung reflektiver Messmodelle erfolgt auf Basis der allgemein verbreiteten Ansätze zur Skalenentwicklung unter Berücksichtigung der Kriterien der klassischen Testtheorie (Churchill, 1979; Homburg & Giering, 1996). Das Vorgehen erschließt sich aus den bisherigen und folgenden Ausführungen (siehe insbesondere Abschnitte 4.3, 4.4 sowie 6.5) und wird nicht näher dargestellt.

2.6.4.1 Konstruktdefinition

Im ersten Schritt erfolgt analog zum Vorgehen bei Churchill (1979) eine Abgrenzung und sorgfältige Definition des Umfangs und Inhalts des zu untersuchenden latenten Konstrukts. Da unter der Annahme eines formativen Messmodells die latente Variable durch die Indikatoren bestimmt wird, ist die Bedeutung für den darauf folgenden Prozessschritt der Definition der Indikatoren jedoch ungleich größer. Die Facetten eines Konstrukts bestimmen sowohl den inhaltlichen Umfang als auch die Reichweite der Aussagen, die im Rahmen der Untersuchung gewonnen werden können. Somit hat die Definition des Konstruktes unmittelbare Wirkung auf die nachfolgende Definition der Indikatoren (Nunally & Bernstein, 1994, S. 333). Nach Rossiter (2002) erweist es sich als sinnvoll, zunächst eine Definition des Bewertungsobjekts, der Bewertungsattribute sowie eine Abgrenzung der Zielgruppe vorzunehmen.

Die Entscheidung bezüglich der Spezifikationsart lässt sich anhand verschiedener unterstützender Fragen ableiten, die in Tabelle 2-3 in drei Kategorien dargestellt werden. Die erste Gruppe bezieht sich auf die Richtung der Kausalität zwischen Konstrukt und Indikatoren. Im Falle eines formativen Messmodells geht die Wirkungsbeziehung zwischen Konstrukt und Indikatoren von den Indikatoren aus, wohingegen bei reflektiven Messmodellen die Kausalität vom Konstrukt ausgeht. Die Fragen der zweiten Gruppe zielen auf die Austauschbarkeit von Indikatoren. Während die Indikatoren im Falle eines formativen Konstrukts nicht austauschbar sein müssen, ist dies bei reflektiven Konstrukten notwendigerweise erforderlich.

Abschließend dient das Kriterium der Korrelation zwischen Indikatoren und Konstrukt der Unterscheidung formativer und reflektiver Messmodelle. Bei einem reflektiven Messmodell ist zwingend eine hohe Korrelation zwischen den Indikatoren zu erwarten, da diese von demselben Konstrukt reflektiert werden. Im Falle eines formativen Messmodells ist dies, wenn auch möglich, nicht notwendigerweise erforderlich.

Tab. 2-3: Entscheidungshilfen zur Frage, ob ein Konstrukt reflektiv oder formativ ist (Eigene Darstellung in Anlehnung an Christophersen & Grape, 2006; Eberl, 2004; Jarvis, MacKenzie & Podsakoff, 2003)

<i>Richtung der Kausalität zwischen Konstrukt und Indikatoren</i>	
⇒ Definieren die Indikatoren in ihrer Kombination das Konstrukt?	→ ja: <i>formativ</i>
⇒ Sind die Indikatoren Ausdruck des Konstrukts?	→ ja: <i>reflektiv</i>
⇒ Stehen die Indikatoren zeitlich vor dem Konstrukt?	→ ja: <i>formativ</i>
⇒ Steht das Konstrukt zeitlich vor den Indikatoren?	→ ja: <i>reflektiv</i>
<i>Elimination und Austauschbarkeit der Indikatoren</i>	
⇒ Verändert sich bei Elimination eines Indikators die inhaltliche Aussage des Konstrukts?	→ ja: <i>formativ</i>
⇒ Sind die Indikatoren untereinander austauschbar?	→ ja: <i>reflektiv</i>
⇒ Besitzen die Indikatoren denselben oder einen ähnlichen inhaltlichen Kern?	→ ja: <i>reflektiv</i>
<i>Korrelation zwischen Indikatoren und Konstrukt</i>	
⇒ Besteht zwischen den Indikatoren teilweise oder insgesamt keine hohe Korrelation?	→ ja: <i>formativ</i>
⇒ Bewirkt die Veränderung der Ausprägung eines von mehreren Indikatoren notwendigerweise eine Veränderung aller übrigen Indikatoren in gleicher Weise?	→ ja: <i>reflektiv</i>
⇒ Bewirkt eine Veränderung eines oder mehrerer Indikatoren notwendigerweise eine Veränderung des Konstrukts?	→ ja: <i>formativ</i>
⇒ Bewirkt eine Veränderung des Konstrukts notwendigerweise eine Veränderung aller Indikatoren?	→ ja: <i>reflektiv</i>

2.6.4.2 Bestimmung der Indikatoren

Die Auswahl der formativen Indikatoren hat anderen Anforderungen zu entsprechen als im Falle reflektiver Indikatoren. Da formative Indikatoren unterschiedliche Facetten eines Konstrukts erheben ist über die *Indikatorauswahl* eine vollständige Erfassung aller definitorischen Bestandteile des zu untersuchenden Konstrukts nötig (Bollen & Lennox, 1991). Die Nichtberücksichtigung eines formativen Indikators führt zur Vernachlässigung einer inhaltlichen Facette und damit zu einer Änderung des konzeptionellen Inhalts des zu operationalisierenden Konstrukts. Eine nachträgliche Elimination von Items kommt bei formativen Konstrukten daher nicht in Frage. Eine Überprüfung von Reliabilitäten auf Basis von Item-to-Total-Korrelationen ist nicht zielführend. Vielmehr ist die Berücksichtigung aller

inhaltlichen Gesichtspunkte erforderlich, so dass eine valide Abbildung des Bedeutungsinhalts des Konstrukts gewährleistet wird.

Konstrukte können sich allerdings als dermaßen komplex und schwer zu beschreiben erweisen, dass nicht alle Facetten durch Indikatoren abgebildet werden können. In diesem Fall ist eine Auswahl an formativen Indikatoren zusammenzustellen, die das Konstrukt in seiner umfassenden Bedeutung angemessen abdecken. Ebenso kann eine Skalenbereinigung durch Ausschluss von Indikatoren im Zuge einer Abwägung durchaus sinnvoll sein, wenn der ergänzende Beitrag zur Erklärung des Konstrukts und die Steigerung der Messgüte marginal sind und den zusätzlichen Erhebungsaufwand nicht rechtfertigen (Christophersen, 2007).

Dabei ist der Anspruch an die Qualität der Indikatoren hoch anzusetzen. Bei der Formulierung gilt es Aspekte zu beachten, durch die ein *Method Bias* möglichst gering gehalten wird (Podsakoff et al., 2003). So ist dafür Sorge zu tragen, dass:

- ⇒ die Items für die Antwortenden verständlich formuliert sind und unklare Begriffe möglichst vermieden oder erläutert werden,
- ⇒ sind vage Konzepte doch erforderlich, so sind sie zu erläutern oder durch veranschaulichende Beispiele zu hinterlegen,
- ⇒ Items sollten in ihrer Formulierung möglichst kurz und spezifisch ausfallen,
- ⇒ ungewöhnliche und mehrdeutige Formulierungen sind zu vermeiden,
- ⇒ jedes Einzelitem sollte sich stets nur auf einen möglichst konkreten Beurteilungsaspekt beziehen.

Dabei sollten zu große Ähnlichkeiten in den Formulierungen der Items vermieden werden. Ansonsten besteht die Gefahr, dass diese, auch wenn sie unterschiedliche Bewertungsaspekte adressieren, als inhaltlich zusammengehörig wahrgenommen werden können und in der Folge möglicherweise ungerechtfertigt hohe Korrelationen in den Bewertungen aufweisen.

Dennoch können sich Indikatoren in der konkreten Anwendung als ungeeignet zur Konstrukt-messung erweisen, “because they are found to either not tap their intended constructs or also tap other, unintended constructs in the set” (Anderson & Gerbing, 1991). Zur Sicherstellung der inhaltlichen Validität, kann es daher sinnvoll sein, die Indikatoren einem Pre-Test zu unterziehen (Anderson & Gerbing, 1991; Eggert & Fassot, 2003; Rossiter, 2002). Dabei werden fachliche Experten oder repräsentative Rater gebeten, Indikatoren den ihrer Meinung nach zugehörigen Konstrukten zuzuordnen und bezüglich ihrer Verwendbarkeit zu beurteilen. Idealerweise sollte eine empirische Voruntersuchung stattfinden. Dabei ist es von Bedeutung, eine ausreichend große Stichprobe zu Grunde zu legen. Mendoza, Stafford und Stauffer (2000) geben für eine Testneuentwicklung als Mindestgrenze $N = 100$ an, da erst ab einer solchen Stichprobengröße die Reliabilität eines Tests zuverlässig geschätzt werden könne. Hierbei ist entsprechend der Spearman-Brown-Korrekturformel der Einfluss der Itemanzahl auf die Reliabilität zu berücksichtigen. Auf der Grundlage vorliegender Daten können die

Indikatoren dann hinsichtlich der *Anteile fehlender Werte* sowie bezüglich ihrer *Itemschwierigkeiten* überprüft und angepasst werden.

2.6.4.3 Multikollinearität

Multikollinearität beschreibt den Grad der linearen Abhängigkeit zwischen den formativen Indikatoren (Schneider, 2006). Formative Messmodelle beruhen auf dem Prinzip der multiplen Regressionsanalyse, der die Annahme zugrundeliegt, dass die unabhängigen Variablen voneinander linear unabhängig sind (Wang, 1996). Im Falle von Multikollinearität lässt sich ein formativer Indikator jedoch als Linearkombination der anderen Indikatoren darstellen. Entsprechend stellt Multikollinearität bei einer hohen linearen Abhängigkeit zwischen den formativen Indikatoren ein Problem dar (Backhaus et al., 1996, S. 33). Mit zunehmender linearer Abhängigkeit zwischen den Indikatoren werden die Standardfehler der Gewichte (γ_i) größer und damit deren Schätzungen ungenauer.

Eine mögliche Herangehensweise zur Verringerung von Multikollinearität besteht darin, betroffene Indikatoren zu eliminieren (Bollen & Lennox, 1991; Wang, 1996). Basis dieser Vorgehensweise ist die Überlegung, dass einer von mehreren hochkorrelierten Indikatoren ausreicht, um die Varianz der latenten Variable zu erklären. Dieses Vorgehen ist aufgrund möglicher inhaltlicher Auswirkungen auf das Gesamtmodell und dessen inhaltliche Validität jedoch als kritisch zu bewerten. Der Ausschluss eines Indikators aufgrund seines Anteils an der Varianzaufklärung ohne Betrachtung inhaltlicher Aspekte kann dazu führen, dass vom ursprünglichen Modell abgewichen wird und eine Fehlspezifikation vorliegt (Steffen, 1994, S. 28).

Albers und Hildebrand (2006) empfehlen, um Multikollinearität zu vermeiden, die untereinander multikollinearen Indikatoren nicht zu eliminieren, sondern diese zu einem Index zusammenzufassen. Dieser ist im Rahmen des Messmodells wie ein Indikator zu handhaben. So können Schwierigkeiten bei der Auswahl zu eliminierender Indikatoren und der Verlust möglicherweise wichtiger Facetten des Konstrukts vermieden werden (Grapentine, 1997). Die Verfahren zum Nachweis von *Multikollinearität* werden in Abschnitt 6.8 näher dargestellt.

2.6.4.4 Schätzung des formativen Messmodells

Im Falle formativer Konstrukte ist die Schätzung eines isolierten Messmodells, wie dies bei einem reflektiven Konstrukt möglich ist, aufgrund einer statistischen Unteridentifizierung nicht durchführbar (Bollen & Lennox, 1991). Vielmehr ist es erforderlich, das formative Schätzmodell in ein Strukturgleichungsmodell einzubinden. Dies kann beispielsweise in einem sogenannten *Zwei-Konstrukt-Modell* erfolgen (Bollen, 1989).

Dabei wird eine reflektive LV hinzugezogen, bei der zu erwarten ist, dass sie vom interessierenden formativen Konstrukt beeinflusst wird. Dieses wird in einem Strukturgleichungsmodell als Prädiktor zur endogenen reflektiven Variable in Beziehung gesetzt (Diamantopo-

lous & Winkelhofer, 2001). Ein Beispiel für ein Zwei-Konstrukt-Modell wird in Abbildung 2-9 dargestellt. Von besonderer Bedeutung ist dabei die Ausprägung des Pfadkoeffizienten β_{21} , der die Stärke des Zusammenhangs zwischen den zwei Konstrukten beschreibt. Damit das Modell identifizierbar ist, ist es erforderlich, die Fehlervarianz der latenten formativen Variable η_1 auf null zu setzen und je einen Indikatorpfad der beiden latenten Variablen zu fixieren. In der vorliegenden Untersuchung soll die Usability sowohl reflektiv als auch formativ erfasst werden. In diesem Fall kann die reflektive Usability-Skala als abhängige LV in einem Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala herangezogen werden.

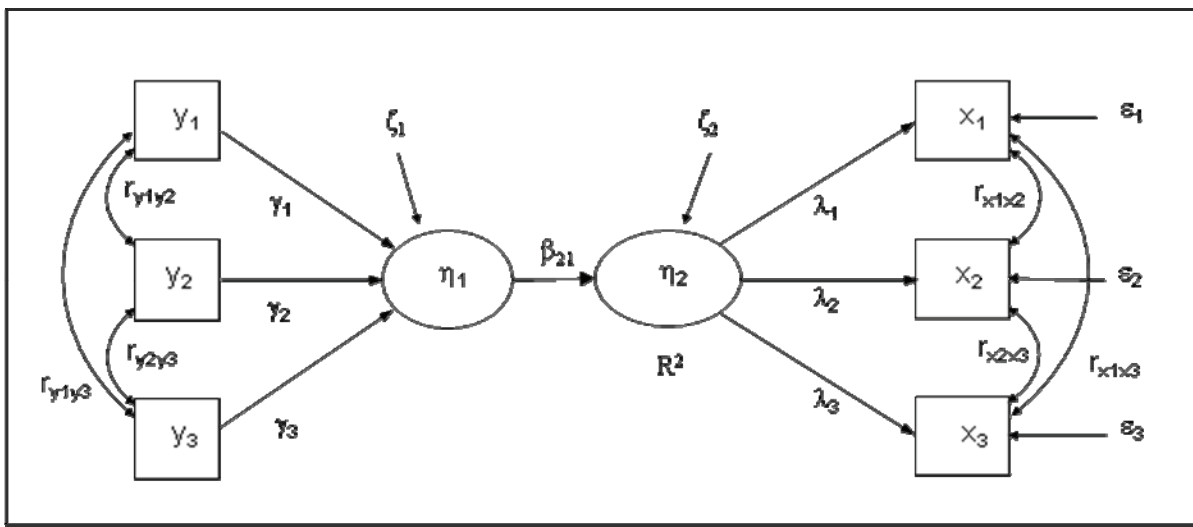


Abb. 2-9: Zwei-Konstrukt-Modell zur Bestimmung der Validität einer formativen LV (Diamantopoulos & Winkelhofer, 2001)

Anmerkung: Latente Variable (η), Gewicht (γ), Faktorladung (λ), formativer Indikator (y), reflektiver Indikator (x), Messfehler auf Ebene der latenten Variable (ζ), Messfehler auf Indikatorebene (ε), Korrelation zwischen den Indikatoren (r), Pfadkoeffizient (β), Determinationskoeffizient (R^2).

2.6.4.5 Indexberechnung

Nach der Bestimmung der Gewichte (γ) für die formativen Indikatoren durch PLS, kann der Index und damit die Werte der formativen LV berechnet werden. Die Berechnung wird anhand der in Abschnitt 2.6.2 bereits dargestellten Formel

$$\eta = \gamma_1 y_1 + \gamma_2 y_2 + \dots + \gamma_n y_n + \zeta \quad (3)$$

zur Spezifikation formativer Messmodelle vorgenommen. Hierfür werden die Werte der Indikatorvariablen mit den zugehörigen Gewichten multipliziert und aufaddiert. Der Wert für den Messfehler ζ wird auf Null gesetzt und in der Berechnung nicht berücksichtigt. Die Werte für die latente formative Variable werden durch die PLS-Software berechnet und ausgewiesen.

2.7 Usability

Im folgenden Kapitel soll das Konstrukt Usability näher betrachtet werden. Nach einem Überblick zur Thematik (siehe Abschnitt 2.7.1) wird in Abschnitt 2.7.2 dargestellt, wie der Begriff Usability im Rahmen der DIN EN ISO 9241 gefasst wird. In den Abschnitten 2.7.3 und 2.7.4 werden allgemeine Prinzipien einer nutzungsfreundlichen Gestaltung interaktiver Anwendungen vorgestellt, die in der Literatur vielfach zur theoretischen Fundierung des Konzeptes angeführt werden. Abschließend wird in Abschnitt 2.7.5 die Bedeutung der Usability als Erfolgsfaktor von Webangeboten dargestellt.

2.7.1 Begriffliche und konzeptionelle Basis

Für Usability, nach Dzida & Wandke (2006, S. 464) auch Gebrauchstauglichkeit oder Benutzbarkeit, existiert keine einvernehmliche Definition (vgl. Fitzpatrick & Higgins, 1998). Während Nielsen (1993, S. 26 ff.) die fünf Kriterien *Effizienz*, *Fehlerrate*, *Erinnerbarkeit*, *Lernbarkeit* und *subjektive Zufriedenheit* einbezieht, nennt Shackel (1991) als vier Komponenten der Usability *Effektivität*, *Erlernbarkeit*, *Flexibilität* sowie *Einstellung*. Er definiert Usability als "... capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfil the specified range of tasks, within the specified range of environmental scenarios ... the capability to be used by humans easily and effectively" (Shackel, 1991, S. 24) bzw. zu späterem Zeitpunkt als „... an emergent quality of an optimum design, which is reflected in the effective and satisfying use of the IT" (Sweeny, Maguire & Shackel, 1993, S. 590). In der ISO-NORM 9241 werden hingegen *Effektivität*, *Effizienz* und die *Zufriedenheit des Anwenders* als Komponenten der Usability benannt.

Usability kann konzeptionell als Dimension der Systemakzeptanz eingeordnet werden (Shackel, 1991). Der Autor bezeichnet die Akzeptanz von technischen Systemen als eine Funktion der drei unabhängigen Dimensionen Usability, Utility und Likeability, wobei Systemakzeptanz vorliegt, wenn Anwender "... gegenüber dem System eine positive Einstellung im Sinne einer grundsätzlichen Anwendungsbereitschaft aufweisen, und gleichzeitig eine aufgabenbezogene Nutzung des System zu beobachten ist" (Rengelshausen, 1999, S. 73).

Auch Nielsen (1993) benennt Usability und Utility (Nützlichkeit) als wichtige Faktoren der Systemakzeptanz, da sie die grundsätzliche Anwendbarkeit bzw. Brauchbarkeit (Usefulness) eines technischen Systems determinieren. Usefulness, und damit Utility und Usability, ist neben den Systemkosten, der Systemkompatibilität zu weiteren Systemen und der Systemzuverlässigkeit eines von vier Systemattributen, die die praktische Akzeptanz eines technischen Systems bestimmen. Soziale Systemakzeptanz zielt hingegen auf die Kongruenz zwischen Systemeigenschaften und Werte- und Normenstrukturen des Individuums.

Sowohl Shackel (1991) als auch Nielsen (1993) grenzen Usability von der Utility eines technischen Systems ab. So definiert Nielsen (1993, S. 25) wie folgt: „... utility is the question

of whether the functionality of the system in principle can do what is needed, and usability is the question of how well users can use that functionality."

Utility zielt demnach darauf ab, inwieweit eine E-Commerce-Anwendung über den zur Erreichung des Nutzungszieles erforderlichen Umfang und Qualität der technischen Funktionalitäten, Services und Inhalte verfügt. Beispielsweise sollten Websites von Krankenkassen neben allgemeinen Navigationselementen und Suchfunktionen weitere Funktionalitäten und Inhalte bereitstellen, die den Nutzungsprozess unterstützen, damit der Besucher der Website sein angestrebtes Ziel erreichen kann.

In Abgrenzung zur Usability zielt *Accessibility* (Barrierefreiheit) auf die gebrauchstaugliche Nutzung von interaktiven Systemen (einschließlich der Webanwendungen) für Benutzer, die in ihrer Mobilität, ihrem Seh- oder Hörvermögen eingeschränkt sind. Accessibility ist eine vom Gesetzgeber geforderte harte Anforderung nach dem Behindertengleichstellungsgesetz (BGG §§ 4, 11; 2002) und weist gegenüber der Usability ein klareres Anforderungsprofil auf. In Bezug auf interaktive Systeme wird Accessibility häufig auf technische Qualitätsmerkmale reduziert, so beispielsweise in den Empfehlungen der „Web Accessibility Initiative (WAI)“, des World Wide Web Consortiums (W3C).

2.7.2 Europäische NORM EN ISO 9241

Die europäische Norm EN ISO 9241 „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ umfasst in einer eher operativ geprägten Definition Effektivität, Effizienz und Zufriedenheit als Komponenten der Usability: "Usability of an application refers to the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments" (DIN-ISO 9241-11).

Auf diese Definition, die auch als Grundlage für das weitere Vorgehen im Rahmen der vorliegenden Arbeit gewählt wird, wird sowohl in der Usability-Praxis als auch in der Literatur zur Human-Computer Interaction häufig zurückgegriffen (u.a. Dzida & Wandke, 2006, S. 479; Frokjaer, Herztum & Hornbaek, 2000; Jordan, Thomas & McClelland, 1996; MacLeod, Bowden, Bevan & Curson, 1997). Mit dieser Arbeitsdefinition geht die Prämisse einher, dass das Auffinden von (Produkt-) Informationen und die Durchführung von Transaktionen das primäre Anliegen der Besucher einer Website ist. Die Ausführung dieses Anliegens wird durch die Funktionen, Inhalte und Services der Website ermöglicht. Dabei können die Effektivität und Effizienz der Durchführung gesteigert werden, wenn die Anforderungen, die sich aus der spezifischen Aufgabenstellung ergeben, durch die Ausgestaltung der Funktionalitäten, Services und Inhalte unterstützt werden.

Die Anforderung an eine Website besteht somit darin, einen effektiven, effizienten und zufriedenstellenden Zugang zu Informationen und Leistungsangeboten, sowie eine effektive, effiziente und zufriedenstellende Durchführung von Transaktionen sicherzustellen (Harms & Schweibenz, 2000, S. 62).

Unter *Effektivität* im Sinne der Norm ISO 9241-11 (Anforderungen an die Gebrauchstauglichkeit) ist zu verstehen, mit welchem Grad an Präzision und wie vollständig ein Nutzer seine individuellen Ziele in der Interaktion mit dem Angebot erreichen kann (vgl. Dzida & Wandke, 2006, S. 465). Indikatoren für Effektivität können beispielsweise die Qualität der Aufgabenerledigung oder die Fehlerrate sein. Die ISO 9241-11 zielt auf die Gebrauchstauglichkeit aus Anwenderperspektive ab und berücksichtigt dabei funktionale und softwaretechnische Merkmale nicht bezüglich ihrer technischen Güte, sondern in Bezug auf ihre Wirkungen auf den Benutzer. In Ergänzung hierzu werden in der ISO/IEC 12119 softwaretechnische Eigenschaften von Softwareprodukten definiert, die als Voraussetzung für deren Benutzbarkeit und die Erreichung der Nutzungsziele gegeben sein müssen. In vielen Fällen kann der konkrete Nutzungskontext, für den ein Softwareprodukt entwickelt wird, nicht genau beschrieben werden. Bei den Anforderungen zur Prüfung der Benutzbarkeit nach ISO/IEC 12119 wird deshalb kein konkreter, sondern ein intendierter Nutzungskontext, z.B. hinsichtlich der beabsichtigten Aufgabenstellung, zu Grunde gelegt.

Die Relation zwischen dem Präzisionsgrad und der Vollständigkeit der Aufgabenlösung einerseits und den hierfür benötigten Ressourcen andererseits bringt die *Effizienz* der Interaktion nach ISO 9241-11 und ISO 9241-110 (Grundsätze der Dialoggestaltung) zum Ausdruck. Eine mögliche Operationalisierung der Effizienz ist beispielsweise die benötigte Zeit zur Aufgabenbewältigung. Die *Zufriedenheit* nach ISO 9241-2 (Anforderungen an die Arbeitsaufgaben) und ISO 9241-11 berücksichtigt die subjektiv wahrgenommene Beeinträchtigungsfreiheit und positive Einstellung zur Produktnutzung durch den Nutzer. Frokjaer et al. (2000) konnten aufzeigen, dass Effektivität, Effizienz und Zufriedenheit voneinander unabhängige Dimensionen sind.

Bezogen auf den Nutzungskontext einer Website liegt Usability im Sinne der Norm ISO 9241 folglich vor, wenn Besucher der Website vor dem Hintergrund ihrer individuellen Erfahrungen und Fähigkeiten die angebotenen Funktionen, Services und Inhalte innerhalb eines gegebenen Kontextes als effektiv, effizient und zufriedenstellend wahrnehmen und zur Zielerreichung nutzen können (vgl. Dzida & Wandke, 2006, S. 469). Die Usability eines Webangebots umfasst den Umstand, wie schnell und erfolgreich die Besucher ihre individuellen Ziele erreichen und wie subjektiv befriedigend sie die Interaktion und das Ergebnis empfinden. Voraussetzung hierfür ist eine tatsächliche Nutzung des Webangebots. Hiervon abzugrenzen ist die „augenscheinliche“ Usability, die wahrgenommene Einfachheit der Nutzung auf Grundlage eines ersten visuellen Eindrucks der Anwendung (Caplan, 1994).

Die *Norm ISO 9241-10* (Grundsätze der Dialoggestaltung) behandelt die ergonomische Gestaltung von Software und beschreibt allgemeine ergonomische Grundsätze. Diese werden ohne Bezug auf die Arbeitssituation, Anwendungen, Umgebungen und Dialogtechnik dargestellt und sollen bei der Leistungsbeschreibung, Gestaltung und Bewertung von Dialogsystemen angewandt werden. Unter Dialog wird dabei eine Interaktion zwischen einem

Benutzer und einem Dialogsystem verstanden, die erfolgt, um ein bestimmtes Ziel zu erreichen. Ein Benutzer wird definiert als ein Mensch, der mit dem Dialogsystem arbeitet. Die folgenden sieben Grundsätze sind für die Gestaltung und Bewertung eines Dialogs als wichtig benannt worden:

Aufgabenangemessenheit

Ein Dialog ist dann aufgabenangemessen, wenn er den Benutzer darin unterstützt, eine Arbeitsaufgabe effektiv und effizient zu erledigen.

Selbstbeschreibungsfähigkeit

Selbstbeschreibungsfähigkeit ist gegeben, wenn jede einzelne Interaktion durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Benutzer auf Anfrage erläutert wird.

Steuerbarkeit

Ein Dialog ist steuerbar, wenn es dem Benutzer möglich ist, den Dialogablauf zu starten sowie Richtung und Geschwindigkeit der Interaktion zu beeinflussen, bis das Ziel erreicht ist.

Erwartungskonformität

Ein Dialog ist erwartungskonform, wenn er auf die Merkmale des Benutzers abgestimmt ist, z.B. seinen Kenntnissen aus dem Arbeitsgebiet, seiner Ausbildung und seiner Erfahrung entspricht sowie allgemein anerkannte Konventionen berücksichtigt.

Fehlertoleranz

Fehlertoleranz ist gegeben, wenn das beabsichtigte Arbeitsergebnis trotz fehlerhafter Eingaben entweder mit keinem oder mit geringem Korrekturaufwand durch den Benutzer erreicht werden kann.

Individualisierbarkeit

Ein Dialog ist individualisierbar, wenn Anpassungen an die Erfordernisse der Arbeitsaufgabe des Benutzers, sowie seine individuellen Vorlieben und Fähigkeiten durch das Dialogsystem zugelassen werden. Obwohl es grundsätzlich wünschenswert ist, dem Benutzer individualisierbare Dialogfunktionen zur Verfügung zu stellen, ist dies kein Ersatz für ergonomisch gestaltete Dialoge.

Lernförderlichkeit

Ein Dialog ist lernförderlich, wenn der Benutzer in der Interaktion mit dem Dialogsystem beim Erlernen der Anwendung unterstützt und angeleitet wird.

Diese Grundsätze können bei der Leistungsbeschreibung, Gestaltung und Bewertung von Dialogsystemen als allgemeine Leitlinien angewandt werden. Bezüglich der Beziehung zwischen den Grundsätzen gilt, dass sie nicht unabhängig voneinander sind. Vielmehr kann es notwendig sein, die Vorteile eines Grundsatzes gegenüber denen eines anderen abzuwägen. Die Anwendbarkeit der Grundsätze und ihre jeweilige Bedeutung hängen vom speziellen Anwendungsfall, der Arbeitssituation, von der gewählten Dialogtechnik und insbesondere

von den Merkmalen der Benutzergruppen ab. Hier werden in der Norm ISO 9241-10 insbesondere die Aufmerksamkeitsspanne, Grenzen des Kurzzeitgedächtnisses, Lerngewohnheiten, Erfahrungen bezüglich der Arbeit und im Umgang mit dem Dialogsystem sowie das mentale Modell des Benutzers bezüglich Struktur und Möglichkeiten des Dialogsystems angeführt.

2.7.3 Heuristiken nach Nielsen

In Anbetracht der Fülle an Erkenntnissen zur Gestaltung interaktiver Anwendungen, insbesondere von Websites und Online-Shops, stellt sich die Frage nach einer geeigneten Strukturierung. Eine gängige Unterscheidung erfolgt im Sinne einer produktorientierten Interpretation der Usability nach den einzelnen Gestaltungsebenen interaktiver Produkte wie z.B. Site Design, Page Design, Content Design etc. Ein solcher „*Product-centred view of usability*“ fokussiert auf die websitespezifischen Ausprägungsmerkmale des Angebots. Formale und inhaltliche Eigenschaften der spezifischen Website stehen im Vordergrund (Yom, 2003, S. 30f.). Steht hingegen die vom Anwender wahrgenommene Interaktionsqualität der Mensch-Maschine-Interaktion und damit verbunden effizientes und effektives Nutzungsverhalten im Fokus der Usabilityinterpretation, so sprechen Bevan & Macleod (1994) von einem „*Quality of use view*“. Diese Perspektive ist insbesondere im Rahmen der nutzerbasierten Usabilityevaluation und -optimierung angemessen. Eine *kontextorientierte Interpretation* der Usability ist dahingehend umfassender, dass Usability im globalen Zusammenspiel individueller personenbezogener Faktoren, allgemeiner websitespezifischer Ausprägungsmerkmale sowie im Kontext der spezifischen Aufgabe bzw. Nutzungsintention sowie weiterer situativer Faktoren betrachtet wird.

Die heuristischen Regeln von Nielsen (1993, 1994) beschreiben allgemeine Prinzipien einer nutzungsfreundlichen Gestaltung interaktiver Anwendungen und berücksichtigen dabei das globale Zusammenspiel von personenbezogenen, websitespezifischen und situativen Einflußfaktoren in Bezug auf die wahrgenommene Usability. Die Heuristiken sind vornehmlich auf die formale Gestaltung von Benutzerschnittstellen ausgerichtet. Tabelle 2-4 gibt einen Überblick über die heuristischen Regeln von Nielsen (1994) sowie über Ergänzungen der Heuristiken nach Jones und Hewitt (1998) sowie Muller et al. (1995). Nachfolgend werden die heuristischen Regeln inhaltlich dargestellt und voneinander abgegrenzt.

Tab. 2-4: Übersicht der Heuristischen Regeln nach Nielsen (1994)

Heuristische Regeln nach Nielsen (1994, S. 30)	
H1: Aesthetic and minimalist design	H6: User control and freedom
H2: Match between system and the real	H7: Flexibility and efficiency
H3: Recognition rather than recall	H8: Help users recognise, diagnose, and recover from errors
H4: Consistency and standards	H9: Error prevention
H5: Visibility of system status	H10: Help and documentation

Ergänzung der Nielsen Regeln nach Muller (1995, S. 115 f.)	
H11. Respect the user and her/his skills	H13. Support quality work
H12. Pleasurable experience with the system	
Ergänzung der Nielsen Regeln nach Jones und Hewitt (1998, S. 25 ff.)	
H14. Minimise disorientation	H16. Make widely accessible
H15. Support effectual navigation	H17. Minimise download times

Heuristik 1: Aesthetic and minimalist design

"Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in dialogue competes with the relevant units of information and diminishes the relative visibility." (Nielsen, 1994, S. 30).

Diese Heuristik zielt auf alle Aspekte der optischen Präsentation der Informationen ab. Dabei geht es um die Reduzierung der Informationskomplexität durch Optimierung der Informationsdarbietung. Ziel ist es, die Informationsverarbeitungsprozesse des Nutzers bestmöglich zu unterstützen und zu entlasten. Desweiteren umfasst die Heuristik Aspekte der vom Konsumenten wahrgenommenen Ästhetik und deren Einfluss auf die Einschätzung der Usability. Auf diesen Zusammenhang wird in Abschnitt 2.8.4 noch ausführlicher eingegangen, weshalb an dieser Stelle eine weitergehende Darstellung nicht erfolgen soll.

Heuristik 2: Match between System and the Real World

"The System should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order" (Nielsen, 1994, S. 30).

Diese Heuristik bezieht sowohl syntaktische als auch strukturell-mentale Aspekte ein. Erstere umfassen die Verwendung anwenderorientierter Begriffe. Dies umfasst ebenso die Vermeidung unnötiger Fachbegriffe und technischer Terminologien als auch die Erwartungskongruenz von Linkbezeichnungen in Bezug auf die dahinter liegenden Informationen. Es ist zu berücksichtigen, welche Assoziationen beziehungsweise welches Verständnis bestimmte Linkbezeichnungen bei den Nutzern auslösen, die wiederum die Erwartungen an die hinter dem Link stehenden Inhalte leiten (Yom, 2003, S. 33).

So stellten Bopp und Wörmann (2002) in einer repräsentativen Online-Befragung mit deutschen Internetnutzern fest, dass bis zu 45% der Befragten die bei vielen Websites gängigen Linkbezeichnungen wie FAQs, Guided Tour oder Sitemap nicht verstehen. Neben der Verständlichkeit von Linkbezeichnungen sind Eindeutigkeit und Erwartungskongruenz von Linkbezeichnungen von hoher Bedeutung für die wahrgenommene Orientierung bei der Nutzung von Websites. Otter und Johnson (2000) identifizierten uneindeutige bzw. zu viele Links als eine der Hauptursachen für die Desorientierung von Websitenutzern während der

Navigation auf der Website und begründen dies vor allem mit dem Umstand, dass die Handlungskonsequenzen bei Aktivierung eines Links nur begrenzt vom Nutzer antizipierbar sind. Nach Conklin (1987) führt dieses Phänomen zu einer informationellen Kurzsichtigkeit des Anwenders. Nielsen schlägt deshalb als pragmatische Möglichkeit die Verwendung von Linktiteln für die verwendeten Navigationselemente und Funktionen vor, die Informationen darüber enthalten, auf welche Seite der Link führt, welche Art von Information den Nutzer erwartet und in welcher Beziehung diese zur aktuellen Seite stehen (Nielsen, 1998, S.1).

Die strukturell-mentale Ebene bezieht Aspekte einer anwenderorientierten Darbietung von Informationen ein. Die Informationen sollen in einer für den Nutzer natürlichen und logisch nachvollziehbaren Anordnung präsentiert werden (Yom, 2003, S. 33). Natürliche Anordnung der Informationen bedeutet in diesem Zusammenhang, dass dem Anwender die Navigation durch das Angebot aufgrund der Informationsarchitektur so einfach wie möglich gemacht wird. Im Vordergrund steht die optimale Abstimmung zwischen dem mentalen Informationsmodell des Nutzers und der Struktur der Informationsdarbietung im Interface.

Eine natürliche und logische Anordnung der Informationen setzt auch voraus, dass nur die für die aktuell zu bewältigende Aufgabenstellung relevanten Informationen und Operationen, wie z.B. Navigationselemente und Funktionen, zum richtigen Zeitpunkt an der richtigen Stelle präsentiert werden. Die richtige Platzierung der Informationen und Operationen hängt von der Erwartungskongruenz der Informationsarchitektur und des Screen-Layouts ab, wobei sich letzteres auf das Zusammenspiel von Text-, Navigations- und Graphikelementen als visuelle Einheit bezieht (Thissen, 2001, S. 123).

Heuristik 3 & 5: Recognition rather than Recall & Visibility of System Status

"Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate." (Nielsen, 1994, S. 30).

Diese Heuristik zielt auf Möglichkeiten der Gestaltung, die den Nutzer eines interaktiven Angebots durch eine Minimierung der zu erbringenden Gedächtnisleistungen unterstützen (Yom, 2003, S. 34f.). Die Regel basiert auf den kognitionspsychologischen Erkenntnissen über den Einfluss der Gedächtnisleistung auf Informationsverarbeitungsprozesse. Hier nimmt die Funktionsweise des Arbeits- bzw. Kurzzeitgedächtnisses eine wichtige Rolle ein, da seine Kapazität mit $7 (\pm 2)$ sog. Chunks, als variable Einheiten der Wissensrepräsentation, begrenzt ist (Miller, 1956).

Menschen sind nur begrenzt in der Lage Informationen gleichzeitig zu verarbeiten, die dann für Entscheidungsprozesse zur Verfügung stehen. So können Informationen nach Aufnahme ins Arbeitsgedächtnis maximal innerhalb von ca. 15 bis 30 Sekunden abgerufen werden. Werden sie innerhalb dieser Zeitspanne nicht ins Langzeitgedächtnis überführt, verblassen sie und gehen verloren. Die Gedächtnisleistung kann verbessert werden, wenn Personen bei der

Reproduktion durch relevante aktivierende Informationen unterstützt werden (recognition), als wenn Gelerntes frei reproduziert werden muss (recall) (Nielsen, 1993, S. 128).

In Bezug auf den Anwendungsbereich von Websites bedeutet dies, dass der Websitenutzer während einer Internetsitzung die wahrgenommenen Informationen großteils nicht dauerhaft speichert. Für die Navigation wichtige Aspekte, z.B. der Ort bestimmter Informationen und dessen Bezeichnung, das Aussehen zugehöriger Elemente und bereits besuchte Seiten werden vergessen, wodurch das (Wieder-)Auffinden von Informationen erschwert wird. Da eine erfolglose Suche in vielen Fällen zur Verärgerung des Nutzers oder gar zum Abbruch der Interaktion führen kann, sollten Navigations- und Handlungsoptionen sowie während der Interaktion vollzogene Aktionen sichtbar gemacht werden. Rückmeldungen über getätigte Aktionen sowie Informationen zur Lagebestimmung entlasten den Nutzer kognitiv durch eine Reduzierung der zu erbringenden Gedächtnisleistung, indem sie ihm bereits genutzte oder noch offen stehende Navigationspfade aufzeigen bzw. an bereits durchgeführte Handlungen erinnern (Yom, 2003, S. 64ff.).

Heuristik 4: Consistency and Standards

"Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions" (Nielsen, 1994, S. 30). Konsistenz bezieht sich auf die einheitliche Verwendung von Bezeichnungen wie Funktions- und Rubrikentiteln, Benennungen von Navigationselementen und zentraler Funktionen, aber auch von Layoutelementen etc., um Irritationen des Nutzers zu vermeiden (Yom, 2003, S. 35).

Neben der angebotsinternen Konsistenz, die auf eine einheitliche Verwendung von Begriffen, Layouts und von Interaktionssequenzen innerhalb einer Website abzielt, steht bei der angebotsübergreifenden Konsistenz die Verwendung von branchen- bzw. webüblichen Standards im Vordergrund (Wilhelm, Yom & Wohlfahrt, 2002).

Heuristik 6: User control and freedom

"Users often choose system functions by mistake and will need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialogue. Support undo and redo." (Nielsen, 1994, S. 30).

Diese Heuristik ist für Websites von relativ geringer Relevanz, da Internetbrowser ihren Nutzern über die Navigationsfunktionen (Back, Forward, Exit, URL-Eingabe etc.) in der Regel eine Bandbreite von Notausgängen im Sinne der Heuristik anbietet.

Heuristik 7: Flexibility and efficiency of use

"Accelerators - unseen by the novice user - may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions." (ebda S. 30).

Nielsen entwickelte seine Heuristiken ursprünglich nicht explizit für Webangebote, sondern allgemein für Offline-Anwendungen wie beispielsweise Textverarbeitungssoftware, Terminal-Banking-Applikationen etc. Im Gegensatz zu Internetangeboten sind die Nutzungsintentionen und -situationen für die angeführten Offline-Anwendungen stark festgelegt, sodass eine Reduzierung der Analyseperspektive auf die personenbezogene Determinante der „Erfahrung mit der Anwendung“ angemessen scheint. Auch bei Webangeboten ist die Erfahrung des Nutzers im Umgang mit dem Medium von Bedeutung. So werden diese von einer großen Bandbreite von Personen mit unterschiedlicher Web- und Shoppingkompetenz (Novize vs. Experte), unterschiedlichen sitespezifischen Erfahrungen (Erstbesucher vs. Wiederholnutzer) und unterschiedlichem themenspezifischem Wissen (Laie vs. Spezialist) genutzt (Yom, 2003, S. 36). Zusätzlich variieren, wie oben angeführt, die situativen Faktoren wie beispielsweise die Nutzungsintention und damit verbundene Informationsbedürfnisse bei der Nutzung von Webangeboten im Vergleich zu Offline-Anwendungen stärker und nehmen Einfluss auf das Surf- und Nutzungsverhalten des Anwenders, weshalb sie durch die Anwendung unterstützt werden müssen. Sowohl der gezielt nach bestimmten Informationen suchende Nutzer als auch der lediglich generell am Thema interessierte Surfer müssen in ihren Nutzungszielen optimal unterstützt werden (vgl. Cove & Walsh, 1988).

Die Heuristik zielt einerseits auf die Gewährleistung einer generellen effizienten Nutzung der Anwendung und andererseits auf die Flexibilität der Anwendung. Während das Effizienzkriterium schlecht abgrenzbar ist zu den anderen Heuristiken, da diese allesamt darauf abzielen, eine effiziente Nutzung der Anwendung zu gewährleisten, ist die Flexibilität der Anwendung ein gut abgrenzbarer Aspekt. Flexibilität ist dahingehend zu verstehen, dass die Gestaltung der Website die gesamte Bandbreite möglicher Nutzungssituationen zu unterstützen vermag.

Die Wahrscheinlichkeit, dass der Nutzer mit seinen spezifischen Bedürfnissen und individuellen Ressourcen optimal unterstützt wird, ist daher umso größer, je variabler die Möglichkeiten des Informationszugriffs und je flexibler die Suchfunktionen gestaltet sind und je optionaler und modularer die Informationseinheiten dargestellt werden (vgl. Esch, Hardiman & Langner, 2000). In Bezug auf Websites stehen im Vordergrund dieser Heuristik also Erkenntnisse darüber, wie den unterschiedlichen Informationsbedürfnissen und Fähigkeiten verschiedener Nutzergruppen in unterschiedlichen Nutzungssituationen optimal Rechnung getragen werden kann.

Heuristik 8 & 9: Help Users Recognize, Diagnose, and Recover from Errors & Error Prevention

Wie bereits angeführt, entwickelte Nielsen seine Heuristiken ursprünglich für die Anwendung auf Offline-Anwendungen. Insbesondere bei komplexeren Computerapplikationen sind verständliche und unterstützende Fehlermeldungen von Relevanz für die Bedienbarkeit, weshalb ihre Berücksichtigung in einer eigenen Heuristik gerechtfertigt erscheint.

Bei Webangeboten sind Fehlermeldungen im weiteren Sinne von Bedeutung, wenn eine vom Nutzer spezifizierte Anfrage an das Angebot gestellt wird und das System ein darauf angepasstes Feedback erzeugt, wie beispielsweise die Rückmeldung auf eine Suchmaschinenanfrage, dass keine passenden Informationsinhalte bzw. kein passendes Produkt gefunden wurden (Yom, 2003, S. 37). Der Einsatz von unterstützenden Fehlermeldungen im engeren Sinne ist insbesondere im Zusammenhang mit Registrierungs-, Buchungs- und Bestellprozessen von Bedeutung, da unklare Rückmeldungen über falsche oder fehlende Angaben zur Unzufriedenheit des Nutzers und zum Abbruch des Registrierungsvorganges bzw. einer Bestellung führen können. Die Fehlermeldung sollte deshalb möglichst präzise die Ursachen des Fehlers benennen und konstruktive Hilfestellungen für die Fehlerbehebung geben, damit der Anwender seinen Fehler schnell und mühelos korrigieren kann.

Im Rahmen von Bestell- oder Registrierungsprozessen bieten sich sogenannte Forcing Functions an. Dies sind Pflichtfelder für die Dateneingabe, die den Nutzer vor fehlerhaften Interaktionen bewahren sollen, bzw. im Falle einer fehlerhaften Interaktion ihn leiten, sein Verhalten zu korrigieren (Lewis & Norman, 1995, S. 690).

Heuristik 10: [Provide] Help and Documentation

Sellen und Nicol (1995) weisen darauf hin, dass allgemein jede Form von Feedback, bspw. Fehlermeldungen, Informationen zur Lagebestimmung oder zu durchgeführten bzw. noch offenstehenden Aktionen etc., Hilfen zu einer effektiven, effizienten und zufriedenstellenden Nutzung der Anwendung darstellen können. Sie unterscheiden zwischen impliziten Hilfen, welche im weitesten Sinne durch die Gestaltung der Benutzerschnittstelle determiniert werden, und expliziten Hilfen wie z.B. Online-Hilfefunktionen oder ein Online-Handbuch. Allerdings konnte Grayling (1998, S. 172) aufzeigen, dass explizite Online-Hilfefunktionen häufig erst als letztes Mittel zur Problemlösung genutzt werden.

Die Nutzer suchen häufig zunächst über einen iterativen Prozess des Versuchs und Irrtums nach einer Lösung des Problems, bevor sie die angebotenen Hilfen in Anspruch nehmen (Stellen & Nicol 1995). Die Autoren begründen dies u.a. mit der häufig schlechten Auffindbarkeit und Verständlichkeit unterstützender Informationen.

Yom und Pothe (2001) konnten dieses Phänomen in einer Monitoring-Studie von Online-Shops bestätigen. Sie stellten fest, dass ein Großteil der untersuchten Online-Shops zwar über eine Hilfefunktion verfügte, diese jedoch nur bei etwa der Hälfte der Online-Shops als gut gestaltet bezeichnet werden konnte.

2.7.4 Ergänzende Heuristiken

Muller et al. (1995) schlagen ergänzend die Regel „Pleasurable experience with the system“ vor. Übertragen auf die Nutzung von Websites trägt diese Forderung der Erkenntnis Rechnung, dass die wahrgenommene Freude bzw. das Vergnügen während der Nutzung eine

wichtige Determinante des Nutzungsverhaltens ist (vgl. u.a. Bost, 1987; Hassenzahl, Kekez & Burmester, 2002; Weinberg & Besemer, 1999). Der Zusammenhang zwischen usability und Nutzungsvergnügen während der Interaktion mit einer Website ist Gegenstand des Abschnittes 2.11.3 und wird deshalb hier nicht weitergehend betrachtet.

Jones und Hewitt (1998, S. 26) schlagen des Weiteren die Heuristik "Minimise disorientation" vor. Diese Regel ist aus dem Phänomen des „Lost in hyperspace“, von Conklin (1987) für Hypertext-Systeme identifiziert, abgeleitet. Um Desorientierung zu minimieren, fordern sie: „... users should know where they are within the information, or how to find this out.“ Es geht bei dieser Regel darum, dass der Anwender optimal darin unterstützt wird, sich jederzeit zu orientieren und seinen aktuellen „Stand(-ort)“ während der Nutzung der Website zu bestimmen (vgl. Gwizdka & Spence, 2007). Dies korrespondiert mit den Zielsetzungen der Heuristiken drei und fünf „Recognition rather than Recall & Visibility of System Status“ nach Nielsen (1994).

Die von Muller et al. (1995) vorgeschlagene ergänzende Regel "Respect the user and her/his skills" sowie die ergänzenden Heuristiken von Jones und Hewitt (1998) "Support effectual navigation" und "Make widely accessible" sind in ihrer Intension in der fünften Regel „Visibility of system status“ sowie in der siebten Regel "Flexibility and efficiency of use" nach Nielsen (1994) bereits enthalten.

Ergänzend zu den dargestellten Heuristiken sind die IBM Ease of Use Guidelines G1 „Product Information“ und G5 „Trust“ anzuführen. Die Guidelines wurden von IBM Usability-Spezialisten basierend auf den Erfahrungen zahlreicher Tests mit eigenen Internet-Präsenzseiten des Konzerns sowie Erkenntnissen der Usability- und E-Commerce-Forschung speziell für die nutzerzentrierte Entwicklung von Webshops und sites entwickelt (IBM, 2001). Während sich die Guideline G1 auf alle Aspekte der Produkt- und Leistungsdarstellung und diesbezüglicher Information bezieht, befasst sich die Guideline G5 vornehmlich mit den Aspekten der Datensicherheit und des Datenschutzes sowie deren Auswirkungen auf das Vertrauen des Nutzers in die Anwendung und den Anbieter der Website.

2.7.5 Usability als Erfolgsfaktor von Webangeboten

Lee, Katerattanakul und Hong (2005) identifizieren bei Web-Nutzern drei Kategorien von Gründen für die Nutzung von Webanwendungen, die Suche nach Informationen, die Abwicklungen ökonomischer Transaktionen sowie das Vergnügen während der Nutzung eines interaktiven Angebots. Entsprechend formulieren die Autoren die Erwartung an ein effektives Webdesign, dass dieses den Anwender in der Interaktion mit Webangeboten diesbezüglich zu unterstützen habe.

Über 60% der Webnutzer nehmen jedoch häufig oder gelegentlich Probleme während der Internetsitzung wahr, so z.B. aufgrund von unübersichtlichen Homepages, aufwändiger Informationssuchen oder aufgrund eines langsamen Seitenaufbaus (vgl. Dzida & Wandke,

2006, S. 483; Van Eimeren, Gerhard & Frees, 2002). Nielsen (2001) untersuchte das Problemlösungsverhalten von Testpersonen auf vornehmlich größeren Webangeboten und stellte fest, dass nur 56% der Lösungsversuche erfolgreich waren.

Dabei sind die Erfahrungen beim ersten Besuch eines Webangebots für die Entwicklung vom Erstbesucher zum Wiederholnutzer besonders kritisch (Lohse, Bellman & Johnson, 2000). Entsprechend ist für die Neukundengewinnung eine Reduzierung der Abbruchquoten ein bedeutsames operatives Ziel. Tatsächlich gelingt es den Betreibern von Webangeboten jedoch nur in eingeschränktem Maße, Erstbesucher einer Website erfolgreich als Nutzer und langfristig als Online-Stammkunden zu gewinnen (vgl. Gräf, 1999, S. 145ff.). Die komfortable, effiziente und flexible Nutzung eines Webangebots und damit verbunden eine Reduzierung des nutzerseitigen Interaktionsaufwands ist vor diesem Hintergrund von besonderer Bedeutung.

Dies ist umso wichtiger, da sich zeigt, dass Online-Kunden durchaus eine hohe Anbietertreue zeigen. So konnte A.T. Kearney (2000) in Studien nachweisen, dass positive Erfahrungen mit einem Webanbieter zu Wechselbarrieren führen und damit verbunden die Anbietertreue steigern (ebenso Scholz, 2001, S. 27).

Chen, Gillenson und Sherrell (2004) identifizierten fünf kritische Erfolgsfaktoren für die Akzeptanz von Webangeboten. Neben der Usability des Webangebots sind dies die Qualität des Produktangebots, die Informationsqualität, die wahrgenommene Servicequalität sowie das Vertrauen in das Webangebot und den Anbieter (vgl. Alba et al., 1997; Gefen, 2000; Gefen & DeVine, 2001; Gillenson, Sherrell & Chen, 2000; Jarvenpaa & Todd, 1997a).

Bhattacharjee (2001) stellte einen positiven Zusammenhang zwischen Usability und Nutzerzufriedenheit einerseits und der Online-Shop-Präferenz der Konsumenten andererseits fest. Die Art und Weise, wie attraktiv Informationen inhaltlich und visuell dargestellt werden und erwartungskonform zugänglich gestaltet sind, beeinflussen das Online-Einkaufserlebnis maßgeblich (Lohse, 2000) und fördern die Kaufneigung in Webanwendungen (Liao & Cheung, 2001). Aspekte des Interface-Designs werden damit zu einem wichtigen Parameter für den kommerziellen Erfolg eines Online-Shops. Christophersen (2007) wies einen signifikanten Einfluss der über einen Fragebogen erhobenen Usability eines Online-Shops auf die Kaufintention und die Kaufentscheidung nach, wobei der Autor die Messung sowohl mittels reflektiver als auch formativer Usability-Skalen vornahm (ebenso Christophersen & Konradt, 2008).

Darüber hinaus gibt es Untersuchungen, die einen unmittelbaren Zusammenhang zwischen Gestaltungsmerkmalen und ökonomischen Zieldimensionen aufzeigen. So konnten Lohse und Spiller (1998) in einer Untersuchung über 28 Online-Shops feststellen, dass die Varianz des monatlichen Umsatzes der untersuchten Online-Shops zu über 60% durch die Gestaltungsqualität der Produktlisten erklärt wurde. Des Weiteren konnten Lynch und Ariely (2000) in einer Studie anhand von zwei Online-Weinshops aufzeigen, dass die Preiselastizität mit

zunehmender Usability abnimmt. Die angeführten Erkenntnisse machen deutlich, dass Aspekte der Nutzerfreundlichkeit wichtige Erfolgsfaktoren von Online-Shops sind.

In Bezug auf Websites ist dabei zu berücksichtigen, dass die Qualität der dargebotenen Informationen und die Art und Weise der Darbietung neben der Marke für den Besucher wesentliche Kriterien zur Beurteilung des Anbieters und der Dienstleistung darstellen. So zeigten Gierl und Bambauer (2001) auf, dass die Höhe des Informationsgehaltes und der optische Eindruck die Einstellung des Besuchers zur Website wesentlich beeinflussen und dass darüber hinaus, vermittelt über die Einstellung zur Website, ein positiver Effekt auf die Einstellung zum Anbieterunternehmen festgestellt werden kann. In eine ähnliche Richtung gehen die Erkenntnisse von Fogg et al. (2001). Sie zeigten auf, dass eine einfache Bedien- und Navigierbarkeit einer Website eine positive Wirkung auf die durch den Nutzer wahrgenommene Glaubwürdigkeit des Website-Anbieters hat.

In Bezug auf Websites von Krankenkassen ist zu beachten, dass es sich bei deren Leistungsangebot u.a. um Serviceleistungen mit Erfahrungs- und Vertrauenseigenschaften handelt. Deren Qualität kann, wenn überhaupt, erst zum Zeitpunkt der Inanspruchnahme festgestellt werden. Das persönliche oder telefonische Gespräch mit dem Kundenberater wird im Falle einer Website durch eine Mensch-Computer-Interaktion ersetzt, die das ausgeprägte Spektrum an Informationsbedürfnissen befriedigen soll. Gerade bei beratungsintensiven Dienstleistungen wie Versicherungsdienstleistungen ist das Potenzial für einen Vertrauensaufbau über eine technische Schnittstelle nur eingeschränkt vorhanden. Zugleich ist das Vertrauen des Kunden in die Fähigkeit und Bereitschaft des Anbieters, sein individuelles Bedürfnis möglichst vorteilhaft zu erfüllen, jedoch eine Voraussetzung für den Aufbau und die Weiterentwicklung der Kundenbeziehung (Grabner-Kräuter & Kaluscha, 2003).

2.8 Methoden der Usabilityevaluation

Die Evaluation der Usability kann mit Hilfe unterschiedlicher Verfahren erfolgen (Hamborg & Gediga, 2006; Yom, 2003). Diese lassen sich hinsichtlich verschiedener Kriterien wie der Breite der Evaluation, der Expertise der Anwender, des Zeitbedarfs und damit verbundener Kosten, des Zeitpunkts und des Ortes der Messung unterscheiden (Oppermann & Reiterer, 1997). Während in Abschnitt 2.8.1 zwei unterschiedliche Zielsetzungen der Usabilityevaluation abgegrenzt werden, folgt in Abschnitt 2.8.2 eine Abgrenzung von deskriptiven und prädiktiven Evaluationsmethoden, um dem Leser die Einordnung der fragebasierten Evaluation zu erleichtern.

2.8.1 Mikroevaluation und summarische Evaluation

Werden Daten des Interaktionsverhaltens und -erlebens der Testperson während der Nutzung einer Software erfasst, so sprechen Mangold, Reese, Klauck und Stanulla (2000) von *Mikroevaluation*. Gegenstand der Beurteilung ist die Wahrnehmung der Funktionen, einzelner

Elemente und Inhalte, etc. während der Interaktion mit der Zielsetzung, Usability-Probleme in der Interaktion zu identifizieren, zu beschreiben und Vorschläge zur Verbesserung abzuleiten. Methoden der interaktionsbegleitenden Mikroevaluation geben daher Antwort auf die Frage nach Mängeln einer Software und zielen bspw. auf Schwachstellen eines Prototyps als Basisinformationen für ein Re-Design. Sie sind daher für eine (weiter-)entwicklungsorientierte Evaluation von Software im Entwicklungsstadium geeignet.

Die *summarische Evaluation* stellt hingegen eine abschließende Beurteilung der Software als Ganzes dar. Sie gibt Antworten auf die Fragen, wie gut eine Software ist bzw. welche Software besser ist. Die Evaluation zielt bspw. auf einen Vergleich von Gestaltungsalternativen oder auf die Bewertung eines Softwareendproduktes bzw. auf die Überprüfung auf Konformität mit Normen.

2.8.2 Deskriptive vs. prädiktive Usability-Evaluationsmethoden

Die unterschiedlichen Methoden zur Evaluation der Usability lassen sich nach Hamborg und Gediga (2006, S. 509 ff.) in zwei Hauptkategorien, in deskriptive und prädiktive Evaluationsmethoden unterscheiden. Deskriptiv erfolgt die Bewertung einer Software, wenn sie im Zusammenhang mit ihrer Nutzung vorgenommen wird. Prädiktiv erfolgt sie analytisch in Form von Begutachtungen. In der nachfolgenden Übersicht (s. Abbildung 2-10) wird eine Systematisierung verschiedener Evaluationsmethoden entsprechend dieser Klassifikation dargestellt.

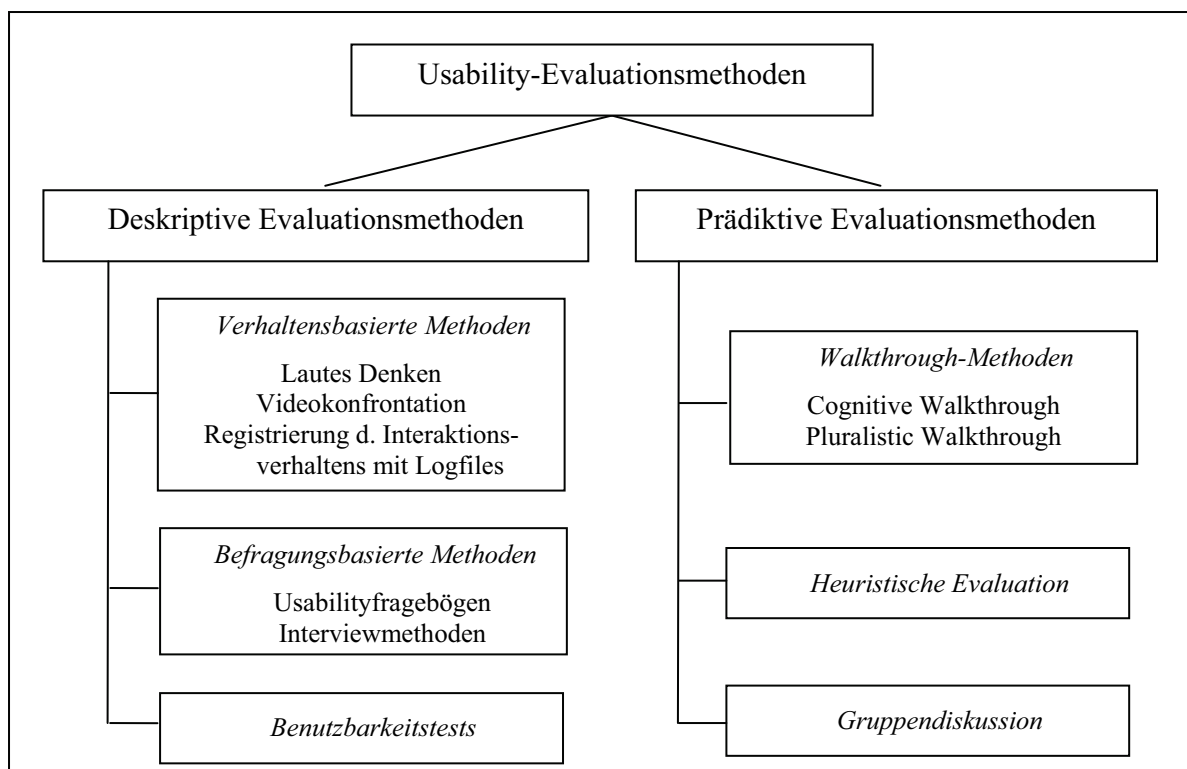


Abb. 2-10: Klassifikation von Methoden zur Usabilityevaluation (in Anlehnung an Hamborg & Gediga, 2006; Yom, 2003)

2.8.2.1 Deskriptive Evaluationsmethoden

Deskriptive Usability-Evaluationsmethoden verfolgen das Ziel, die ergonomische Qualität einer Software objektiv, reliabel und valide aus Benutzersicht zu beschreiben, also den Zustand und akute Problembereiche der Software zu erfassen (Hamborg & Gediga, 2006, S. 509 ff.). Sie lassen sich nach drei Ansätzen in verhaltensbasierte und befragungsbasierte Evaluationsmethoden sowie Benutzbarkeitstests unterscheiden.

Verhaltensbasierte Evaluationsmethoden:

Verhaltensbasierte Evaluationsmethoden erfassen konkretes Verhalten von Nutzern in der Interaktion mit Software durch Beobachtung. Aus dieser wird anhand von Variablen wie beispielsweise Fehlerdaten, Bearbeitungszeiten oder Anzahl der erforderlichen Operationen zur Aufgabenerledigung auf Benutzbarkeitsprobleme der Software geschlossen (Kent & Panizzi, 2006). Als wesentliche verhaltensbasierte Evaluationsmethoden werden von Hamborg und Gediga (2006, S. 509 f.) Videokonfrontationen, Analysen des Interaktionsverhaltens anhand von Logfiles sowie die Methode des lauten Denkens angeführt (vgl. Hamborg, 2007).

Befragungsbasierte Evaluationsmethoden:

Im Gegensatz zu verhaltensbasierten Methoden erfolgt bei den befragungsbasierten Methoden die Evaluation nicht über die Beobachtung konkreten Nutzerverhaltens, sondern anhand von Benutzerurteilen, die durch unterschiedlich standardisierte mündliche oder schriftliche Befragungsformen erhoben werden (Hamborg & Gediga, 2006, S. 511). Vorteile von Befragungen im Vergleich zu Verhaltensbeobachtungen bestehen darin, dass sich Daten erheben lassen, die nicht oder nur eingeschränkt beobachtbar sind, wie bspw. Einstellungen oder Befindlichkeiten, der Einsatz von Fragebogen i.d.R. kostengünstiger ist und statistische Gütekriterien überprüfbar sind (vgl. Dzida & Wandke, 2006, S. 465; Nielsen, 1993).

- *Interviewmethoden:* Interviews in Form unstandardisierter, offener Nachbefragungen oder in halbstandardisierter Form eignen sich aufgrund ihrer spezifischen Fokussierung auf eine Stimuluskonstellation, in diesem Falle das Webangebot, gut zur explorierenden Identifizierung von Usability-Problemen (Diekmann, 1998, S. 375).
- *Usability-Fragebögen:* Die fragebogenbasierte Usabilityevaluation ist Gegenstand dieser Arbeit und wird vornehmlich im Abschnitt 2.9 dargestellt. Da ein Ziel dieser Arbeit u.a. in der Überprüfung und Weiterentwicklung von Usability-Skalen liegt, wird auf diesen methodischen Ansatz zur Usabilityevaluation fokussiert.

Benutzbarkeitstests (usability tests): Benutzbarkeitstest basieren auf einer Kombination von Beobachtungen bzw. Aufzeichnungen des Interaktionsverhaltens und befragungsorientierten Methoden und dienen der systematischen Erkennung konkreter Schwachstellen einer Software (Hassenzahl & Seewald, 2004; Rubin, 1994). Ziel ist nicht, wie in experimentellen

Untersuchungen, die Ableitung verallgemeinerbarer Erkenntnisse, sondern die Aufdeckung von Schwachstellen und darauf aufbauend die Ableitung von Verbesserungsvorschlägen. Die Durchführung von Benutzbarkeitstests erfolgt zumeist in einem Labor, wobei der Zeitaufwand aufgrund der hohen Betreuungsintensität vergleichsweise hoch ist; entsprechend hoch sind auch die Kosten dieser Form der Usabilityevaluation.

2.8.2.2 Prädiktive Evaluationsmethoden

Während deskriptive Usability-Evaluationsmethoden das Ziel verfolgen, den Zustand einer Software beschreibend darzustellen, zielen prädiktive Evaluationsmethoden direkt auf die Identifikation von Schwachpunkten einer Software und die Formulierung von Gestaltungsempfehlungen (Hamborg & Gediga, 2006, S. 512). Entsprechend spielen Usability-Experten mit ihrem Expertenwissen eine zentrale Rolle bei der Anwendung dieser Verfahren, die häufig bereits in einer frühen Phase im Entwicklungszyklus einer Software zum Einsatz kommen. Als wesentliche prädiktive Evaluationsmethoden werden von Hamborg und Gediga (2006, S. 512ff.) die Walkthrough-Methode, die Heuristische Evaluation und die Gruppendiskussion angeführt.

2.9 Fragebogenbasierte Usabilityevaluation

Für die Evaluation der Usability von Software existieren unterschiedliche standardisierte *Fragebogenverfahren*, die die Zufriedenheit der Nutzer mit der Software bzw. die Bewertung der Gebrauchstauglichkeit aus Sicht der Nutzer ermöglichen.

Durch die folgenden Fragebogen werden sowohl Aspekte der Gebrauchstauglichkeit als auch der Zufriedenheit erfasst: *Post Study System Usability Questionnaire* (PSSUQ, Lewis, 1995), *Computer System Usability Questionnaire* (CSUQ, Lewis, 1993), *Web Analysis and Measurement Inventory* (WAMMI, Kirakowski & Claridge, 2001), *IBM After-Scenario Questionnaire* (ASQ, Lewis, 1995), *Software Usability Measurement Inventory* (SUMI, Kirakowski, 1997), *Questionnaire for User Interface Satisfaction* (QUIS, Chin, Diehl & Norman, 1988; Harper, Slaughter & Norman, 1997), *Attrakdiff* (Hassenzahl, Burmester & Koller, 2003) und *Usability Fragebogen für Online-Shops* (ufosV2, Christophersen, 2007).

Mithilfe des Fragebogens zur *Wahrgenommenen Orientierung in Online-Shops* (WOOS, Yom, 2001) werden die durch den Nutzer wahrgenommene Qualität der Orientierung während der Nutzung eines Online-Shops sowie affektive Komponenten des Shoppingerlebens erhoben. Die Bewertung von Software nach den Gestaltungsgrundsätzen spezieller Normen und Standards ermöglichen der ISONorm- (Prümper & Anft, 1993) und der IsoMetrics-Fragebogen (Gediga, Hamborg & Düntsch, 1999). Diese beiden letztgenannten Inventare wie auch der *Purdue Usability Testing Questionnaire* (PUTQ, Lin, Choong & Salvendy, 1997) finden in der weiteren Darstellung allerdings keine Beachtung, da die Instrumente vornehmlich auf Items zur Beurteilung konkreter Softwaremerkmale basieren und eine Übertragbarkeit auf die Besonderheiten von Webangeboten, die im Vordergrund

dieser Arbeit stehen, nicht sinnvoll erscheint. Für eine detaillierte Darstellung sei hier auf gesonderte Veröffentlichungen (Diekmann, 1998; Gediga & Hamborg, 2002; Hamborg, Gediga & Hassenzahl, 2003; Schnell, Hill & Esser, 1999) oder die Originalliteratur verwiesen.

Standardisierte Fragebogenverfahren ermöglichen den Zugang zu nicht oder nur eingeschränkt beobachtbaren subjektiven Daten, wie z.B. Befindlichkeiten oder Einstellungen der Testpersonen (Nielsen, 1993). Darüber hinaus sind sie häufig kostengünstiger als befragungsorientierte Evaluationsmethoden oder Benutzbarkeitstests, können mit begrenztem Aufwand im Feld eingesetzt werden und die Ergebnisse sind anhand statistischer Gütekriterien überprüfbar.

Für die nachfolgend dargestellten Fragebogeninventare gibt es computergestützte Versionen, die auf Arbeitsplatzrechnern oder per Intra-/Internet dargeboten werden. Durch computergestützte Befragungen, die weitgehend vergleichbare Ergebnisse zu Papier-Bleistift-Versionen liefern (Richter, 1999; Slaughter, Harper & Norman, 1994), lässt sich die Datenerhebung und -auswertung ökonomischer gestalten.

Schwächen von Fragebogenverfahren bestehen darin, dass die Beantwortung der Fragen stark vom Wahrnehmungs- und Erinnerungsvermögen der Befragten abhängt. Die erhobenen Daten sind somit anfällig für unwillkürliche Fehler und Verzerrungen wie auch für bewusste Verfälschungen (Hamborg & Gediga, 2006, S. 512).

2.9.1 Questionnaire for User Interface Satisfaction (QUIS)

Der von Chin, Diehl und Norman (1988) entwickelte "Questionnaire for User Interface Satisfaction" dient der Erhebung der subjektiven Anwenderzufriedenheit mit spezifischen Aspekten einer Benutzerschnittstelle. In der Online-Version 7.0 werden wie in der ursprünglichen Version anhand von sechs Items globale Zufriedenheitsurteile erhoben, sowie bezüglich vier spezifischer Aspekte der Interface-Gestaltung (Screendesign, Terminologie/Systemfeedback, Erlernbarkeit und Systemfähigkeiten) objektbezogene Beurteilungen vorgenommen. Ergänzend werden in der Version 7.0 fünf weitere Aspekte zur Evaluation spezifischer Systemkomponenten „technical manuals and online help, online tutorials, multimedia, teleconferencing and software installation“ optional zur Verfügung gestellt (Harper, Slaughter & Norman, 1997). Die Itemabfrage wird über neunstufige, bipolare Adjektivskalen zuzüglich einer Ausweichkategorie vorgenommen und ist ergänzt um Raum zur freien Kommentierung der spezifischen Aspekte der Interface-Gestaltung.

Messages which appear on screen:	confusing		Clear	
		1 2 3 4 5 6 7 8 9		NA
Instructions for commands or choice:	confusing		Clear	
		1 2 3 4 5 6 7 8 9		NA

Abb. 2-11: Zwei Items aus dem QUIS-Fragebogen 7.0 (Harper, Slaughter & Norman, 1997)

Gütenachweise

Die globale interne Konsistenz des QUIS-Inventars 7.0 ist mit einem Cronbachs Alpha in Höhe von 0,95 hoch (vgl. Weise, 1975, S. 219). Die Konstruktvalidität wurde durch einen Abgleich der Messwerte in der aktuellen Version mit den sechs Items zur Erhebung der globalen Zufriedenheitswerte aus der in vorhergehenden Studien validierten Version 5.0 signifikant geprüft. Die Korrelationen zwischen den Konstrukten und der gemessenen Globalzufriedenheit variieren zwischen Werten von 0,49 und 0,61, was auf eine angemessene Übereinstimmung zwischen den neuen ergänzten Aspekten des QUIS und den globalen Zufriedenheitsurteilen hinweist. Die Kriteriumsvalidität des Inventars wurde anhand einer Aufteilung der Stichprobe in zwei Gruppen, entsprechend der Selbstauskunft der Teilnehmer vorgenommen, ob sie die bewertete Software "mögen" oder "nicht mögen". Bei der Überprüfung der Testwerte der beiden Gruppen wurden signifikante Unterschiede in Bezug auf die globalen Zufriedenheitsurteile sowie teilweise auch in Bezug auf die Items der o.g. Dimensionen Systemfähigkeiten sowie Erlernbarkeit festgestellt. Es stellt sich allerdings die Frage, ob das gewählte Kriterium "Mögen" ein adäquates Merkmal für die Abgrenzung der Gruppen darstellt, zumal die Operationalisierung nicht dargestellt wird.

Anwendbarkeit des Inventars bei Websites

Die Items von QUIS erfassen sowohl Beurteilungen von softwarespezifischen als auch von softwareunspezifischen Attributen, wie z.B. „Lesbarkeit am Bildschirm“, „die Konsistenz von Begriffen“, und muten teilweise eher wie eine Checkliste für Experten an. Die Anwendung auf Websites setzt zum Teil sprachliche Modifikationen voraus, wobei die Reliabilität und Validität dann erneut zu prüfen wären, und scheint nur begrenzt sinnvoll.

2.9.2 Software Usability Measurement Inventory (SUMI)

Das Befragungsinventar SUMI misst die Usability von Softwareanwendungen auf den fünf Dimensionen Efficiency, Affect, Helpfulness, Control und Learnability (Kirakowski, 1997, S. 9). Es wurde anhand mehrerer größerer Stichproben von Softwareanwendern, die kommerzielle Softwareprodukte aus ihrer beruflichen Erfahrung heraus beurteilten, entwickelt und geprüft. Die umfangreiche Datenbasis diente den Autoren als Grundlage für die sogenannte Item Consensual Analysis, die eine Identifizierung von konkreten Usability-Problemen durch

Abgleich geschätzter Antwortmuster mit tatsächlichen Itemausprägungen unterstützt (vgl. ebda S. 12).

In der Version (3.9) umfasst das Inventar 50 Statements mit dreistufigen Antwortskalen (agree, undecided, disagree). Mithilfe einer exploratorischen Faktorenanalyse wurden fünf Dimensionen mit hoher interner Konsistenz identifiziert. Die nachfolgend dargestellte Faktorenstruktur konnte in späteren Studien bestätigt werden (vgl. ebda S. 11):

Affect: Die Items dieser Dimension beziehen sich auf allgemeine emotionale Reaktionen und konkrete Erfahrungen im Rahmen der Nutzung.

Efficiency: Items dieser Dimension messen, inwieweit die Softwareanwendung den Befragten nach seinem subjektiven Empfinden unterstützt, bzw. inwieweit die Interaktion mentale Belastungen verursacht.

Helpfulness: Diese Dimension misst inwieweit die Softwareanwendung aufgrund eindeutig verständlicher Bezeichnungen und Instruktionen vom Befragten als selbsterklärend wahrgenommen wird sowie die wahrgenommene Güte des System-Feedbacks.

Control: Diese Dimension zielt auf die vom Anwender wahrgenommene Kontrolle während der Nutzung der Anwendung.

Learnability: Items dieser Dimension messen neben Aspekten der Erinnerbarkeit und der subjektiv empfundenen Qualität der Dokumentation, wie einfach bzw. anstrengend das Erlernen der Funktionalitäten der Anwendung empfunden wird.

	Agree	Undecided	Disagree
1. This software responds too slowly to inputs.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. I would recommend this software to my colleagues.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 2-12: Zwei Items aus dem SUMI-Fragebogen 3.9 (Kirakowski, 1997)

Gütenachweise

Die globale interne Konsistenz der Version 3.9 konnte in einer Studie über 150 Softwareanwendungen mit einem Stichprobenumfang von $N = 1.100$ mit einem Cronbachs Alpha von 0.92 als hoch nachgewiesen werden (vgl. Weise, 1975 S. 219). Sie variiert für die Sub-Skalen der fünf benannten Dimensionen von 0.71 (Kontrolle) bis 0.85 (Affektion). Die Validität des Befragungsinventars wurde in mehreren Laborstudien signifikant geprüft. Neben der empirischen Überprüfung sprechen auch theoretische Überlegungen für die Konstruktvalidität des Inventars, da zwischen vier der identifizierten Sub-Skalen (Efficiency, Helpfulness, Control

und Learnability) deutliche inhaltliche Übereinstimmungen mit den Prinzipien der ISO 9241-10 Norm bestehen (ebda S. 21).

Anwendbarkeit des Inventars bei Websites

Die Items des SUMI-Inventars sind zu einem großen Teil inhaltlich wie auch sprachlich stark auf Funktionalitäten und Anforderungen von Software-Anwendungen abgestimmt. Es werden zwar Aussagen im Inventar verwendet, die für Web-Anwendungen relevant sein können, etwa Statements der affektiven Dimension wie "I enjoy my sessions with this software" oder "Working with this software is satisfying". Andere Items orientieren sich jedoch an der Nutzung von Softwareanwendungen wie z.B. "Getting data files in and out of the system is not easy" oder "It takes too long to learn the software commands."

So berichten Kirakowski, Claridge und Whitehand (1998, S. 1) selbst von erfolglosen Bemühungen, das SUMI-Inventar an die Spezifika von Webanwendungen anzupassen. Sie entwickelten in der Folge mit dem Web Analysis and Measurement Inventory ein neues Inventar, das speziell an den Erfordernissen von Websites ausgerichtet ist.

2.9.3 Web Analysis and MeasureMent Inventory (WAMMI)

Basis für die Entwicklung der verwendeten Itemformulierungen waren die Aussagen und Meinungen von Internetnutzern, Webdesignern und Webmastern bezüglich positiver und negativer Erfahrungen während der Nutzung von Websites. Die Aussagen wurden inhaltlich ausgewertet und in die fünf Dimensionen des SUMI-Inventars einbezogen. Im Ergebnis entstand ein 60 Items umfassender Fragebogen. Die fünf WAMMI-Faktoren zur Erfassung der "Zufriedenheit mit der Website" umfassen jeweils 12 Items, die auf einer siebenstufigen Ratingskala zu bewerten sind und werden wie folgt beschrieben (Kirakowski, Claridge & Whitehand, 1998, S. 2 f.):

Attraktivität (Attractiveness): Die Items dieser Dimension beziehen sich auf die anwenderseitig subjektiv empfundene Gefälligkeit der Interaktion mit dem Angebot und messen, inwieweit das Angebot für den Anwender von Interesse ist und dieser die Interaktion als angenehm empfindet.

Kontrolle (Control): Diese Dimension zielt ähnlich wie im SUMI-Inventar auf die Güte der Struktur und Organisation der Website und darauf, wie leicht dem Anwender die Navigation auf der Website fällt.

Effizienz (Efficiency): Items dieser Dimension zielen auf die Einschätzung des Anwenders, wie leicht und schnell die Website die gesuchten Informationen bereitstellt und ob diese in einer akzeptablen Zeit abrufbar sind.

Unterstützung (Helpfulness): Diese Dimension bezieht sich auf die Einschätzung des Anwenders, inwieweit die Website ihn bei seiner Problemlösung mittels einer zielführenden Navigation und hilfreicher Informationen unterstützt.

Erlernbarkeit (Learnability): Die Items dieser Dimension messen, inwieweit es dem Anwender möglich ist, die Website bereits beim Erstbesuch mit einem Minimum an Einweisung und Erklärung zu nutzen und den Umgang mit den angebotenen Funktionalitäten schnell zu erlernen.

Aus den Statements mit den höchsten Reliabilitätswerten wurde eine 20-Item Kurzversion mit fünfstufigen bipolaren Ratingskalen gebildet, die auf insgesamt 25 Websites mit 1.500 Nutzern eingesetzt wurde.

	Strongly Agree							Strongly Disagree
This web site has much that is of interest to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
It is difficult to move around this web site.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Abb. 2-13: Zwei Items aus dem WAMMI-Fragebogen (Kirakowski & Claridge, 2001)

Gütenachweise

Die globale interne Konsistenz des WAMMI-Inventars ist mit einem Cronbachs Alpha in Höhe von 0.92 hoch. Für die Sub-Skalen der fünf genannten Dimensionen variiert der Wert zwischen 0.70 (Kontrolle) und 0.90 (Attraktivität). Eine erste Validitätsprüfung des WAMMI-Inventars erfolgte im Rahmen einer Feldstudie zur Überprüfung von fünf informationsorientierten Webangeboten, von denen zwei mit Hilfe von HCI-Experten entwickelt worden waren. Ein Vergleich der WAMMI-Werte für die mit bzw. ohne Hilfe von HCI-Experten entwickelten Webangebote zeigte signifikante Unterschiede ($p < .05$) für alle fünf Sub-Skalen (Kirakowski, Claridge & Whitehand, 1998, S. 8).

Für die 20-Item Kurzversion beträgt das Cronbachs Alpha 0.92, die Werte der Subskalen variieren etwas niedriger zwischen 0.63 (Effizienz) und 0.74 (Erlernbarkeit) (Kirakowski & Claridge, 2001). In einer Validitätsprüfung des Inventars im Rahmen einer Laborstudie anhand von zwei informationsorientierten Websites konnten zwar deutliche Unterschiede in den Performanewerten der beiden Websites gemessen werden, allerdings wurde nicht ausgewiesen, ob diese Ergebnisse signifikant waren. Eine quantitative Einschätzung der Kriteriumsvalidität ist daher nicht möglich.

Anwendbarkeit des Inventars bei Websites

Das WAMMI-Inventar wurde speziell für die Evaluierung von Websites entwickelt, was für die Anwendbarkeit für Web Usability-Tests spricht. Insbesondere die Items der Kurzversion sind sprachlich und inhaltlich an den Anforderungen von informationsorientierten Websites ausgerichtet. Neben der unzureichenden Dokumentation der Validität des Inventars ist

einschränkend anzumerken, dass die veröffentlichten Ergebnisse der Validierungsstudien vornehmlich auf den Beurteilungen von informationsorientierten Websites basieren.

2.9.4 IBM After-Scenario Questionnaire (ASQ)

Der After-Scenario Questionnaire (ASQ) wurde von IBM zur Erhebung der Nutzerzufriedenheit bei szenariobasierten Usability-Tests von Software-Anwendungen entwickelt (Lewis, 1995). Mithilfe eines 3-Item Kurzfragebogens, der direkt nach jedem Szenario eingesetzt wird, wird die Nutzerzufriedenheit mit der Einfachheit der Aufgabenerfüllung, mit dem benötigten Zeitbedarf für die Aufgabenlösung und den unterstützenden Informationen erhoben. Die Items sind als Statements formuliert und werden auf siebenstufigen bipolaren Ratingskalen (strongly agree vs. strongly disagree) zzgl. einer Ausweichkategorie (not applicable) erhoben, ergänzt um Raum zur freien Kommentierung (ebda S. 62).

For each of the questions below, circle the answer of your choice.

1. Overall, I am satisfied with the ease of completing the tasks in this scenario.

Strongly agree		Strongly disagree	Not applicable
1	2	3	4
5	6	7	N/A

Comments:

Abb. 2-14: Auszug aus dem After-Scenario Questionnaire (Lewis, 1995)

Gütenachweise

Die globale interne Konsistenz des ASQ-Inventars ist mit einem Cronbachs Alpha in Höhe von 0.93 hoch. Für die Kriteriumsvalidität des ASQ sprechen die Ergebnisse eines Labortests, in dem die gemessenen ASQ-Werte signifikant ($p < .01$) mit den Beobachtungswerten bzgl. des Erfolges bzw. Misserfolges der Versuchspersonen bei der Aufgabenlösung korrelierten. Versuchspersonen, die die Aufgaben erfolgreich lösten, stimmten den Statements stärker zu und wiesen damit geringere ASQ-Werte aus als Versuchspersonen, die die Aufgaben nicht erfolgreich lösten. Auch diskriminieren die Items signifikant zwischen den verschiedenen getesteten Systemen und Szenarien. So konnten in Bezug auf die Befragungswerte ein signifikanter Haupteffekt der Szenarien ($p < .001$) und ein signifikanter Szenario x System Interaktionseffekt ($p < .05$) festgestellt werden (ebda S. 62 ff.).

Anwendbarkeit des Inventars bei Websites

Insgesamt ist die Aussagekraft des Inventars in Bezug auf die Usability der getesteten Anwendung aufgrund der begrenzten Itemanzahl nur begrenzt. Auch wenn das Inventar ursprünglich für die Anwendung auf Software-Applikationen entwickelt wurde, sind die ersten beiden Items (die Anwenderzufriedenheit in Bezug auf den wahrgenommenen Schwierigkeitsgrad der Aufgabe und den wahrgenommenen Zeitbedarf für die Aufgabenlösung) global formuliert, so dass sie auch für die szenariobasierte Evaluation von Websites geeignet sind.

Das dritte Statement, das die Zufriedenheit mit den unterstützenden Informationen abfragt, bezieht sich gemäß seiner ursprünglichen Intention auf die für Software-Anwendungen üblichen Online-Dokumentationen und Handbücher und ist für Websites nicht geeignet.

2.9.5 IBM Post-Study System Usability Questionnaire (PSSUQ) & Computer System Usability Questionnaire (CSUQ)

Wie der ASQ wurde auch der Post-Study System Usability Questionnaire (PSSUQ) für den Einsatz in szenariobasierten Usability-Tests entwickelt (Lewis, 1993). Beim Computer System Usability Questionnaire (CSUQ) handelt es sich um eine leicht modifizierte Version des PSSUQ, bei dem die Items lediglich sprachlich auf eine Benutzbarkeit außerhalb des szenariobasierten Usability-Tests angepasst wurden. Im Gegensatz zum ASQ, der nach jeder gelösten Aufgabe bearbeitet wird, werden diese Inventare einmalig nach Beendigung des gesamten Tests ausgefüllt. CSUQ und PSSUQ bestehen aus jeweils 19 Statements mit siebenstufigen, bipolaren Antwortskalen (strongly agree bis strongly disagree) zuzüglich einer Ausweichkategorie (not applicable). Die Items wurden von einer Gruppe Usability-Experten entwickelt und basieren auf einem praxisorientierten subjektiven Verständnis der usabilityrelevanten Dimensionen "Einfache Nutzung" (ease of use), "Leichte Erlernbarkeit" (ease of learning), "Einfachheit" (simplicity), "Effektivität" (effectiveness), "Informationen" (information) und "Benutzerschnittstelle" (user interface). Die Durchführung einer exploratorischen Faktoranalyse ergab eine drei faktorielle Struktur mit den Faktoren „System Usefulness“, „Information Quality“ und „Interface Quality“. An diesem Vorgehen ist als kritisch einzuschätzen, dass keine an der Theorie und Forschung ausgerichtete konzeptionelle Einbettung der Zielkonstrukte erfolgte.

Gütenachweise

Mit einem globalen Cronbachs Alpha in Höhe von 0.97 und Werten für die Sub-Skalen zwischen 0.91 (Information- und Interface Quality) und 0.96 (System Usefulness) kann dem PSSUQ-Inventar eine sehr gute interne Konsistenz bescheinigt werden. Auch die interne Konsistenz des CSUQ-Inventar ist mit einem Cronbachs Alpha von 0.95 und Werten für die Sub-Skalen von 0.89 (Interface Quality) bis 0.93 (System Usefulness) hoch.

Lewis (1993) prüfte die Validität über die Korrelation der PSSUQ-Werte mit dem Anteil der erfolgreich gelösten Testaufgaben sowie mit den ASQ-Werten. Für die Gesamtskala ergab sich eine signifikante Korrelation ($r = -.40, p < .05$) mit den Beobachtungswerten; ebenso für die Sub-Skala System Usefulness ($r = -.40, p < .01$). Diese Werte sprechen für eine mittlere Validität (vgl. Weise, 1975, S. 219). Für die Sub-Skala Interface Quality konnte lediglich eine Tendenz ($r = -.29, p < .10$) und für die Sub-Skala Information Quality keine signifikante Korrelation mit den Beobachtungsdaten festgestellt werden. Bezüglich der ASQ-Werte konnte eine hohe signifikante Korrelation ($r = .80, p < .0001$) mit den PSSUQ-Werten festgestellt werden. Ein unterstützendes Indiz für die Validität des Inventars ist die signifikante Diskriminierung der Gesamtskala sowie aller drei Sub-Skalen-Werte zwischen den drei Nutzergruppen ($p < .10$).

Anwendbarkeit des Inventars bei Websites

Die Statements von PSSUQ und CSUQ sind applikationsunspezifisch formuliert, so dass insbesondere die Items des CSUQ nach sprachlichen Modifikationen und einer Übersetzung ins Deutsche für die standardisierte Beurteilung von Websites in Frage kommen. Ergebnisse einer diesbezüglichen Überprüfung der Faktorstruktur, Reliabilitätswerte und Validitäten sind dem Verfasser nicht bekannt.

2.9.6 AttrakDiff 2

Der Fragebogen AttrakDiff 2 dient nicht primär der Erfassung der Usability, sondern der Messung von Qualitätsaspekten, die über die reine Gebrauchstauglichkeit hinausgehen (Hassenzahl, Burmester & Koller, 2003). Die Autoren weisen darauf hin, dass Gebrauchstauglichkeit inzwischen zwar ein anerkanntes und breit gefordertes Qualitätsmerkmal sei, dass es daneben allerdings weitere Qualitätsanforderungen zu berücksichtigen gelte (Hassenzahl, Platz, Burmester & Lehner, 2000; Burmester, Hassenzahl & Koller, 2002). So verbinden Personen mit der Nutzung eines Produktes neben der Erwartung einer effektiven und effizienten Zielerreichung auch die Bedürfnisse, dass das Produkt stimulierend und identitätsstiftend sei (Hassenzahl, 2001).

Produkte können stimulierend wirken, indem sie Nutzer in ihrer persönlichen Entwicklung, in der Verbesserung ihrer Kenntnisse und Fähigkeiten unterstützen. Ebenso kann ein Produkt identitätsstiftend sein, indem es eine gewünschte Identität oder Facetten derer kommuniziert, sodass der Nutzer durch das Objekt auch sein Selbst zum Ausdruck bringt (Prentice, 1987).

Hassenzahl et al. (2003) unterscheiden im Hinblick auf Software eine *pragmatische Qualität*, die bezeichnet, inwiefern ein interaktives Produkt geeignet ist, seine Umwelt zu manipulieren und vom Nutzer auch entsprechend wahrgenommen wird, und eine *hedonische Qualität*. Diese liegt vor, wenn ein interaktives Produkt aufgrund seiner visuellen Gestaltung oder neuartiger Interaktionsformen stimuliert oder eine gewünschte Identität kommuniziert.

Der AttrakDiff 2, ein Fragebogen im Format eines semantischen Differentials, umfasst 21 siebenstufige Items, deren Endpunkte jeweils durch ein gegensätzliches Adjektiv beschrieben werden (z.B. „außergewöhnlich üblich“, „verwirrend übersichtlich“). Der Fragebogen dient der Erhebung der Konstrukte der *pragmatischen Qualität* (PQ), der *hedonischen Qualität-Stimulation* (HQ-S) sowie der *hedonischen Qualität-Identität* (HQ-I). Die Entwicklung des AttrakDiff 2 erfolgte auf Grundlage des AttrakDiff 1, der nach Aussage der Autoren den Nachteil hatte, dass er es nicht vermochte, zwischen den beiden Aspekten der hedonischen Qualität, Identität und Stimulation zu differenzieren. Der Itempool für die neue Skala setzte sich aus 50 von Usability-Experten gesammelten Items zur Erfassung der drei Komponenten HQ-S, HQ-I und PQ zusammen. Mit Hilfe einer Hauptkomponentenanalyse (Varimax-Rotation) wurden 21 verbleibende Items identifiziert, auf Grund derer in einer erneuten HKA unter Anwendung des Kaiser-Kriteriums die drei Komponenten HQ-S, HQ-I und PQ bestätigt werden konnten.

Beurteilung des Produkts Demo - A

Bitte geben Sie mit Hilfe der folgenden Wortpaare Ihren Eindruck zu **Demo - A** wieder.
Bitte klicken Sie in jeder Zeile eine Position an!

menschlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	technisch
isolierend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	verbindend

Abb. 2-15: Zwei Items aus dem AttrakDiff 2-Fragebogen (www.attrakdiff.de)

Gütenachweise

In einer Untersuchung wurden 4 unterschiedliche graphische Darbietungsformen, sogenannte „Skins“, derselben Software zum Abspielen von MP3-Musikdateien getestet. Die interne Konsistenz ist mit Werten für Cronbachs Alpha für die drei Skalen in Höhe von 0.73 bis 0.90 gut bis sehr gut. Zwischen den beiden Subskalen zur hedonischen Qualität HQ-S und HQ-I ergibt sich eine deutliche Korrelation von $r = .55$. Unterschiede in den Skalenwerten für die vier unterschiedlichen Darbietungsformen interpretierten die Autoren als Hinweis für Konstruktvalidität.

Anwendbarkeit des Inventars bei Websites

Wie bereits ausgeführt, stellt der Fragebogen AttrakDiff 2 kein klassisches Verfahren zur Erfassung der Usability dar. Das Inventar wurde für die Messung der hedonischen und pragmatischen Qualität interaktiver Produkte entwickelt. In dieser Hinsicht sind die 21 Items sowohl sprachlich als auch inhaltlich auch auf Websites anwendbar. Einschränkend kann angemerkt werden, dass die veröffentlichten Ergebnisse der Validierungsstudien auf den Beurteilungen von interaktiven Softwareprodukten basieren.

2.9.7 Wahrgenommene Orientierung in Online-Shops (WOOS)

Das Inventar zur Erfassung der wahrgenommenen Orientierung in Online-Shops (WOOS) befasst sich mit dem Aspekt der durch den Nutzer wahrgenommenen Qualität der Orientierung während der Nutzung eines Online-Shops (Yom, 2001; Yom & Wilhelm, 2004). Das zugrundeliegende Phänomen der Desorientierung beschreibt Conklin (1987) mit "Lost in Hyperspace" und wird von Roselli (1991, S. 42) definiert als "... the problem of [not] knowing precisely at which point in the network he is at any given moment and how to reach a specific node within it."

Begründet werden die Ursachen für Desorientierung in einem zusätzlichen "Cognitive Load" während der Informationsaufnahme und -verarbeitung in hypermedialen Kontexten. Im Gegensatz zur Vorgehensweise bei linearer Präsentation der Informationsinhalte muss der Nutzer diese in einer hypermedialen Umgebung durch Navigationsentscheidungen auswählen und zu einer zielführenden Struktur vernetzen. So muss der Anwender permanent Selektionsentscheidungen treffen, welche er aus der Vielzahl der angebotenen Informationen nutzen möchte und kann dadurch von seinem eigentlichen Ziel abgelenkt werden. In Anlehnung an die oben genannte Definition beschreibt Yom (2003) Orientierung in Online-Shops als die durch den Nutzer subjektiv empfundene Leichtigkeit, während der Nutzung eine Lagebestimmung durchzuführen, die hierfür zur Verfügung stehenden unterstützenden Navigations-elemente zu erkennen und zielführend nutzen zu können.

Grundlage für die Entwicklung der ersten Version waren verschiedene Operationalisierungen des Konstruktes Orientierungsfreundlichkeit klassischer Offline-Ladenwelten aus der Konsumentenforschung (Bost, 1987; Gröppel, 1988), sowie von Otter und Jonson (2000) eingesetzte Items. Usability-Experten entwickelten darauf aufbauend, an den Anforderungen von Online-Shops orientiert, zehn Statements, die auf einer sechsstufigen Ratingskala (1 trifft überhaupt nicht zu bis 6 trifft voll und ganz zu) zu bewerten sind. Die Anzahl der Items wurde im Rahmen einer empirischen Untersuchung von Online-Reiseshops auf Grundlage der Faktorenstruktur einer Hauptkomponenten- sowie Reliabilitätsanalyse von zehn auf sieben Items reduziert.

Gütenachweise

Die Überprüfung der internen Konsistenz, Konstrukt- und Kriteriumsvalidität erfolgte im Rahmen einer laborexperimentellen Untersuchung von Online-Buchshops. Ein Cronbachs Alpha Wert von 0.90 für das Inventar der Version 2 weist insgesamt auf eine sehr gute interne Konsistenz hin und konnte gegenüber der Version 1 trotz Herausnahme der Items W08, W09 und W10 von ehemals 0.87 nochmals gesteigert werden.

Zur Überprüfung der Konstruktvalidität wurden als abhängige Variablen die wahrgenommene Orientierung in der 7-Item-Version und das Shoppererleben erhoben. Die Operationalisierung des Shoppererlebens erfolgte anhand einer von Yom (2001) in Anlehnung an

Donovan und Rossiter (1982) entwickelten bipolaren Vier-Item-Skala zur Erfassung der affektiven Komponente des Shopperlebens. Die Erhöhung der Varianz in der Orientierung erfolgte durch den Einsatz einer Guided Tour zur Unterstützung der User. Die Kriteriumsvalidierung erfolgte anhand der abhängigen Variablen *benötigte Anzahl der Clicks*, *benötigte Zeit für die Produktsuche und bestellung* sowie *globale Beurteilung der Benutzungs-freundlichkeit des Online-Shops*. Auf Grundlage einer signifikanten Korrelation zwischen WOOS-Werten und den dargestellten AVs konnte die Konstrukt- und Kriteriumsvalidität des WOOS-Inventars bestätigt werden (Yom & Wilhelm, 2004). Diese Ergebnisse korrespondieren mit der Annahme von Otter und Johnson (2000, S. 15), dass Orientierung eine Facette der wahrgenommenen Usability sei.

Signifikante Korrelationen der WOOS-Werte mit der benötigten Anzahl der Clicks für die Produktsuche ($r = -.60, p < .01$) sowie für die Produktbestellung ($r = -.34, p < .05$) unterstützen diese Ergebnisse. Bezüglich der benötigten Zeit für die Produktsuche konnte ebenfalls eine signifikante Korrelation ($r = -.56, p < .01$), für die benötigte Zeit im Bestellvorgang allerdings nur eine Tendenz ($r = -.27, p < .10$) festgestellt werden. Bei Betrachtung der kumulierten Werte bzgl. Clickanzahl bzw. benötigter Zeit für Produktsuche und Bestellvorgang, ergeben sich signifikante Korrelationen der WOOS-Werte mit der Gesamtzahl benötigter Clicks ($F = -0,63, p < .01$) sowie mit der benötigten Gesamtzeit. Die Kriteriumsvalidität von WOOS ist damit als mittelmäßig bis hoch zu beurteilen (vgl. Weise, 1975, S. 219).

Bezüglich der Wirkung der Guided Tour auf die wahrgenommene Orientierung konnte ein signifikanter Effekt ($t = 2,09, df = 38; p < .05$) gemessen werden, was für die Konstruktvalidität der WOOS-Skala spricht. Ein signifikanter Effekt der Guided Tour auf das Shoppererleben konnte nicht nachgewiesen werden.

Anwendbarkeit des Inventars bei Websites

Das WOOS-Inventar wurde explizit in Hinblick auf die spezifischen Anforderungen von Websites entwickelt und ist somit für die standardisierte Beurteilung von Webangeboten geeignet. Allerdings zielt das Inventar weniger auf die globale Erfassung der Usability einer Website als vielmehr auf die Qualität der Orientierung des Nutzers innerhalb der Website als spezifischer Facette der Usability.

2.9.8 Usability Fragebogen für Online-Shops (ufosV2)

Der Fragebogen ufosV2 stellt eine Weiterentwicklung des ufosV1 zur Erfassung des Konstruktes Usability von Online-Shops aus Kundensicht dar. Christophersen (2007) zeigt dabei auf, dass dieses Konstrukt sowohl unter Zugrundelegung eines reflektiven als auch eines formativen Messmodells erfasst werden kann und dass im Falle einer Fehlspezifikation negative Auswirkungen auf die Güte der Erfassung zu erwarten sind (ebenso Christophersen & Konradt, 2008). Bei der Entwicklung von Fragebögen zur Erfassung der Usability fand diese Unterscheidung der beiden Messmodelle bis dato keine Berücksichtigung. So wurden in

der Vergangenheit nach Einschätzung von Christophersen (2007) Skalen zur Usability-Messung häufig auf Basis einer reflektiven Messmodellannahme konstruiert, obwohl die Konstrukte eigentlich formative Messmodelle erforderten. Ziel der Entwicklung des ufosV2 war es, eine Fehlspezifikation im Messmodell des Usability-Fragebogens für Online-Shops (ufosV1) von Konradt et al. (2003) auszuräumen.

Wie in Abschnitt 2.6. dargestellt, wird in formativen Messmodellen im Gegensatz zu reflektiven Messmodellen die latente Variable durch die Indikatoren verursacht. In der Folge müssen nicht zwingend hohe Korrelationen zwischen den formativen Indikatoren eines Konstruktes auftreten. Folglich ist die Anwendung von an der klassischen Testtheorie orientierten Ansätzen zur Skalenentwicklung, wie bspw. der Einsatz der Faktorenanalyse oder das Cronbachs Alpha bei formativen Indikatoren nicht gerechtfertigt. Bei reflektiven Indikatoren hingegen lässt die Höhe der Korrelationen zwischen den Indikatoren auf die Güte der Skalenreliabilität schließen. Nach Einschätzung des Autors der vorliegenden Arbeit liegt nahezu allen in diesem Abschnitt betrachteten Usability-Fragebögen ein fehlspezifiziertes Messmodell zugrunde. Grundsätzlich können zwei Arten der Fehlspezifikation unterschieden werden. Einerseits können in einer Skala ausschließlich formative Usability-Indikatoren Verwendung finden, während als Gütekriterium das Maß Cronbachs Alpha herangezogen wird. Andererseits können sich in derselben Skala sowohl reflektive als auch formative Indikatoren finden. Diese Form der Fehlspezifikation liegt bei den betrachteten Inventaren QUIS, SUMI, WAMMI, ASQ, CSUQ und WOOS vor.

Auf Basis des Itempools des ufosV1 sowie einer Literatursichtung wurden sowohl eine formative Skala (ufosV2f) mit 46 Items als auch eine reflektive Skala (ufosV2r) mit 8 Items zur Erfassung der Usability von Online-Shops aus Kundensicht entwickelt. Es erfolgte eine Überprüfung und Modifikation der Items des ufosV1 anhand der Analyse fehlender Werte, der Optimierung der Itemschwierigkeiten sowie einer Prüfung auf inhaltliche Redundanz unter den Items. Zur Überprüfung der Items auf Verständlichkeit und allgemeine Eignung wurde eine Voruntersuchung unter Beteiligung von Usability-Experten durchgeführt. Im Rahmen dieser Vorstudie erfolgte auch die Bestimmung des zugrundeliegenden Messmodells (formativ vs. reflektiv). Die Erfassung der Items erfolgt anhand siebenstufiger Ratingskalen mit den Polen „stimme gar nicht zu“ bis „stimme voll zu“, zuzüglich einer Ausweichkategorie mit der Bezeichnung „keine Angabe“.

Die Preise der einzelnen Produkte sind sofort ersichtlich.
stimme gar nicht zu stimme voll zu keine Angabe

Der Warenkorb ist übersichtlich.
stimme gar nicht zu stimme voll zu keine Angabe

Abb. 2-16: Beispielitems aus dem ufosV2-Fragebogen (Christophersen, 2007)

Gütenachweise

Zur Überprüfung der Gütekriterien der Skalen wurde ein Laborexperiment mit $N = 378$ Teilnehmern durchgeführt. Die Teilnehmer beurteilten jeweils zwei von 35 Online-Shops bezüglich ihrer Usability und weiterer kundennaher Erfolgsfaktoren.

Der Autor berichtet von einem Cronbachs Alpha der reflektiven Usability-Skala ufosV2r mit einem Wert von $\alpha = 0.94$. Die Reliabilität der ufosV2r-Skala zur Erfassung dieses Kernkonstruktes kann damit als sehr gut gewertet werden. Die Überprüfung der Validität der ufosV2r-Skala erfolgte mithilfe *Direct Oblim* rotierter Hauptachsenanalysen (HAA) sowie einer konfirmatorischen Faktorenanalyse unter Anwendung der Asymptotically Distribution-Free Methode (ADF). Aus Tabelle 2-5 gehen die Ausprägungen der relevanten Gütemaße der durchgeführten CFA hervor.

Tab. 2-5: Evaluation des Messmodells der CFA (Christophersen, 2007, S. 182)

1. Globalkriterien:						
χ^2	= 256,27	GFI	= 0,95	SRMR	= 0,067	
df	= 87	AGFI	= 0,92	RMSEA	= 0,05	
χ^2/df	= 2,95	CFI	= 0,93	(LO 90 = 0,4; HI 90 = 0,6)		
NFI	= 0,90					
2. Detailkriterien:						
Konstrukt	Indikator	Indikator-reliabilität $rel(x_i)$	Konvergenzvalidität		Diskriminanzvalidität	
			DEV(ξ_j)	$rel(\xi_j)$	DEV(ξ_j)	φ_{ij}^2
ξ_1 Usability (ufosV2r)	Ur1	0,764	0,670	0,942	0,670	> 0,39 (Vertrauen) > 0,27 (W. Produktanz.) > 0,50 (Ästhetik)
	Ur2	0,509				
	Ur3	0,631				
	Ur4	0,690				
	Ur6	0,667				
	Ur7	0,872				
	Ur8	0,596				
	Ur9	0,659				

Anmerkung: Faktor j (ξ_j), Indikatorreliabilität ($rel(x_i)$), Chi-Quadrat-Wert (χ^2), Freiheitsgrade (df), durchschnittlich erfasste Varianz eines Faktors $DEV(\xi_j)$, quadrierte Korrelation mit LVn (φ_{ij}^2).

Die Überprüfung, inwieweit die reflektive Skala ufosV2r in der Lage ist, Unterschiede zwischen den untersuchten Objekten zu identifizieren, erfolgte anhand einer mehrdimensionalen einfaktoriellen Varianzanalyse auf einem Niveau von $p < .001$ signifikant.

Die Validierung des formativen Usability-Maßes ufosV2f erfolgte mithilfe eines PLS-Zwei-Konstrukt-Modells, im Rahmen dessen die Skala als Prädiktor für das reflektive Messmodell ufosV2r eingebunden wurde. Ein hoher Pfadkoeffizient von β 0,88 sowie ein Determinationskoeffizient von R^2 0,77 lassen nach Auffassung des Autors darauf schließen, dass das Set der formativen ufosV2f-Indikatoren das Globalurteil der Usability, welches durch die reflektive Skala ufosV2r erfasst wird, offensichtlich sehr gut abdeckt.

2.10 Nomologisches Modell

Wie in Abschnitt 2.5 benannt, bietet es sich an, die zu entwickelnden Usability-Skalen zu evaluieren, indem diese in einem komplexen Modell in einen nomologischen Zusammenhang¹ zu weiteren Variablen gebracht werden. Nachfolgend werden daher die neuen Usabilitymaße in einem nomologischen Modell zu weiteren kundennahen Erfolgsfaktoren kommerzieller Websites in Beziehung gesetzt. In Abbildung 2-17 wird das nomologische Modell zur Untersuchung der Einflusskriterien auf Nutzungs- und Auswahlentscheidungen bei Dienstleistungen im Kontext von Webangeboten dargestellt. Als Erfolgsgrößen dienen die Variablen *Nutzerzufriedenheit*, *Nutzungsintention* sowie *Auswahl*, die im Abschnitt 2.5.4 dargestellt und diskutiert wurden. Gemäß der Annahme der Theorie des geplanten Verhaltens dient die

¹ Die nomologische Validität besteht in dem Maße, wie sich theoretisch angenommene Zusammenhänge zu weiteren Konstrukten in einem Modell bestätigen (Fritz, 1995, S. 138 f.).

Erfolgsgröße Nutzungsintention als Prädiktor für die Auswahlentscheidung (Ajzen, 1991). Neben der Usability werden die kundennahen Erfolgsfaktoren Vertrauen, Ästhetik, Nutzungsvergnügen, das Involvement sowie die Reputation als Prädiktoren herangezogen. Diese Konstrukte werden in den nachfolgenden Abschnitten 2.10.1. bis 2.10.4 vorgestellt.

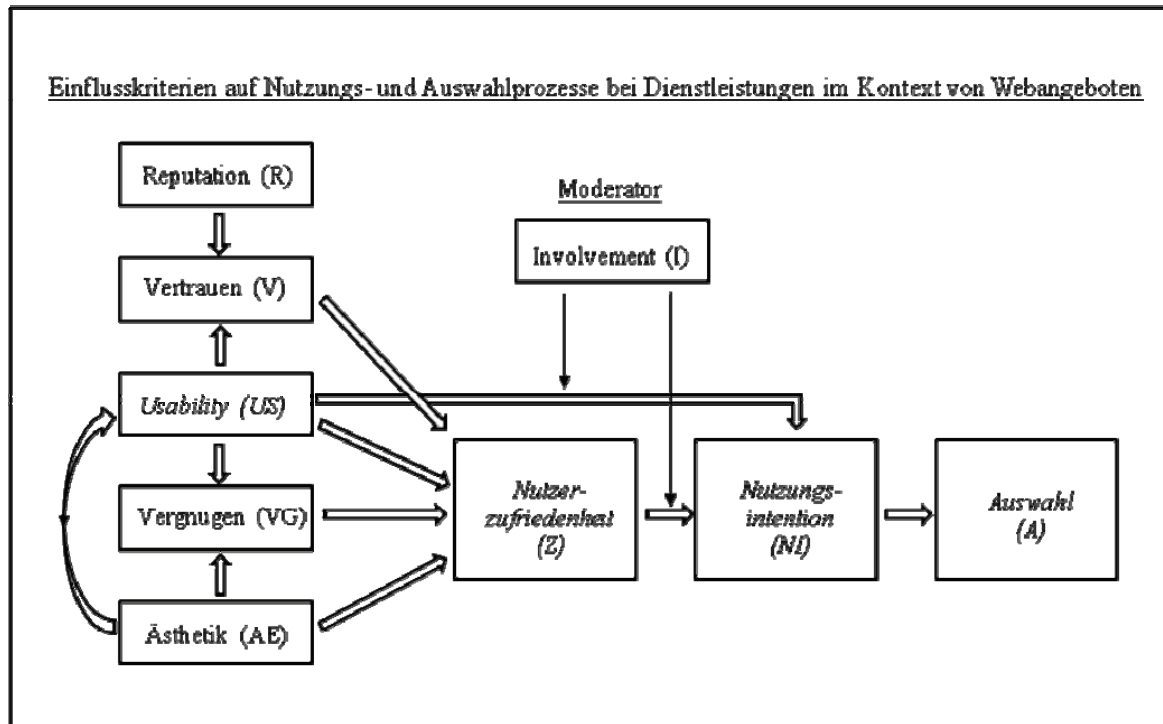


Abb. 2-17: Nomologisches Modell. Die Konstrukte des Technology Acceptance Model (TAM) sind kursiv dargestellt.

Als theoretische Grundlage wird das Technology Acceptance Model (Davis, 1989, S. 320ff.; vgl. auch Adams, Nelson & Todd, 1992; Davis, 1993, S. 452ff.; Jackson, Chow & Leitch, 1997) herangezogen. Davis (1989) stellte anlässlich der Einführung eines E-mail Systems und eines File-Editors bei IBM Canada einen signifikanten Einfluss der Usability auf die anwenderseitig berichtete Nutzung der Systeme bzw. in einer daran anschließenden Studie auf die Nutzungsintention fest. Der positive Zusammenhang zwischen Intention und Handeln wird durch die Theorie des überlegten Handelns (Fishbein & Ajzen, 1975) sowie die Theorie des geplanten Verhaltens (Ajzen, 1985) ausführlich beschrieben. Mahmood, Hall und Swanberg (2001) konnten darüber hinaus einen signifikanten Einfluss der Usability auf die Nutzerzufriedenheit feststellen. Diese wiederum stellt eine wesentliche Einflussgröße für Technologieakzeptanz dar (DeLone & McLean, 2003; Wixom & Todd, 2005).

Das TAM wurde in weiterentwickelter Form für eine große Bandbreite von Technologien und Nutzergruppen validiert, beispielsweise für den Anwendungsbereich der PC-Nutzung (Brown, Massey, Montoya-Weiss, & Burkmann, 2002; Igarria, Zinatelli, Cragg & Cavaye, 1997), bei Desktop Video Konferenzen in virtuellen Teams (Townsend, Demarie & Henrickson, 2001) für die Nutzung von E-Mailsystemen (Gefen, 1997; Straub, Keil & Brenner, 1997), von Internet Anwendungen (Venkatesh, 1999), verschiedener Informationstechnologie- und

Softwaresysteme (Dillon & Morris, 1996; Fang, Chan, Brzezinski & Xu, 2005; Karahanna, Straub & Chervany, 1999; Legris, Ingham & Colletere, 2003; Shih, 2006; Shin, 2007) sowie zur Vorhersage der Nutzung von Arbeitsplatz-Intranet-Anwendungen und Internetsystemen (Al-Gahtani & King, 1999; Anandarajan, Simmers & Igarria, 2000; Horton, Buck, Waterson & Clegg, 2001; Konradt, Christophersen & Schäffer-Kuelz, 2006; Szajna, 1996).

Verschiedene empirische Studien untermauern die Validität der Anwendung des TAM auf den Anwendungsbereich kommerzieller Websites (Chen, Gillenson & Sherrell, 2002; Gefen & Straub, 2000; Gefen, Karahanna & Straub, 2003; Koufaris, 2002; Lederer et al., 2000; Moon & Kim, 2001; Pavlou & Fygenson, 2006). Auch wenn das TAM ursprünglich begründet wurde, um die Nutzung neuer Technologien am Arbeitsplatz zu untersuchen, so kann das Modell verlässlich genutzt werden, um Konsumentenverhalten im E-Commerce vorherzusagen. Pavlou (2003) bezeichnet die TAM-Variablen als Schlüsselgrößen für die Akzeptanz von Anwendungen des E-Commerce und begründet dies damit, dass diese stark technologiegetrieben sind.

Chen, Gillenson und Sherrell (2004) unterscheiden bezüglich E-Commerce-Anwendungen zwei Ebenen der Systemnutzung. So werden Webangebote einerseits für die Informationssuche und andererseits zur Abwicklung von Onlinekäufen genutzt. Die Akzeptanz einer Website ist in dem Maße begründet, wie der Nutzer beabsichtigt, diese zur Informationsgewinnung bzw. für Kauftransaktionen zu nutzen. Pavlou und Fygenson (2006) unterscheiden entsprechend bei der Betrachtung der TAM-Variablen hinsichtlich der Anwendungen Informationssuche und Produktkauf.

Das TAM in seiner ursprünglichen Ausgestaltung wurde mehrfach erweitert, um den Einfluss weiterer Variablen auf die Nutzungsintention und das Nutzungsverhalten für den Anwendungsbereich des E-Commerce zu untersuchen. So wurde das ursprüngliche Modell aufgrund der Unsicherheit von Anwendungen des E-Commerce mehrfach um den Einfluss des Vertrauens auf die Akzeptanz des Systems ergänzt (Chiravuri & Nazareth, 2001; Dahlberg, Mallat & Oorni, 2003; Pavlou, 2003; Suh & Han, 2003). Vertrauen erfolgt dabei in dem Maße, in dem ein Anwender eine E-Commerce-Anwendung als sicher bzw. den Anbieter als vertrauenswürdig betrachtet. Pavlou (2003) erweitert das Modell um die Kontrollvariable Anbieterreputation (vgl. auch Keat & Mohan, 2004). Die Reputation eines Webanbieters ist eine wichtige Einflussgröße auf das Vertrauen des Anwenders (Dooney & Cannon, 1997) und beeinflusst die Kaufintention (Jarvenpaa, Tractinsky, Saarinen & Vitale, 1999). Venkatesh, Cheri und Morris (2002) unterstellen einen positiven Einfluss intrinsischer Motivation auf die wahrgenommene Usability und schlagen deshalb vor, Einflussfaktoren der intrinsischen Motivation, wie Vergnügen und Freude während der Nutzung, zu berücksichtigen (vgl. auch Dahlberg et al., 2003).

2.10.1 Vertrauen

Vertrauen spielt in sozialen Beziehungen, die von Unsicherheit und Abhängigkeit geprägt sind, eine große Rolle. Online-Transaktionen sind häufig durch Unsicherheit aufgrund der Anonymität und einen Mangel an Kontrolle sowie potentiell opportunistischem Anbieterverhalten gekennzeichnet, sodass das Anwendervertrauen einen entscheidenden Erfolgsfaktor im E-Commerce darstellt (Einwiller & Will, 2001).

Luhmann (1989) bezeichnet Vertrauen als einen Mechanismus, um Komplexität in Situationen, in denen Menschen mit Unsicherheit umgehen müssen, zu reduzieren. Eine wesentliche Ursache für mangelndes Vertrauen im E-Commerce ist der durch den Konsumenten gegenüber traditionellen Einkaufssituationen höhere Grad an subjektiv wahrgenommener Unsicherheit elektronischer Transaktionen (Pavlou, 2003). Grabner-Kräuter und Kaluscha (2003) unterscheiden diesbezüglich *systemabhängige* und *transaktionsspezifische Unsicherheit*. Systemabhängige Unsicherheit entsteht aufgrund dynamischer Entwicklungen exogener nicht vom Anwender beeinflussbarer Faktoren und zunehmender Komplexität der Umwelt (Hirshleifer & Riley, 1979). Übertragen auf den Kontext des E-Commerce bedeutet dies Unsicherheit aufgrund technologischer Risiken und Fehlerquellen (Brielmaier & Diller, 1995). Transaktionsspezifische Unsicherheit ist begründet in den unterschiedlichen Informationsständen und dem Verhalten der beteiligten Akteure (Weiber & Adler, 1995). Die Frage, inwieweit der Anbieter willens und in der Lage ist, das angebotene Versprechen einzulösen, kann Unsicherheit beim Kunden erzeugen. Dabei ist das Vertrauen des Kunden in die Fähigkeit und Bereitschaft eines Online-Anbieters, sein Bedürfnis möglichst vorteilhaft zu erfüllen, eine Voraussetzung für den Aufbau und die Weiterentwicklung einer Kundenbeziehung (Grabner-Kräuter & Kaluscha, 2003). Dieses Vertrauen kann im E-Commerce beeinträchtigt sein, wenn der im Dienstleistungsbereich oftmals wichtige persönliche Beratungskontakt durch eine technische Schnittstelle ersetzt wird (Coughlan, Macredie & Patel, 2007; Winand & Pohl, 2000, S. 270). Gerade bei beratungsintensiven Dienstleistungen wie zum Beispiel Krankenversicherungsdienstleistungen ist das Potenzial für einen Vertrauensaufbau über eine technische Schnittstelle nur eingeschränkt vorhanden.

Nach Mayer, Davis und Schoorman (1995, S. 712) ist Vertrauen begründet in der Bereitschaft des Käufers, sich den Aktionen des Anbieters verwundbar auszusetzen, unabhängig von der Möglichkeit dessen Verhalten zu beeinflussen. Neben diesem personellen Vertrauen in den Anbieter ist beim Systemvertrauen die Technology, das Internet selbst, Objekt des Vertrauens (Luhmann, 1989), während das Konstrukt des dispositionalen Vertrauens situationsunabhängig die grundsätzliche individuelle Bereitschaft zu vertrauen kennzeichnet (Mayer et al., 1995).

Das Vertrauen in eine Website nimmt wesentlichen Einfluss auf die Aufenthaltsdauer eines Konsumenten auf dieser (Dahui, Browne & Wetherbe, 2006; Lin, 2007). Entsprechend stellt ein Mangel an Vertrauen einen der am häufigsten von Konsumenten angeführten Gründe dar,

eine Seite zu verlassen und nicht von einem Internetverkäufer zu kaufen (Lee & Turban, 2001, S 75; Pavlou & Gefen, 2005). In verschiedenen empirischen Studien konnte ein positiver Einfluss des Vertrauens in den Anbieter auf die Kauf- bzw. Nutzungsintention bestätigt werden (Chiravuri & Nazareth, 2001; Gefen & Straub, 2000; Nicolaou & McKnight, 2006; Pavlou, 2003). Koufaris und Hampton-Sosa (2002) konnten darüber hinaus einen positiven Einfluss auf die Wiederwahlintention der Konsumenten feststellen, während Mukherjee und Nath, (2007) das Vertrauen als zentrale Einflussgröße für den Aufbau einer langfristigen Kundenbeziehung im Onlinebereich identifizieren konnten. Angesichts der hier aufgezeigten Befunde wird geschlossen:

Es besteht ein positiver Zusammenhang zwischen dem Vertrauen, das Nutzer einem Online-Anbieter entgegenbringen und der Intention, dessen Angebote zu nutzen.

2.10.2 Reputation

Mit Reputation verbindet man das Ansehen bzw. die Qualität einer Person, eines Produktes, einer Organisation oder allgemein einer Institution (Schwalbach, 2004). Die durch Konsumenten wahrgenommene Reputation einer Organisation entspricht dem Maße, in dem diese davon überzeugt sind, dass die Organisation sich ehrlich verhält und um das Wohl ihrer Kunden bemüht (Doney & Cannon, 1997).

Für den Anwendungsbereich von Konsumgütern konnte festgestellt werden, dass die langfristige Reputation eines Anbieters bedeutsamer für den Erfolg ist als eine kurzfristige Veränderung der Produktqualität (Eberl & Schwaiger, 2005; Landon & Smith, 1997). Eine hohe Reputation vermittelt dem Konsumenten Sicherheit über die Leistungsfähigkeit, Integrität und das Wohlwollen des Anbieters. Dies gilt insbesondere, wenn der Konsument nicht auf Erfahrungen mit dem Anbieter zurückgreifen kann (McKnight, Cummings & Chervany, 1998). So konnten Ward und Lee (1999) feststellen, dass Neukunden eher dazu tendieren, sich bei Kaufentscheidungsprozessen im E-Commerce auf die Reputation des Anbieters zu stützen, während erfahrene Kunden ihre Entscheidung auch auf andere Faktoren begründen.

Der Aufbau einer hohen Reputation erfordert langfristige Investitionen in den Aufbau und die Pflege der Kundenbeziehungen und stellt eine wichtige immaterielle Ressource im Prozess des Erringens, der Verbesserung und der Verteidigung der Wettbewerbsposition sowie bei der Erhöhung des Unternehmenswerts dar (Hall, 1992). Je besser die Reputation eines Anbieters und je höher die damit in der Vergangenheit verbundenen Investitionen ausgefallen sind, umso empfindlicher fällt die Bestrafung durch Schädigung der Reputation bei Verletzung des Konsumentenvertrauens aus. Eine gute Reputation signalisiert dem Konsumenten, dass der Anbieter in der Vergangenheit Schwächen auf Seiten der Konsumenten nicht opportunistisch ausgenutzt hat (Smith & Barclay, 1997).

Tatsächlich bemühen sich viele Anbieter im E-Commerce ihre Reputation zu stärken, indem sie beispielsweise Kundenurteile, Besuchszahlen oder durch unabhängige Institutionen vergebene Qualitätssiegel auf ihrer Website veröffentlichen (MacInnes, Li & Yurcik, 2005).

Neben der Reputation eines Anbieters lässt sich auch die Reputation einer Website beurteilen. Nach Rafiei und Mendelzon (2000) entspricht sie dem Anteil der Seiten, welche einen Link auf diese Seite gesetzt haben, an der Gesamtheit aller Seiten, die einen Treffer bei Eingabe eines Suchwortes in eine Suchmaschine anzeigen.

Im Folgenden wird unter Reputation die Reputation des Anbieters verstanden. Dabei nimmt die Reputation in dem Maße zu, in dem Konsumenten davon überzeugt sind, dass der Anbieter sich ehrlich verhält und um das Wohl seiner Kunden bemüht ist. Angesichts der hier aufgezeigten Befunde wird geschlossen:

Es besteht ein positiver Zusammenhang zwischen der Reputation eines Online-Anbieters und dem Vertrauen der Kunden in den Online-Anbieter.

2.10.3 Ästhetik

Für die Optimierung der Informationspräsentation im Internet sind aufgrund der Fülle und Komplexität der präsentierten Informationen die zugrundeliegenden Prozesse der visuellen Aufmerksamkeit von Bedeutung. Eine Anpassung der Benutzerschnittstelle an die kognitiven Limitationen potentieller Nutzer ist erforderlich, damit diese eine Interaktionsqualität wahrnehmen, die sie in ihrer Zielerreichung effizient und effektiv unterstützt. In diesem Zusammenhang ist die Erkenntnis aus der Kognitionspsychologie wichtig, dass die Kapazität des menschlichen Informationsverarbeitungssystems begrenzt ist und diese Ressourcen für die Verarbeitung ausgewählter Informationen genutzt werden, die in dem Moment im Focus der Aufmerksamkeit stehen (Anderson, 1996, S. 10f.). Dabei sind aktivierungstheoretische Erkenntnisse von Nutzen, um die Wirkung der Aktivierung auf die Prozesse der Informationsaufnahme, -verarbeitung und -speicherung zu erklären (Kroeber-Riel, 1992, S. 75). Visuelle emotionale Reize können das Aktivierungsniveau beeinflussen und die Informationsverarbeitung stimulieren (vgl. ebda S. 68).

Die Gestaltung eines Webangebots muss in der informationellen Konkurrenz des Internets so attraktiv und ansprechend gestaltet sein, dass die Aufmerksamkeit des Nutzers auf die wichtigsten Informationen und Funktionen gelenkt wird. So ist häufig der erste Eindruck mitentscheidend, ob ein Webangebot weiter exploriert wird (Jennings, 2000). Webangebote, die ausschließlich nach Usability-Prinzipien gestaltet werden, laufen Gefahr, sich sehr ähnlich zu sehen und aufgrund einer reduzierten äußerlichen Erscheinung Nutzerbedürfnisse zu vernachlässigen.

Insbesondere bei der Gestaltung kommerzieller Webangebote ist nicht nur auf eine gute Usability zu achten, sondern es kann durch eine sorgfältige graphische Ausgestaltung ein distinkter Auftritt geschaffen werden, der in der Erinnerung des Nutzers haften bleibt (Cyr,

2008; Wiesner, 2004, S. 104). Schenkman und Jönsson (2000) identifizierten in einer Studie über 13 Websites aus unterschiedlichen Branchen die wahrgenommene Attraktivität des Angebots als stärkste Einflussgröße auf den Gesamteindruck einer Website.

Hall und Hanna (2004) variierten in einem Experiment systematisch die farbliche Gestaltung von Internetseiten. Im Ergebnis konnten sie feststellen, dass die von Nutzerseite eingeschätzte Ästhetik einer Website positiven Einfluss auf die Kaufintention nimmt. Die dargestellten Befunde legen folgenden Schluss nahe:

Es besteht ein positiver Zusammenhang zwischen der von Nutzerseite eingeschätzten Ästhetik eines Webangebots und deren Intention, dieses Webangebot wieder zu nutzen.

2.10.4 Nutzungsvergnügen

In der Kaufverhaltensforschung werden affektive Faktoren wie die wahrgenommene Freude bzw. Lust beim Einkauf (Babin, Darden & Griffin, 1994; Bost, 1987; Donovan & Rossiter, 1982; Diller & Kusterer, 1986), emotionale Eindrücke (Gröpel, 1988) oder auch Gefallen, Aktivierung und Lebendigkeit (Weinberg & Besemer, 1999) als wichtige Determinanten des Shoppererlebens mit wesentlichem Einfluss auf das Einkaufsverhalten identifiziert. So entsteht Kundenzufriedenheit nicht alleine durch den Kauf eines Produktes, sondern auch aufgrund emotionaler Aspekte während des Kaufvorgangs (Bloch & Richins, 1983).

Im Rahmen einer traditionellen Einkaufssituation führt durch den Käufer empfundenes Vergnügen zu einer längeren Aufenthaltsdauer im Geschäft, einer Ausweitung der Produktkäufe und Ausgaben sowie einer höheren Rückkehrbereitschaft. Je nachdem, ob der Einkauf emotional erfahrungsgetrieben oder zielgerichtet und rational erfolgt, können Unterschiede im Einkaufsverhalten festgestellt werden. Konsumenten zeigen in einer emotionalen und erfahrungsgetriebenen Einkaufssituation tendenziell eine positivere Stimmung, größere Zufriedenheit mit dem Einkaufserlebnis und neigen eher zu Spontankäufen (Babin, Darden & Griffen, 1994).

Diverse Studien im Bereich Mensch-Maschine-Interaktion weisen darauf hin, dass hedonische Aspekte von wesentlichem Einfluss auf das Nutzungsverhalten und die Akzeptanz von Softwareanwendungen sind (Hassenzahl et al., 2000; Hassenzahl, Schöbel & Trautmann, 2008; Hudlicka, 2003; Mundorf, 1993). Nach Logan (1994) erfüllen hedonische Aspekte einer Software die Bedürfnisse der Anwender nach persönlichem Wachstum und sozialem Status. So konnten Igbaria, Schiffmann und Wickowski (1994) bei einer Untersuchung der Einflussfaktoren auf die Akzeptanz von Softwareanwendungen einen ähnlich starken Effekt für die Größen Perceived Fun und Perceived Usefulness aufzeigen. Bezogen auf die Nutzerzufriedenheit wirkt Perceived Fun stärker als Perceived Usefulness.

Bezogen auf Webanwendungen kann die Schlussfolgerung gezogen werden, dass neben einer effizienten und effektiven Informationsdarbietung affektiv und kognitiv stimulierende Einkaufserlebnisse von Einfluss auf die Nutzerzufriedenheit und das Nutzungs- sowie Kaufver-

halten sind. Diverse empirische Studien untermauern den Einfluss des Einkaufsvergnügens für den Bereich des Online-Shoppings. So identifizierte Henderson, Rickwood und Roberts (1998) das Einkaufsvergnügen während eines Online-Einkaufsvorgangs als eine starke Determinante für die Wiedernutzungsabsicht der Website (ebenso Koufaris, 2002). In weiteren Studien zum Online-Einkaufsverhalten konnte ein Einfluss des Einkaufsvergnügens auf die Zufriedenheit während des Einkaufsvorgangs und die Kaufintention nachgewiesen werden (Eighmey, 1997; Jarvenpaa & Todd, 1997a, 1997b). Auf Grundlage der angeführten Befunde wird zusammenfassend die Annahme getroffen:

Es besteht ein positiver Zusammenhang zwischen dem Vergnügen, das die Nutzer bei der Nutzung einer Website erfahren, und der Nutzerzufriedenheit mit dieser Website.

2.11 Zusammenhänge innerhalb des nomologischen Modells

In dem in Kapitel 2.10 dargestellten nomologischen Modell werden zwischen einigen der einbezogenen Konstrukten Zusammenhänge angenommen. Diese Beziehungen werden in den folgenden Abschnitten konzeptionell und empirisch begründet und dargestellt.

2.11.1 Usability und Vertrauen

Das Vertrauen in den Anbieter einer Website ist neben der Usability als eine wesentliche Erfolgsgröße im E-Commerce anzusehen (siehe Abschnitt 2.10.1). In der Literatur wird vielfach die These vertreten, dass zwischen den beiden Größen ein Zusammenhang in ihrer Wirkung auf das Konsumentenverhalten bestehe (Chiravuri & Nazareth, 2001; Dahlberg et al., 2003; Featherman & Pavlou, 2002; Pavlou, 2003).

Keat und Mohan (2004) führten eine Metaanalyse anhand von Untersuchungen zum Einfluss des Nutzervertrauens auf die Akzeptanz von Anwendungen des E-Commerce durch. Auf Basis der Befunde schlagen sie eine Integration des Konstruktes Vertrauen in das Technology Acceptance Model von Davis (1989) in Abhängigkeit der Größen Perceived Usefulness und Perceived Ease of Use vor. Der Einfluss der Usability auf das Vertrauen in den Anbieter einer Website kann damit begründet werden, dass der Eindruck der Gebrauchstauglichkeit den Nutzern als Beurteilungsgrundlage diene, inwiefern dem Anbieter vertraut werden könne (Egger, 2001). So stellten Silberer und Yom (2001, S. 443) in einer Untersuchung des Kaufverhaltens von Internet-Neulingen fest, dass grundsätzlich die Bekanntheit der Online-Marke zur Risikoreduzierung bei den Konsumenten eine große Rolle spielt. Kennen die Konsumenten den Anbieter allerdings nicht oder nur wenig, gewinnen Kriterien der Usability, wie bspw. eine gute Benutzerführung sowie gut aufbereitete Produktinformationen an Bedeutung für das Vertrauen der Konsumenten (Hampton-Sosa & Koufaris, 2005). Die Autoren untersuchten das Vertrauen von Erstbesuchern einer Website und fanden heraus, dass Perceived Usefulness und Perceived Ease of Use einen starken mediierenden Einfluss auf das Vertrauen in den

Anbieter der Website haben und dass die Variable Vertrauen selbst signifikant die Kauf- und Wiederwahlabsicht der Konsumenten beeinflusst.

Der Zusammenhang zwischen der Usability und dem Vertrauen findet weitere empirische Bestätigung bei Dahlberg et al. (2003), Featherman und Pavlou (2003), Koufaris und Hampton-Sosa (2002), Pavlou (2003) sowie Pavlou und Fygenson (2006). Auf Grundlage der angeführten Untersuchungen kann geschlossen werden:

Es besteht ein positiver Zusammenhang zwischen der Usability einer Website und dem Vertrauen, das die Nutzer dem Anbieter der Website entgegen bringen.

2.11.2 Reputation und Vertrauen

Die vom Konsumenten wahrgenommene Reputation einer Organisation wird neben ihrer wahrgenommenen Größe häufig als wesentlicher Einflussfaktor auf das Konsumentenvertrauen benannt (Doney & Cannon, 1997; Ganesan, 1994). Die Theorie des wahrgenommenen Risikos geht davon aus, dass Konsumenten ab einer bestimmten Toleranzschwelle Risikoreduzierungsstrategien einsetzen, um das Risiko zu mindern (Lingenfelder, 2001, S. 380). Silberer und Yom untersuchten derartige Risikoreduzierungsmaßnahmen bei Internet-Neulingen und stellten fest, dass die Bekanntheit der Marke einen großen Einfluss auf das Vertrauen nimmt (Silberer & Yom, 2001, S. 442).

De Ruyter et al. (2001) konnten in empirischen Studien bestätigen, dass eine hohe organisationale Reputation das Vertrauen der Nutzer in E-Services signifikant erhöht. Pavlou (2003) bestätigte eine positive Wirkung der Reputation auf das Vertrauen in den Online-Anbieter. Jarvenpaa et al. (1999, 2000) stellten einen positiven Einfluss der wahrgenommenen Reputation und der wahrgenommenen Größe eines Online-Shops auf das Konsumentenvertrauen in den Online-Shop fest, wobei die wahrgenommene Reputation einen deutlich stärkeren positiven Effekt auf das Konsumentenvertrauen ausübte.

Quelch und Klein (1996) sowie Lohse und Spiller (1998) vermuten, dass eine positive Reputation die Wahrnehmung einer Website positiv beeinflusst. Entsprechend veröffentlichen E-Commerce-Anbieter teilweise Kundenreferenzen auf ihrer Website, die ihnen eine positive Reputation bestätigen. Ebenso wird mit Hilfe von Reputation in Online-Auktionen wie eBay ein effizientes Marktgleichgewicht angestrebt, indem Nachfrager die Möglichkeit haben, eine Verkäuferbewertung abzugeben, die von allen Marktteilnehmern einsehbar ist (Dellarocas, 2002; Gregg & Scott, 2006). Entsprechend dieser Befunde kann die Aussage getroffen werden:

Es besteht ein positiver Zusammenhang zwischen der von Nutzerseite wahrgenommenen Reputation eines E-Commerce-Anbieters und dem Vertrauen der Kunden in den Anbieter.

2.11.3 Usability und Nutzungsvergnügen

Eine hohe Nutzungsfreundlichkeit, die den Nutzer aufgrund eines effizienten und effektiven Informationszugangs darin unterstützt, sein Ziel leicht, schnell und mühelos zu erreichen, ist eine notwendige und wichtige Bedingung für einen positiv erlebten Umgang mit einer Website. Nielsen (2002, S. 2 f.) charakterisiert die Rolle der Usability prägnant wie folgt: "... greatest joy of using computers comes through user empowerment and engagement... Such engagement requires usability. If users can't master the interface, they'll feel oppressed rather than empowered, and are unlikely to explore or use anything beyond the absolute minimum".

Die Konstrukte Usability und Vergnügen basieren auf teilweise widersprüchlichen Grundlagen. Während erstere Einfachheit und Konsistenz in der Gestaltung einer Website begünstigt, unterstützt letztere durchaus ein gewisses Maß an Komplexität und Heterogenität (Carroll & Thomas, 1988). Andererseits kann eine gut ausgeprägte Usability einer Website ein positives Einkaufserlebnis unterstützen. So konnte Koufaris (2002) in einer Studie im Online-Buchhandel einen positiven Einfluss nutzenstiftender Suchmechanismen auf das während der Nutzung einer Website wahrgenommene Vergnügen feststellen. Van der Heijden und Verhagen (2004) bestätigen im Rahmen einer Fragebogenuntersuchung die Hypothese, wonach die Variablen Perceived Usefulness und Perceived Ease of Use signifikanten Einfluss auf das wahrgenommene Einkaufsvergnügen nehmen (vgl. Hassenzahl & Hartwig, 2005).

Aufgrund der angeführten Befunde wird auf einen Zusammenhang zwischen der wahrgenommenen Usability einer Website und dem während der Nutzung erlebten Vergnügen geschlossen:

Es besteht ein positiver Zusammenhang zwischen der von Nutzerseite wahrgenommenen Usability einer Website und dem Vergnügen, das die Nutzer bei der Nutzung der Website erfahren.

2.11.4 Ästhetik und Usability

Im Rahmen ihrer Forschung im Bereich Mensch-Maschine-Interaktion untersuchten Kurosu und Kashimura (1995) den Zusammenhang zwischen Ästhetik und Usability bei der Gestaltung von Software-Interfaces. Dabei konnten Sie feststellen, dass die durch den Nutzer wahrgenommene augenscheinliche, apriori Usability („apparent usability“) einer Benutzeroberfläche viel stärker mit der Einschätzung der Ästhetik der Oberfläche korreliert als mit ihrer tatsächlichen Usability („inherent usability“). Die Autoren schließen daraus, dass weniger die tatsächliche Usability einer Oberfläche als vielmehr ästhetische Aspekte ausschlaggebend dafür sind, wie benutzerfreundlich eine Oberfläche zunächst eingeschätzt wird (vgl. Schrepp, Held & Laugwitz, 2006). Nach Caplan (1994) ist die augenscheinliche Usability als die vom Nutzer wahrgenommene Einfachheit der Nutzung zu verstehen, die auf einem ersten visuellen Eindruck und nicht auf einer tatsächlichen Nutzung basiert.

In Folgestudien konnten Tractinsky (1997) sowie Brady und Phillips (2003) die Untersuchungsergebnisse von Kurosu und Kshimura mit Stichproben aus anderen Kulturkreisen bestätigen. Zudem gelang es Tractinsky, den Einwand zu entkräften, dass der Einfluss der Ästhetik auf die Usability nur während des ersten Eindrucks der Nutzer einer Benutzeroberfläche bestehe. Vielmehr nimmt der erste Eindruck eines Nutzers Einfluss auf die längerfristige Beurteilung eines Software-Interfaces (Tractinsky, 1997). In einer späteren Untersuchung blieben die oben angeführten Zusammenhänge auch dann bestehen, als die Testpersonen die tatsächliche Benutzerfreundlichkeit erfahren haben (Tractinsky, Katz & Ikar, 2000).

Matthews (1999, S. 12) beschreibt ebenfalls einen Zusammenhang zwischen Ästhetik und Usability. Als die Ästhetik unterstützende Gestaltungsmerkmale benennt er neben der Schlichtheit auch Hervorhebung, Unterscheidbarkeit und Klarheit. Angesichts dieser Befunde wird die folgende Annahme getroffen:

Es besteht ein positiver Zusammenhang zwischen der Ästhetik einer Website und der von Nutzerseite eingeschätzten Usability der Website.

2.11.5 Ästhetik und Nutzungsvergnügen

Den Ergebnissen einer Studie von Jordan (1998) zum Nutzungsverhalten bei elektronischen Gebrauchsgütern zufolge besteht ein hoher Zusammenhang zwischen der Einschätzung der Ästhetik eines Produktes und dem Vergnügen, das mit der Produktnutzung einher geht. Verschiedene Untersuchungen unterstützen die These, dass es sich ebenfalls positiv auf das Vergnügen während der Nutzung einer Software auswirkt, wenn die Schnittstelle vom Benutzer als ästhetisch gestaltet wahrgenommen wird. Tractinsky et al. (2000) untersuchten die Zusammenhänge der nutzerseitig wahrgenommenen Ästhetik und Usability der Benutzerschnittstelle von Geldautomaten und kamen zu dem Schluss, dass die Ästhetik beim Nutzer affektive Prozesse auslöst, die sich auf die Stimmung des Nutzers und letztendlich auf seine Gesamtbeurteilung des Systems auswirken.

Bezogen auf den Anwendungsbereich kommerzieller Websites stellen Tractinsky und Rao (2001) einen positiven Einfluss der Ästhetik auf die Variablen Usability und Einkaufsvergnügen fest. Die Autoren führen diesen Zusammenhang auf einen motivationalen Einfluss der Ästhetik auf das Konsumentenverhalten aufgrund ihrer Wirkung auf Kundenzufriedenheit und involvement zurück (vgl. Webster & Martocchio, 1992).

In einer umfangreichen experimentellen Untersuchung stellten Kim, Lee und Choi (2003) Zusammenhänge zwischen den Designfaktoren einer Website und den emotionalen Reaktionen der Nutzer fest. In mehreren Studien identifizierten sie Schlüsseldesignfaktoren, die bei der Gestaltung emotional ansprechender Websites häufig eingesetzt werden sowie 13 emotionale Dimensionen, die dabei bei Nutzern ausgelöst werden können. In einer Onlineuntersuchung konnten sie die Wirkungen der Designfaktoren auf die emotionalen Dimensionen nachweisen.

Diese Ergebnisse legen den Schluss nahe, dass ein enger Zusammenhang zwischen der ästhetischen Gestaltung einer Website und dem affektiven Erleben durch die Nutzer besteht. Angesichts der dargestellten Befunde kann die folgende Annahme getroffen werden:

Es besteht ein positiver Zusammenhang zwischen der von Nutzerseite eingeschätzten Ästhetik einer Website und dem Vergnügen, das die Nutzer bei der Nutzung der Website erfahren.

2.11.6 Moderierende Einflüsse

Neben den in den vorangegangenen Abschnitten dargestellten Zusammenhängen im nomologischen Modell werden moderierende Einflüsse des Involvement auf ausgewählte Beziehungen zwischen den Konstrukten angenommen. Nach einem Überblick zur Thematik und empirischen Begründung der angenommenen moderierenden Einflüsse (siehe Abschnitt 2.11.6.1) werden in Abschnitt 2.11.6.2 verschiedene Methoden zur Analyse latenter Moderatormodelle vorgestellt und eine Auswahl für die vorliegende Arbeit begründet.

2.11.6.1 Begriffsklärung

Eine systematische Beziehung zwischen einer Prädiktorvariablen X und einer Kriteriumsvariablen Y kann durch die Berücksichtigung einer weiteren Prädiktorvariablen Z nicht nur ergänzt werden, sondern auch moderiert und damit interaktiv verändert werden.

Eine Moderatorbeziehung besteht dann, wenn eine qualitative Variable (z.B. personenbezogene, eigenschaftsbezogene oder situationsbezogene Variablen wie Entwicklungsstand, Denkmotivation, Involvement oder soziale Erwünschtheit), oder eine quantitative Variable (z.B. Alter, Gehalt, Wettbewerbsintensität) die Richtung oder die Stärke einer Beziehung zwischen einer Prädiktorvariable X und einer Kriteriumsvariablen Y beeinflusst (Baron & Kenny, 1986). Dabei verändert sich die Intensität der Beziehung zwischen den Größen X und Y mit der Ausprägung der Drittvariablen Z . Die Variable Z wird dabei als Moderatorvariable bezeichnet.

Moderatormodelle werden in der Regel theoriegeleitet zur Erklärung solcher Beziehungen eingesetzt. Dabei ist die Hypothese über einen Moderatoreffekt keine einfache Wirkhypothese, sondern eine Strukturhypothese bezüglich des Zusammenwirkens verschiedener Prädiktorvariablen (Klein, 2000, S. 222). Von der Moderatorvariable zu unterscheiden ist die Mediatorvariable. Diese überträgt im Gegensatz zur Moderatorvariablen den Effekt einer Prädiktorvariablen X auf eine von ihr abhängige Kriteriumsvariable Y vermittelnd (Baron & Kenny, 1986). So ändert sich von Stufe zu Stufe des Mediators Z zwar der Mittelwert von Y , jedoch nicht die Stärke des direkten Effekts von X auf Y . Im Falle der Moderation hingegen ändert sich die Stärke des Effekts von X auf Y .

Eine Moderatorvariable in einem Regressionsmodell zu berücksichtigen, kann inhaltlich aus unterschiedlichen Gründen hilfreich sein. Einerseits kann eine Erhöhung der differentiellen Vorhersagbarkeit im Vordergrund stehen (Amelang & Bartussek, 1970; Borkenau, 1981).

Andererseits kann eine Moderatorhypothese auch zum Zwecke einer Effektgeneralisierung genutzt werden, indem über den Ausschluss einer Drittvariablenwirkung die externe Validität einer Variable erhöht wird (Dalbert & Schmitt, 1986). Schließlich lassen sich durch die Berücksichtigung von Moderatorvariablen Effekte im Hinblick auf Personen- oder Situationsmerkmale differenzieren (Emmons & Diener, 1986; Mummendey, 1979).

So konnten Parasuraman und Grewal (2000) für den Anwendungsbereich neuer Technologien eine moderierende Wirkung der Variable *Aufgeschlossenheit gegenüber Technologien* auf die Beziehung zwischen der Zufriedenheit mit einer neuen Technologie und deren privater wie auch beruflicher Akzeptanz feststellen. Mittal und Kamakura (2001) stellten starke moderierende Effekte *demographischer Kriterien* wie Geschlecht, Alter, Familienstatus und Bildungsstand auf die Wirkungsbeziehung zwischen Nutzungszufriedenheit und Wiederkaufabsicht bzw. -verhalten fest. In Abhängigkeit des individuellen *Involvement* einer Person können dabei die Aufmerksamkeit und die Intensität der Bemühungen zur Informationsgewinnung im Kaufentscheidungsprozess variieren (Gabott & Hogg, 1999). Bettman (1974) beschreibt den Grad des Involvement als eine Mediatorvariable in Bezug auf die Intensität der Suche nach relevanten Informationen in Kaufentscheidungsprozessen. Ranaweera, McDougall und Bansal (2004, 2005) leiten für den Anwendungsbereich von Webangeboten einen Moderatoreffekt des Involvement auf die Beziehung zwischen Nutzungszufriedenheit und Nutzungsabsicht bzw. Nutzungsverhalten ab. Des Weiteren konnten Suh und Yi (2006) einen moderierenden Effekt des Involvement auf die Beziehung zwischen der Zufriedenheit und der Kundenloyalität aufzeigen.

2.11.6.2 Analyse latenter Moderatormodelle

Bevor Moderatoreffekte mithilfe von Strukturgleichungsmodellen untersucht werden konnten stellten „klassische“ Verfahren wie die Regressions- und die Varianzanalyse das Instrumentarium der Wahl.

2.11.6.2.1 Regressionsanalytischer Ansatz

Im Bezugsrahmen dieser Verfahren hat sich die Verbreitung eines regressionsanalytischen Moderatorbegriffs herausgebildet (Aiken & West, 1991; Jaccard, Wan & Turrisi, 1990; O'Connor, 2006). Dabei gehen diese Verfahren von der Prämisse aus, dass die in Moderatorbeziehung stehenden Prädiktor- und Moderatorvariablen des Modells manifeste, fehlerfreie Messungen der untersuchten Konstrukte darstellen. Der regressionsanalytische Ansatz umfasst die Modellierung linearer (Bartussek, 1970) und nichtlinearer Moderatoreffekte (Moosbrugger, 1983). Der Moderatoreffekt wird durch Regressionsanalyse auf Grundlage eines moderierten Regressionsmodells der Form

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon \quad (4)$$

festgestellt. Demnach äußert sich bei einem linearen Regressionszusammenhang zwischen einem Prädiktor X und der Kriteriumsvariablen Y ein Moderatoreffekt darin, dass sich von Stufe zu Stufe der Moderatorvariablen Z die Steigung der Regressionsgeraden verändert. Die Hinzunahme einer Moderatorvariablen Z in ein lineares Regressionsmodell wirkt sich aus, indem an die Stelle eines konstanten Regressionskoeffizienten β_1 des Prädiktors X eine von Z abhängige Moderatorfunktion $[\beta_1 + \beta_3 Z]$ tritt und sich damit die Steigung der Regressionsgeraden zwischen Y und X auf den Stufen von Z verändert:

$$Y = \beta_0 + [\beta_1 + \beta_3 Z]X + \beta_2 Z + \varepsilon . \quad (5)$$

Zeigt sich nach Durchführung der Regressionsanalyse ein signifikanter Anstieg des Determinationskoeffizienten R^2 gegenüber einem linearen Modell ohne Einbeziehung des Regressors XZ , so gilt ein Moderatoreffekt als empirisch nachgewiesen. Die Variable Z stellt nur dann zu Recht eine Moderatorvariable dar, wenn die Regressionsgleichung unter Einbeziehung des Regressors XZ zu einer signifikant besseren Varianzaufklärung führt (Aiken & West, 1991).

Bei der regressionsanalytischen Überprüfung eines theoretisch postulierten Moderatoreffekts wird unabhängig von der Aufnahme eines Produktterms XZ in die Regressionsgleichung implizit vorausgesetzt, dass die erfassten Regressoren messfehlerfreie Messungen der durch sie repräsentierten Konstrukte darstellen (Heise, 1986). Liegen jedoch keine perfekt reliablen Regressorvariablen vor, so impliziert dies inkonsistente und gering effiziente Schätzungen der Regressionsparameter mit erhöhtem β -Fehler des Tests auf Moderation (Moosbrugger, Klein, Frank & Schermelleh-Engel, 1996). Diese Problematik verschärft sich für die Analyse eines nichtlinearen Moderatormodells gegenüber einem linearen Modell aufgrund der Einbeziehung eines Produktterms zweier Prädiktoren, da die Reliabilität eines Produktterms XZ bei Korrelation der Ausgangsvariablen unter der Reliabilität der Ausgangsvariablen X und Z liegt (Busemeyer & Jones, 1983). In einem Strukturgleichungsmodell mit latenten Variablenbeziehungen wird ein potentieller Moderatoreffekt im Rahmen einer Regressionsanalyse daher umso mehr unterschätzt, je weniger reliabel die Prädiktor- und Moderatorvariablen erfasst werden.

2.11.6.2.2 Analysemethoden für Strukturgleichungsmodelle

Die Beziehungen zwischen latenten, nicht direkt messbaren Variablen können mit Hilfe von Verfahren zur Untersuchung von Strukturgleichungsmodellen geschätzt und überprüft werden. Diese Verfahren tragen dem Aspekt der Fehlerbehaftetheit latenter Modelle ausdrücklich Rechnung und stellen die geeigneten Methoden für die Analyse latenter Moderatoreffekte dar (Aiken & West, 1991). Der Aufbau und die Besonderheiten linearer Strukturgleichungsmodelle werden in Abschnitt 6.6 ausführlich beschrieben, weshalb an dieser Stelle nicht weiter ausgeführt werden soll. Die Analyseverfahren für Strukturgleichungsmodelle kombinieren faktorenanalytische sowie regressionsanalytische Techniken für Mess- sowie Strukturmodell in Form einer simultanen (Kovarianzstrukturanalyse), einer mehrstufigen (Zweistufige

Kleinstquadrateschätzung) oder einer auf Resamplingverfahren beruhenden (Bayes-Statistische Analyse latenter Modelle) Schätzung der Modellparameter (Klein, 2000, S. 22).

In den verbreiteten Programmsystemen LISREL, EQS oder AMOS erfolgt die Analyse eines linearen Strukturgleichungsmodells durch eine Analyse der Kovarianzstruktur der beteiligten Indikatorvariablen (Browne & Arminger, 1995). Dabei ist zu beachten, dass die Theorie der Kovarianzstrukturanalyse von Strukturgleichungsmodellen primär für die Analyse linearer Variablenzusammenhänge konzipiert wurde. Latente Moderatormodelle stellen hingegen aufgrund der nichtlinearen Funktion der durch den Moderatoreffekt beeinflussten Prädiktorvariablen nichtlineare Strukturgleichungsmodelle dar. Desweiteren erweist sich für den Einsatz kovarianzbasierter Schätzverfahren die Spezifikation des Modells hinsichtlich einer formativen Konstruktion latenter Variablen als kritisch (Herrmann, Huber & Kressmann, 2006; MacCallum & Browne, 1993; McDonald & Ho, 2002). Daher kann bereits ohne eine weitergehende Betrachtung von Methoden der Kovarianzstrukturanalyse eine Anwendung auf Moderatormodelle a priori als möglicherweise nur suboptimal effizient eingeschätzt werden (Klein, 2000, S. 31).

Bei der Bayes-Statistischen Analyse handelt es sich um eine von Wittenberg und Arminger (1997) entwickelte Schätzmethode für nichtlineare latente Modelle. Es handelt sich allerdings um ein verteilungsgebundenes Verfahren, das normalverteilte Prädiktor- und Fehlervariablen voraussetzt. Das Verfahren basiert auf dem Einsatz spezieller Resampling-Verfahren und ist als sehr rechenintensiv einzuschätzen, was die Einsatzmöglichkeiten dieser Methode erheblich limitiert (Arminge & Muthen, 1998).

PLS, das auf der Methode der Kleinst-Quadrate-Schätzungen beruht, stellt hingegen geringere Anforderungen hinsichtlich der statistischen Verteilungsvoraussetzungen sowie an den Stichprobenumfang und ermittelt explizit Werte für die latenten Variablen (Henseler, 2005; Marcoulides & Saunders, 2006). Mit PLS lassen sich Strukturgleichungsmodelle sowohl mit formativen als auch mit reflektiv konstruierten Messmodellen schätzen. Die für kovarianzbasierte Verfahren angeführten Einschränkungen besitzen für gemischte Modelle, die mittels PLS-Verfahren geschätzt werden, keine Bedeutung, da deren iterative Berechnungsweise mit Hilfe gewichteter Konstruktwerte auch nicht eindeutig identifizierte Gleichungssysteme zulässt (Herrmann et al., 2004).

Somit erscheint PLS zur Berechnung von Strukturgleichungsmodellen die sowohl formativ als auch reflektiv operationalisierte Konstrukte beinhalten und bei denen Moderatormodelle untersucht werden als das geeignete Verfahren.

3. Forschungsfragen

Ausgehend von den Darstellungen des vorangehenden Kapitels werden nachfolgend zunächst die Ziele der Untersuchung dargestellt und der konkrete Untersuchungsbedarf abgeleitet (Abschnitt 3.1). Anschließend erfolgt in Abschnitt 3.2 eine Beschreibung der inhaltlichen Hypothesen.

3.1 Ziele der Untersuchung

In der vorliegenden Arbeit wird die Bedeutung der Usability für die Akzeptanz und Nutzung von Websites im Kontext von E-Commerce-Anwendungen im Dienstleistungsbereich untersucht. Ein zentrales Anliegen der Arbeit besteht darin, die Anwendbarkeit der Usability-Skala uvosV2 von Christophersen (2007) zur Erfassung des Konstruktes Usability auf den Kontext von E-Commerce-Anwendungen im Dienstleistungsbereich zu überprüfen. Christophersen entwickelte das Instrument dahingehend, dass eine genaue Unterscheidung zwischen einem formativen und einem reflektiven Messmodell zur fragebogenbasierten Erfassung der Usability vorgenommen wurde. Es wurde einerseits eine reflektive Skala ufosV2r entwickelt, die das Konstrukt in Form von Globalurteilen erfassbar macht und andererseits eine formative Skala ufosV2f, die die wesentlichen Usabilityaspekte berücksichtigt, um das Konstrukt valide abzudecken.

Ziel der vorliegenden Arbeit ist es, die Einflussfaktoren auf die Auswahl- und Nutzungsentscheidungen von E-Commerce-Anwendungen im Dienstleistungsbereich zu bestimmen und die ufosV2-Skalen hinsichtlich der Anforderungen dieses Anwendungsbereiches weiterzuentwickeln¹. Die Überprüfung der angepassten Skalen soll am Beispiel der gesetzlichen Krankenkassen erfolgen. Eine besondere Berücksichtigung soll hierbei der Einfluss des Involvement auf die betrachteten Entscheidungsprozesse der Nutzer erfahren.

Ranaweera et al. (2004, 2005) stellen für Kaufprozesse im E-Commerce eine moderierende Wirkung des Kundeninvolvement auf Kaufentscheidungsprozesse dar (siehe Abschnitt 2.4.5). Sie zeigen für Onlinekunden in Abhängigkeit ihres Involvement eine Verknüpfung des Online-Kaufverhaltens mit der Kundenzufriedenheit auf und schreiben dem Involvement eine moderierende Rolle auf die Beziehungen zwischen Kundenzufriedenheit und Online-Kaufverhalten bzw. Usability und Kundenzufriedenheit zu. Mit der vorliegenden Arbeit sollen diese Zusammenhänge für den Bereich der Online-Dienstleistungen untersucht werden. Ziel ist es, mögliche moderierende Einflüsse des Involvement auf die Zusammenhänge zwischen Nutzerzufriedenheit und Nutzungsintention bzw. Usability und Nutzungsintention zu untersuchen.

¹ Die hinsichtlich der Anforderungen von E-Commerce-Anwendungen im Dienstleistungsbereich weiterentwickelten ufosV2-Skalen werden im weiteren Verlauf durch ein angehängtes „a“ gekennzeichnet.

Eine weitere Zielsetzung ist es, zu überprüfen, wie sich Veränderungen in der Abfragereihenfolge der zu bewertenden Variable Usability auf die Untersuchungsergebnisse auswirken. Es gilt zu untersuchen, ob diese durch die Reihenfolge der Konstrukte in der Abfrage, hier am Beispiel der Usability, beeinflusst werden. Von Interesse ist es dabei, die determinierenden Variablen zu bestimmen, die diese Effekte erklären können, um daraus Möglichkeiten abzuleiten, wie diese Effekte im Rahmen einer Befragung berücksichtigt und potentielle Beeinflussungseffekte verringert werden können.

Die weiterentwickelten Usability-Skalen sollen hinsichtlich ihrer Güte auf Grundlage einer Datenerhebung für den Bereich der Websites gesetzlicher Krankenkassen beurteilt werden. Die diesbezüglichen Güteanforderungen werden in Form inhaltlicher Hypothesen im nächsten Abschnitt dargestellt.

3.2 Inhaltliche Hypothesen

Die Gütekriterien der weiterentwickelten reflektiven Usability-Skala ufosV2ra können gemäß der klassischen Testtheorie überprüft werden. Einerseits ist zu fordern, dass das Maß eine reliable Erfassung der latenten Variable Usability ermöglicht. Zur Überprüfung der Reliabilität lässt sich das Cronbachs Alpha als Maß für die interne Konsistenz heranziehen. Entsprechend lässt sich die erste inhaltliche Hypothese der Untersuchung wie nachfolgend formulieren:

H1: Die reflektive Skala ufosV2ra zur fragebogenbasierten Erfassung der Usability von Websites erweist sich als reliabel im Sinne einer hohen internen Konsistenz.

Desweiteren lässt sich entsprechend des Vorgehens der klassischen Testtheorie die faktorielle Validität der Skala ufosV2ra überprüfen (Bühner, 2004, S. 32). Es kann somit die folgende Hypothese aufgestellt werden:

H2: Die reflektive Skala ufosV2ra erweist sich gegenüber reflektiven Skalen, die weitere nutzerorientierte Erfolgsfaktoren erfassen, als faktoriell valide.

Wie in Abschnitt 2.6.1 dargestellt, werden durch eine reflektive Usability-Skala globale Einschätzungen zur Gebrauchstauglichkeit einer Benutzerschnittstelle erfasst. Als summatives Evaluationsinstrument ist sie nicht geeignet, um einzelne Usability-Facetten zu erheben. Ihr Einsatz als summatives Evaluationsverfahren bietet sich vielmehr an, wenn es darum geht, verschiedene Benutzerschnittstellen zu vergleichen, oder eine Website im zeitlichen Verlauf wiederholt zu evaluieren. Gebrauchstaugliche Benutzerschnittstellen sollen im Rahmen einer Evaluation mit Hilfe der reflektiven Usability-Skala ufosV2ra besser bewertet werden, als weniger gebrauchstaugliche Anwendungen. Es ist daher zu fordern, dass die reflektive Usability-Skala ufosV2ra geeignet ist, zwischen unterschiedlichen Websites zu differenzieren (siehe Abschnitt 2.8.1). Entsprechend lässt sich bezüglich der Diskriminationsfähigkeit der reflektiven ufosV2ra-Skala folgende Hypothese formulieren:

H3: Mit Hilfe der reflektiven Skala ufosV2ra kann zwischen mehreren Websites differenziert werden.

Neben der reflektiven Usability-Skala ufosV2ra soll auch eine formative Skala zur Erfassung der Usability von Websites im Dienstleistungsbereich entwickelt werden. Es bietet sich an, zur Validierung der formativen Skala ein Vorgehen anhand eines Zwei-Konstrukt-Modells zu wählen. Dabei dient die formative latente Usability-Variable als Prädiktor für die reflektive latente Usability-Variable (siehe Abschnitt 2.6.4.4). Auf Inhaltsvalidität des formativen Maßes kann geschlossen werden, wenn sich ein hoher positiver Zusammenhang zwischen den beiden Variablen zeigt. In Abbildung 3-1 wird diese Annahme dargestellt. Entsprechend lässt sich folgende Hypothese formulieren:

H4: Zwischen der formativen Usability-Skala ufosV2fa und der reflektiven Usability-Skala ufosV2ra zeigt sich in einem Zwei-Konstrukt-Modell ein signifikanter positiver Zusammenhang.

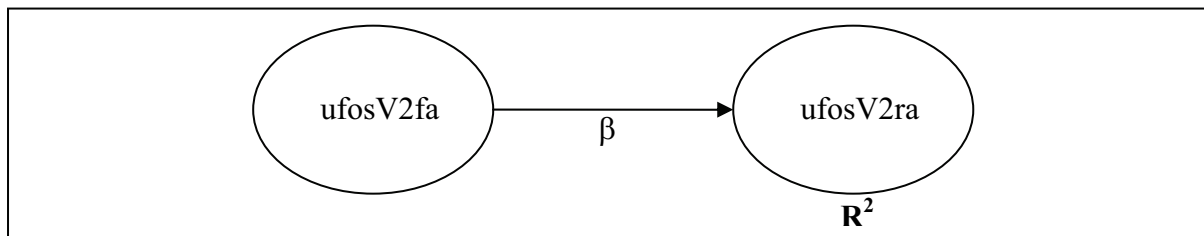


Abb. 3-1: Zwei-Konstrukt-Modell zur Validierung der formativen Skala ufosV2fa anhand der reflektiven Usability-Skala ufosV2ra

Anmerkung: Pfadkoeffizient (β), Determinationskoeffizient (R^2).

In Abschnitt 2.11 wurden die Zusammenhänge zwischen der Usability und weiteren Variablen dargestellt. Diese bilden ein nomologisches Modell, das als komplexes Validierungsmodell herangezogen werden kann. Erweisen sich die in dem Modell angenommenen Zusammenhänge zwischen der Usability und den weiteren kundennahen Erfolgsfaktoren als signifikant, so spricht dies für eine nomologische Validität der beiden ufosV2a-Maße (Judd, Smith, & Kidder 1991, S.56f.). Erweist sich die Usability im Rahmen des Modells als signifikanter Prädiktor für die Nutzerzufriedenheit (Hypothese 5) bzw. Nutzungsintention (Hypothese 6), so spricht dies darüber hinaus für eine konkurrente Kriteriumsvalidität der Messung.

Wie in Abschnitt 2.7.5 gezeigt wurde, wirkt sich die Gebrauchstauglichkeit einer Webanwendung auf deren tatsächliche Nutzung aus. Es ist daher neben den angeführten Zusammenhängen auch zu fordern, dass die ufosV2a-Skalen eine Vorhersage für die Auswahlentscheidung zwischen unterschiedlichen Webangeboten ermöglichen (Hypothese H7). Diese Annahme betrifft die prädiktive Validität der beiden ufosV2a-Maße (Rossiter, 2002).

In Tabelle 3-1 werden alle Hypothesen des nomologischen Modells aufgeführt, die das Konstrukt Usability betreffen. Der letzten Spalte kann der jeweilige Abschnitt im zweiten

Kapitel entnommen werden, in dem der theoretische und empirische Hintergrund der Hypothesen erläutert wurde. Die den Hypothesen zugrunde liegenden Zusammenhänge zwischen der Usability und den weiteren kundennahen Erfolgsfaktoren sollen für beide ufosV2a-Skalen erfüllt werden. Demzufolge werden für jede Konstruktbeziehung zwei Hypothesen aufgestellt.

Tab. 3-1: Inhaltliche Hypothesen im nomologischen Modell zur Güte der ufosV2a-Skalen (H5 bis H10)

<i>Nr.</i>	<i>Annahme</i>	<i>Hintergrund</i>
H5	<p>Je gebrauchstauglicher eine Website aus Sicht der Nutzer gestaltet ist, umso höher wird deren Zufriedenheit mit der Nutzung der Website ausfallen.</p> <p>H5a: Zwischen der reflektiven Usability-Skala ufosV2ra und der Zufriedenheit der Nutzer besteht ein signifikanter positiver Zusammenhang.</p> <p>H5b: Zwischen der formativen Usability-Skala ufosV2fa und der Zufriedenheit der Nutzer besteht ein signifikanter positiver Zusammenhang.</p>	2.5.4
H6	<p>Je gebrauchstauglicher eine Website aus Sicht der Nutzer gestaltet ist, umso höher wird deren Intention ausfallen, diese Website zukünftig wieder zu nutzen.</p> <p>H6a: Zwischen der reflektiven Usability-Skala ufosV2ra und der Nutzungsintention von Nutzern besteht ein signifikanter positiver Zusammenhang.</p> <p>H6b: Zwischen der formativen Usability-Skala ufosV2fa und der Nutzungsintention von Nutzern besteht ein signifikanter positiver Zusammenhang.</p>	2.5.4
H7	<p>Die Gebrauchstauglichkeit einer Website besitzt einen signifikanten Einfluss auf die Auswahlentscheidung des Nutzers.</p> <p>H7a: Die reflektive Skala ufosV2ra besitzt in signifikantem Maße Vorhersagekraft für die Auswahlentscheidung des Nutzers.</p> <p>H7b: Die formative Skala ufosV2fa besitzt in signifikantem Maße Vorhersagekraft für die Auswahlentscheidung des Nutzers.</p>	2.5.4
H8	<p>Je gebrauchstauglicher eine Website gestaltet ist, umso größer ist das Vertrauen der Nutzer in den jeweiligen Websiteanbieter.</p> <p>H8a: Zwischen der reflektiven Usability-Skala ufosV2ra und dem Vertrauen in den Websiteanbieter besteht ein signifikanter positiver Zusammenhang.</p> <p>H8b: Zwischen der formativen Usability-Skala ufosV2fa und dem Vertrauen in den Websiteanbieter besteht ein signifikanter positiver Zusammenhang.</p>	2.11.1

H9	Je gebrauchstauglicher eine Website gestaltet ist, umso mehr Vergnügen bereitet deren Nutzung den Nutzern. H9a: Zwischen der reflektiven Usability-Skala und dem Nutzungsvergnügen besteht ein signifikanter positiver Zusammenhang. H9b: Zwischen der formativen Usability-Skala und dem Nutzungsvergnügen besteht ein signifikanter positiver Zusammenhang.	2.11.3
H10	Je ästhetischer Nutzer die Gestaltung einer Website empfinden, umso höher schätzen sie die Ausprägung der Usability der Website ein. H10a: Zwischen der Ästhetikbewertung und der Usabilityeinschätzung anhand der reflektiven Usability-Skala ufosV2ra besteht ein signifikanter positiver Zusammenhang. H10b: Zwischen der Ästhetikbewertung und der Usabilityeinschätzung anhand der formativen Usability-Skala ufosV2fa besteht ein signifikanter positiver Zusammenhang.	2.11.4

Anmerkung: In der Spalte „Hintergrund“ wird der Abschnitt im zweiten Kapitel angeführt, in dem die Befunde zur jeweiligen Annahme des nomologischen Modells dargestellt werden.

In Abbildung 3-2 werden die Hypothesen den zugrundeliegenden Pfadbeziehungen im nomologischen Modell zugeordnet.

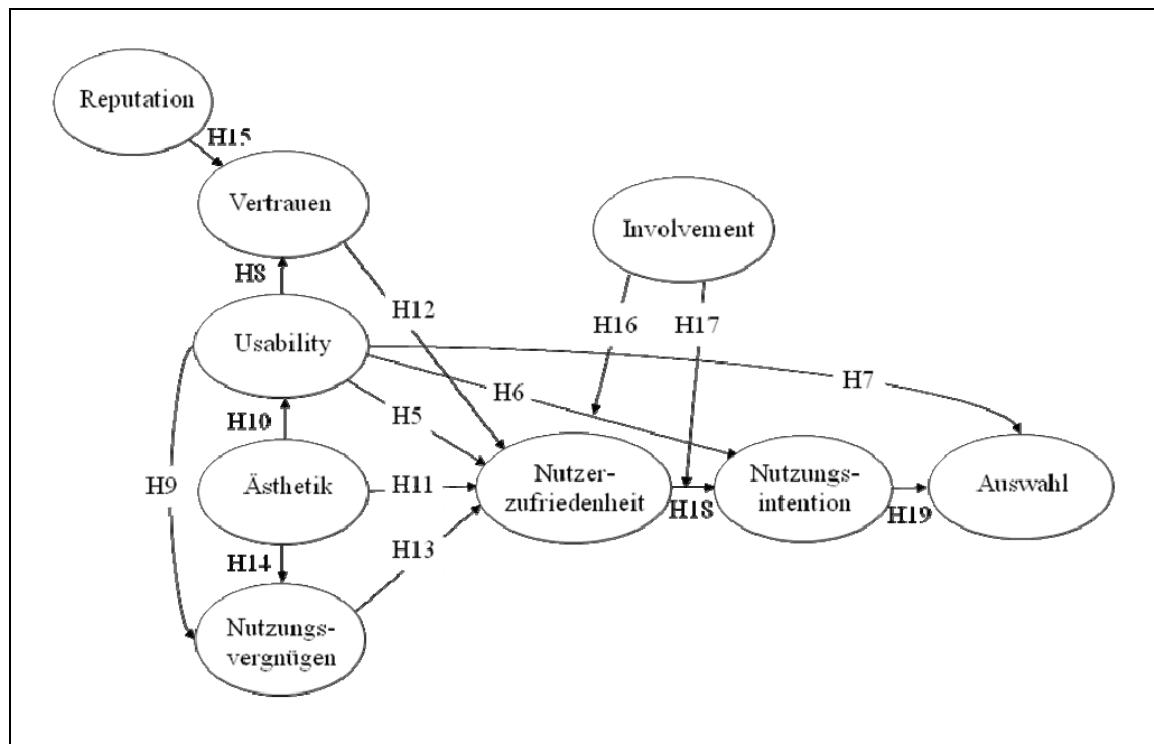


Abb. 3-2: Zuordnung der Hypothesen im nomologischen Modell

Anmerkung: Den Pfadbeziehungen wird jeweils die Nummer der inhaltlichen Hypothese zugeordnet.

Neben den inhaltlichen Hypothesen, die die Usability betreffen, werden im nomologischen Modell weitere Hypothesen zu den Zusammenhängen zwischen den übrigen Konstrukten abgeleitet. Diese werden in Tabelle 3-2 als Hypothesen H11 bis H19, unter Verweis auf den zugehörigen Abschnitt in Kapitel zwei, dargestellt.

Tab. 3-2: Inhaltliche Hypothesen im nomologischen Modell zu den Zusammenhängen der weiteren Konstrukte (H11 bis H19)

<i>Nr.</i>	<i>Annahme</i>	<i>Hintergrund</i>
H11	Je höher Nutzer die Ästhetik einer Website einschätzen, umso höher fällt ihre Zufriedenheit mit der Nutzung der Website aus.	2.10.3
H12	Je stärker Nutzer dem Anbieter einer Website vertrauen, umso höher fällt ihre Zufriedenheit mit der Nutzung der Website aus.	2.10.1
H13	Je höher das Nutzungsvergnügen ausfällt, desto höher ist die Zufriedenheit der Nutzer mit der Nutzung.	2.10.4
H14	Je ästhetischer eine Website aus Sicht der Nutzer gestaltet ist, umso mehr Vergnügen empfinden sie bei der Nutzung derselben.	2.11.5
H15	Je höher die Reputation des Anbieters einer Website durch die Nutzer eingeschätzt wird, umso mehr Vertrauen bringen sie dem Anbieter entgegen.	2.11.2
H16	Die Wirkung der Usability auf die Nutzungsintention wird durch das Involvement des Nutzers moderiert.	2.11.6
H17	Die Wirkung der Nutzerzufriedenheit auf die Nutzungsintention wird durch das Involvement des Nutzers moderiert.	2.11.6
H18	Je zufriedener Nutzer mit der Nutzung einer Website sind, umso höher wird deren Intention ausfallen, diese Website zukünftig wieder zu nutzen.	2.5.4
H19	Die Intention, eine Website zu nutzen, besitzt einen signifikanten Einfluss auf die Auswahlentscheidung eines Nutzers.	2.5.4

Anmerkung: In der Spalte „Hintergrund“ wird der Abschnitt im zweiten Kapitel angeführt, in dem die Befunde zur jeweiligen Annahme des nomologischen Modells dargestellt werden.

4. Untersuchungsplanung

Im folgenden Kapitel wird die Vorgehensweise der quasi-experimentellen Studie zur Untersuchung der Fragestellungen und Überprüfung der Hypothesen dargestellt. In den Abschnitten 4.1 bis 4.5 wird erläutert, wie die unabhängige Variable (UV), die den Erfolg bestimmenden Einflussfaktoren, die abhängigen Variablen (AVn) sowie weitere Variablen operationalisiert werden. Die Herleitung und Entwicklung des Versuchsszenarios ist Gegenstand von Abschnitt 4.6. In Abschnitt 4.7 werden mögliche Formen von Reihenfolgeeffekten bei der Fragebogenbeantwortung sowie die Vorgehensweise erläutert, um deren Einfluss zu kontrollieren. Abschließend wird in Abschnitt 4.8 der Versuchsplan vorgestellt.

4.1 Unabhängige Variable Website

Wie bereits in Abschnitt 3.1 dargestellt, bestand eine Zielsetzung der vorliegenden Arbeit darin, die Usability-Skala uvosV2 von Christophersen (2007) auf den Anwendungsbereich von E-Commerce-Anwendungen im Dienstleistungsbereich zu überprüfen und falls erforderlich anzupassen. Der Autor hatte einen branchenübergreifenden Usability-Fragebogen entwickelt und in einem Quasi-Experiment anhand mehrerer Online-Shops für verschiedene Produktarten validiert (siehe Abschnitt 2.9.8). In der vorliegenden Untersuchung soll das Fragebogeninventar nun hinsichtlich der spezifischen Anforderungen von Websites aus dem Dienstleistungsbereich angepasst werden und bezüglich seiner Validität in einem Quasi-Experiment überprüft werden. Die verwendeten Websites stellen die Ausprägungen der systematisch variierten UV dar.

4.1.1 Websites gesetzlicher Krankenkassen als Untersuchungsobjekt

Wie bereits in Abschnitt 2.2.5 dargestellt wurde, bietet die Nutzung von Online-Angeboten den Nachfragern von Gesundheitsleistungen vielfältige Möglichkeiten. So können auf Websites von Krankenkassen Informationen zu gesundheitsspezifischen und gesundheitsnahen Themen eingesehen werden sowie Versicherungsdienstleistungen und weitere Services in Anspruch genommen werden. Die Krankenversicherungsunternehmen agieren wie bereits in Abschnitt 2.2.5 dargestellt als Anbieter gesundheitsspezifischer Contents, insbesondere aus den Themenfeldern Prävention und Versorgungsleistung und unterstützen ihre Versicherten bspw. beim Zugang zu Qualitätsinformationen in der stationären und ambulanten Versorgung im Internet. So können sie ihre Versicherten in der Rolle eines Navigators durch das Überangebot an Gesundheitsinformationen führen und in deren Entscheidungsprozessen qualitativ unterstützen (Poensgen & Larsson, 2001, S.7f).

Andererseits erfüllen Electronic-Commerce-Anwendungen für Krankenversicherungsunternehmen auch die Funktion digitaler Vertriebsformen, die sowohl Anbietern als auch

Konsumenten „ganzheitliche Bequemlichkeit“ und Kosteneinsparungen versprechen (Parente, 2000; Chain, Sarasohn-Kahn & Wayne, 2000). Krankenkassen nutzen digitale Vertriebsformen im Rahmen von Electronic-Commerce-Anwendungen insbesondere in der Versichertengewinnung sowie im Vertrieb von sogenannten Ergänzungs- und Wahltarifen. So rechnen Vertreter großer Krankenkassen wie der Techniker Krankenkasse oder auch der Kaufmännischen Krankenkasse damit, dass bis zu 20% der extern geworbenen Versichertenzugänge über die Websites der Krankenkassen gewonnen werden (Roland Berger, 2006). In einem maßgeblich durch Verdrängungswettbewerb gekennzeichneten Wettbewerbsumfeld erhalten die Krankenkassen durch das Internet und die damit verbundenen Möglichkeiten im Rahmen der Versichertengewinnung und Versichertenbindung ein Instrumentarium zur Stärkung ihrer Wettbewerbsfähigkeit und relativen Wettbewerbsposition.

4.1.2 Festlegung der Websites

Für die Untersuchung wurden die Websites von sieben gesetzlichen Krankenkassen ausgewählt. Die Auswahl erfolgte auf Grundlage einer Voruntersuchung zur Verwendbarkeit der Websites im Rahmen der Hauptuntersuchung unter Berücksichtigung nachfolgender Kriterien:

1. *Kriterium:* Die Websites sollten über eine ausreichende Bandbreite an Informationen und Angeboten verfügen, damit alle Aufgaben im Rahmen der Versuchsdurchführung durch die Versuchsteilnehmer ausgeführt werden konnten. Um dies sicherzustellen, wurden die Websites mit Hilfe eines *Pretests* auf Grundlage einer Stichprobe von $N = 25$ Versuchspersonen hinsichtlich ihrer Eignung überprüft. Die Versuchspersonen wurden instruiert, die gestellten Aufgaben auf jeweils zwei Websites auszuführen, diese bezüglich der Durchführbarkeit zu kommentieren und die benötigte Zeit zu erfassen. Im Ergebnis mussten zwei Websites aufgrund mangelnder Eignung vom Versuch ausgeschlossen werden. Auf eine ausführliche Darstellung des *Pretests* soll aus Platzgründen verzichtet werden.
2. *Kriterium:* Um einen Einfluss aus früheren Erfahrungen mit den Websites so gering wie möglich zu halten, war es wünschenswert, dass die Versuchsteilnehmer die Websites möglichst nicht aus früheren Besuchen kannten. Im Rahmen der Voruntersuchung konnte festgestellt werden, dass die Websites der ausgewählten Krankenkassen den Teilnehmern des *Pretests* vorab nicht oder nur wenig bekannt waren. Dies war insbesondere dann der Fall, wenn die Versuchsperson nicht bei der Krankenkasse versichert war. Selbst im Versicherungsfall war in der Regel eine aktive Auseinandersetzung mit der Website in der Vergangenheit nicht erfolgt, was die Vermutung eines grundsätzlich geringen Interesses der Versuchspersonen an dem Themenfeld Krankenversicherung bestätigt.

3. *Kriterium:* Um das Model und die Hypothesen angemessen testen zu können, insbesondere wie gut die Skalen des ufosV2a Fragebogens zwischen unterschiedlichen Websites differenzieren, war es erforderlich, dass die Websites sich bezüglich ihrer Usability unterscheiden. Die Einschätzung der Benutzerfreundlichkeit wurde zunächst durch den Autor auf Grundlage der Heuristiken von Nielsen (1994) vorgenommen. Die Einschätzung wurde später im Rahmen des oben angeführten Pretests mit Hilfe frei formulierter Kommentierungen der Versuchspersonen überprüft. Als weitere Angabe konnten Unterschiede in der Bearbeitungsdauer zur Bestätigung der Einschätzung herangezogen werden.
4. *Kriterium:* Angesichts der in der Literatur anzutreffenden Diskussion bezüglich einer Vereinbarkeit von Usability und Design bei der Ausgestaltung von Webangeboten wurde darauf geachtet, dass Unterschiede hinsichtlich der grundsätzlichen Gestaltung der Websites bestehen (Cloninger, 2001; Wiesner, 2004). Ähnlich wie das dritte Kriterium zielte auch dieses Kriterium auf eine möglichst hohe Varianz in der Usabilityeinschätzung ab. Einerseits wurden Websites ausgewählt, deren Gestaltung eher nüchtern, schlicht und wenig grafikintensiv war, und andererseits wurden Websites mit eher innovativer, ästhetisch ansprechender und designorientierter Gestaltung der Oberfläche einbezogen.
5. *Kriterium:* Desweiteren sollte bei der Auswahl der Krankenkassen sichergestellt sein, dass diese in den Regionen, in denen der Versuch durchgeführt wurde als Anbieter präsent und den Versuchspersonen bekannt waren. Websites von Krankenkassen, die in den Versuchsregionen nicht am Markt verfügbar waren, wie beispielsweise regional tätige Betriebskrankenkassen, wurden deshalb nicht berücksichtigt.
6. *Kriterium:* Es ist davon auszugehen, dass größere, bekanntere Anbieter mehr und qualifiziertere Ressourcen für eine kundenorientierte Ausgestaltung ihres Webangebots einsetzen können. Damit verfügen diese grundsätzlich über die erforderlichen Potentiale, um in Bezug auf die Variablen Usability, Vertrauen, Ästhetik und Benutzervergnügen eine besser ausgestaltete Benutzerschnittstelle zur Verfügung stellen zu können. Aus diesem Grunde wurden Websites von Krankenkassen unterschiedlicher Größe und Bekanntheit herangezogen.

Insgesamt konnten sieben Krankenkassen ausgewählt werden, deren Websites den angeführten Kriterien genügten. Die ausgewählten Krankenkassen vereinen nahezu 80% der Versicherten in der gesetzlichen Krankenversicherung, wobei die „AOK Rheinland und Hamburg“ stellvertretend für das AOK-System⁴ ausgewählt wurde. In Tabelle 4-1 werden die

⁴ Die Websites der regionalen AOKs werden durch den AOK-Bundesverband gehostet und unterscheiden sich hinsichtlich ihrer Struktur und Layout in den betrachteten Bereichen nicht voneinander.

sieben im Rahmen der Untersuchung berücksichtigten Websites aufgelistet.⁵

Tab. 4-1: Im Rahmen der Untersuchung eingesetzte Websites

<i>Lfd. Nr.</i>	<i>Krankenkasse</i>	<i>Code</i>	<i>URL</i>
1	AOK Die Gesundheitskasse	1	www.aok.de
2	BARMER	2	www.barmer.de
3	DAK	3	www.dak.de
4	Deutsche BKK	4	www.deutschebkk.de
5	GEK Gmünder Ersatzkasse	5	www.gek.de
6	KKH Die Kaufmännische	6	www.kkh.de
7	Techniker Krankenkasse	7	www.tk-online.de

4.2 Anpassung des Itempools der ufosV2a-Skalen

In diesem Kapitel werden die Items für die formative Skala ufosV2fa und die reflektive Skala ufosV2ra zur Erfassung der *Usability von Websites* aus Nutzersicht festgelegt. Dabei wird auf die in der Untersuchung von Christophersen (2007) entwickelten und eingesetzten Skalen ufosV2f und ufosV2r zurückgegriffen (ebenso Christophersen & Konradt, 2008). Der Autor entwickelte die Skalen, um die *Usability von Online-Shops* aus Nutzersicht zu messen. Während mit der reflektiven Skala ufosV2r Gesamturteile zur Usability aus Nutzersicht erfasst wurden, setzt sich die formative Skala ufosV2f aus verschiedenen wesentlichen Einzelaspekten der Gebrauchstauglichkeit von Online-Shops aus Nutzersicht zusammen. Die Eigenschaften und Unterschiede reflektiver und formativer Skalen werden ausführlich in Abschnitt 2.6 dargestellt, weshalb an dieser Stelle keine weitere Erörterung folgen soll.

4.2.1 Vorüberlegungen

In der vorliegenden Arbeit soll die Anwendbarkeit der Usability-Skalen ufosV2 auf Websites von Krankenkassen überprüft werden. Diese unterscheiden sich in der Zielsetzung und damit auch im Aufbau und im Funktionsumfang von Online-Shops. Online-Shops zielen auf den Warenvertrieb über das Internet ab, indem sie Waren und digitale Produkte im Internet zum Verkauf bereitstellen (Schmid, 2001). Entsprechend orientiert sich der Funktionsumfang eines Online-Shopsystems vorrangig an Onlinekaufprozessen. Krankenkassen hingegen nutzen ihre Webangebote einerseits als digitale Vertriebsform im Rahmen der Versicherungsgewinnung und im Vertrieb von sogenannten Ergänzungs- und Wahltarifen. Andererseits agieren sie, wie in Abschnitt 2.2.5 dargestellt, als Anbieter gesundheitspezifischer Contents und unterstützen ihre Versicherten beim Zugang zu gesundheitspezifischen Informationen. Hieraus ergeben sich teilweise andere Anforderungen an den Funktionsumfang einer Website und damit auch an den Itemkatalog zur Messung der Gebrauchstauglichkeit.

⁵ Es bestand im Zusammenhang mit der Untersuchung keine Zusammenarbeit mit den beteiligten Krankenkassen.

In der Untersuchung von Christophersen (2007) wurden Online-Shops für die fünf Produktbereiche CDs, DVDs, Druckerpatronen, Bücher und Veranstaltungstickets herangezogen. Für die vorliegende Untersuchung war es deshalb zunächst erforderlich, die Skalen auf ihre Eignung zur Untersuchung der Gebrauchstauglichkeit des spezifischen Gegenstandsbereichs von Krankenkassenwebsites zu überprüfen und gegebenenfalls anzupassen.

Die Überprüfung der Items, sowie ihre Anpassung an die Erfordernisse von Krankenkassenwebsites stehen im Vordergrund des Abschnittes 4.2.2. In Abschnitt 4.2.3 wird der Prozess der Ergänzung und Vervollständigung des Itempools beschrieben. In Abschnitt 4.2.4 werden mögliche Störeffekte aufgrund der Itemformulierung sowie Maßnahmen dargestellt, um verzerrenden Antworttendenzen entgegenzuwirken. Abschnitt 4.2.5 behandelt das in der Befragung verwendete Antwortformat der ufosV2a-Skalen. Das Vorgehen im Rahmen der Voruntersuchung zur Verwendbarkeit der Usability-Items wird in Abschnitt 4.2.6 dargestellt. Abschließend werden in Abschnitt 4.2.7 die im Rahmen der Untersuchung verwendeten Items der reflektiven Skala uvosV2ra und der formativen Skala uvosV2fa dargestellt.

4.2.2 Überprüfung der Eignung und Modifikation der Items des ufosV2

Als Grundlage für das weitere Vorgehen diene die in Abschnitt 2.9.8 vorgestellte Untersuchung zur Entwicklung des Inventars ufosV2 (Christophersen, 2007). In Anbetracht des systematischen und umfassenden Vorgehens bei der Entwicklung des Itempools von ufosV2 sowie angesichts der guten Messwerte zur Validität der beiden Skalen wurden die acht Items der reflektiven Skala ufosV2r sowie die 46 Items der formativen Skala ufosV2f, die in der ursprünglichen Untersuchung zum Einsatz kamen, als Ausgangspunkt herangezogen.

In den Tabellen A-1.1 und A-1.2 im Anhang A werden die Ursprungsitems aus der Untersuchung von Christophersen (2007) dargestellt. Im nachfolgenden Abschnitt wird dargelegt, welche Items von der Verwendung in der vorliegenden Untersuchung ausgeschlossen bzw. angepasst wurden.

4.2.2.1 Analyse der Anteile fehlender Werte

Fehlende Werte im Umfang von bis zu 10% fehlender Items, gerechnet als Datensätze mit fehlenden Werten in Relation zu allen vorhandenen Datensätzen, liegen nach Schnell, Hill und Esser (1999, S. 430) im Rahmen der Erwartung. Höhere Missingquoten lassen unter Umständen auf eine mangelnde Verständlichkeit des Indikators oder auf eine fehlende Beurteilungsgrundlage schließen, sodass keine Einschätzung vorgenommen werden konnte. Um eine möglichst hohe Verwendbarkeit der ufosV2-Skalen sicherzustellen, wurde in der Ursprungsuntersuchung als Cut-Off-Kriterium für den Verbleib der Items in der Skala festgelegt, dass der Anteil fehlender Werte 10% nicht überschreiten darf (Christophersen, 2007). In der Folge wurden Items mit einem höheren Anteil fehlender Werte unter Berücksichtigung ihrer inhaltlichen Relevanz bereits durch den Autor ausgeschlossen, sodass

eine Elimination einzelner der insgesamt 54 vorliegenden Items aufgrund eines zu hohen *Anteils fehlender Werte* ausschied.

4.2.2.2 Analyse der inhaltlichen Verwendbarkeit

Eine Überprüfung der Indikatoren auf ihre inhaltliche Anwendbarkeit bei der Beurteilung von Krankenkassenwebsites ergab, dass diese nicht immer gegeben ist. Insbesondere Items, die speziell auf Kaufprozesse und deren Abwicklung abzielen, konnten nicht in der aktuellen Untersuchung berücksichtigt werden. So wurden Items, die sich auf die Prozesse des Bestellvorgangs und des Versands von Waren beziehen, eliminiert. Ebenso wurden Items, die den Warenkorb, die Zahlungsmodalitäten und die allgemeinen Geschäftsbedingungen betreffen, ausgeschlossen, da diese Funktionalitäten bei den aktuellen Untersuchungsobjekten nicht in vergleichbarer Form existieren. In der Folge wurden die Items uf28, uf35, uf36, uf38, uf39c, uf40, uf41, uf42, uf43, uf44, uf45, uf46, uf47, uf48, uf50, uf51, uf53, uf54c und uf57c aus dem Itempool ausgeschlossen. Desweiteren wurden die Items uf25 (*„Ich kann sehr gut auf die Suchfunktion zugreifen.“*) sowie uf56c (*„Bei der Benutzung des Shops treten störende technische Probleme auf.“*) ausgeschlossen, da beide Sachverhalte im Vorfeld der Untersuchung objektiv und mit geringem Aufwand durch Experten überprüft werden konnten.

Die acht Items der reflektiven Skala ufosV2r dienen der Erhebung von Globalurteilen zur Usability aus Nutzersicht. Sie zielen nicht auf spezielle Funktionalitäten und Prozesse eines Online-Shops ab und sind inhaltlich bei der Beurteilung von Krankenkassenwebsites anwendbar. In der Folge wurden sie in den Itempool der aktuellen Untersuchung aufgenommen.

4.2.2.3 Umformulierung von Items

Bei der Formulierung von Items gilt es Aspekte zu beachten, durch die ein Method Bias möglichst gering gehalten wird (Podsakoff, MacKenzie & Lee, 2003). Es ist dafür Sorge zu tragen, dass jeder Indikator verständlich formuliert wird und die Facette, die er erfassen soll, auch wirklich trifft. Dabei sollten zu hohe Ähnlichkeiten in den Formulierungen der Indikatoren vermieden werden, um ungerechtfertigt hohe Korrelationen in den Bewertungen zu unterbinden (siehe auch Abschnitt 2.6.4.2).

Aufgrund der Änderung des Untersuchungsobjektes war bei den acht reflektiven Items sowie bei 22 der insgesamt 27 übernommenen formativen Items eine Anpassung der Formulierung erforderlich. In 25 Fällen wurden lediglich untersuchungsobjektspezifische Begriffe ersetzt. So wurde der Begriff *„Shop“* durch *„Website“* ersetzt, *„Einkauf“* wurde zu *„Nutzung“* und *„Produkte“* zu *„Angebote“*. Eine weitergehende sprachliche Abwandlung erfuhren diese Items jedoch nicht. Das reflektive Item ur1 sowie die vier formativen Items uf13c, uf15, uf23 und uf24 erfuhren hingegen eine weitergehende Anpassung, um eine konkrete Bezugnahme auf den Bewertungsgegenstand der Website einer Krankenkasse zu erreichen. Dabei wurde darauf geachtet, dass der Sinngehalt des jeweiligen Items beibehalten wurde. So wurde

beispielsweise aus dem Item uf23 („*Es ist schnell erkennbar, welche Produkte ich in dem Online-Shop kaufen kann und welche nicht.*“) das Item uf18a („*Es ist schnell erkennbar, welche Inhalte auf der Website angeboten werden.*“). Insgesamt konnten die fünf formativen Items uf9, uf18c, uf27, uf32, uf34 in ihrer ursprünglichen Formulierung übernommen werden.

4.2.2.4 Analyse inhaltlich redundanter Indikatoren

Eine hohe Ähnlichkeit in der Beantwortung formativer Items kann auf eine inhaltliche Redundanz der Items hinweisen. Um den Item-Umfang einer formativen Skala nicht unnötig groß zu gestalten, kann in Erwägung gezogen werden, inhaltlich redundante Indikatoren zu eliminieren. Als Kriterium für die Identifikation redundanter Indikatoren eignen sich die bivariaten Pearson-Korrelationen zwischen den Items, wobei mindestens ein $r \geq 0,6$ vorliegen sollte, bevor ein Ausschluss in Betracht gezogen werden kann. In der aktuellen Untersuchung schien eine Elimination von Indikatoren der formativen Skala uvosV2f auf Basis der bivariaten Korrelationen aus der Voruntersuchung angesichts der Veränderung in den Untersuchungsobjekten und der daraus resultierenden sprachlichen Anpassung der Items nicht angemessen. Eine Elimination einzelner Items der formativen Skala uvosV2f aufgrund inhaltlicher Redundanzen wurde deshalb nicht vorgenommen.

4.2.3 Ergänzung des Itempools

Um sicherzustellen, dass alle wichtigen Aspekte der Usability bezüglich des Untersuchungsobjektes „*Website einer Krankenkasse*“ in der formativen Usability-Skala berücksichtigt wurden, wurden jeweils drei Experten und Anwender von Krankenkassenwebsites in den Prozess der Identifikation noch offener, nicht adressierter Usability-Aspekte einbezogen.

Im Ergebnis wurde deutlich, dass vor dem Hintergrund der Untersuchung von Websites, die beim Zugang zu gesundheitspezifischen Informationen unterstützen sollen, Aspekte der Güte und Bereitstellung der Informationen noch stärker einzubeziehen waren. Zusätzlich zu den ausgewählten Items der ufosV2 Skalen wurden deshalb vier weitere formative Items in den Itempool aufgenommen. Zwei der Items (Itemcodes uf28a und uf30a) gehen auf Items zurück, die Chen, Gillenson und Sherell (2004) verwendeten, um die „*information richness*“ der untersuchten Online-Shops zu erfassen. Das Item uf27a entstammt dem Itempool uvosV1 aus der Untersuchung von Konradt et al. (2003) und bezieht sich auf die *Übersichtlichkeit* der auf der Website verwendeten *Themenlisten*. Das Item uf29a entstammt einer Veröffentlichung von Lowengart und Tractinsky (2002) und beschreibt die Anwenderzufriedenheit mit der *Güte und dem Umfang der Angebote* auf der Website.

Alle vier Items waren in ihrer Ursprungsformulierung für die Anwendung auf Online-Shops entwickelt worden. Sie wurden im Hinblick auf die Anwendung bei Websites umformuliert, ohne eine Veränderung im Sinngehalt zu erfahren. Die Items uf27a, uf28a und uf29a waren in der Originalfassung englischsprachig formuliert. Es wurden mehrere Alternativübersetzungen

der Originalitems angefertigt, wobei die Festlegung auf die ausgewählte deutsche Fassung konsensual unter den Experten erfolgte.

4.2.4 Berücksichtigung von Störeffekten aufgrund der Polung bei der Itemformulierung

Die Formulierung der Items kann großen Einfluss auf die fragebogengestützte Bewertung von Web Interfaces nehmen (Chignell, 1990). So findet sich in der Literatur teilweise die Auffassung, dass die Aufnahme negativ formulierter, entgegengesetzt codierter Items, eine geeignete Maßnahme darstellen, um verzerrenden Antworttendenzen in einer Befragung entgegenzuwirken (vgl. Hinkin, 1995; Idaszak & Drasgow, 1987). Begründet wird dieser Effekt damit, dass entgegengesetzt codierte Items den Antwortenden stärker zu einem kognitiv kontrollierten Antwortverhalten bewegen. Die Aufnahme sowohl negativ als auch positiv formulierter Items kann somit einer Antworttendenz vorbeugen.

Dem widersprechend konnte in mehreren Studien gezeigt werden, dass die Verwendung negativ formulierter Items zu erhöhten Messfehlern führen kann (vgl. Harvey, Billings & Nilan, 1985), die sich bei einer positiven Formulierung nicht zeigen (vgl. Idaszak & Drasgow, 1987). Dies kann darin begründet sein, dass die Verwendung negativ formulierter Items gegenüber positiv formulierten Items ein höheres Maß an kognitiver Beanspruchung von Seiten des Antwortenden erforderlich macht und diesen überfordert (vgl. Motl & DiStefano, 2002). Schmitt und Stults (1986) beschreiben als mögliche Ursache von Störeffekten die Herausbildung von Antwortmustern bei den Antwortenden, die diese darin behindern, negativ gepolte Items zu erkennen. In der Folge können diese Effekte zu erhöhten Messfehlern führen. Entsprechend wurde in der vorliegenden Untersuchung darauf geachtet, den Umfang negativ gepolter Items möglichst zu begrenzen und nur dann negative Formulierungen zu verwenden, wenn deren Beantwortung kein hohes Maß an kognitiver Beteiligung von Seiten der Antwortenden erfordert. In der vorliegenden Arbeit waren acht der formativen Items und ein reflektives Item negativ formuliert. Die Items sind in Tabelle 4-2 und Tabelle 4-3 durch ein „c“ gekennzeichnet, das dem Itemcode angehängt wurde.

4.2.5 Antwortformat der Usability-Items

Zur Erhebung der Gebrauchstauglichkeit mit Hilfe des Inventars uvosV2 wurde eine Likert-Skala mit siebenfach gestuften Rating-Skalen eingesetzt. Likert-Skalen dienen der Einstellungsmessung. Diese erfolgt anhand von Items, die von den Respondenten jeweils auf einem Kontinuum von extrem positiv bis extrem negativ bewertet werden (Stier, 1999, S.80f.). In der vorliegenden Untersuchung lauteten die Polbezeichnungen „stimme gar nicht zu“ bis „stimme voll zu“. Das Konzept, Einstellungen anhand mehrerer Items zu messen, verfolgt das Ziel, den Messfehler der Skala zu minimieren, indem sich die Messfehler der Einzelitems gegenseitig aufheben (Schumann, 1997, S. 74). Das Vorgehen anhand von siebenfach gegliederten Rating-Skalen entspricht der Empfehlung von Jäpel (1985, S.151),

der zur groben Orientierung als Faustregel bei Itembatterien die Verwendung von sieben +/- zwei Kategorien als geeignet vorschlägt (vgl. auch Schumann, 1997, S. 74).

Eine ungerade Anzahl von Kategorien ermöglicht es dem indifferenten oder ambivalenten Respondenten, eine neutrale Mittelposition zu wählen. Gleichzeitig birgt diese Art von neutraler Position die Gefahr, von Versuchspersonen genutzt zu werden, wenn diese aufgrund von Unkenntnis eine indifferente Meinung zu einem Merkmal haben (Bortz & Döring, 2002, S.179). Infolgedessen ist es ratsam, eine explizite Ausweichkategorie als weitere von der eigentlichen Skala optisch getrennte Ausweichkategorie aufzunehmen. Versuchspersonen, die aufgrund von Unkenntnis des adressierten Aspektes indifferent sind, werden somit nicht zu einem Urteil gezwungen (Stier, 1999, S. 69). Neben der siebenstufigen Skala wurde im uvosV2 deshalb die Antwortoption „keine Angabe“ angeboten, um einen potenziellen „*Uninformed Response Error*“ zu minimieren. Dieser hat seine Ursache darin, dass Respondenten eines Fragebogens Bewertungen zu Aspekten abgeben, obwohl sie über keinerlei relevantes Hintergrundwissen verfügen (Hawkins & Coney, 1981; Schneider & Johnson, 1994). Ein solcher Störeffekt kann beispielsweise darin begründet sein, dass Respondenten eine Bewertung zu einem Usability-Aspekt abgeben, obwohl sie keine entsprechende Erfahrung mit dem Interface gemacht haben, die durch das Item abgefragt wird.

Gemäß den Erfahrungen in der Untersuchung von Christophersen (2007) waren die Versuchspersonen gut in der Lage, die siebenstufigen Likert-Skalen zu nutzen und differenzierte Usability-Bewertungen abzugeben. Dies korrespondiert mit den Ergebnissen von Lewis (1993), wonach die Verwendung siebenstufiger Skalen bei Usability-Fragebogen im Vergleich zu fünfstufigen Skalen zu besseren Ergebnissen führt.

Anlässlich dieser positiven Befunde und aufgrund der Zielsetzung, die Anwendbarkeit der Skalen uvosV2 auf die Untersuchung der Gebrauchstauglichkeit von Krankenkassenwebsites zu überprüfen, wurde das Antwortformat aus der Untersuchung von Christophersen (2007) für die aktuelle Untersuchung übernommen.

4.2.6 Voruntersuchung zur Verwendbarkeit der Items

Um zu gewährleisten, dass die ausgewählten Items für die ufosV2a-Skalen verständlich formuliert und verwendbar waren, wurde eine Voruntersuchung auf Basis einer Stichprobe von $N = 48$ Teilnehmern durchgeführt. Bei der Überprüfung eines Fragebogenentwurfs ist es von Bedeutung, eine ausreichend große Stichprobe zu Grunde zu legen. Mendoza, Stafford und Stauffer (2000) geben für eine Testneuentwicklung als Mindestgrenze von $N = 100$ an, da erst ab einer solchen Stichprobengröße die Reliabilität eines Tests zuverlässig geschätzt werden könne, wobei entsprechend der Spearman-Brown-Korrekturformel der Einfluss der Itemanzahl auf die Reliabilität zu berücksichtigen ist. Im vorliegenden Fall wurde allerdings auf bestehende Skalen aus der Untersuchung von Christophersen (2007) und aus der Literatur

zurückgegriffen, sodass lediglich eine Überprüfung der Verständlichkeit und inhaltlichen Anwendbarkeit erforderlich war, wofür ein Stichprobenumfang von $N = 48$ als ausreichend erachtet wurde.

Die Versuchspersonen wurden instruiert, die Usability-Items hinsichtlich ihrer Verständlichkeit sowie ihrer Verwendbarkeit zu überprüfen. Hierfür hatten die Teilnehmer drei Aufgaben auf der Website einer ausgewählten Krankenkasse zu bearbeiten. Daran anschließend sollten sie die 31 formativen Items sowie die acht reflektiven Items der *uVosV2*-Skalen mit Hilfe eines strukturierten Fragebogens auf siebenstufigen Likertskalen beantworten. Darüber hinaus wurden die Teilnehmer um eine Kommentierung der Verwendbarkeit der Items gebeten. Bei den vier formativen Items *uf10*, *uf20*, *uf34* und *uf49* überschritt der Anteil fehlender Items die 10%-Grenze mit Werten von 15-20% deutlich. Ergänzend wurden die Kommentierungen der Items durch die Versuchspersonen zu deren Beurteilung herangezogen. Im Ergebnis wurde das Item *uf34* von der weiteren Verwendung in der Skala ausgenommen. Die drei weiteren Items *uf10*, *uf20* und *uf49* wurden im Itempool belassen, um die ursprüngliche Skala *ufosV2f* möglichst weitgehend nachzubilden.

4.2.7 Darstellung der verwendeten formativen und reflektiven Items

In den Tabellen 4-2 und 4-3 werden die reflektiven und formativen Items dargestellt, die in der Hauptuntersuchung zur Anwendung kamen, um die *Usability von Krankenkassenwebsites* zu erfassen. Die Codes der reflektiven Items beginnen mit „ur“ für *Usability reflektiv*, die formativen mit „uf“ für *Usability formativ*. Zur Unterscheidung von den Ursprungsitems werden sie mit einem „a“ gekennzeichnet. In der zweiten Spalte wird jeweils angegeben, auf welches *ufosV2* Ursprungsitem das Item zurückgeht, oder ob es sich um ein neues Item handelt.

Tab. 4-2: Reflektive Indikatoren zur Erfassung der Usability

Neuer Code	Code <i>uVosV2</i>	Itemformulierung
ur1a	ur1	Diese Website macht das Suchen nach Krankenversicherungsinformationen im Netz leicht.
ur2ac	ur2c	Es ist mir zu kompliziert, diese Website zu nutzen (R).
ur3a	ur3	Man kann sich schnell einen Überblick über die Website verschaffen
ur4a	ur4	Der Umgang mit der Website ist leicht zu erlernen.
ur5a	ur6	Auf dieser Website kann eine erfolgreiche Suche zügig erledigt werden.
ur6a	ur7	Insgesamt bin ich mit der Benutzerfreundlichkeit dieser Website zufrieden.
ur7a	ur8	Die Website bietet alle Möglichkeiten, die ich mir von einer Krankenversicherungswebsite wünsche.
ur8a	ur9	Ich kann diese Website so benutzen, wie ich es erwarte.

Tab. 4-3: Formative Indikatoren zur Erfassung der Usability

<i>Neuer Code</i>	<i>Code uvosV2</i>	<i>Itemformulierung</i>
uf1a	uf1	Bei der Nutzung der Website kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.
uf2ac	uf5c	Ich muss mir viel merken, wenn ich auf dieser Website etwas suche.
uf3a	uf7	Es ist einfach, zwischen den verschiedenen Seiten hin und her zu springen.
uf4a	uf8	Zentrale Seiten (z.B. die Startseite) sind jederzeit schnell zu erreichen.
uf5a	uf9	Ich kann jede Aktion wieder rückgängig machen.
uf6a	uf10	Mir ist bei der Nutzung dieser Website stets klar, welche Handlungen ich als nächstes durchführen kann.
uf7ac	uf11c	Ich weiß manchmal nicht, wo ich mich innerhalb der Website befinde.
uf8a	uf12	Diese Website ist gut strukturiert.
uf9ac	uf13c	Es sind zu viele Schritte nötig, um das Gesuchte zu finden.
uf10a	uf14	Ich finde mich auf den einzelnen Seiten gut zurecht.
uf11a	uf15	Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.
uf12ac	uf16c	Es werden auf dieser Website häufig zu viele Informationen dargestellt
uf13ac	uf18c	Die Texte sind manchmal nur schwer lesbar.
uf14ac	uf19c	Das optische Design der Website stört mich bei der Nutzung.
uf15a	uf20	Wenn ich auf dieser Website auf etwas klicke, passiert genau das, was ich erwarte.
uf16ac	uf21c	Manchmal finde ich die verwendeten Texte inhaltlich unverständlich.
uf17a	uf22	Die Benennungen sind so gewählt, dass ich die Website gut benutzen kann.
uf18a	uf23	Es ist schnell erkennbar, welche Inhalte auf der Website angeboten werden.
uf19a	uf24	Ich finde leicht, was ich suche.
uf20a	uf27	Die Suchfunktionen liefern mir sinnvolle Ergebnisse.
uf21a	uf29	Die Angebote sind in sinnvolle Kategorien eingeteilt.
uf22a	uf31	Ich bin mit den Beschreibungen der Angebote zufrieden.
uf23a	uf32	Ich finde die verwendeten Bilder und Grafiken hilfreich.
uf24a	uf33	Die Website empfiehlt mir interessante Angebote, nach denen ich ursprünglich nicht gesucht habe.
uf25ac	uf49c	Es dauert zu lange, bis die Website auf meine Eingaben reagiert.
uf26a	uf55	Die Startseite stellt einen guten Ausgangspunkt für meine Nutzung der Website dar.
uf27a	neu	Die Themenlisten auf der Website sind übersichtlich.
uf28a	neu	Insgesamt verfügt die Website über ein reichhaltiges Informationsangebot.
uf29a	neu	Ich bin mit den Informationen, die mir der Anbieter zur Verfügung stellt, sehr zufrieden.
uf30a	neu	Diese Website ist eine Quelle hochwertiger Angebote.
uf31a	uf52	Ich erhalte nützliche Informationen zum Unternehmen, das die Website betreibt.

4.3 Operationalisierung der Variablen auf Prädiktorebene

Im folgenden Kapitel wird dargestellt, wie die Konstrukte auf Prädiktorebene operationalisiert wurden. Die Skalen zur Erhebung der Konstrukte wurden weitestgehend auf Basis von in früheren Untersuchungen verwendeten Skalen entwickelt oder aus der Literatur übernommen. Die Operationalisierung der Nutzungsabsicht basiert auf den Vorschlägen von Ajzen (1991). Für die Skalen zur Messung des Vertrauens, des Vergnügens und des wahrgenommenen Preises wurde auf Skalen von Koufaris und Hampton-Sosa (2002) sowie Christophersen (2007) zurückgegriffen. Die Operationalisierung der Ästhetik erfolgte auf Grundlage der

Skala von Mathwicka, Malhotra und Rigdon (2001), die Messung des Involvement auf Basis des semantischen Differentials von McQuarrie und Munson (1992) und die Messung der Reputation basiert auf der Skala von Cannon und Doney (1997).

Die Originalitems, die den im Fragebogen verwendeten Items zugrunde liegen, werden im Anhang C aufgeführt. Bis auf die Items zur Erfassung des Involvement kamen wie bereits im Falle der Usability, bei allen Items siebenstufige Likert-Skalen mit den Polen „stimme gar nicht zu“ und „stimme voll zu“ sowie einer zusätzlichen Antwortoption „keine Angabe“ zum Einsatz.

Für die Operationalisierung der Konstrukte Ästhetik, Vertrauen, Involvement, Vergnügen, und Reputation wurden mehrere Alternativübersetzungen der englischen Originalitems angefertigt. Die Festlegung auf die ausgewählte deutsche Fassung erfolgte auf Grundlage einer Expertenbefragung.

4.3.1 Erfassung der Ästhetik

Der Operationalisierung des Konstruktes Ästhetik (siehe Abschnitt 2.10.3) liegt die *Visual Appeal Skala* von Mathwicka, Malhotra und Rigdon (2001) zugrunde (vgl. auch Lowengart & Tractinsky, 2001). Die Skala umfasst drei englischsprachige Items. Nach Angaben der Autoren weist sie eine interne Konsistenz von α 0,92 auf. Um den Umfang des Fragebogens zu begrenzen, wurden in der Hauptuntersuchung nur zwei der drei Items verwendet. Die beiden Items zur Erfassung der Ästhetik lauten:

Item ae1: *Ich finde es ansprechend, wie diese Website ihre Informationen darbietet.*

Item ae2: *Die Website ist ästhetisch ansprechend gestaltet.*

4.3.2 Erfassung des Vertrauens

Die Erfassung des reflektiv angenommenen Konstruktes Vertrauen (siehe Abschnitt 2.10.1) erfolgte mit Hilfe von drei Items aus der Untersuchung von Christophersen (2007). In dieser wurden die Items v1 bis v3 in verbal leicht abgewandelter Form mit Hilfe einer siebenstufigen Likert-Skala abgefragt. Bei den Items handelte es sich um Übersetzungen englischsprachiger Items von Koufaris und Hampton-Sosa (2002). Diese basieren wiederum auf einer Skala von Jarvenpaa et al. (2000). In der Untersuchung von Koufaris und Hampton-Sosa (2002) ergab sich mit einem Cronbach's Alpha von α 0,87 als Maß für die interne Konsistenz der Skala ein befriedigender Wert. In der Untersuchung von Christophersen (2007) zeigte sich für eines der Items („*Ich bin mir sicher, dass der Online-Shop seine Versprechen und Verpflichtungen einhalten wird*“) im Vergleich zu den anderen Items der Skala eine leicht höhere Itemschwierigkeit von p_m 0,69. Durch eine Umformulierung des Items wurde versucht, dessen Verständlichkeit zu optimieren (siehe Item v3 unten). Die beiden anderen Items erreichten in der Untersuchung von Christophersen (2007) befriedigende Werte zwischen p_m 0,606 und p_m 0,639.

Um die interne Konsistenz der Skala zu erhöhen, wurde mit dem Item V4 auf ein weiteres Item von Koufaris und Hampton-Sosa (2002) zurückgegriffen. Das Item, das auf der Skala von Hampton-Sosa (2002) basiert, ist auf die besonderen Anforderungen einer Website ausgerichtet und wurde mit dem Item-Code v4 bezeichnet. Die vier verwendeten Items lauteten:

Item v1: *Dieser Anbieter ist vertrauenswürdig.*

Item v2: *Ich vertraue darauf, dass dieser Anbieter meine Interessen berücksichtigt.*

Item v3: *Dieser Anbieter hält seine Versprechen und Verpflichtungen ein.*

Item v4: *Ich vertraue in die Informationen, die dieser Anbieter mir zur Verfügung stellt.*

4.3.3 Erfassung des Involvement

Die Erfassung der Variable *Involvement* (siehe Abschnitt 2.4.2) erfolgte mit Hilfe einer von McQuarrie und Munson (1992) entwickelten Skala, dem *Revised Product Involvement Inventory (RPII)*. Das RPII basiert auf dem *Product Involvement Inventory (PII)* von Zaichkowsky (1985), wobei die ursprüngliche Skala zur Verbesserung der Anwendbarkeit von zwanzig Items auf zehn angepasst wurde, ohne dass sich dadurch die Skalenreliabilität signifikant verringerte. Das RPII ist ein kontextfreies Messinstrument, um Involvement bei Objekten und in Kaufsituationen mit Hilfe eines semantischen Differentials zu messen. Die Autoren nutzen dabei im Gegensatz zum ursprünglichen PII zwei Subskalen, *Importance* und *Interest*. In der Untersuchung von McQuarrie und Munson (1992) wurde das RPII auf neun verschiedene Produktkategorien in jeweils zwei unterschiedlichen Nutzungssituationen angewandt. Dabei ergaben sich in mehr als drei von vier Fällen Werte für Cronbach's Alpha von $\alpha > 0,90$ als Maß für die interne Konsistenz, sowohl für die Gesamtskala als auch für die beiden Subskalen *Importance* und *Interest*. Die Anforderungen an die Reliabilität bei bipolaren Skalen waren somit erfüllt (Nunally, 1978). Die beiden Subskalen *Importance* und *Interest* korrelierten mit der Messung der Einstellung gegenüber den Objekten mit $r = 0,74$ und $r = 0,65$.

Unabhängig vom Einsatzbereich ist Involvement allgemein Ausdruck einer hohen persönlichen Relevanz eines Objektes für die Bedürfnisse und Werte einer Person (Greenwald & Leavitt, 1984). In der vorliegenden Arbeit geht es um Involvement als Ausdruck von persönlicher Relevanz des Themenkomplexes Krankenversicherung für die Bedürfnisse und Werte der Versuchsperson.

Die Subskala *Importance* deckt inhaltlich die Kriterien ab, die im Rahmen der Untersuchung von Bedeutung waren (vgl. Koufaris, 2002; McQuarrie & Munson, 1992; Zaichkowsky, 1985, 1994). Die Subskala *Interest* wurde hingegen aufgrund der mangelnden Anwendbarkeit ihrer Kriterien (*exciting, dull, fun*) auf den Bereich von Krankenversicherungsthemen nicht berücksichtigt. Mit der Subskala *Importance* kommt eine Skala zum Einsatz, die im

Gegensatz zu den sonst im Fragebogen eingesetzten siebenstufigen Likert-Skalen mit Hilfe bipolarer Rating-Skalen auf einem semantischen Differential beantwortet wird. Die Pole werden durch gegensätzliche Begriffspaare gekennzeichnet. Dies bietet den Vorteil, dass sie sich über diese Begriffspaare gegenseitig definieren (Trommsdorff, 1995, S. 1067 ff.).

Vor dem Hintergrund der Bewährung der Skala in früheren Untersuchungen und der Eignung semantischer Differentiale zur Messung des Involvement im Allgemeinen wurde bewusst von einer Übertragung der Items auf ein Likert-Skalenformat abgesehen (vgl. Lastovicka & Gardner, 1978; Traylor, 1981; Zaichkowsky, 1985, 1994). Die Erhebung des Involvements erfolgte mit Hilfe des nachfolgenden semantischen Differentials:

„Wir würden gerne mehr über Ihre Einschätzung der Bedeutung von Krankenversicherungsthemen erfahren. Bitte nutzen Sie dafür nachfolgende Beschreibungen. Kreuzen Sie auf der Skala von 1 (im ersten Beispiel: wichtig) bis 7 (im ersten Beispiel: unwichtig) Ihre Bewertung an. Nutzen Sie hierfür die sieben Auswahlfelder von 1 bis 7. Falls eine Aussage nicht bewertet werden kann, wählen Sie die Antwortmöglichkeit „keine Angabe“.

„Die Beschäftigung mit Krankenversicherungsthemen ist für mich ...

<i>wichtig</i>	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	1	2	3	4	5	6	7	<i>unwichtig</i>	<i>keine Angabe</i>
1	2	3	4	5	6	7				
<i>bedeutet mir viel</i>	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	1	2	3	4	5	6	7	<i>bedeutet mir nichts</i>	<i>keine Angabe</i>
1	2	3	4	5	6	7				
<i>ist mir viel wert</i>	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	1	2	3	4	5	6	7	<i>ist mir egal</i>	<i>keine Angabe</i>
1	2	3	4	5	6	7				
<i>von großem Belang</i>	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	1	2	3	4	5	6	7	<i>nicht von Belang</i>	<i>keine Angabe</i>
1	2	3	4	5	6	7				
<i>relevant</i>	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	1	2	3	4	5	6	7	<i>irrelevant</i>	<i>keine Angabe</i>
1	2	3	4	5	6	7				

4.3.4 Erfassung des Nutzungsvergnügens

Zur Erfassung des reflektiven Konstruktes Nutzungsvergnügen (siehe Abschnitt 2.10.4) wurden drei reflektive Indikatoren eingesetzt. Dabei wurde auf ein Item („Es macht wirklich Spaß, bei diesem Online-Shop einzukaufen.“) aus der Untersuchung von Christophersen (2007) zurückgegriffen. Das Item wurde umformuliert und auf die Anforderungen von Websites angepasst (siehe Wortlaut Item vg3 unten). Zusätzlich wurden zwei weitere Indikatoren mit den Item-Codes vg1 und vg2 aus einer Untersuchung von Koufaris und Hampton-Sosa (2002) einbezogen, die auf einer Skala von Ghani, Supnick und Rooney (1991) basieren. In der Untersuchung von Koufaris und Hampton-Sosa (2002) ergab sich mit einem Cronbach's Alpha von α 0.95 als Maß für die interne Konsistenz der Skala ein sehr guter Wert (Nunnally, 1978).

Die drei Items, die für die Untersuchung herangezogen wurden, lauteten:

Item vg1: *Die Website ist unterhaltsam.*

Item vg2: *Die Website zu nutzen ist reizvoll.*

Item vg3: *Es macht wirklich Spaß, diese Website zu nutzen.*

4.3.5 Erfassung der Reputation

Die Reputation eines Anbieters entspricht dem Ausmaß, in dem Konsumenten überzeugt sind, dass der Anbieter sich ehrlich verhält und um das Wohl seiner Kunden bemüht ist (Doney & Cannon, 1997). Zur Erfassung des reflektiven Konstruktes *Reputation des Anbieters* (siehe Abschnitt 2.10.2) wurde auf zwei reflektive Items aus der Untersuchung von Pavlou (2003) zurückgegriffen. Dabei wurde das in seiner ursprünglichen Form negativ formulierte Item („This Web retailer has a bad reputation in the market.“) zur Erhöhung der Verständlichkeit positiv formuliert (siehe Wortlaut zu Item r1 unten). Bei Pavlou (2003) lässt sich keine konkrete Angabe zur Reliabilität der Skala finden. Es wird jedoch angeführt, dass das Cronbach's Alpha als Maß für die interne Konsistenz für alle Skalen gut ausfällt (vgl. Jarvenpaa, 2000). Doney und Cannon (1997) auf die die Skala ursprünglich zurückgeht, berichten ein Cronbach's Alpha von α 0.78. Die beiden Items, die zur Erfassung der Reputation des Anbieters herangezogen wurden, lauten:

Item r1: *Dieser Anbieter hat einen guten Ruf bei den Kunden.*

Item r2: *Dieser Anbieter ist dafür bekannt, zuverlässig zu sein.*

4.4 Operationalisierung der abhängigen Variablen

In den Abschnitten 4.4.1, 4.4.2 sowie 4.4.3 wird dargestellt, wie die abhängigen Variablen (AVn) Nutzerzufriedenheit, Nutzungsintention und Auswahlentscheidung operationalisiert wurden. Dabei ist grundsätzlich zu beachten, dass eine Verzerrung der Untersuchungsergebnisse aufgrund eines *Common Method Bias* in Studien wahrscheinlich ist, in denen die Datenerhebung sowohl für die UVn als auch für die AVn von denselben Personen, im selben Erhebungskontext und mit ähnlichen Itemcharakteristiken erfolgt (Podsakoff et al., 2003).

Die AVs Nutzerzufriedenheit und Nutzungsintention werden in gleicher Weise wie die Prädiktoren über Fragebogenitems anhand von siebenstufigen Likert-Skalen erhoben. Diese Vorgehensweise hat den Vorteil, dass aufgrund des Intervallskalenniveaus multivariate statistische Analysemethoden zur Anwendung gebracht werden konnten. Andererseits kann aufgrund des Vorgehens ein Common Method Bias nicht ausgeschlossen werden. Um einen verzerrenden Einfluss auf die AVs zu reduzieren, wurden die AVs Nutzerzufriedenheit und Nutzungsintention in der Erhebung den Prädiktoren vorangestellt.

Die AV Auswahlentscheidung wurde als Auswahl der Teilnehmer zwischen drei Optionen erfasst. Da die verschiedenen Erfolgsfaktoren mit Hilfe von siebenstufigen Likert-Skalen

erhoben wurden, unterscheidet sich die Operationalisierung der Variablen auf abhängiger und unabhängiger Ebene. Es kann somit davon ausgegangen werden, dass ein potentieller Common Method Bias aufgrund des Messkontextes reduziert wurde.

4.4.1 Reflektive Erfassung der Nutzerzufriedenheit

In der Literatur besteht weitgehend Einigkeit, dass die Zufriedenheit mit der Nutzung einer neuen Technologie als eine wesentliche Einflussgröße für ihre Akzeptanz angesehen werden kann (DeLone & McLean, 1992, 2003; Shankar, Smith & Rangaswamy, 2003; Wixom & Todd, 2005). Die Operationalisierung der Variable *Nutzerzufriedenheit* (siehe Abschnitt 2.5.4.1) erfolgte mit Hilfe einer von Wixom und Todd (2005) entwickelten Skala zur Messung der Zufriedenheit mit der Nutzung von Informationsverarbeitungssystemen. Die Skala umfasst zwei reflektive Indikatoren, die auf einer sieben-stufigen Likert-Skala beantwortet wurden. Die Autoren berichten auf Basis einer Studie zur Nutzung von Datawarehouse-Anwendungen mit einem Stichprobenumfang von $N = 465$ ein Cronbach's Alpha von $\alpha = 0.92$ als Maß für die interne Konsistenz der Skala. Nach Nunnally (1978) lässt dieser Wert auf eine sehr gute interne Konsistenz schließen. Eines der beiden Originalitems („All things considered, I am very satisfied with ...“) wurde analog des in der Untersuchung von Konradt, Christophersen und Schaeffer-Kuelz (2006) verwendeten Items („Overall, how satisfied are you with ...“) zur Frage umformuliert, um eine ausreichende sprachliche Differenzierung der beiden verwendeten Items zu gewährleisten. Hintergrund der Änderung war, dass die beiden Items in der Originalformulierung nach Wixom und Todd (2005) zu ähnlich waren und dies dazu führen kann, dass die Items unabhängig von ihrer semantischen Bedeutung ähnlich beantwortet werden (Podsakoff, MacKenzie & Lee, 2003). Die beiden Items, die zur Erfassung der Nutzerzufriedenheit herangezogen wurden, lauteten:

Item z1: *Insgesamt: Wie zufrieden sind Sie mit der Website?*

Item z2: *Mit der Nutzung der Website bin ich zufrieden.*

4.4.2 Reflektive Erfassung der Nutzungsintention

Bei der Erfassung der AV *Nutzungsintention* wurde auf Indikatoren zurückgegriffen, die bereits in der Untersuchung von Christophersen (2007) zum Einsatz kamen. Der Autor verwendete vier reflektive Items zur Erfassung der Kaufintention. Drei der Indikatoren gehen auf Items von Jarvenpaa et al. (2000) zurück, die ihrerseits auf der Grundlage der Empfehlungen von Fishbein und Ajzen (1975) zur Erfassung der Verhaltensintention entwickelt wurden. In der Untersuchung von Christophersen (2007) wurden die Items nicht wie in der Ursprungsversion in Frageform sondern in Aussageform unter Verwendung siebenstufiger Likert-Skalen formuliert. Zusätzlich verwendete der Autor ein weiteres selbstentwickeltes Item zur Erfassung einer zeitlich nicht näher spezifizierten Kaufintention. Als Maß für die interne Konsistenz der Skala wurde nach Elimination des Items ki3 („Es ist recht wahrscheinlich, dass ich im Rahmen des laufenden Versuchs in diesem Shop bestelle.“)

ein Cronbach's Alpha von α 0.91 gemessen. Die Trennschärfe sowie die Itemschwierigkeiten der drei verbleibenden Indikatoren liegen laut Angabe des Autors im angemessenen Bereich, sodass die Messgüte der reflektiven Skala zur Erhebung der Kaufintention positiv bewertet wurde.

Die verbleibenden drei Items wurden für die vorliegende Untersuchung umformuliert und sprachlich auf die Anforderungen von Websites angepasst. Ergänzend wurde ein weiteres Item (Item-Code ni3) eingeführt. Es stellt eine leichte inhaltliche Abwandlung zu Item ni2 dar, indem es auf den Bedarf an Informationen zum Thema Krankenversicherung abzielt. Zwei der Items (Item-Code ni2 und ni3) zielen auf eine generelle, zeitlich nicht spezifizierte Nutzungsabsicht ab. Die beiden anderen Items (Item-Code ni1 und ni4) beziehen sich nicht direkt auf die Intention, die Website zu nutzen, sondern auf die Intention, diese nach dem Versuch wieder aufzusuchen. Die Items wurden vor dem Hintergrund aufgenommen, dass die Intention, eine Website zu besuchen, eine Voraussetzung für einen tatsächlichen Besuch und dieser wiederum Voraussetzung für eine erneute Nutzung ist. Die vier in der Untersuchung verwendeten Items lauteten:

Item ni1: *Wenn ich eine neue Krankenkasse suche, werde ich diese Website wieder aufsuchen.*

Item ni2: *Wenn ich mich über Krankenkassen informieren möchte, werde ich diese Website nutzen.*

Item ni3: *Wenn ich wieder einmal Informationen zum Thema Krankenversicherung benötige, werde ich diese Website nutzen.*

Item ni4: *Ich werde diese Website nach dem Versuch vermutlich wieder aufsuchen.*

4.4.3 Operationalisierung der Auswahlentscheidung

Als zentrale Erfolgsgröße der Untersuchung wurde die *Auswahlentscheidung* der Versuchsteilnehmer herangezogen. Die Auswahlentscheidung bezog sich auf die Auswahl zwischen den beiden Krankenkassenwebsites, die den Versuchspersonen präsentiert wurden. Jede Vp wurde zum Abschluss des Versuchs aufgefordert, eine Auswahl zu treffen, welche der beiden Krankenkassenwebsites sie bevorzugt aufsuchen würde, um sich zu informieren, wenn sie eine neue Krankenkasse suchen würde. Die Aufnahme der Auswahlentscheidung als Indiz für die Präferenz der Vp für eine der beiden präsentierten Websites erfolgte inhaltlich als Folge der Abfrage der Nutzungsintention. Eine tatsächliche Auswahlentscheidung zwischen den beiden Krankenkassen als Versicherungsgeber ist im Rahmen des Versuchs nicht praktikabel durchführbar. Alternativ wäre unter der Annahme, dass die Vp einen neuen Versicherungsgeber suche, die Abfrage einer fiktiven Auswahlentscheidung zwischen den Krankenkassen denkbar. Es ist allerdings davon auszugehen, dass in einer realistischen Auswahlentscheidung mehr als nur zwei Krankenkassen in den Auswahlprozess einbezogen würden. Vielmehr wäre

ein ganzes Set an konkurrierenden Anbietern in den Vergleich einzubeziehen (vgl. Lowengart et al., 2001).

Die Einbeziehung von mehr als zwei Websites zur Bewertung durch die Vp war jedoch aufgrund der zeitlichen Restriktion bei der Versuchsdurchführung nicht möglich. Eine differenzierte Bewertung einer Website im Rahmen der Versuchsdurchführung ist nur möglich, wenn inhaltlich wie auch zeitlich eine ausreichende Auseinandersetzung mit der Website vorangegangen ist. In einer Vorstudie zur Praktikabilität des Untersuchungsdesigns zeigte sich, dass die Phasen des Kennenlernens und der Bewertung einer Website durchschnittlich ca. 20 Minuten beanspruchen. Die Einbindung weiterer Websites in einen Versuchsdurchgang hätte die Versuchsdauer in einem Umfang erweitert, der für die Teilnehmer als unzumutbar angesehen wurde. In der Folge könnte es zu Selbstselektionseffekten kommen, indem bei zunehmender Versuchsdauer nur Nutzer mit besonders hohem Interesse an wissenschaftlichen Untersuchungen dieser Art sämtliche Aufgaben vollständig bearbeiten. Des Weiteren könnte im Falle des Auftretens von Ermüdungseffekten bei den Vpn die Qualität der Versuchsdurchführung leiden. Um den Teilnehmern dennoch eine weitere Auswahloption zu eröffnen, wurde mit der Variante „keine Angabe“ eine Option zur Wahl gestellt, sich gegen eine Auswahl zwischen den beiden alternativen Websites zu entscheiden.

Zusammenfassend wurde die Auswahlentscheidung als Auswahl zwischen den drei Optionen *Auswahl der Website 1*, *Auswahl der Website 2* und *keine Angabe* operationalisiert.

4.5 Erhebung weiterer Variablen

Neben den zentralen Variablen der Untersuchung, den Variablen auf Prädiktorenebene sowie den abhängigen Variablen Nutzerzufriedenheit, Nutzungsintention und Auswahlentscheidung, wurden weitere Variablen identifiziert, bei denen ein Einfluss auf die abhängigen Variablen zu erwarten ist. Als Kontrollvariablen wurden im Rahmen des Quasi-Experiments neben der Vorerfahrung in den Bereichen PC- und Internetnutzung (Abschnitt 4.5.1), die Bekanntheit des Webanbieters (Abschnitt 4.5.2), der aktuelle Versicherungsgeber (Abschnitt 4.5.3), der wahrgenommene Preis (Abschnitt 4.5.4), verschiedene demographische Merkmale der Versuchspersonen (Abschnitt 4.5.5) sowie Kontrollvariablen zur Erhebungssituation (Abschnitt 4.5.6) erfasst.

4.5.1 Vorerfahrung im Bereich PC-/Internetnutzung und Online-Shopping

Studien haben gezeigt, dass in der Vergangenheit gezeigtes Verhalten Einfluss auf die zukünftige Verhaltensabsicht sowie auf zukünftiges tatsächliches Verhalten nimmt (Bagozzi et al., 1992; Conner & Armitage, 1988). Des Weiteren konnte aufgezeigt werden, dass die Vorerfahrung im Bereich der Internetnutzung eine wesentliche Einflussgröße auf das Onlineverhalten einer Person darstellt (Hoffman et al. 1999). Dies kann damit erklärt werden, dass Anwender sich aufgrund von Vorerfahrungen mit dem Internet im allgemeinen oder mit

Websites im besonderen Wissen und Fähigkeiten aneignen, die sie in der weiteren Nutzung unterstützen und ihre Einstellung gegenüber der weiteren Nutzung beeinflussen können (Jarvenpaa et al., 2000). So berichtet Romita (2001) über einen hohen Zusammenhang zwischen der Häufigkeit, der in der Vergangenheit online getätigten Einkäufe und der Intention, zukünftig online einzukaufen.

Aus diesem Grunde wurden im Rahmen der vorliegenden Untersuchung die Vorerfahrungen in den Bereichen PC-Nutzung, Nutzung des Internet sowie Online-Shopping als Kontrollvariablen erhoben. Die Vorerfahrungen in den Bereichen PC-Nutzung und Internet-Nutzung wurden analog zur Untersuchung von Christophersen (2007) jeweils durch drei Items erfasst, die die Nutzungserfahrungen in Jahren, Tagen je Monat und Stunden pro Tag abfragen. Zusätzlich wurde für die Erhebung der Häufigkeit, mit der das Internet zur Informationsrecherche genutzt wird (Item-Code *recherch*), auf ein Item aus einer Untersuchung von Pavlou (2003) zurückgegriffen. Für die aktuelle Untersuchung wurde das ursprünglich als Aussage formulierte Item zur Frage mit Bezug auf den Aspekt der Nutzung des Internet für Recherchen umformuliert. Das Maß an Erfahrung im Bereich Online-Shopping wurde mit Hilfe der Übersetzung eines englisch-sprachigen Items aus einer Studie von Pavlou et al. (2006) erhoben. Die Items lauten im Einzelnen:

⇒ Items zur Erfassung der Vorerfahrung in der Nutzung von PCs:

Item *pc_j*: *Seit wie vielen Jahren nutzen Sie einen PC? (freies Eingabefeld)_____*

Item *pc_t*: *Wie häufig nutzen Sie einen PC momentan? _____ Tag(e) im Monat*

Item *pc_h*: *(zusätzliches Eingabefeld zur PC-Nutzung) _____ Stunde(n) pro Tag*

⇒ Items zur Erfassung der Vorerfahrung in der Nutzung des Internet:

Item *www_j*: *Seit wie vielen Jahren nutzen Sie das Internet? _____*

Item *www_t*: *Wie häufig nutzen Sie das Internet momentan? _____ Tag(e) im Monat*

Item *www_h*: *(zusätzliches Eingabefeld zur Internet-Nutzung) _____ Stunde(n) pro Tag*

Item *recherch*: *Wie oft nutzen Sie das Internet für Informationsrecherchen? _____ mal im Monat.*

⇒ Item zur Erfassung der Vorerfahrung im Bereich Online-Shopping

Item *shopperf*: *Wie oft haben Sie während des letzten Jahres im Internet eingekauft? (Bitte beziehen Sie Online-Auktionen wie EBay mit ein.) _____*

4.5.2 Bekanntheit des Webanbieters

Das Ausmaß des Vertrauens seitens der Anwender wird neben der Reputation auch durch die Bekanntheit des Anbieters eines E-Commerce-Angebots beeinflusst (Jarvenpaa et al., 2000).

Doney und Cannon (1997) argumentieren, eine hohe Reputation und ein hoher Bekanntheitsgrad werde von potentiellen Anwendern als Ausdruck des Vertrauens Dritter in den Anbieter gewertet. Um die Bekanntheit der hinter den untersuchten Websites stehenden Krankenkassen zu erfassen, wurde auf ein einzelnes Item („*This store is well known*“) aus der Untersuchung von Jarvenpaa et al. (2000) zurückgegriffen. Das Item erfasst, inwiefern die Versuchsperson die Einschätzung teile, dass der Anbieter der Website allgemein sehr bekannt sei. Das herangezogene Item lautet:

Item b1: *Dieser Anbieter ist allgemein sehr bekannt.*

4.5.3 Aktueller Versicherungsgeber

In der Literatur herrscht weitgehend Einigkeit, dass sich das Konsumentenvertrauen in einen Anbieter mit zunehmender Dauer und Kontakthäufigkeit der Anbieter-Konsumenten-Beziehung entwickelt (Jarvenpaa, 2000). Jarvenpaa argumentiert, dass beide Parteien in den Aufbau der Beziehung investieren und dieses Investment im Falle nicht vertrauenswürdigen Verhaltens gefährdet sei. Zudem leiten Konsumenten aus zunehmender Erfahrung mit einem Anbieter Rückschlüsse auf dessen zukünftig zu erwartendes Verhalten ab. Um eine mögliche Beeinflussung der Befragungsergebnisse durch Vorerfahrungen mit dem Versicherungsgeber aufgrund eines aktuell bestehenden Versicherungsverhältnisses berücksichtigen zu können, wurde zu Beginn der Befragung der aktuelle Krankenversicherungsgeber erfragt. Das Item, das zur Erhebung des aktuellen Versicherungsgebers herangezogen wurde, lautet:

Aktuell bin ich bei folgender Krankenversicherung versichert: [freies Eingabefeld]

4.5.4 Wahrgenommener Preis

Die Auswahl von Produkten und Leistungen auf Webangeboten ist unter anderem abhängig vom Preis (Ba & Pavlou, 2002). Für den Bereich der Krankenkassen besteht die Besonderheit, dass aufgrund des Gesetzes zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (in der Fassung vom 26.03.2007) ab Januar 2009 die Beitragssätze der Krankenkassen weitestgehend vereinheitlicht werden. Eine Preisdifferenzierung wird nur noch über die Erhebung von Zusatzbeiträgen möglich sein, die maximal 1% der beitragspflichtigen Einnahmen des Mitglieds betragen können. Vor dem Hintergrund dieser Entwicklung und um den Einfluss der Beitragssatzhöhe bei der Beurteilung der Websites möglichst gering zu halten, wurden die Versuchsteilnehmer instruiert, die Webangebote unter der Annahme zu beurteilen, dass die Beiträge für alle Krankenkassen bereits gleich wären.

Dennoch wurde, um einen möglichen Preiseffekt kontrollieren zu können, das Preisniveau der Versicherungsleistung einer Krankenkasse, die subjektiv wahrgenommene Höhe des Beitragssatzes, mit Hilfe eines einzelnen Items erfasst. Hierfür wurde auf ein Item („*Die Preise in diesem Shop sind mir zu hoch.*“) aus der Untersuchung von Christophersen (2007) zurückgegriffen, das für die Anwendung bei Krankenkassen umformuliert wurde.

Da sich das Konstrukt „wahrgenommener Preis einer Krankenkasse“ auf einen konkreten Bewertungsgegenstand bezieht und nicht über verschiedene Facetten verfügt, kann der Argumentation von Rossiter (2002) gefolgt werden, wonach in diesem Falle ein einzelner Indikator zur Erhebung des Konstruktes ausreicht. Verschiedene empirische Studien zeigen sowohl für Einstellungs- als auch bei Verhaltensskalen, dass Single-Item-Messungen von ausreichender Reliabilität und Validität möglich sind (Nagy, 2002; Wanous & Hudy, 2001). Dies belegen auch Befunde im Zusammenhang mit Usability-Skalen (Christophersen & Konradt, in press). Das Item, das zur Erfassung des wahrgenommenen Preises herangezogen wurde, lautete:

Item subp: *Der Beitragssatz dieser Krankenversicherung ist mir zu hoch.*

4.5.5 Demographische Merkmale der Teilnehmer

Als weitere Kontrollgrößen wurden die demographischen Merkmale Alter, Geschlecht, Schulabschluss, berufliche Tätigkeit, Nettoeinkommen, Haushaltsgröße und Deutschkenntnisse abgefragt. Die Stufen der Variable *Schulabschluss* (im Sinne des höchsten Abschlusses) lauten: 0 *Abitur*, 1 *Mittlere Reife*, 2 *Hauptschule*, 3 *keiner* sowie 4 *Sonstiges*. Im Falle der *beruflichen Tätigkeit* wurden die Kategorien: 0 *Angestellter/Arbeiter*, 1 *Selbständig*, 2 *Beamter*, 3 *Student*, 4 *Auszubildender*, 5 *Schüler*, 6 *Keine* sowie 7 *Sonstiges* herangezogen. Für die Erhebung der rangskalierten Variable *Netto-Einkommen pro Monat* wurden die folgenden Stufen vorgegeben: 0 *bis <300 Euro*, 1 *300 bis <750 Euro*, 2 *750 bis <1500 Euro*, 3 *1500 bis <2000 Euro*, 4 *2000 bis <2500 Euro* sowie 5 *2500 Euro und mehr*.

4.5.6 Kontrollvariablen zur Erhebungssituation

Da die Untersuchung nicht im Labor, sondern mit Hilfe einer Online-Versuchsumgebung durchgeführt wurde, war es nicht möglich, Störeffekte aufgrund unterschiedlicher Zugangsvoraussetzungen auszuschließen. Um mögliche Einflüsse der Erhebungssituation auf die Ergebnisse identifizieren und statistisch kontrollieren zu können, erfolgte eine automatische Erfassung der *Bildschirmauflösung* der Versuchsperson, der genutzten *Browsersoftware* sowie der für den Versuch benötigten *Durchführungsdauer*.

Desweiteren wurden zwei Items zur Einschätzung der Zufriedenheit mit den im Versuchsverlauf aufgesuchten Inhalten der Website erfasst.

Item co1: *Die Information zum Zahnersatz entspricht meinen Vorstellungen.*

Item co2: *Die Angebote zur Stressbewältigung entsprechen meinen Vorstellungen.*

4.6 Versuchsszenario

Bei einem szenariobasierten Test nutzen die Versuchspersonen die Website unter bestimmten ausgewählten Problemstellungen (Yom 2003, S.126). Durch vorgegebene Aufgabenstellung-

en, die für die Nutzung der Website relevant sind, werden die Versuchspersonen dabei zur Ausführung definierter Abläufe motiviert. Dabei sollten die für die Nutzung der Website wesentlichen typischen Interaktionsprozesse abgedeckt werden. Idealerweise basiert die Auswahl typischer Problemstellungen auf den Ergebnissen einer Kontextanalyse. Diese dient neben der systematischen Identifikation typischer Problemsituationen auch der Bestimmung typischer Nutzerprofile und Nutzungsintentionen. Bei der Ausgestaltung des Versuchsszenarios ist darauf zu achten, dass die Versuchspersonen in der Interaktion mit der Website ausreichende Erfahrungen sammeln, um daran anschließend eine umfängliche Bewertung der verschiedenen Bewertungsaspekte vornehmen zu können.

Entsprechend sollte die Interaktion der Probanden mit den zu untersuchenden Websites nicht zu kurz ausfallen. Zudem galt es, die grundlegenden Schritte einer typischen Nutzung einer Krankenkassenwebsite abzubilden, um eine möglichst hohe externe Validität der Untersuchungsergebnisse zu erreichen und dadurch eine Übertragbarkeit der Ergebnisse zu rechtfertigen. Um sicherzustellen, dass die wesentlichen Nutzungsintentionen von Onlinekunden in der Interaktion mit Krankenkassenwebsites unterstützt wurden, wurden zunächst typische und von den Versicherten häufig gewählte Nutzungsszenarien eruiert. Auf deren Grundlage erfolgte die Auswahl und Definition der im Rahmen der onlinegestützten Erhebung gestellten Aufgaben.

Empirische Grundlage der verwendeten Nutzungsszenarien sind zwei verfahrensgleiche Untersuchungen von TNS Emnid/Infratest aus den Jahren 2004 und 2005 zur Nutzung von Websites der gesetzlichen Krankenversicherungsunternehmen (TNS Infratest, 2005). Die Erhebungen wurden im November 2004 und Mai 2005 im Rahmen einer Online-Panel-Befragung auf Basis einer Stichprobe von 3.200 Teilnehmern unterschiedlicher Krankenkassen onlinegestützt durchgeführt. Als häufigste Gründe für den Besuch einer Krankenkassenwebsite wurden von nahezu 50% der Befragten Fragen zu Versorgungsleistungen genannt, dicht gefolgt von dem Wunsch nach Informationen über die Krankenkasse (von knapp 40% bei Mitgliedern und rund 55% bei Nichtmitgliedern der Krankenkasse genannt) sowie generellem Interesse an Gesundheitsinformationen. Weitere häufig genannte Gründe für den Besuch einer Website waren die Suche nach dem nächsten Geschäftstellenstandort (34%) sowie Fragen zu Gesundheitsthemen (29%) und zum Beitragssatz.

Darüber hinaus gaben rund 40% der Befragten als Anlass für den Besuch der Website die Nutzung ihres persönlichen geschützten Bereichs an. Der am häufigsten genannte Grund für die Nutzung des persönlichen geschützten Bereichs liegt in der Anwendung von Online-Formularen. Darauf folgen das Herunterladen von Auslandskrankenscheinen, die Änderung persönlicher Daten und allgemeines Interesse an aktuellen Informationen. Der geschützte persönliche Bereich dient der Abwicklung versichertenindividueller Anliegen und ist nicht frei einsehbar, sondern nur den jeweils zugangsberechtigten Versicherten einer Krankenkasse

zugänglich. Er eignet sich damit nicht für eine repräsentative Erhebung und wird nicht in die vorliegende Untersuchung einbezogen.

Entsprechend der am häufigsten angeführten Gründe für den Besuch einer Krankenkassenwebsite wurden folgende Aufgaben für die Online-Untersuchung ausgewählt:

Erste Aufgabe: Die Versuchspersonen wurden entsprechend des Wunsches nach allgemeinen Informationen über die Krankenkasse sowie einem generellen Interesse an Gesundheitsinformationen gebeten, sich einen *allgemeinen Überblick über die Website* zu verschaffen. Die Versuchsteilnehmer erhielten für diese Phase der freien Exploration 5 Minuten Zeit. Sie diente einer ersten Orientierung über Inhalte, Ausgestaltung und Handhabung der untersuchten Website.

Zweite Aufgabe: In der zweiten Aufgabe geht es um *Zahnersatz*. Zahnersatz ist ein klassisches Thema aus dem Bereich der gesetzlichen Regelversorgungsleistungen (SGB V, §§55, 56). Er gehört zu dem Block der am häufigsten genannten Gründe für den Besuch einer Krankenkassenwebsite und eignet sich aufgrund des relativ hohen Fallvolumens. So betrug beispielsweise im Jahr 2004 in Baden Württemberg der Anteil der Leistungsausgaben für Zahnersatz an den gesamten Leistungsausgaben rund 3% (Pristl & Reuther, 2005). Umfang und Voraussetzungen der Leistungsübernahme sind gesetzlich geregelt und für alle gesetzlichen Krankenkassen gleich, werden jedoch unterschiedlich dargestellt. Die Versuchsteilnehmer wurden in der zweiten Aufgabe instruiert, die Regelungen zur Kostenübernahme aufzusuchen und sich einen groben Überblick darüber zu verschaffen, inwieweit die Krankenkasse die Kosten für die kostspielige Behandlung übernimmt.

Dritte Aufgabe: Als weiterer häufig genannter Grund für den Besuch einer Krankenkassenwebsite (34% der befragten Personen) wurde die Suche nach dem nächsten Geschäftstellenstandort der Krankenkasse angegeben. In einer durch TNS-EMNID durchgeführten Untersuchung zur Nutzung von Krankenkassengeschäftsstellen aus dem Jahre 2002 wurde jedoch deutlich, dass diese eher von Rentnern und Familien mit Kindern bis 14 Jahren aufgesucht werden. Befragte im Alter bis 25 Jahre gaben als präferierten Kontaktweg, um ihre Anliegen mit ihrer Krankenkasse abzuwickeln, das Internet an (TNS-EMNID, 2003). Da im Rahmen der vorliegenden Untersuchung das Nutzungsverhalten von Studenten untersucht werden sollte, wurde keine Aufgabe zur Geschäftsstellensuche in die Untersuchung aufgenommen.

Das Interesse an Gesundheitsthemen (von 29% der befragten Personen genannt) umfasst Fragen zur Erhaltung der Gesundheit durch gesundheitsbewusstes Verhalten. So äußerten sich die zu Gesundheits- und Krankenversicherungsthemen befragten Personen in einer Studie der forsa-Gesellschaft für Sozialforschung und statistische Analyse mbH aus dem Jahre 2006 als sehr interessiert an den Themen Prävention (53%) und Entspannungstechniken (28%). Entsprechend wurden die Versuchsteilnehmer in der dritten Aufgabe gebeten, die

Informationen zum Themenkomplex „Gesundheit durch Fitness und Bewegung“ aufzusuchen und sich einen Überblick über die Angebote der Krankenkasse zu verschaffen.

Vierte Aufgabe: Wie bereits angeführt, wurde in der vorliegenden Untersuchung vorrangig die Zielgruppe der Studenten betrachtet, da für diese Zielgruppe bei Ausscheiden aus der Familienversicherung die Erfordernis besteht, sich mit dem Thema Krankenversicherung auseinander zu setzen und sich selbst zu versichern (§§ 5 Abs.1 Nr.9; 10 Abs. 2 SGB V). Entsprechend wurden die Versuchsteilnehmer im Rahmen der vierten Aufgabe gebeten, die *Voraussetzungen für die kostenlose Familienversicherung für Studenten* auf der aktuellen Website zu finden.

Fünfte Aufgabe: Mit der Umstellung der Vergütung auf Festzuschüsse ist beim Zahnersatz der Anteil der Regelleistungen deutlich gesunken: Von 67 Prozent aller Leistungsfälle im Jahr 2004 auf 45 Prozent im Jahr 2005. Dafür stieg der Anteil der so genannten andersartigen Versorgung - dazu gehören beispielsweise Implantate. Bei dieser Versorgungsform wie auch bei der 2005 neu hinzugekommenen gleichartigen Versorgung rechnen Zahnärzte einen großen Teil der Leistungen nach der privaten Gebührenordnung ab (Spitzenverbände der GKV, 2006). Angesichts dieser Entwicklung wurden die Versuchspersonen aufgefordert, sich über die Möglichkeiten einer *weitergehenden Kostenabsicherung für Zahnersatz mittels einer privaten Zusatzversicherung* zu informieren.

Durch die Vorgabe konkreter Aufgaben im Quasi-Experiment wird der Besuch der Website zu einem gewissen Grad standardisiert. Alternativ wäre auch denkbar gewesen, den Versuchsteilnehmern eine generelle Anweisung zu geben, sich innerhalb eines vorgegebenen Zeitraums die Webseiten anzuschauen und anschließend zu bewerten. Dies hätte jedoch aufgrund der unterschiedlichen Wahrnehmungen zu nicht vergleichbaren Bewertungsgrundlagen und damit zu nicht vergleichbaren Bewertungen geführt.

4.7 Berücksichtigung möglicher Reihenfolgeeffekte

In verschiedenen Untersuchungen konnte gezeigt werden, dass die Präsentationsreihenfolge von Items einen Einfluss auf die Ergebnisse fragebogenbasierter Erhebungen hat (Konradt, Christophersen & Schäffer-Külz, 2006; Moore, 2002; Schwarz, Strack & Mai, 1991; Sudman, Bradburn & Schwarz, 1996). Vorausgehende Items können aufgrund der kognitiven Auseinandersetzung der Versuchsteilnehmer einen Einfluss auf die Beantwortung später präsentierter Items haben (Schuman & Presser, 1981, S. 24). Dies lässt sich durch den Einfluss verfügbarer Informationen aus der Beantwortung vorhergehender Items erklären (Sudman, Bradburn & Schwarz, 1996, S.56). Vorhergehende Items können für die Beantwortung nachfolgender Items demnach einen Bezugsrahmen bilden, indem sie für den Beurteilungsprozess relevante Gedächtnisinhalte zugänglich machen (Tourangeau & Rasinski, 1988). Bei der Abfrage der AVn Nutzerzufriedenheit und Nutzungsintention ist

deshalb darauf zu achten, inwiefern und in welchem Ausmaß dem Zeitpunkt der Abfrage positive oder negative Informationen vorausgehen.

Schwarz und Hippler (1995) konnten für mailgestützte Erhebungen vergleichbare verzerrende Effekte aufgrund der Informationen aus nachfolgenden Items auf die Beantwortung vorhergehender Items feststellen. Die Autoren erklären dies mit der Möglichkeit, nachfolgende Items einzusehen und die Information bei der Beantwortung vorhergehender Items zu berücksichtigen. Zur Vermeidung dieses Effektes wäre es erforderlich, die Items in der Befragung, vergleichbar der Situation in einem persönlichen oder telefonisch geführten Interview, einzeln zur Beantwortung anzuführen. Aufgrund des Umfangs der verwendeten Items wurde diese Möglichkeit als für die Versuchspersonen unzumutbar verworfen.

In der vorliegenden Untersuchung wurden potentielle Störeffekte aufgrund der Präsentationsreihenfolge unter mehreren Aspekten berücksichtigt. In Abschnitt 4.7.1 werden Effekte aufgrund der Abfolge der Websitedarbietung beschrieben. Die Abschnitte 4.7.2 und 4.7.3 setzen sich mit Effekten aufgrund der Abfolge in der Usabilityabfrage bzw. in der Konstruktabfrage auseinander.

4.7.1 Reihenfolge der Websitedarbietung

Jede Versuchsperson lernt im Verlauf ihrer Versuchsteilnahme nacheinander zwei Websites kennen und bewertet diese. Dabei ist ein Einfluss der Präsentationsreihenfolge der Websites auf die Bewertung der Variablen nicht auszuschließen. So kann die Information aus der Bewertung der ersten Website einen Bezugsrahmen für kognitive Vergleichsprozesse im Rahmen der Bewertung der zweiten Website bilden. Um einen Einfluss der Reihenfolge der Websites auszuschließen, galt es, die UV Website dergestalt auszubalancieren, dass alle möglichen Reihenfolgen im Versuchsablauf berücksichtigt wurden (Huber 1995, S.154f.). In Folge dessen wurde im Versuchsplan einerseits jede mögliche Paarung der sieben Websites berücksichtigt und andererseits die Reihenfolge der Präsentation für jede Kombination systematisch variiert.

4.7.2 Reihenfolge der Usabilityabfrage reflektiv versus formativ

Wie in Abschnitt 4.2.2.2 dargestellt, wurde durch die reflektive Usability-Skala ein Globalurteil zur Benutzerfreundlichkeit einer Website abgefragt. Die formative Usability-Skala hingegen diente der Erfassung der verschiedenen wesentlichen Einzelaspekte der Benutzerfreundlichkeit. Dabei ist aufgrund der oben beschriebenen möglichen Reihenfolgeeffekte zu erwarten, dass die formativen Einzelaspekte zur Usability einer Website unterschiedlich bewertet werden, je nachdem ob der Bewertung eine Abfrage von Globalurteilen vorangeht oder nicht. Ebenso kann auch die Abgabe von Globalurteilen durch eine vorhergehende Auseinandersetzung mit den formativen Einzelaspekten der Usability beeinflusst werden. Infolgedessen wurde um derartige verzerrende Effekte zu vermeiden, die

Präsentationsreihenfolge der reflektiven und formativen Usability-Skalen systematisch variiert.

4.7.3 Reihenfolge der Konstruktabfrage

Die Präsentation der Items innerhalb des Fragebogens erfolgte wie bereits in der Ursprungsuntersuchung von Christophersen (2007) entsprechend der Variablen geblockt. Der Autor sah mit diesem Vorgehen von einer randomisierten Präsentation aller Items ab. Er begründete dies damit, dass eine zufällige Abfragereihenfolge der Items ein höheres Maß an kognitiver Auseinandersetzung bei den Versuchspersonen erfordert hätte, da diese sich bei jedem Item mit einem neuen Bewertungsaspekt hätten auseinandersetzen müssen. Desweiteren sollte vermieden werden, dass Teilnehmer verärgert reagieren, wenn bei einer zufälligen Abfragereihenfolge der Eindruck entsteht, dass sich Items wiederholen und sie diese erneut beantworten sollten.

Bei einer geblockten Präsentation der Skalen stellte sich die Frage, in welcher Reihenfolge die Itemblöcke präsentiert werden sollten. Abbildung 4-1 gibt einen Überblick über den schematischen Aufbau des Fragebogens zur Bewertung von Krankenkassenwebsites.

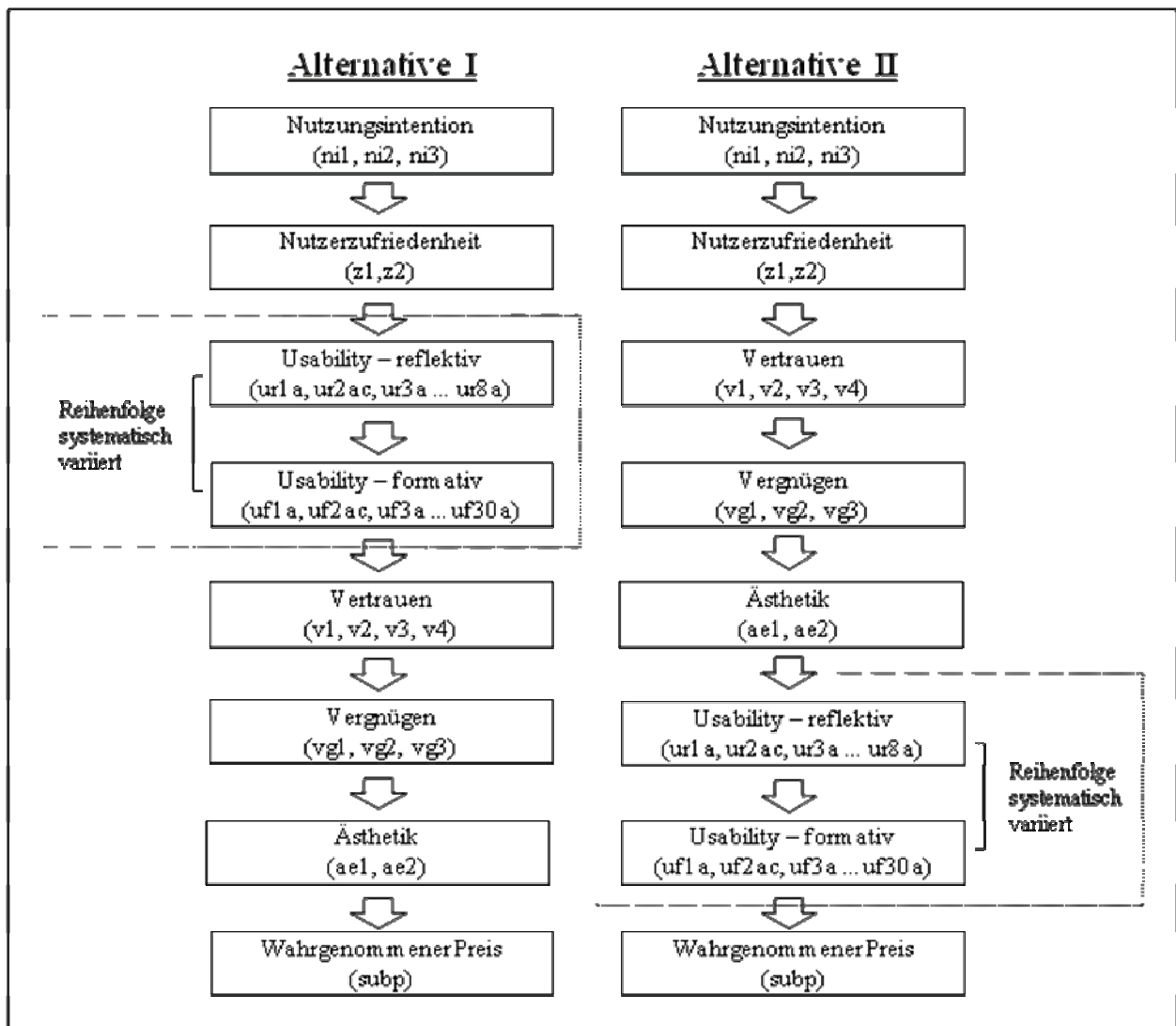


Abb. 4-1: Schematischer Aufbau des Fragebogens zur Bewertung der Websites

Die Reihenfolge der Konstruktabfrage orientiert sich an den Zusammenhängen zwischen den Konstrukten, wie sie in Abschnitt 2.11 dargestellt wurden. Aufgrund der Annahme, dass die Nutzungsintention als zentrale AV in der Untersuchung direkt oder indirekt von den anderen Variablen beeinflusst wird, wurde sie als erste in der Abfrage präsentiert. Bei einer Positionierung an späterer Stelle hätte damit gerechnet werden müssen, dass aufgrund der vorhergehenden kognitiven Auseinandersetzung mit den beeinflussenden Variablen möglicherweise ein verzerrender Effekt auf die Ergebnisse eintritt. Entsprechend wurde im Anschluss an die Nutzungsintention die zweite AV Nutzerzufriedenheit erhoben.

Analog zu dieser Argumentation wurden entsprechend den Annahmen bezüglich der Zusammenhänge zwischen den Prädiktorvariablen die Skalen zu den latenten Variablen Vertrauen, Vergnügen und Ästhetik in dieser Reihenfolge abgefragt. Eine Ausnahme wurde bei der Positionierung der beiden Usability-Skalen gemacht.

Der bisherigen Argumentation folgend wären die Usability-Skalen aufgrund der in Abschnitt 2.10.3 dargestellten Einflussnahme auf das Nutzungsvergnügen in der Abfragereihenfolge im

Anschluss an die Skala zur Erfassung des Vergnügens zu positionieren. Allerdings kommt der Betrachtung der Usability in dieser Studie die höchste Priorität zu. Neben den kausalen Wirkungszusammenhängen bei Auswahl- und Nutzungsprozessen im Electronic-Commerce von Dienstleistungen sollten auch die Auswirkungen von Veränderungen in der Abfrage-reihenfolge der zu bewertenden Variablen auf die Untersuchungsergebnisse untersucht werden. So konnten Konradt et al. (2005) in einer Untersuchung von Reihenfolgeeffekten im Zusammenhang mit TAM feststellen, dass die Positionierung der AV Benutzerzufriedenheit vor bzw. nach den Fragebogenitems zu den Prädiktorvariablen Einfluss auf deren Bewertung nimmt. Entsprechend wurden die Usability-Skalen in der Abfolge der Konstrukte einmal am Anfang des Fragebogens im Anschluss an die AVs platziert (Alternative I) und einmal am Ende des Fragebogens vor der Preiseinschätzung (Alternative II). Der wahrgenommene Preis wurde bei beiden Alternativen zum Abschluss des Fragebogens erhoben, um einen verzerrenden Einfluss auf die Bewertung der übrigen Variablen zu vermeiden.

4.8 Versuchsplan

Wie bereits dargestellt, stellen die Websites der sieben ausgewählten gesetzlichen Krankenkassen (siehe Abschnitt 4.1.2) die Varianten der UV dar. Jeder Teilnehmer lernte im Rahmen eines Versuchsdurchlaufes zwei Websites kennen, sodass bei sieben zur Verfügung stehenden Websites 21 unterschiedliche Paarungen von Websites möglich waren. Um einen potentiellen Reihenfolgeeffekt aufgrund der Präsentationsreihenfolge der Websites zu kontrollieren, wurde zusätzlich die Reihenfolge der 21 Paarungen variiert, sodass sich die Anzahl der zu berücksichtigenden Paare auf 42 verdoppelte (siehe Abschnitt 4.7.1). Eine weitere Verdopplung der Versuchsbedingungen auf 84 Varianten ergab sich aus der Erfordernis, die Reihenfolge der reflektiven und formativen Skalen durch systematische Variation auszubalancieren. Die Berücksichtigung der beiden Alternativen I und II bezüglich der Abfolge der Konstrukte im Fragebogen verdoppelte die Anzahl nochmals auf 168 unterschiedliche Versuchsbedingungen.

Bei einem Stichprobenumfang von $N = 256$ Teilnehmern war also sichergestellt, dass jede Versuchsbedingung abgedeckt werden konnte. Da jeder Teilnehmer zwei Websites bewertete ergaben sich insgesamt 512 Einzelbewertungen, bei sieben verschiedenen Websites im Durchschnitt 73 Bewertungen je Website.

Bei dem hier vorgestellten Design handelt es sich um einen quasi-experimentellen Versuchsplan. Wesentliche Merkmale sind einerseits die systematische Variation relevanter UVs, im vorliegenden Fall die Websites der ausgewählten Krankenkassen, und andererseits der Umstand, dass aufgrund des unvollständigen Versuchsdesigns nicht sicherzustellen ist, den Einfluss aller Störvariablen auszuschließen bzw. zu kontrollieren (Sarris, 1992, S.145ff.; Rack & Christophersen, 2006, S.23).

Abbildung 4.2 stellt den Aufbau des Teilnehmercodes dar. Dieser soll am Beispiel des Teilnehmercodes 97-2-7-1-1 erläutert werden. In diesem Fall wurde der Versuchsperson mit der Teilnehmernummer 97 zunächst die Website der Barmer (2. Zahl 2), dann die Website der Techniker Krankenkasse (3. Zahl 7) präsentiert. Die Usability-Skalen wurden entsprechend Alternative I am Anfang des Fragebogens präsentiert (4. Zahl 1). Dabei wurden die Items der reflektiven Usability-Skala vor den Items der formativen Usability-Skala abgefragt (5. Zahl 1).

1. Zahl	2. Zahl	3. Zahl	4. Zahl	5. Zahl
Fortlaufende <u>Teilnehmernummer</u>	Nummer der <u>ersten</u> präsentierten Website	Nummer der <u>zweiten</u> präsentierten Website	Reihenfolge der <u>Konstruktabfrage</u>	Reihenfolge der <u>Usabilityabfrage</u>
1	1 Website 1	1 Website 1	1 Alternative I	1 zunächst
2	2 Website 2	2 Website 2	2 Alternative II	reflektiv, dann
3	3 Website 3	3 Website 3		formativ
.	4 Website 4	4 Website 4		2 zunächst
.	5 Website 5	5 Website 5		formativ, dann
.	6 Website 6	6 Website 6		reflektiv
.	7 Website 7	7 Website 7		

Abb. 4-2: Zusammensetzung des Teilnehmercodes

5. Untersuchungsdurchführung

Nachfolgend wird die Durchführung des Quasi-Experiments beschrieben, das sich über den Zeitraum von April bis Oktober 2007 erstreckte. Zunächst wird in Abschnitt 5.1 die Teilnehmerstichprobe hinsichtlich ihrer Rekrutierung und Zusammensetzung beschrieben. In Abschnitt 5.2 wird die Online-Versuchsumgebung vorgestellt, mit deren Hilfe das Quasi-Experiment durchgeführt wurde. Abschließend wird in Abschnitt 5.3 der Ablauf des Versuchs dargestellt.

5.1. Teilnehmerstichprobe

Wie bereits angeführt, wurde in der vorliegenden Untersuchung um die praktische Relevanz der Ergebnisse zu steigern, vorrangig die Zielgruppe der Studenten betrachtet. Für diese besteht nach Abschluss des Studiums bei Aufnahme eines sozialversicherungspflichtigen Beschäftigungsverhältnisses, spätestens jedoch nach Ausscheiden aus der beitragsfreien Familienversicherung, die Anforderung, sich selbst zu versichern (§§ 5 Abs.1 Nr.9, 10 Abs. 2 SGB V). Die Notwendigkeit, sich mit der Thematik Krankenversicherung auseinanderzusetzen, kann daher mit zunehmendem Alter der Studenten vorausgesetzt werden.

In ähnlicher Weise gilt dies auch für Auszubildende. Für diese besteht mit Aufnahme der Berufsausbildung die Pflicht der eigenen beitragspflichtigen Krankenversicherung, aufgrund der Einkommenshöhe i.d.R. in der gesetzlichen Krankenversicherung. Zu diesem Zeitpunkt besteht also wie bereits im Falle der Studenten angeführt, erstmals die Anforderung, sich mit der Thematik einer eigenen Krankenversicherung auseinanderzusetzen und einen Versicherungsgeber zu wählen. Entsprechend intensiv werden Studenten und Auszubildende von den gesetzlichen Krankenkassen als Zielgruppen umworben.

5.1.1. Rekrutierung der Teilnehmer

Bei der Rekrutierung der Versuchsteilnehmer wurden mehrere Wege verfolgt. Diese werden nachfolgend in Abschnitt 5.1.1.1 dargestellt. In Abschnitt 5.1.1.2 wird anschließend das Vorgehen bei der Rekrutierung beschrieben.

5.1.1.1 Wege der Rekrutierung

Die Werbung der studentischen Teilnehmer wurde an sechs deutschen Universitäten und Hochschulen in den Städten Berlin, Bremen, Hamburg, Kiel und München durchgeführt. Sie erfolgte einerseits über Aushänge, die gezielt an den „schwarzen Brettern“ von Lehrstühlen der Fakultäten für Betriebswirtschaftslehre, Informatik, verschiedenen Ingenieurwissenschaften, Pädagogik und Psychologie angebracht wurden (siehe Abb. D-1 in Anhang D). Ebenso wurden Fachschaften eingebunden, die im Rahmen ihrer Kommunikationsmittel für die Studie warben. Um eine weitere studienfachbezogene Durchmischung der Teilnehmer zu

erreichen, wurden die Aushänge teilweise auch in den Mensen der Hochschulen ausgelegt. Zudem wurde unter den Studenten der Fachhochschule Kiel (University of Applied Sciences) in verschiedenen betriebswirtschaftlichen Vorlesungen für die Teilnahme an der Untersuchung geworben.

Die Rekrutierung der Auszubildenden erfolgte durch Anschreiben mehrerer ausbildender Unternehmen. Die Anschreiben waren an die Personalabteilungen der Unternehmen mit der Bitte gerichtet, den Auszubildenden die Information über die Untersuchung zur Verfügung zu stellen.

5.1.1.2 Ablauf der Rekrutierung

Die Interessenten meldeten sich auf die Aushänge per Listeneintrag oder Email bei der Versuchsleitung und wurden daraufhin per Email begrüßt (siehe Abb. D-1 in Anhang D). Die Email enthielt neben dem Zugangscode zur Online-Versuchsumgebung eine Einweisung, wie man zur Versuchsumgebung gelangt sowie einen Hinweis zur Vertraulichkeit und Wahrung des Datenschutzes. In einer Anlage zur Email wurden die Teilnehmer gebeten, zwischen einer Aufwandsentschädigung in Höhe von 15,- Euro oder einer Bestätigung über die Versuchsteilnahme in Höhe von 1,5 Versuchspersonenstunden als Belohnung für die Versuchsteilnahme zu wählen (siehe Abb. D-1 in Anhang D)⁶.

Da die Versuchsteilnahme onlinegestützt erfolgte und damit Ort und Zeitpunkt der Durchführung durch die Teilnehmer frei bestimmt wurden, konnte eine begleitende Überprüfung der Qualität der Teilnahme nicht vorgenommen werden, wie dies bei einem Laborversuch möglich ist. Zur Unterstützung der Qualität bei der Ausfüllung des onlinegestützten Fragebogens wurde deshalb in den Unterlagen darauf hingewiesen, dass die eingegebenen Daten bezüglich ihrer Qualität und Konsistenz systematisch überprüft werden und eine Auszahlung erst in Folge eines positiven Prüfergebnisses vorgenommen werde.

5.1.1.3 Probleme bei der Teilnehmerrekrutierung

Insgesamt führten die oben angeführten Rekrutierungswege zu einer zufrieden stellenden Resonanz. So konnten innerhalb der Versuchslaufzeit rund 350 Interessenten gewonnen werden, von denen letztendlich 256, also rund 75%, am Versuch teilnahmen. Der Anspruch, alle 168 Versuchsbedingungen abzubilden, wurde damit erfüllt.

Die Hintergründe, weshalb Interessenten ihr Interesse an der Versuchsteilnahme kurzfristig wieder aufgaben, konnten nicht ermittelt werden. Bei den nicht teilnehmenden Interessenten wurde 3 Wochen nach Versand der ersten Email mittels einer zweiten Email nachgefasst, wodurch es gelang, einige weitere Teilnehmer zu aktivieren. Von weiteren Nachfassaktionen

⁶ Ein t-Test für unabhängige Stichproben mit einem Datensatz mit zwei Berechnungen je Fall zeigte keine signifikanten Unterschiede in den zentralen Variablen Usability, Nutzerzufriedenheit und Nutzungsintention.

wurde abgesehen, da nach Einschätzung des Versuchsleiters selbst bei Versuchsteilnahme nur ein geringes Interesse am Versuch zu erwarten gewesen wäre.

5.1.2 Beschreibung der Stichprobe

Entsprechend der Rekrutierungswege konnten 92,4% der Versuchsteilnehmer aus den Zielgruppen der Studenten und Auszubildenden gewonnen werden. Aus Tabelle 5-1 kann die Verteilung nach Geschlecht und Alter entnommen werden. Die Gruppe der 18 bis 29 Jahre alten Teilnehmer umfasst erwartungsgemäß 87,9% der Versuchsteilnehmer. Im Vergleich dazu umfasst diese Altersgruppe in der Gesamtpopulation der deutschen Bevölkerung rund 19% (AGOF, 2006). Der Anteil der männlichen Teilnehmer liegt mit 61,3% deutlich über dem Anteil in der Population mit 55,7%, und spiegelt die Auswahl der Studiengänge wieder.

Tab. 5-1: Geschlecht und Alter der Versuchsteilnehmer

Alter	Geschlecht		Total
	männlich	weiblich	
18-29	142	83	225
30-39	8	15	23
40-49	5	0	5
50-59	1	1	2
> 60	1	0	1
Total	157 (61,3%)	99 (38,7%)	256

5.2 Computergestützte Durchführung des Versuchs

Die Durchführung des Quasi-Experiments erfolgte online mit Hilfe einer computergestützten Versuchsumgebung, unter Nutzung eines Servers des Rechenzentrums der Christian-Albrechts-Universität zu Kiel. Bei der Versuchsumgebung handelt es sich um eine in PHP programmierte datenbasierte Webapplikation. Als Datenbank kam MySQL zum Einsatz. Der Zugriff war mit Hilfe gängiger Browser-Software möglich.

Da die computergestützte Abfrage der Items sich bereits in Voruntersuchungen als unproblematisch erwiesen hatte (Christophersen, 2007; Konradt et al., 2003) erfolgte die Präsentation des Fragebogens ebenfalls eingebettet in eine Hypertext-Versuchsumgebung. In mehreren vergleichenden Studien konnte nachgewiesen werden, dass bei der Erhebung des Usability Fragebogens QUIS die Beurteilungen nicht durch die Wahl des Formats Online versus Paper-Pencil beeinflusst werden (Slaughter et al., 1994; Harper et al., 1997). Vergleichbare Ergebnisse konnten in einer späteren Vergleichsstudie auch für den Usability Fragebogen WAMMI, ein Inventar zur Erhebung der Zufriedenheit mit Websites (siehe Abschnitt 2.6.7), festgestellt werden (Kirakowski, Claridge & Whitehand, 1998).

In Abbildung 5-1 wird beispielhaft eine Seite der Online-Versuchsumgebung dargestellt. Weitere Seiten des sequentiell aufgebauten Hypertext-Dokumentes werden im Anhang D als Screenshots aufgeführt. Vor Einsatz der Versuchsumgebung wurde diese umfangreich getestet, sodass sichergestellt werden konnte, dass sowohl die Seitendarstellung, als auch die Verlinkung zu den zu bewertenden Websites sowie die Speicherung der Datenfiles problemlos funktionierten. Der Zugang zur Versuchsumgebung wurde den Teilnehmern durch einen persönlichen Versuchspersonencode ermöglicht (siehe Abschnitt 4.8).



Einführung → Vorbefragung → Website 1 → Bewertung 1 → Website 2 → Bewertung 2 → Auswahl → Ende

Überblick über den Versuchsablauf

Liebe(r) Versuchsteilnehmer(in),

vielen Dank, dass Sie an diesem Versuch zum Einkaufen im Internet teilnehmen.

Sie werden sich während dieser Untersuchung die Websites von zwei Krankenkassen anschauen. Zu jeder Website erhalten Sie fünf Aufgaben. Dadurch werden Sie die Website genauer kennen lernen. Im Anschluss an die Bearbeitung der Aufgaben bitten wir Sie, jede Website durch die Beantwortung eines Fragebogens zu bewerten. Insgesamt wird der Versuch ca. 60 min. dauern.

Im Rahmen des Versuchs werden nicht Sie, sondern die Websites getestet. Sie können also nichts "falsch machen". Außerdem werden natürlich alle Angaben vertraulich behandelt.

[Ich möchte fortfahren.](#)

Abb. 5-1: Screenshot einer Seite der Hypertext -Versuchsumgebung

Die Instruktionen zur Versuchsdurchführung wurden auf den Seiten der Versuchsumgebung dargestellt. Dabei wurde auf eine verständliche, übersichtliche und einfach zu handhabende Darstellung geachtet. So wurde beispielsweise der Umfang der einzelnen Seiten so begrenzt, dass eine Nutzung der Scrollfunktion der Browser-Software nicht erforderlich war, um die Informationen einer Seite zu erfassen. Die Vorwärts-Navigation erfolgte anhand der Links, die jeweils am Ende einer Seite stehen. Mit Hilfe einer fortlaufenden Orientierungshilfe am oberen Rand der Seiten konnte der Teilnehmer jederzeit erkennen, in welcher Phase des Versuchsablaufs er sich befindet.

Die insgesamt 53 Items des Fragebogens zur Erfassung der Konstrukte und Kontrollvariablen wurden auf zwei Seiten dargestellt. Abbildung 5-2 zeigt einen Ausschnitt aus dem verwendeten Fragebogen mit drei Items zur formativen Erfassung der Usability.

The image shows a portion of a questionnaire with three items, each followed by a Likert scale. The scales are labeled 'stimme gar nicht zu' (I completely disagree), 'stimme voll zu' (I completely agree), and 'keine Angabe' (no answer). Each scale consists of seven square boxes, each containing a small white circle. The first item is 'Diese Website ist gut strukturiert.' with the first three boxes filled. The second item is 'Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.' with the first six boxes filled. The third item is 'Wenn ich auf dieser Website auf etwas klicke, passiert genau das, was ich erwarte.' with the first five boxes filled.

Item	stimme gar nicht zu	stimme voll zu	keine Angabe
Diese Website ist gut strukturiert.	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Wenn ich auf dieser Website auf etwas klicke, passiert genau das, was ich erwarte.	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>

Abb. 5-2: Ausschnitt aus dem Fragebogen

Als Alternative zur gewählten Darstellung der Items bot sich die Möglichkeit an, die Items jeweils auf einer einzelnen Seite darzustellen. So argumentieren Sudman, Bradburn und Schwarz (1996, S.123), dass bei einer gruppierten Anordnung der Items auf einer Einzelseite die Gefahr eines Common Method Bias aufgrund erhöhter Korrelationen zwischen den Indikatoren bestehen könne. Andererseits konnte in verschiedenen Studien festgestellt werden, dass die Gesamtantwortzeiten bei Einzel- gegenüber Multiitem-Darbietung deutlich höher ausfallen (Couper, Traugott & Lamias, 2001). Angesichts der großen Anzahl zu beantwortender Items und der Sorge vor Unzufriedenheit bei den Versuchsteilnehmern aufgrund einer zu langen Bearbeitungsdauer wurde die Entscheidung getroffen, die Items auf zwei Seiten gruppiert darzustellen.

Die erhobenen Daten wurden je Teilnehmer automatisch in einem Ergebnisfile festgehalten und auf dem Server des Rechenzentrums gespeichert. Mit Hilfe einer weiteren Anwendung konnten die Ergebnisfiles für eine Verwendung in der Statistik-Software SPSS aufbereitet werden.

5.3 Versuchsablauf

Nachfolgend wird der Ablauf des Quasi-Experiments dargestellt (siehe Abbildung 5-3). Der Versuchsablauf gliedert sich in 8 Phasen, die in den folgenden Abschnitten erläutert werden.

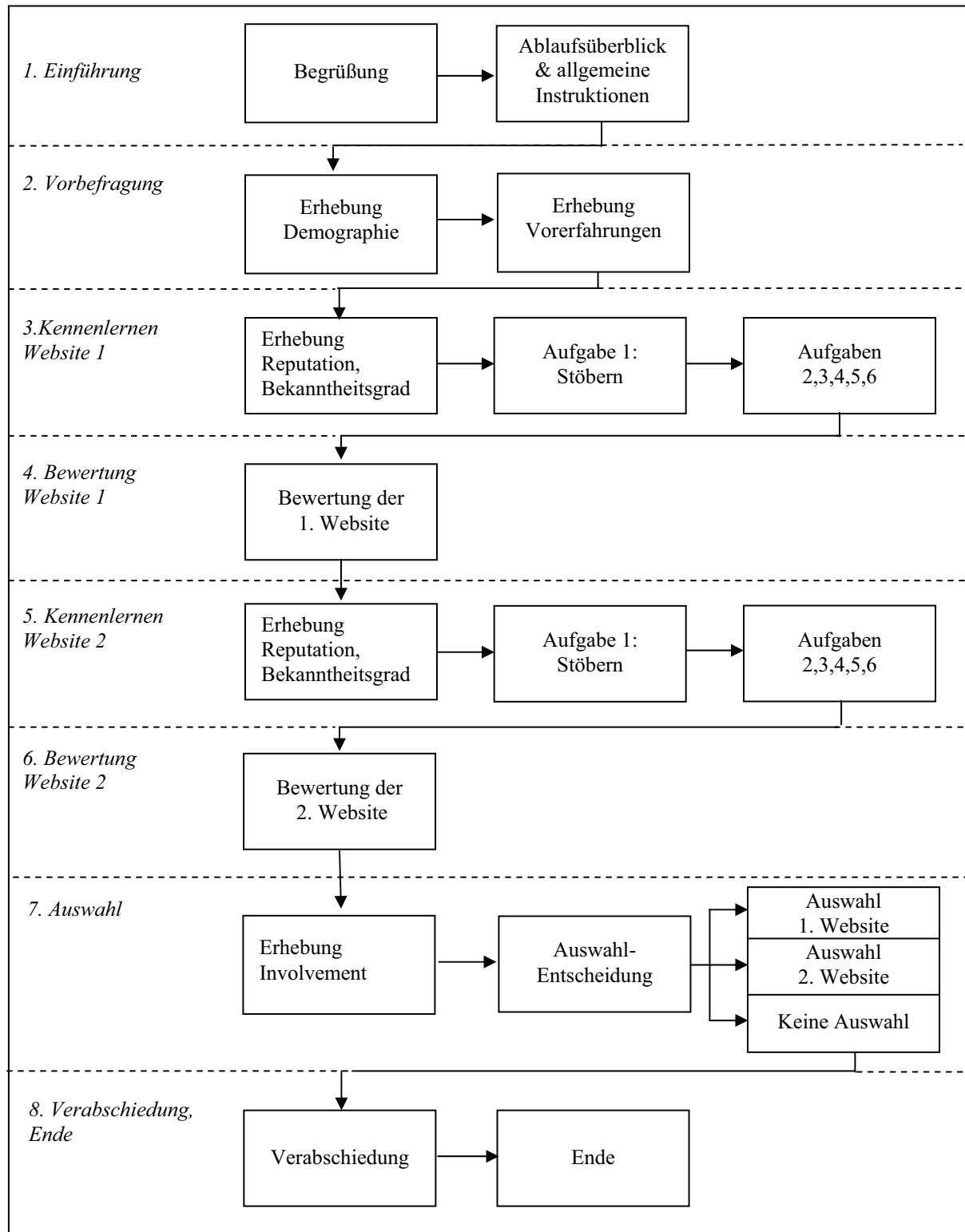


Abb. 5-3: Schematische Darstellung des Versuchsablaufs

Phase 1: Einführung

Die erste Phase geht über drei Einzelseiten der Onlineversuchsumgebung. Auf der ersten Seite wird der Teilnehmer begrüßt und erhält einen groben Überblick über den Ablauf des Versuchs sowie die ungefähre Versuchsdauer. Er wird außerdem darauf hingewiesen, dass seine Angaben vertraulich behandelt werden und im Rahmen des Versuchs nicht er getestet wird, sondern die beiden Websites.

Auf der zweiten Seite wird die Ausgangssituation beschrieben. Der Teilnehmer sollte von der Annahme ausgehen, er sei auf der Suche nach einer neuen Krankenkasse und verschaffe sich einen Eindruck über die Websites verschiedener Krankenkassen. Bezüglich des Beitragssatzes wurde er darauf hingewiesen, dass dieser ab Januar 2009 vereinheitlicht werde und er diesen deshalb bereits jetzt im Rahmen seiner Beurteilung vernachlässigen solle. In einer zweiten Alternative wurden die Teilnehmer auf einen Krankenkassen-Test der Zeitschrift FOCUS-MONEY aus dem November 2006 aufmerksam gemacht. Als wesentliche Ergebnisse wurde dargestellt, dass die Krankenkassen sich - obwohl 90% der Leistungen gesetzlich vorgeschrieben sind - in ihren Angeboten und im tatsächlich erbrachten Leistungsumfang teilweise erheblich unterscheiden.

Insbesondere in Bezug auf Zusatzleistungen, integrierte Versorgungsmodelle, alternative Medizin, Serviceleistungen und Bonusprogramme stellten die Tester Unterschiede bei den Krankenkassen fest, die in der Bewertung des PreisLeistungsverhältnisses zu Unterschieden von bis zu 50% führten. Mittels dieser Darstellungen sollte bei den Versuchspersonen die Bedeutung einer kritischen Auseinandersetzung mit der Thematik Krankenversicherung und damit deren Involvement während der Versuchsteilnahme gesteigert werden. Die Zuordnung der beiden Alternativen zu den Versuchspersonen erfolgte zufällig.

Auf der dritten Seite der Einführungsphase erfolgen weitere Instruktionen zur Vorgehensweise und Handhabung der Versuchsumgebung.

Phase 2: Vorbefragung

Die *Phase 2* des Quasi-Experiments dient der Vorbefragung und umfasst zwei Hypertextseiten. Auf der ersten Seite werden verschiedene Indikatoren zur Demographie (Alter, Geschlecht, Schulabschluss, Beruf, Einkommen, Haushaltsgröße, derzeitige Krankenkasse; siehe Abschnitt 4.5.5) erhoben. Auf der zweiten Seite werden Items zur Vorerfahrung in der PC-Anwendung (pc_j, pc_t, pc_h), zur Vorerfahrung in der Anwendung des Internet (www_j, www_t, www_h), zur Erfahrung mit Einkäufen im Internet (shopperf) sowie zur Nutzung des Internet für Informationsrecherchen (recherch; siehe Abschnitt 4.5.1) abgefragt.

Phase 3: Kennenlernen der Website 1

In der dritten Phase sollen die Teilnehmer die erste Website ausführlich kennen lernen. Dabei sollten die für die Nutzung der Website einer Krankenkasse typischen Interaktionsprozesse abgedeckt werden. Bei der Ausgestaltung des Versuchsszenarios wurde darauf geachtet, dass die Versuchspersonen in der Interaktion mit der Website ausreichende Erfahrungen sammeln, um eine Bewertung bezüglich der verschiedenen Bewertungsaspekte vornehmen zu können.

Bevor die Teilnehmer die Website kennen lernten wurden zunächst drei Items zur Einschätzung der Bekanntheit und Reputation des Websiteanbieters erhoben (b1, r1, r2; siehe Abschnitt 4.5.2 und 4.3.5). Daran anschließend wurden fünf Aufgaben gestellt, die eine intensive Beschäftigung mit der Website erforderte.

Die *erste Aufgabe* diente der freien Exploration der Website. Die Versuchspersonen wurden gebeten, sich einen allgemeinen Überblick über die Website zu verschaffen und erhielten hierfür fünf Minuten Zeit. Sie diente einer ersten Orientierung über Inhalte, Ausgestaltung und Handhabung der untersuchten Website. Die Instruktion lautete:

Verschaffen Sie sich einen allgemeinen Überblick über die Website zu Ihrer Orientierung. Nehmen Sie sich hierfür 5 Minuten Zeit.

Durch Anklicken des Links mit der Bezeichnung „Ich möchte jetzt mit dem Versuch fortfahren“ wurde zusätzlich zum aktuellen Fenster der Versuchsumgebung in einem zweiten Browserfenster die Startseite der ersten Website aufgerufen. Die Versuchsteilnehmer wurden instruiert, nach der Bearbeitung der ersten Aufgabe selbständig wieder zur Versuchsumgebung zurück zu kehren. Dazu war es erforderlich, in der Taskleiste auf das mit „Versuch“ bezeichnete Kästchen zu klicken.

In der *zweiten Aufgabe* wurden die Versuchspersonen instruiert, auf der Website die Regelungen zur Kostenübernahme von Zahnersatz aufzusuchen und sich einen groben Überblick darüber zu verschaffen, inwieweit die Krankenkasse die Kosten für die Behandlung übernimmt. Der Wortlaut der Instruktion war:

Suchen Sie die Regelungen über die Zuschüsse der Krankenkasse zum Zahnersatz auf. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.)

Nach Bearbeitung der Aufgabe sollten die Versuchspersonen selbständig wieder zur Versuchsumgebung zurückkehren, um mit der nächsten Aufgabe fortzufahren.

In der *dritten Aufgabe* sollten die Teilnehmer die Informationen zum Thema Gesundheit durch Fitness und Bewegung auffinden. Wie bereits in der vorhergehenden Aufgabe war nicht gefordert, die inhaltliche Qualität der Angebote zu erfassen und zu bewerten. Es sollten vielmehr die betreffenden Informationen gesucht und gefunden werden. Der Wortlaut der Instruktion war wie folgt:

Finden Sie auf der aktuellen Website Informationen und Angebote zum Thema Gesundheit durch Fitness und Bewegung. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.)

Die vierte Aufgabe bezog sich auf die Familienversicherung der Studenten. Die Versuchspersonen sollten auf der Website die Voraussetzungen für die kostenlose Familienversicherung für Studenten finden. Die Teilnehmerinstruktion lautete:

Finden Sie die Voraussetzungen für die kostenlose Familienversicherung für Studenten auf der aktuellen Website.

Die fünfte Aufgabe betraf die Möglichkeit der zusätzlichen Absicherung von Kosten für Zahnersatz durch eine private Krankenzusatzversicherung. Die Teilnehmer sollten auf der aktuellen Website eine private Zusatzversicherung auffinden, über die Aufwendungen für den gesetzlichen Eigenanteil bei Zahnersatz zusätzlich abgesichert werden können. Der Wortlaut der Teilnehmerinstruktion war:

Finden Sie auf der aktuellen Website eine Zusatzversicherung, mit der Sie die Kosten für Zahnersatz zusätzlich absichern können.

Phase 4: Bewertung der Website 1

Im Anschluss an die Phase der Aufgabenbearbeitung erfolgte die Bewertung der Website. Diese wurde anhand der Indikatoren zur Erfassung der Konstrukte Nutzungsintention (ni1, ni2, ni3), Nutzerzufriedenheit (z1, z2), Usability reflektiv (ufosV2ra: ur1a, ur2ac, ur3a, ..., ur8a), Usability formativ (ufosV2fa: uf1a, uf2ac, uf3a, ..., uf31a), Nutzungsvergnügen (vg1, vg2, vg3), Vertrauen (v1, v2, v3, v4), Ästhetik (ae1, ae2), sowie wahrgenommener Preis (subp) vorgenommen. Die insgesamt 53 Items wurden über zwei Seiten der Versuchsumgebung dargestellt. Beantwortete ein Teilnehmer ein Item auf einer Seite nicht, wurde eine Fehlermeldung in roter Schrift angezeigt, sobald der Hyperlink zur nächsten Seite angeklickt wurde. Die Teilnehmer konnten den Versuch erst fortsetzen, nachdem sie das ebenfalls in roter Schrift gekennzeichnete Item beantwortet hatten.

Während des Bewertungsvorgangs war es den Versuchspersonen möglich, zur Unterstützung ihrer Entscheidungsfindung bezüglich einzelner Bewertungsaspekte zur Website zurückzukehren. Die Indikatoren wurden bei der Speicherung der Daten zur Kennzeichnung der Bewertung auf Grundlage der ersten Website mit dem vorangestellten Zusatz „w1_“ codiert (z.B. w1_ni1).

Phase 5: Kennenlernen der Website 2

Nach Abschluss der Bewertung der ersten Website lernten die Teilnehmer in Phase 5 die zweite Website kennen. Die Versuchspersonen hatten dieselben fünf Aufgaben zu bearbeiten, wie bereits in Phase 3 beschrieben. Zu Beginn wurden zunächst wieder die Items zur Bekanntheit und Reputation des Anbieters erhoben.

Phase 6: Bewertung der Website 2

Die Bewertung der zweiten Website erfolgte identisch zur Vorgehensweise bei der ersten Website. Um die gespeicherten Ergebnisdaten der zweiten Website von den Datensätzen der ersten Website unterscheiden zu können, wurde jeder Indikator mit dem vorangestellten Zusatz „w2_“ codiert (z.B. w2_ni1).

Phase 7: Auswahl

Die siebte Phase umfasst zwei Seiten. Nachdem die Versuchsperson die beiden Websites in den Phasen 3 bis 6 kennen gelernt und bewertet hatte, wurde in Phase 7 auf der ersten Seite das Involvement der Versuchsperson und auf der zweiten Seite ihre Auswahlentscheidung erfasst (siehe Abschnitte 4.3.3 und 4.4.3).

Die Erfassung der Variable Involvement als Ausdruck persönlicher Bedeutung des Themenkomplexes Krankenversicherung für die Versuchsperson erfolgte mit Hilfe eines semantischen Differentials anhand von fünf Items (inv1, inv2, ..., inv5). Die Teilnehmerinstruktion lautete:

„Wir würden gerne mehr über Ihre Einschätzung der Bedeutung von Krankenversicherungsthemen erfahren. Bitte nutzen Sie dafür nachfolgende Beschreibungen. Kreuzen Sie auf der Skala von 1 (im ersten Beispiel: wichtig) bis 7 (im ersten Beispiel: unwichtig) Ihre Bewertung an. Nutzen Sie hierfür die sieben Auswahlfelder von 1 bis 7. Falls eine Aussage nicht bewertet werden kann, wählen Sie die Antwortmöglichkeit „keine Angabe“.“

Die Auswahlentscheidung bezog sich auf die Auswahl zwischen den beiden Krankenkassenwebsites, die den Versuchspersonen präsentiert wurden. Jede Versuchsperson wurde zum Abschluss des Versuchs aufgefordert, eine Auswahl zu treffen, welche der beiden Websites sie bevorzugt aufsuchen würde, um sich zu informieren, wenn sie eine neue Krankenkasse suchen würde.

Zu Kontrollzwecken wurde, ausgehend von der Annahme, dass die Versuchsperson einen neuen Krankenversicherungsgeber suche, ergänzend eine fiktive Auswahlentscheidung zwischen den beiden Krankenkassen abgefragt. Dabei wurde im Vorgriff auf die Regelungen des GKV-Wettbewerbsstärkungsgesetzes (in der Fassung vom 26.03.2007) unterstellt, dass die Beitragssätze vergleichbar und damit als Auswahlkriterium bedeutungslos wären.

Phase 8: Verabschiedung und Ende

Der Versuch endete mit der achten Phase. Die Versuchspersonen wurden verabschiedet und erhielten den Hinweis, sich im Falle von Fragen zum Versuch unter der angegebenen E-Mail-Adresse an den Versuchsleiter wenden zu können.

6. Statistische Auswertung der Daten

In diesem Kapitel werden zunächst die methodischen Schritte erläutert, die zur Aufbereitung der Daten im Vorfeld der Analyse erforderlich sind. Im Anschluss daran werden die bei der Analyse der Daten eingesetzten statistischen Verfahren sowie ihre Voraussetzungen und Gütekriterien dargestellt.

Zunächst wird das Vorgehen bei der Analyse und der Umgang mit fehlenden Werten erläutert. Es wird dargestellt, wie mit hohen Anteilen fehlender Daten auf Fallebene umgegangen wird, wie der Fehlendmechanismus bestimmt wird und welche Verfahren zur Ersetzung fehlender Werte angewandt werden (Abschnitt 6.1). In Abschnitt 6.2 wird das Vorgehen zur Identifikation potenziell problematischer Teilnehmer anhand der Analyse der Varianz und Konsistenz im Antwortverhalten vorgestellt. Anschließend werden in Abschnitt 6.3 die Kriterien zur Prüfung der Verteilung der eingesetzten Indikatoren erläutert. Zur statistischen Untersuchung, inwiefern ein Common Method Bias in dem vorliegenden Datensatz eine Rolle spielt, wird in Abschnitt 6.4 der Harman's Single-Factor Test erläutert. Das Verfahren zur Überprüfung der Güte der reflektiven Maße auf Prädiktorenebene orientiert sich an den Empfehlungen von Homburg und Giering (1996) und wird in Abschnitt 6.5 dargestellt. Das Verfahren verbindet Elemente der klassischen Itemanalyse (Abschnitt 6.5.2) mit der Durchführung einer explorativen sowie einer konfirmatorischen Faktorenanalyse. Deren Voraussetzungen und Gütekriterien werden in den Abschnitten 6.5.1 und 6.5.3 erläutert. In Abschnitt 6.6 werden die Gütekriterien des Strukturgleichungsverfahrens *Partial Least Squares* (PLS) dargestellt, mit Hilfe dessen zwei der wesentlichen Fragestellungen dieser Arbeit untersucht werden. Zum einen dient das Verfahren der Analyse des konzeptuellen Erfolgsmodells zur Vorhersage der Nutzungsintention und zum anderen wird ein Zwei-Konstrukt-Modell zur Überprüfung des angewandten formativen Usability-Maßes berechnet. Die Vorhersage des nominalskalierten Kriteriums Auswahlentscheidung soll mit Hilfe der *binomialen logistischen Regression* (LR) erfolgen. Deren Voraussetzungen und Gütekriterien sind Abschnitt 6.7 zu entnehmen. Desweiteren werden in den Abschnitten 6.8.1 bis 6.8.3 die Voraussetzungen und Kriterien des t-Tests, der mehrdimensionalen einfaktoriellen Varianzanalyse sowie des Levene-Tests erläutert, die im Rahmen der Datenanalyse eingesetzt werden.

Um die Bezüge zwischen den korrespondierenden Abschnitten zu verdeutlichen, wird in Tabelle 6.1 dargestellt, welches statistische Verfahren zur Untersuchung der einzelnen Hypothesen zur Anwendung kam. Für jede Hypothese werden die korrespondierenden Abschnitte zum theoretischen Hintergrund (Kapitel 2), zur Operationalisierung (Kapitel 4), zu den angewandten statistischen Verfahren (Kapitel 5) sowie zu den Ergebnissen (Kapitel 7) aufgeführt.

Tab. 6-1: Übersicht über die korrespondierenden Abschnitte zu den Hypothesen

Nr.	Hypothese (siehe Abschnitt 3.2)	Hintergrund	Operationalisierung		Statistische Verfahren	Ergebnis	
H1	Interne Konsistenz der Skala ufosV2ra	2.9	4.2		α	7.3.2	
					Composite Reliability	7.5.3	
H2	Faktorielle Validität der Skala ufosV2ra	2.9	4.2		EFA	7.3.1	
					CFA	7.3.3	
					PLS: Komplexes Modell	7.6.1	
H3	Differenzierungsfähigkeit der Skala ufosV2ra zwischen Websites	2.9	4.2		ANOVA	7.3.4	
H4	ufosV2fa → ufosV2ra	2.6.4	4.2		PLS: Zwei-Konstrukt-Modell	7.5.3	
H5a	ufosV2ra → Nutzerzufriedenheit	2.5.4	4.2	4.4.1	PLS: Komplexes Modell	7.6.2	
H5b	ufosV2fa → Nutzerzufriedenheit	2.5.4	4.2	4.4.1			
H6a	ufosV2ra → Nutzungsintention	2.5.4	4.2	4.4.2			
H6b	ufosV2fa → Nutzungsintention	2.5.4	4.2	4.4.2			
H7a	ufosV2ra → Auswahlentscheidung	2.5.4	4.2	4.4.3	LR	7.7	
H7b	ufosV2fa → Auswahlentscheidung	2.5.4	4.2	4.4.3			
H8a	ufosV2ra → Vertrauen	2.11.1	4.2	4.3.2	PLS: Komplexes Modell	7.6.2	
H8b	ufosV2fa → Vertrauen	2.11.1	4.2	4.3.2			
H9a	ufosV2ra → Nutzungsvergnügen	2.11.3	4.2	4.3.4			
H9b	ufosV2fa → Nutzungsvergnügen	2.11.3	4.2	4.3.4			
H10a	Ästhetik → ufosV2ra	2.11.4	4.3.1	4.2	-		
H10b	Ästhetik → ufosV2fa	2.11.4	4.3.1	4.2	-		
H11	Ästhetik → Nutzerzufriedenheit	2.10.3	4.3.1	4.4.1	-		
H12	Vertrauen → Nutzerzufriedenheit	2.10.1	4.3.2	4.4.1	PLS: Komplexes Modell	7.6.2	
H13	Nutzungsvergnügen → Nutzerzufriedenheit	2.10.4	4.3.4	4.4.1			
H14	Ästhetik → Nutzungsvergnügen	2.11.5	4.3.1	4.3.4	-		
H15	Reputation → Vertrauen	2.11.2	4.3.5	4.3.2	-		
H16	Involvement x Usability → Nutzungsintention	2.11.6	4.3.3	4.2	4.4.2	PLS: Komplexes Modell	7.6.2
H17	Involvement x Nutzerzufriedenheit → Nutzungsintention	2.11.6	4.3.3	4.4.1	4.4.2		
H18	Nutzerzufriedenheit → Nutzungsintention	2.5.4	4.4.1	4.4.2			
H19	Nutzungsintention → Auswahlentscheidung	2.5.4	4.4.2	4.4.3	LR	7.7	

Anmerkung: Hypothese (H), Partial Least Squares (PLS), binomiale logistische Regression (LR), einfaktorische Varianzanalyse (ANOVA), explorative Faktorenanalyse (EFA), Cronbachs Alpha (α), konfirmatorische Faktorenanalyse (CFA).

6.1 Analyse und Behandlung fehlender Werte

Datenausfälle in Umfragen, als ein Phänomen der Nichtantwort (Nonresponse) verstanden, können in die Nichtbeantwortung des gesamten Fragebogens (sog. Unit-Nonresponse) oder einzelner Fragen (sog. Item-Nonresponse) unterschieden werden (Grässler, 2000). Bei der Unit-Nonresponse reagiert ein Befragter gar nicht oder verweigert die Beteiligung an der

Untersuchung. Im Falle des Item-Nonresponse antworten die Respondenten nicht auf alle gestellten Fragen, sodass Lücken in den für die Auswertung vorgesehenen Datenmatrizen verbleiben.

Die Analyse fehlender Werte erleichtert es, Probleme zu erkennen, die durch unvollständige Daten entstehen können (Abschnitte 6.1.1 und 6.1.2). So können Fälle mit fehlenden Werten, die sich systematisch von vollständigen Fällen unterscheiden, die Interpretation der Ergebnisse erschweren. Andererseits können fehlende Werte dazu führen, dass die berechneten Statistiken aufgrund reduzierter Information ungenauer werden. Den nachfolgend angewandten statistischen Prozeduren liegt die Annahme zugrunde, dass alle Fälle vollständig sind, sodass eine Imputation der fehlenden Werte vorgenommen wird (Abschnitt 6.1.3).

6.1.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte

Fehlende Werte im Umfang von 1 bis 10 Prozent Item-Nonresponse, gerechnet als Datensätze mit fehlenden Werten in Relation zu allen vorhandenen Datensätzen, können als normal angesehen werden (Schnell, Hill & Esser, 1999, S. 430), Unit-Nonresponse ist hierbei nicht eingeschlossen. Diese Lücken bergen allerdings die Gefahr, dass es bei der Analyse zu Verzerrungen und in der Konsequenz zu Fehlschlüssen und Fehlentscheidungen kommt. In der vorliegenden Untersuchung werden deshalb Datensätze ab einem Anteil von mehr als 30% fehlender Daten von der Analyse ausgeschlossen und ersatzlos eliminiert.

6.1.2 Fehlendmechanismen und ihre Konsequenzen

Der Datenausfall kann in einem Zusammenhang mit dem Untersuchungsinhalt stehen oder aber rein zufällig erfolgen. Das Phänomen des Item-Nonresponse kann in drei Fehlendmechanismen unterschieden werden (Rubin, 1987; Schafer, 1999; Little & Rubin, 2002):

- ⇒ Missing Completely At Random (MCAR). Die fehlenden Werte können als Zufallsstichprobe aus den insgesamt zufällig ausgewählten Werten betrachtet werden. Es handelt sich um vollkommen zufälligen Datenausfall und die beobachteten Daten beinhalten keine Information über einen Fehlendmechanismus.
- ⇒ Missing At Random (MAR, auch: ignorierbare Nichtantwort). Die fehlenden Werte können als Zufallsstichprobe für gegebene beobachtete Werte betrachtet werden. Es handelt sich um bedingt zufälligen Datenausfall, bei dem die beobachteten Werte eine Information über den Fehlendmechanismus beinhalten.
- ⇒ Missing Not At Random (MNAR, auch: nicht ignorierbare Nichtantwort). Zwischen den Antwortwahrscheinlichkeiten und den fehlenden Werten besteht eine Abhängigkeit. Es handelt sich um einen nicht zufälligen Datenausfall, er ist damit verzerrend.

Abgesehen davon, dass durch den Informationsverlust die Analysen ungenauer werden, ist der MCAR-Ausfall sowohl bei der Unit- als auch der Item-Nonresponse unproblematisch. Es

entsteht keine grundsätzliche Verfälschung der späteren Analyseergebnisse. Hingegen führt ein MNAR-Ausfall immer zu verzerrten, nicht die Realität wiedergebenden Resultaten.

Bei einem univariaten Ausfallmuster kann eine Prüfung auf einen MCAR-Ausfall entsprechend eines MCAR-Tests nach Little durchgeführt werden. Im ersten Schritt werden die Datensätze nach vollständigen und unvollständigen getrennt. In den Teildatensätzen werden für die vorhandenen Variablen die Verteilungen ermittelt und schließlich im dritten Schritt zwischen den Teildatensätzen mittels t-Test oder Maximum-Likelihood-Quotienten-Test verglichen. Liegen keine signifikanten Differenzen vor, kann von einem MCAR-Ausfall ausgegangen werden. Komplizierter wird diese Überprüfung bei anderen Fehlendmustern (Little, 1988).

6.1.3 Imputation der fehlenden Daten

Bei der Imputation werden die fehlenden Werte auf individueller Ebene ergänzt. Die einfachste Möglichkeit der Imputation ist die der "einfachen Ergänzung" (engl. "single imputation"), bei der der fehlende Wert durch einen anderen, wie auch immer ermittelten Wert ersetzt wird. Bei der "multiplen Ergänzung" (engl. "multiple imputation") werden jedem fehlenden Wert zwei oder mehrere akzeptabel erscheinende Werte zugeordnet, die meist einer Wahrscheinlichkeitsverteilung folgen.

Da bei formativen Variablen nicht zwingend von einer hohen Korrelation zwischen den Indikatoren ausgegangen werden kann, werden an das Vorgehen der Imputation besondere Anforderungen gestellt. Das Verfahren der Mittelwertergänzung, bei dem die fehlenden Werte durch das arithmetische Mittel (alternativ auch den Modus oder den Median, siehe Bankhofer 1995, S. 106) über die für die Variable verfügbaren Werte ersetzt werden, ist nicht zur Ersetzung fehlender Werte formativer Indikatoren geeignet. Eine alternative Methode zur Ersetzung fehlender Werte bei formativen Messmodellen stellt die multiple Regressionsimputation nach Rubin (1987) dar. Für jeden fehlenden Wert werden pro leerem Datenfeld $m > 1$ Ersetzungswerte mithilfe von Regressionsanalysen geschätzt. Als Regressoren werden dabei Variablen ausgewählt, zu denen ein Zusammenhang vermutet wird und die bei allen Einheiten vollständig vorliegen. Die Zahl m wird willkürlich festgelegt und liegt in der Praxis gewöhnlich zwischen drei und zehn (Schafer, 1999). Auf diese Weise erhält man m vollständige Datentabellen, die parallel mit den Standardmethoden analysiert werden können. Die errechneten m Maßzahlen werden abschließend über m gemittelt (Rubin, 1987).

Im Ergebnis liefern multiple Imputationsverfahren sowohl bei MCAR-Ausfall als auch bei MAR-Ausfall meist verlässliche Ergebnisse (Schafer, 1999). Es darf jedoch nicht verkannt werden, dass selbst die modernen Verfahren einem MNAR-Ausfall, der anerkanntermaßen der wahrscheinlichste Ausfallmechanismus ist (Schafer, 1997, S. 22), nicht wirksam begegnen können (Schnell, Hill & Esser, 1999, S. 431).

6.2 Vorgehen zur Identifikation problematischer Fälle

Eine geringe Sorgfalt von Teilnehmern bei der Beantwortung des Fragebogens kann sich negativ auf die Güte der Beantwortung auswirken und zu höheren Messfehlern der erhobenen Variablen führen. Diese können dahingehend problematisch sein, dass sich vorhandene Zusammenhänge mit Hilfe der angewandten statistischen Verfahren nicht bestätigen lassen (vgl. Cote & Buckley, 1988). Deshalb ist es erforderlich, Teilnehmer zu identifizieren, die die Beantwortung der präsentierten Items nicht sorgfältig handhaben.

Nachfolgend werden zwei Kriterien dargestellt, die als Hinweise auf potentiell problematische Fälle dienen sollen. Werden die anhand dieser Kriterien identifizierten Fälle im Verlaufe der weiteren Analyse als Ausreißer auffällig, so können diese aus der jeweiligen Analyse ausgeschlossen werden. Dem liegt die Annahme zugrunde, dass die Ausreißer nicht der zugrunde gelegten Population angehören (vgl. Tabachnick & Fidell, 2006, S.73).

In Abschnitt 6.2.1 wird das Vorgehen zur Untersuchung der Varianz im Antwortverhalten der Teilnehmer dargestellt. Ein undifferenziertes Antwortprofil über die einbezogenen Items kann als Indiz für eine unreflektierte Bewertung gewertet werden. Abschnitt 6.2.2 beschreibt die Analyse der Konsistenz in der Beantwortung negativ bzw. positiv gepolter Items derselben reflektiven Skala.

6.2.1 Fallweise Analyse der Varianz im Antwortverhalten

Um Teilnehmer zu identifizieren, die sich ohne differenziertes Antwortverhalten durch den Fragebogen klicken, soll eine fallweise Analyse der Antwortvarianz vorgenommen werden. Teilnehmer, die ein besonders gleichförmiges Profil in den Itemausprägungen aufweisen, werden dabei als problematisch bewertet. Hierfür werden für jeden Teilnehmer die Antworten auf alle Items, ausgenommen Demographie und Vorerfahrung, betrachtet. Herangezogen werden die vier Items zur Nutzungsintention (ni_1, ni_2, ni_3, ni_4), die zwei Items zur Zufriedenheit (z_1, z_2), die acht Items zur reflektiven Usability ($ufosV2ra; ur_1, ur_2c, ur_3, \dots, ur_8$), die 31 Items zur formativen Usability ($ufosV2fa; uf_1, uf_2, \dots, uf_{31}$), die vier Items zum Vertrauen (v_1, v_2, v_3, v_4), die drei Items zum Nutzervergnügen, die fünf Items zum Involvement (i_1, i_2, i_3, i_4, i_5), die beiden Items zur Ästhetik (ae_1, ae_2) sowie das Item zur wahrgenommenen Beitragssatzhöhe. Als Kriterium für ein gleichförmiges Profil im Antwortverhalten wird die Streuung über die aufgeführten Indikatoren berechnet, wobei als kritischer Wert für die Streuung ein Wert von $SD \leq 1$ festgelegt wird (vgl. Christophersen, 2007). In diesem Falle wird ein undifferenziertes "Durchklicken" durch den Fragebogen unterstellt.

6.2.2 Konsistenz des Antwortverhaltens

Zur Identifikation weiterer problematischer Fälle wird die Konsistenz in der Beantwortung negativ bzw. positiv gepolter Items derselben reflektiven Skala geprüft. Eine negative

Formulierung von Items wurde weitestgehend vermieden, da die Beantwortung negativ gepolter Items ein höheres Maß an kognitiver Anstrengung erfordert und daraus ein höherer Messfehler resultieren kann (4.2.4). Lediglich die reflektive Skala zur Usability enthält ein negativ gepoltes Item. Als Indikator für eine inkonsistente Bewertung wird deshalb die Abweichung der Bewertung des negativ gepolten Items ur_{2ac} zum Mittelwert der übrigen sieben Indikatoren der reflektiven Skala $ufosV_{2ra}$ herangezogen. Bei einem Betrag der Abweichung von größer 4 wird diese als Hinweis auf ein Durchklicken gewertet. Schließlich erscheint es beispielhaft wenig plausibel, dass der Umgang mit der Website einerseits leicht zu erlernen ist (ur_{4a}) und gleichzeitig die Nutzung zu kompliziert ist (ur_{2ac}).

Weitere negativ gepolte Items finden sich in der formativen Skala $ufosV_{2fa}$. Diese bieten sich für eine entsprechende Analyse allerdings nicht an, da im Falle einer formativen Skala nicht zwingend hohe Korrelationen zwischen den Items erwartet werden können (2.6.7). In der Folge lässt sich durch das dargestellte Vorgehen ein differenziertes Antwortverhalten nicht von einem Durchklicken unterscheiden.

Kritisch ist anzumerken, dass auf diese Weise nur diejenigen Teilnehmer als „Durchklicker“ identifiziert werden können, die zumindest im Falle der reflektiven Usability-Skala eine besonders positive bzw. negative Bewertung abgeben. Bei Teilnehmern, die regelmäßig eine eher mittelmäßige Bewertung abgeben, eignet sich das Verfahren aufgrund der eher als gering zu erwartenden Abweichungen zwischen positiv und negativ gepolten Items nicht, um ein unreflektives Antwortverhalten zu erkennen.

6.3 Verteilungsprüfung

Voraussetzung einiger der verwendeten statistischer Verfahren ist eine Normalverteilung der einbezogenen Variablen. Um zu überprüfen, inwiefern die in der Untersuchung erhobenen Indikatoren normalverteilt sind, soll für jede Variable ein *Kolmogoroff-Smirnow-Anpassungstest* durchgeführt werden sowie die *Schiefte* und der *Exzess* bestimmt werden.

Der *Kolmogoroff-Smirnow-Anpassungstest* (KSA-Test) ist ein statistischer Test auf Übereinstimmung zweier Wahrscheinlichkeitsverteilungen. Das kann der Vergleich der Verteilung zweier Stichproben sein, aber auch der Test darauf, ob eine Stichprobe einer zuvor angenommenen Wahrscheinlichkeitsverteilung folgt. Mit Hilfe des Kolmogoroff-Smirnow-Anpassungstests wird für jede Variable die Nullhypothese getestet, eine Normalverteilung läge zugrunde. Ein signifikantes Ergebnis im Test weist somit darauf hin, dass die Verteilung von der Normalverteilung abweicht (Bortz, Lienert & Boehnke, 2000, S. 319 ff.).

Zur genaueren Beschreibung der Verteilungsform sollen die *Schiefte* und der *Exzess* bestimmt werden (Bortz, 2005, S.45f.). Die Schiefe ist ein Maß für die Symmetrie einer Wahrscheinlichkeitsverteilung zum Mittelwert. Von Schiefe spricht man bei einer eingipfligen assymetrischen Verteilung, wobei man die Verteilung als linkssteil bezeichnet, wenn die Schiefe > 0 ist und als rechtssteil, wenn die Schiefe < 0 ist. Ist die Schiefe $= 0$, so

bezeichnet man die Verteilung als symmetrisch. Da die Gaußsche Normalverteilung die Schiefe Null hat, ist die Schiefe ein geeignetes Instrument, um eine beliebige Verteilung mit der Normalverteilung zu vergleichen. Der Exzess (Ex) beschreibt die Form einer Verteilung dahingehend, wie spitz- oder breitgipflig die Wölbung der Verteilung verläuft. Der Exzess einer Normalverteilung beträgt $Ex = 0,263$. Liegt der Exzess einer Verteilung über diesem Wert, so beschreibt dies eine im Vergleich zur Normalverteilung zu steile Verteilung. Liegt er unter diesem Wert, beschreibt er eine im Vergleich zur Normalverteilung flachere Verteilung. Anhand der z-Werte lässt sich überprüfen, ob eine Verteilung bezüglich Schiefe und Exzess signifikant von der Normalverteilung abweichen.

6.4 Prüfung auf Vorliegen eines Common Method Bias

Ein *Common Method Bias*, teilweise auch *Common Method Variance* genannt, ist in Studien wahrscheinlich, in denen die Datenerhebung sowohl für die Prädiktorvariablen als auch für die Kriteriumsvariable von derselben Person, im selben Erhebungs- und Itemkontext und mit ähnlichen Itemcharakteristiken erfolgt (Podsakoff, MacKenzie & Lee, 2003; Spector, 2006).

Ein *Common Method Bias* kann grundsätzlich durch zwei Herangehensweisen kontrolliert werden, durch das Design des Vorgehens sowie mit Hilfe von statistischen Kontrollen (Podsakoff et al., 2003). Aspekte, die bei der Formulierung der Fragebogenitems zu beachten sind, um einen *Method Bias* möglichst gering zu halten, werden in Abschnitt 2.6.4.2 beschrieben. Möglichkeiten zum Umgang mit potentiellen Effekten aufgrund negativ benannter Items sowie aufgrund der Reihenfolge in der Itemdarstellung werden in den Abschnitten 4.2.4 und 4.7 dargestellt. Mögliche Beeinflussungen, die daraus resultieren, dass eine Testperson die Website des aktuellen Versicherungsgebers bewertet, werden im Versuch durch eine Abfrage desselben kontrolliert. Darüber hinaus können bei dem vorliegenden Untersuchungsobjekt der Websites von Krankenkassen methodische Effekte aufgrund einer durch die Testperson vermuteten sozialen Erwünschtheit bei der Beantwortung einzelner Items oder Konstrukte auftreten.

Zur statistischen Untersuchung, inwiefern ein Common Method Bias in einem Datensatz eine Rolle spielt, kann der *Harman's Single-Factor Test* eingesetzt werden (vgl. Podsakoff et al., 2003). Hierbei wird über alle Indikatoren der Untersuchung eine explorative Faktorenanalyse ohne Rotation der Faktoren durchgeführt. Auf einen Common Method Bias kann geschlossen werden, wenn lediglich ein wesentlicher Faktor extrahiert wird und dieser einen Großteil der Gesamtvarianz des Modells erklärt. Podsakoff et al. (2003) kritisieren diesen Test insofern als eher insensitiv, als dass sich nur dann ein einziger Faktor ergibt, wenn der Common Method Bias die Kovarianzen der Items in aller Gänze beeinflusst, was jedoch eher unwahrscheinlich sei (Podsakoff et al., 2003). Trotz dieser Einschränkung wird in Ermangelung eines geeigneteren Verfahrens in der Versuchsanordnung der Harman's Single-Factor Test angewandt.

6.5 Analyse der Güte reflektiver Maße auf Prädiktorenebene

Die Überprüfung der Güte der reflektiven Skalen, die in der Untersuchung auf Prädiktorenebene herangezogen werden, erfolgt mit Hilfe eines dreistufigen Vorgehens, das sich an den Empfehlungen von Homburg und Giering (1996) orientiert.

Schritt 1: Sämtliche reflektiven Indikatoren werden in eine explorative Faktorenanalyse einbezogen um die angenommene Skalenstruktur zu überprüfen.

Schritt 2: Jede einzelne reflektive Skala wird nach Kriterien der klassischen Testtheorie untersucht.

Schritt 3: Mit Hilfe einer konfirmatorischen Faktorenanalyse (CFA) werden die Hypothesen über die Faktorenstruktur der reflektiven Skalen erneut untersucht.

6.5.1 Explorative Faktorenanalyse zur Analyse der Skalenstruktur

Die theoretisch angenommene Struktur der reflektiven Skalen wird mit Hilfe einer explorativen Faktorenanalyse überprüft. Für die Berechnung der Faktoren steht eine Auswahl an Extraktionsmethoden zur Verfügung, von denen die Hauptachsenanalyse und die Maximum-Likelihood-Analyse (ML) die am häufigsten angewandten sind (Bühner, 2004, S.153). Es wird eine Hauptachsenanalyse bei Verwendung des obliquen Rotationsverfahren Direct Oblimin unter Einbeziehung der Indikatoren zu den Konstrukten Usability (reflektiv), Vertrauen, Vergnügen, Ästhetik, Involvement und Reputation durchgeführt (vgl. Parasuraman, Zeithaml & Malhotra, 2005; McKinney & Zahedi, 2002; Hiltz & Johnson, 1990; Burroughs & Sabherwal, 2002).

6.5.1.1 Voraussetzungen

- a. *Skalierung und Verteilungen*: im Gegensatz zur ML, die multivariate Normalverteilung voraussetzt, ist das Verfahren der Hauptachsenanalyse nicht an bestimmte Verteilungsannahmen gebunden. Normalverteilung und Intervallskalenniveau sind keine Voraussetzung für die Durchführung einer Hauptachsenanalyse. Liegt beides vor, bestehen allerdings optimale Bedingungen für die Durchführung (Backhaus et al. 2003, S.273; Bühner, 2004, S.159).
- b. *Eignung der Korrelationsmatrix*: Die Anwendung einer Faktorenanalyse ist nur dann sinnvoll, wenn die Variablen untereinander substantielle Korrelationen aufweisen. Eine Quantifizierung, ob ausreichende Korrelationen in der Korrelationsmatrix vorliegen, wird mit Hilfe des *Kaiser-Meyer-Olkin-Koeffizienten* (KMO) und des *Bartlett-Tests* vorgenommen. Der KMO-Koeffizient zeigt an, ob substantielle Korrelationen vorliegen, die die Durchführung einer Faktorenanalyse rechtfertigen und gibt damit Anhaltspunkte, ob die Variablenauswahl für eine Faktorenanalyse geeignet ist (Backhaus et al. 2003, S. 276). Faktisch wird der gemeinsame Varianzanteil ($\sum_{i,j} r_{ij}^2$) aller Variablen miteinander

bestimmt und durch die Summe aus gemeinsamem Varianzanteil (r_{ij}^2) zwischen allen Variablen plus den quadrierten Partialkoeffizienten ($r_{ij.z}^2$) geteilt. Die Bewertung des KMO-Koeffizienten wird folgendermaßen vorgenommen (vgl. Kaiser & Rice, 1974): 0,9 bis 1,0 *erstaunlich*; 0,8 bis unter 0,9 *verdienstvoll*; 0,7 bis unter 0,8 *ziemlich gut*; 0,6 bis unter 0,7 *mittelmäßig*; 0,5 bis unter 0,6 *kläglich*; unter 0,5 *untragbar mit der Durchführung*.

- c. Als Entsprechung zum KMO-Maß werden auf Indikatorebene die *MSA-Werte* (Measure of Sampling Adequacy) betrachtet. Bei der Berechnung werden jedoch nur die variablenbetreffenden Korrelationen summiert. Die Bewertung der Wertausprägung erfolgt analog zum KMO-Koeffizienten. Bei einem $MSA < 0,5$ sollte der entsprechende Indikator eliminiert werden (Backhaus et al., 2003, S. 310). Der Bartlett-Test prüft die Nullhypothese, alle Korrelationen seien gleich null (Bühner, 2004, S.171).

6.5.1.2 Bestimmung der Faktorenanzahl und Faktorenqualität

Die Auswahl der Faktoren dient in erster Linie der Gewinnung von aussagekräftigen, gut interpretierbaren Ergebnissen. Einen Anhaltspunkt für die Bestimmung der Faktorenanzahl bietet das Kaiser-Guttman-Kriterium, nach dem die Zahl der zu extrahierenden Faktoren gleich der Zahl der Faktoren mit *Eigenwert* $\lambda > \text{eins}$ ist. Der *Eigenwert* λ ist ein Indikator für den Anteil der Gesamtvarianz aller Variablen, den dieser Faktor aufklärt und ergibt sich aus der Summe aller quadrierten Faktorladungen des Faktors. Ist der Eigenwert eines Faktors größer als Eins, klärt er mehr Varianz auf als eine standardisierte Variable.

Die *Kommunalität* h^2 gibt an, in welchem Maße die Varianz eines Indikators durch alle Faktoren zusammen aufgeklärt wird. Generell nimmt die Stabilität der Faktorenlösung mit wachsender Stichprobengröße zu, da der Standardfehler abnimmt. Zusätzlich erhöht sich die Stabilität, wenn die Kommunalitäten der Indikatoren ansteigen. Ein Indikator mit einer geringen Kommunalität wird durch das Modell insgesamt schlecht vertreten. Bei einer Stichprobengröße ab $N > 300$ bestehen mit *Kommunalitäten* $h^2 \geq 0,5$ gute Voraussetzungen für die Durchführung einer Faktorenanalyse (MacCallum, Wideman, Zhang & Hong, 1999). Bei hohen Ladungen eines Indikators auf mehreren Faktoren (Nebenladungen) ist seine Elimination zu überprüfen.

Es ergibt jedoch, unabhängig von den gewählten Extraktionskriterien, keinen Sinn, eine Anzahl von Faktoren zu extrahieren, die inhaltlich nicht plausibel interpretierbar ist. Nach Fabrigar, Wegener, MacCallum und Strahan (1999) ist eine Überfaktorisierung weniger problematisch als eine Unterfaktorisierung, bei der zu wenige Faktoren extrahiert werden.

Um die Faktorenstruktur besser interpretieren zu können, existieren verschiedene Rotationstechniken. Durch eine geeignete *Rotation* lassen sich die Faktorladungen im Sinne von Thurstones Einfachstruktur so optimieren, dass für jeden Faktor die Bedeutung der einzelnen Indikatoren deutlich wird und dabei möglichst hohe Ladungen eines Indikators auf

einem Faktor und gleichzeitig niedrige Ladungen auf den anderen Faktoren erzielt werden (Bühner, 2004, S.164). Es wird das oblique Rotationsverfahren Direct Oblimin unter SPSS gewählt, da dieses Verfahren korrelierende Faktoren unterstützt. Es wird dabei die Annahme eines Faktorenmodells aufgegeben, dass die Faktoren untereinander keine Korrelationen aufweisen.

6.5.2 Güteprüfung entsprechend der klassischen Testtheorie

Im Folgenden werden die Kriterien des zweiten Prüfschrittes zur Messgüte der reflektiven Skalen nach Homburg und Giering (1996) sowie weitere Kriterien, die im Rahmen der *Item-analyse* verwendet werden, dargestellt (vgl. Bortz & Döring, 2006, S. 217 ff.).

Cronbach's α ist ein Maß für die interne Konsistenz einer aus mehreren Items zusammengesetzten Skala und stellt insbesondere bei unterschiedlichen Trennschärfen bzw. unterschiedlich hohen Faktorladungen ein Mindestmaß der Reliabilität einer Skala dar (MacDonald & Paunonen, 2002). Solange der Alpha-Wert unterhalb von 0,7 liegt, kann derjenige Indikator mit der niedrigsten Item-to-Total-Korrelation eliminiert werden (Churchill 1979, S. 68). Dabei muss allerdings berücksichtigt werden, dass das Konstrukt noch erhalten bleibt. Bei formativen Messmodellen ist dieses Vorgehen der Item-Bereinigung aus konzeptionellen Gründen nicht zulässig, da bei formativen Messmodellen die unterschiedlichen Items nicht miteinander korreliert sein müssen (Christophersen & Grape, 2006). Die interne Konsistenz kann damit nicht als Reliabilitätskriterium bei formativen Konstrukten verwendet werden.

Da zur Messung der Konstrukte Reputation und Ästhetik jeweils nur zwei Indikatoren herangezogen werden, besitzt der Cronbachs Alpha-Wert für diese Skala nur begrenzte Aussagekraft, so dass zusätzlich die Spearman-Brown-Korrelation als Kriterium der internen Konsistenz zu bestimmen ist (Fisseni, 2004, S. 54ff.).

Im Rahmen der Itemanalyse sind nach Homburg und Giering (1996) neben den *Anteilen fehlender Werte*, die *Homogenität*, die *Itemschwierigkeit* (p_m) sowie die *Itemtrennschärfe* zu überprüfen.

Homogenität der Items bezieht sich auf die Forderung, dass die die Skala bildenden Items widerspruchsfrei und eindimensional sein sollten (Peter, 1979).

Die *Itemschwierigkeit* hat einen Wertebereich von 0 bis 1 und sollte im Idealfall bei einer mittleren Schwierigkeit von $p_m = 0,5$ liegen, da Items dann am besten zwischen unterschiedlichen Merkmalsausprägungen differenzieren können und eine gute Voraussetzung für hohe Trennschärfe bieten (Bühner, 2004, S.91). Itemschwierigkeiten im Wertebereich von 0,2 bis 0,8 sind grundsätzlich akzeptabel (Fisseni, 1997, S.124). Außerhalb des Wertebereiches sollten Items eliminiert werden (Bortz & Döring, 1995, S. 199).

Die *Itemtrennschärfe* oder Diskriminationsfähigkeit ist ein Indikator dafür, inwieweit ein Item etwas Ähnliches erfasst wie die restlichen Items der Skala. Bei der Ermittlung des Trennschärfekoeffizienten wird für jedes Item die Korrelation des Items mit der um das jeweilige Item korrigierten Skala ermittelt (Bühner, 2004, S. 87). Der Wertebereich liegt zwischen -1 und +1. A priori sind möglichst hohe positive Trennschärfen wünschenswert. Fisseni (1997, S.124) bezeichnet Werte ab 0,3 als akzeptabel, Werte ab 0,5 als gut. Items mit geringer und negativer Trennschärfe sind zu eliminieren.

6.5.3 Prüfung der Skalenstruktur mit Hilfe konfirmatorischer Faktorenanalyse

Im dritten Schritt werden die Hypothesen über die Faktorenstruktur der reflektiven Skalen zur Erfassung der Prädiktoren einer konfirmatorischen Faktorenanalyse (CFA) unterzogen, um die Übereinstimmung des Modells mit den Daten zu untersuchen (Bühner, 2004, S.197).

6.5.3.1 Voraussetzungen zur Berechnung einer CFA

Die Durchführung einer CFA ist an Voraussetzungen bezüglich Umfang und Qualität der Daten gebunden.

Verteilung: Sowohl die Maximum Likelihood Methode (ML) als auch die Generalized-Least-Squares-Methode (GLS) setzen eine multivariate Normalverteilung sowie ein intervallskaliertes Skalenniveau der in die Analyse einbezogenen Variablen voraus. McDonald und Ho (2002) konnten allerdings zeigen, dass die ML-Methode robust gegenüber Verletzungen der Normalverteilungsannahme ist.

Linearität: Lineare Zusammenhänge zwischen den Variablen sowie das Fehlen verzerrender Ausreißer sind Voraussetzungen für eine sinnvolle Interpretation einer CFA (vgl. Bühner, 2004, S. 208).

Kollinearität: Da hoch korrelierende Items insbesondere im Rahmen von ML-Schätzungen zu Schätzproblemen führen können, sollten keine Items mit bivariaten Korrelationen $r > 0,85$ in der CFA verwendet werden (Bühner, 2004, S. 208).

Stichprobenumfang: Nach Marsh, Hau, Balla und Grayson (1998) können bei kleineren Stichproben häufiger Schätzprobleme auftreten. Es sollte daher eine Stichprobengröße von $n > 200$ angestrebt werden. Nach Kline (1998, S. 112) sollte das Verhältnis von Stichprobengröße und Itemanzahl nicht kleiner als fünf zu eins, besser zehn zu eins sein.

Bei einer Stichprobengröße von $N > 100$ und Werten für Schiefe und Exzess innerhalb der von West, Finch und Curran (1995) dargestellten Grenzen (Schiefe < 2 , Exzess < 7) wird die Durchführung einer CFA nach der ML-Methode empfohlen (Bühner, 2004, S.160; Olsson, Foss, Troje & Howell, 2000).

6.5.3.2 Gütekriterien der CFA

Im Folgenden werden die *Gütekriterien* zur Durchführung einer CFA dargestellt (vgl. Backhaus et al., 2003, S. 371ff; Homburg & Baumgartner, 1998). Zur Beurteilung der einzelnen Modellkomponenten werden häufig die Indikatorreliabilität, die Faktorreliabilität und die durchschnittlich erfasste Varianz als Gütekriterien herangezogen. Während die Indikatorreliabilität ein Mass für die Reliabilität eines einzelnen Indikators darstellt, zeigen die Faktorreliabilität und die durchschnittlich erfasste Varianz eines Faktors, wie gut dieser durch die ihm zugeordneten Indikatoren gemessen wird.

Die *Indikatorreliabilität* gibt für einen einzelnen Indikator den Anteil der durch den zugehörigen Faktor erklärten Varianz an der Gesamtvarianz des Indikators an. Die Bestimmung der Indikatorreliabilität erfolgt nach der Formel (vgl. Homburg & Giering, 1996):

$$(3) \quad rel(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}}$$

wobei: λ_{ij} die geschätzte Faktorladung zwischen dem Indikator x_i und dem zugrunde liegenden Faktor ξ_j ,

ϕ_{jj} die geschätzte Varianz von ξ_j und

θ_{ii} die geschätzte Varianz des Messfehlers δ_i .

Relevanter als die Indikatorreliabilität ist die *Faktorreliabilität*. Diese gibt die Eignung eines Faktors zur Erklärung aller ihm zugeordneten Indikatoren an. Sie wird anhand der folgenden Formel bestimmt:

$$(4) \quad rel(\xi_j) = \frac{\left(\sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj}}{\left(\sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}}$$

Zur Beurteilung der *Diskriminanzvalidität* der Faktoren wird das Fornell-Larcker-Kriterium verwendet (Fornell & Larcker, 1981). Demnach liegt Diskriminanzvalidität zwischen einzelnen Faktoren vor, wenn die durchschnittlich erfasste Varianz eines Faktors $DEV(\xi_j)$ größer ist als jede quadrierte Korrelation dieses Faktors mit einem anderen Faktor. Die durchschnittlich erfasste Varianz gibt die Eignung eines Faktors zur Erklärung aller ihm zugeordneten Indikatoren an.

Die Berechnung der *durchschnittlich erfassten Varianz* eines Faktors ξ_j erfolgt anhand nachfolgender Formel (Fornell & Larcker, 1981, S. 46):

$$(5) \quad DEV(\xi_j) = \frac{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj}}{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}}.$$

Als Schwellenwerte werden genannt für die

- ⇒ *Indikatorreliabilität*: $rel(\chi_i) \geq 0,5$ (Götz & Liehr-Gobbers, 2004).
- ⇒ *Faktorreliabilität*: $rel(\zeta_i) \geq 0,6$ (Bagozzi & Yi, 1988, S. 82).
- ⇒ *Durchschnittlich erfasste Varianz*: $dev(\zeta_i) \geq 0,5$ (Bagozzi & Yi, 1988, S. 82; Homburg & Baumgartner, 1995).

Die *Beurteilung der Gesamtstruktur* erfolgt auf Grundlage eines χ^2 -Tests mit der Nullhypothese, dass das Modell zur Datenstruktur passt. Liegt die Differenz über dem $\chi^2(1)$ -Wert von 3,841 (5 %-Niveau), ist die Differenz signifikant und die Nullhypothese abzulehnen. Dabei steigt mit zunehmender Stichprobengröße die Teststärke, der χ^2 -Test wird sensitiver gegenüber Modellfehlspezifikationen. Das bedeutet, dass auch kleine Abweichungen der Daten vom Modell mit der Ablehnung des Modells bestraft werden (Bühner, 2004, S.203).

Zusätzlich werden globale Fit-Indizes überprüft, um die Konsistenz des Messmodells mit den Daten zu beurteilen (Homburg & Giering, 1996, S. 9 f.). Die Fit-Indizes unterscheiden sich in Ihrer Sensitivität gegenüber Modellfehlspezifikationen und ihrer Abhängigkeit von der Stichprobengröße. Hu und Bentler (1999) empfehlen, bei der Anwendung der ML-Methode neben dem χ^2 -Wert zwei weitere Kennwerte zur Prüfung des Modell-Fit zu kombinieren. So schlagen sie vor, den SRMR (Standardized-Root-Meansquare-Residual) mit dem RMSEA (Root-Mean-Square-Error-of-Approximation) bzw. dem CFI (Comparative-Fit-Index) oder dem RNI (Relative-Noncentrality-Index) zu kombinieren. In der Literatur werden häufig außerdem der GFI (Goodness-of-Fit-Index) und der AGFI (Adjusted-Goodness-of-Fitness-Index) angeführt. Allerdings empfehlen Hu und Bentler (1998) nach umfangreichen Simulationsstudien bei der Anwendung der ML-Methode keinen der beiden Indizes zur Evaluation des globalen Modell-Fits zu verwenden (ebenso Bühner, 2004, S.205). Ein Messmodell ist zu verwerfen, wenn die Globalkriterien keine zufriedenstellende Ausprägung aufweisen.

Die Schwellenwerte anhand derer die Güte des Modells bestimmt werden kann, sind nach Hu und Bentler (1998, 1999) für den CFI und RNI jeweils mit 0,95 bestimmt, wobei ein gutes Modell Werte darüber erzeugt. Für den SRMR fordern sie Werte $\leq 0,11$.

Tab.6-2: Anspruchsniveaus der herangezogenen Gütemaße zur Beurteilung des CFA-Modells (Bagozzi und Yi 1988, S. 82; Bühner, 2004, S. 205; Götz & Liehr-Gobbers, 2004; Homburg & Baumgartner, 1995, S. 170; Hu & Bentler, 1998, 1999)

Bezeichnung des Gütekriteriums	Anspruchsniveau
Indikatorreliabilität	$\geq 0,5$
Faktorreliabilität $rel(\xi_j)$	$\geq 0,6$
Durchschnittlich erfasste Varianz eines Faktors	$\geq 0,5$
Fornell-Larcker-Kriterium	$DEV(\xi_i) > \text{quadr. Korrelation}(\xi_i, \xi_j)$, für alle $i \neq j$
χ^2 -Differenztest (5%-Niveau)	χ^2 -Differenz $\geq 3,841$
SRMR	$\leq 0,11$
RMSEA	$\leq 0,06$ (bzw. $0,08$ bei $N > 250$)
CFI	$\geq 0,95$
RNI	$\geq 0,95$
GFI*	$\geq 0,9$
AGFI*	$\geq 0,9$

Anmerkung: * Bei der Schätzmethode ML wird von einer Anwendung abgeraten (Hu & Bentler, 1998).

6.6 Kausalanalyse mit Hilfe von Partial Least Squares (PLS)

Mit Hilfe der Kausalanalyse auf Basis des *Partial Least Squares* (PLS) Algorithmus wird zum einen das konzeptuelle Erfolgsmodell zur Vorhersage der Nutzungsintention untersucht. Zum anderen wird ein Zwei-Konstrukt-Modell zur Überprüfung des angewandten formativen Usability-Maßes berechnet. Nachfolgend werden zunächst die Voraussetzungen zur Anwendung des Partial Least Squares Algorithmus und anschließend die Gütekriterien vorgestellt.

6.6.1 Voraussetzungen zur Anwendung des Partial Least Squares Algorithmus

Formal wird das PLS-Modell durch zwei Gleichungssysteme bestimmt: Das innere Strukturmodell gibt auf Basis theoretischer Vorüberlegungen die Beziehungen zwischen den latenten Variablen oder auch so genannten Konstrukten des Modells wieder, während das äußere Messmodell die Beziehungen zwischen den beobachtbaren, manifesten Variablen und den nicht beobachtbaren latenten Variablen aufweist. Hinsichtlich des Messmodells berücksichtigt PLS im Gegensatz zu kovarianzbasierten Ansätzen nicht nur reflektive Indikatoren, sondern auch Indikatoren formativer Art. Bei der PLS-Methode handelt es sich um ein iteratives KQ-Schätzverfahren, in dem nur Variablen-Subsets einbezogen werden. Das iterative Schätzverfahren bildet auch den Grund, warum PLS im Gegensatz zum kovarianzbasierten Ansatz, der eine multivariate Normalverteilung voraussetzt, keine Verteilungsannahmen für die Schätzung der Parameter benötigt.

Die zur Schätzung benötigte Größe des Samples richtet sich nach der umfangreichsten Regressionsgleichung des Modells. Diese kann entweder durch die größte Anzahl formativer Indikatoren eines Konstruktes oder durch die größte Anzahl an Konstrukten, die auf ein endogenes Konstrukt des Modells laufen, bestimmt werden. Chin und Newstedt (1999, S.335) empfehlen etwa, mindestens zehnmal so viele Datensätze einzubeziehen, wie maximal Pfade (Strukturpfade oder Indikatoren) in einem Konstrukt zusammenlaufen. Damit dürfte die notwendige Stichprobengröße selten 100 übersteigen. Um die erforderliche Stichprobengröße statistisch fundiert zu bestimmen, eignet sich der Power-Test nach Cohen (1988; vgl. Faul, Erdfelder, Lang & Buchner, 2007). Unter Power wird das Potenzial einer Studie verstanden, Unterschiede bzw. Beziehungen ausfindig zu machen, die tatsächlich in einer Population existieren. Es handelt sich also um die Wahrscheinlichkeit, dass ein statistischer Test einen signifikanten Unterschied feststellt, der tatsächlich existiert. Das Risiko eines Typ-II-Fehlers, bei dem eine tatsächlich existente Wirkung nicht entdeckt wird, kann anhand einer Power-Analyse bestimmt werden. Nach Cohen (1988) liegt das kleinste akzeptable Powerniveau einer Studie bei 0,8. Auf diesem Powerniveau ergibt sich eine Wahrscheinlichkeit von 20% für einen Typ-II-Fehler. Bei diesem vermag eine Studie existente Wirkungen nicht zu entdecken.

6.6.2 Gütebeurteilung von PLS-Modellen

Bei der Schätzung von Kausalmodellen mit dem Partial Least Squares-Verfahren kommt der Modellbeurteilung eine zentrale Bedeutung zu, wobei aufgrund fehlender empirischer Verteilungsannahmen im Vergleich zur Kovarianzstrukturanalyse die Anzahl möglicher Gütemaße wesentlich geringer ist (Ringle, 2004). Die traditionellen, parametrisch ausgerichteten Techniken für Signifikanztests zur Modellbeurteilung eignen sich nicht für die Partial Least Squares-Methode, weshalb Wold (1982) vorschlägt, statt dessen Tests zu verwenden, die dem verteilungsannahmenfreien Charakter des Verfahrens Rechnung tragen. Partial Least Squares-Modelle können nicht über Beurteilungsmaße zur Bestimmung der Anpassungsgüte von Kovarianzen evaluiert werden. Dafür müssen schätzungsorientierte Gütemaße, die zudem nichtparametrisch sind, herangezogen werden (Ringle, 2004). Beispiele für solche Maße zur Ergebnisbeurteilung sind das Bestimmtheitsmaß latenter endogener Variablen, der Stone-Geisser-Test zur Bestimmung der Schätzrelevanz latenter exogener Variablen für das ermittelte Bestimmtheitsmaß latenter endogener Variablen im Strukturmodell sowie die auf Fornell und Larcker zurückgehende faktoranalytische Bestimmung der durchschnittlich extrahierten Varianz. Ferner lässt sich die Stabilität der Schätzung über Verfahren wie Jackknifing oder Bootstrapping zur systematischen Veränderung der empirisch erhobenen Daten feststellen.

Zunächst erfolgt eine Beurteilung des Strukturmodells, dann die Beurteilung reflektiver und formativer Messmodelle, abschließend wird eine Beurteilung des Gesamtmodells vorgenommen.

6.6.2.1 Beurteilung des Strukturmodells

Da PLS keine Verteilungsannahmen zugrunde liegen, müssen nicht parametrische Tests herangezogen werden (Chin, 1998, S. 316). Dabei sind neben den Bestimmtheitsmaßen die Parameter des Modells zu evaluieren. Darüber hinaus sollen die Stabilität und Prognosefähigkeit des jeweiligen Erklärungsmodells untersucht werden. Zur Validierung des Modells können grundsätzlich das Resampling- und das Data-Splitting-Verfahren herangezogen werden.

Bestimmtheitsmaß R^2 : Für das Strukturmodell eines mittels Partial Least Squares-Verfahren geschätzten Kausalmodells lassen sich die vom statistischen Verfahren der Regressionsanalyse bekannten Bestimmtheitsmasse R^2 der abhängigen latenten Variablen des Strukturmodells ermitteln. Die latente endogene Variable ist in einem solchen multiplen linearen Regressionsmodell die abhängige Variable, während die latenten exogenen Variablen die unabhängigen Variablen darstellen und das Bestimmtheitsmaß den über die lineare Regressionsgleichung bestimmten Anteil der erklärten Varianz an der Gesamtvarianz angibt. Sowohl zur Beurteilung des Bestimmtheitsmaßes R^2 als auch für die geschätzten Werte der Regressionskoeffizienten in Partial Least Squares-Modellen liegen unterschiedliche Empfehlungen vor. Chin (1998, S.323) bezeichnet ein Bestimmtheitsmaß von 0,67 als „substantiell“, während er Ergebnisse in Höhe von 0,33 und 0,19 als „durchschnittlich“ bzw. „schwach“ einstuft. Pfadkoeffizienten in einer Höhe ab 0,1 werden von Lohmöller (1989, S.60f.) in Partial Least Squares-Modelle aufgenommen, während Chin (1998, S.324f.) signifikante Werte erst ab 0,2 feststellt.

Schätzrelevanz Q^2 : Die Bestimmung der Schätzrelevanz Q^2 stützt sich auf die Nutzung der Ergebnisse der von Stone und Geisser (Geisser, 1974) entwickelten Technik zur Wiederverwertung von Daten (bzw. sample reuse technique). Die Anwendung dieses Vorgehens auf die Partial Least Squares-Methode erfolgt durch die *Blindfolding-Prozedur*, über die Teile der empirisch erhobenen Daten für einen bestimmten Block manifester Variablen unterdrückt und anschließend mit Hilfe der Partial Least Squares-Ergebnisse geschätzt werden (Chin, 1998, S.317). Mit Hilfe der Schätzrelevanz Q^2 kann überprüft werden, wie gut durch das Model und seine Parameter empirisch erhobene Daten ohne Verlust an Freiheitsgraden rekonstruiert werden können. Anhand der *Blindfolding-Prozedur* wird hierfür die Summe der quadrierten Fehler für die geschätzten Werte (E) ebenso wie die Summe der quadrierten Fehler für den Durchschnittswert der Schätzung ermittelt (O). Mit Hilfe der nachfolgenden Gleichung (6) lässt sich Q^2 als Maß für die Schätzrelevanz eines Blockes manifester Variablen berechnen (Chin, 1998). D bezeichnet den Abstand zwischen zwei auszulassenden und daraufhin zu schätzenden Fällen.

$$\text{Schätzrelevanz } Q^2 = 1 - \frac{\sum_D E_D}{\sum_D O_D} \quad (6)$$

Ist Q^2 größer als Null, so hat das Modell eine Schätzrelevanz. Bei einem Q^2 kleiner als Null liegt eine fehlende Schätzrelevanz eines Blocks manifester Variablen vor und die darüber erfolgte Bestimmung einer latenten Variable ist als unsicher anzusehen (Ringle, 2004).

Effektstärke f^2 : Zudem lässt sich anhand der Effektstärke f^2 untersuchen, ob eine unabhängige (exogene) latente Variable einen substantiellen Einfluss auf abhängige (endogene) latente Variablen ausübt (Chin, 1998, S.316f.). Die Effektstärke f^2 beschreibt den Quotienten aus der R^2 Differenz zwischen dem modifizierten und dem vorangehenden Modell sowie 1 minus R^2 des vorangehenden Modells, also das Verhältnis von zusätzlicher erklärter Varianz und vormals nicht erklärter Varianz (Gefen, Straub & Boudreau, 2000). Entsprechend der für die multiple Regressionsanalyse operationalisierten Definition von Cohen (1988, S. 412ff.) gelten f^2 -Werte von 0,02, 0,15 und 0,35 als Beurteilungsmaß dafür, ob eine exogene latente Variable einen geringen, mittleren oder großen Einfluss auf eine zu ihr in Beziehung stehende endogene Variable ausübt.

Neben der Überprüfung der latenten Konstrukte im Strukturmodell muss festgestellt werden, ob es sich bei deren zueinander bestehenden Beziehungen bezüglich Richtung und Einflussstärke um statistisch signifikante Ergebnisse handelt. Hierfür können Resampling-Ansätze herangezogen werden, zu denen das *Jackknifing* und das *Bootstrapping* gehören, zwei nichtparametrische Verfahren, mit denen sich die Qualität von Partial Least Squares-Schätzergebnissen beurteilen lässt, ohne bestimmte Verteilungsannahmen zu treffen (Bollen & Stine, 1993). Beide Methoden lassen sich dadurch kennzeichnen, dass eine bestimmte Anzahl an Fällen aus den empirisch erhobenen Daten unterdrückt wird, um anschließend das Modell zu schätzen. Mittels eines t-Tests kann die jeweilige Signifikanz der Schätzergebnisse bestimmt werden. Einer solchen Überprüfung von über das Partial Least Squares-Verfahren geschätzter Kausalmodelle fällt eine zentrale Bedeutung zur Feststellung des Vorliegens robuster Parameterergebnisse und damit eines stabilen, reliablen Gesamtmodells zu (Ringle, 2004).

Generalisierbarkeit der Ergebnisse: Die Prognosegüte des Modells kann mit Hilfe von Data-Splitting-Verfahren überprüft werden, bei denen die Modellschätzung und dessen Validierung voneinander getrennt werden. Neben der Möglichkeit der Erhebung zweier unabhängiger Samples bietet es sich aus ökonomischen Gründen an, den vorhandenen Datensatz zu teilen. Hierbei wird zumeist eine zufällige Zuordnung der Fälle zu einem Holdout-Sample vorgenommen, wobei die Größe des Validierungs-Samples circa 25% der Daten betragen sollte (Steckel & Vanhonacker, 1993). Mit Hilfe der Modell-Parameter und den Daten des Holdout-Samples werden pro Fall die Werte der abhängigen latenten Variablen berechnet. Abschließend erfolgt die Überprüfung der Korrelation der errechneten mit den beobachteten Werten. Je höher die Korrelation r ausfällt, desto höher ist der Anteil der erklärten Varianz (r^2). Korrelationen von 0,1, 0,3 bzw. 0,5 können als gering, mittel und hoch bezeichnet

werden. Mit abnehmender Differenz zwischen R^2 und r^2 , nimmt die Prognosekraft des Modells zu (Chin & Todd, 1995).

6.6.2.2 Beurteilung reflektiver Messmodelle

Zur Beurteilung reflektiver Messmodelle werden die Maße zur Beurteilung der Indikator- bzw. Faktorreliabilität sowie der durchschnittlich erfassten Varianz eingesetzt (Homburg & Pflesser, 2000a, S. 420 ff.).

Die *Reliabilität eines reflektiven Indikators* kann anhand seiner Ladung auf das ihm zugeordnete Konstrukt beurteilt werden (Hulland, 1999). Bei einer Ladung (λ) über 0,707 kann ein Indikator im Allgemeinen als reliabel angesehen werden, da in diesem Fall mehr als 50% der Varianz des Indikators durch das zu messende Konstrukt erklärt werden. Die Signifikanzen der Ladungen lassen sich mittels t-Werte abschätzen, wobei der t-Wert $> 1,66$ sein sollte (Herrmann, Huber & Kressmann, 2004). Zur Berechnung von t-Werten für die geschätzten Parameter dienen bei PLS die beiden Prozeduren Bootstrapping und Jackknifing, wobei das Bootstrapping aufgrund eines geringeren Standardfehlers dem Jackknifing vorzuziehen ist (Efron & Gong, 1983). Ergänzend sollten die Kommunalitäten herangezogen werden (Fornell & Cha, 1994, S. 68 f.). Diese geben an, in welchem Maße die Varianz eines Indikators durch alle Faktoren erklärt wird.

Die *Faktorreliabilität* ist ein Maß für die Beurteilung der Eignung eines Faktors zur Erklärung des zu ihm in Beziehung stehenden Blocks reflektiver Indikatorvariablen. Mit Hilfe der *durchschnittlich erfassten Varianz (AVE)* nach Fornell und Larcker wird überprüft, wie hoch der durch einen Faktor erklärte Varianzanteil manifester Variablen im Verhältnis zum Anteil des Messfehlers ist. Die Werte beider Gütemaße liegen zwischen 0 und 1. Werte unter 0,6 für die Faktorreliabilität und unter 0,5 für die durchschnittlich erfasste Varianz werden als nicht mehr ausreichend angesehen (Homburg & Baumgartner, 1998).

Diskriminanzvalidität: Für mittels Partial Least Squares-Verfahren geschätzte reflektive Messmodelle ist außerdem eine Überprüfung der Diskriminanzvalidität erforderlich (Fornell & Cha, 1994, S.69). Fornell und Larcker (1981) schlagen diesbezüglich vor, dass die durchschnittlich erfasste Varianz latenter Variablen größer sein muss als die quadrierten Korrelationen zwischen den latenten Variablen.

6.6.2.3 Beurteilung formativer Messmodelle

Zur Güteprüfung formativer Messmodelle ist zunächst eine Prüfung auf *Multikollinearität* zwischen den formativen Indikatoren vorzunehmen, um Verzerrungen in den Parameterschätzungen zu vermeiden. Zur Feststellung von *Multikollinearität* können mehrere Kriterien herangezogen werden. So deuten Pearson'sche Korrelationskoeffizienten der Indikatoren mit Werten von größer als 0,5 auf ernsthafte Multikollinearität zwischen den Indikatoren hin und ziehen eine Anpassung des Messmodells nach sich (Backhaus et al. 2003, S. 89). Da die

Korrelationskoeffizienten jedoch nur paarweise Abhängigkeiten zwischen den Indikatoren messen, kann auch bei niedrigen Werten eine hohe Multikollinearität im Messmodell bestehen (Backhaus et al. 2003, S. 90). Darüber hinaus lässt sich die Multikollinearität zwischen den formativen Indikatoren mit Hilfe des Variance Inflation Factors untersuchen. Dieser beschreibt den Kehrwert der Toleranz und basiert auf dem Varianzanteil eines Indikators, den die übrigen Konstruktindikatoren erklären können. Ein Wert größer als 10 deutet auf hohe Multikollinearität hin (Chatterjee & Price, 1977, S.182; Gujarati, 2003, S. 362), wobei nach Schneider (2006, S.193) bereits VIF-Werte über 2 tendenziell als problematisch gewertet werden können.

Die Schätzung eines formativen Messmodells kann in einem sogenannten *Zwei-Konstrukt-Modell* erfolgen (Bollen, 1989; Diamantopoulos & Winklhofer, 2001). Hierbei wird die formative LV als Prädiktor (exogene Variable) zu einer endogenen reflektiven LV in Beziehung gesetzt.

Zur Gütebeurteilung auf Strukturmodellebene sind zwei Parameter heranzuziehen. Zum einen ist der Pfadkoeffizient β , der die Stärke des Zusammenhangs zwischen der exogenen formativen LV und der endogenen reflektiven LV angibt, bezüglich Höhe und Signifikanz zu prüfen. Die Signifikanzprüfung lässt sich mit Hilfe einer Resampling-Prozedur wie Bootstrapping vornehmen (Herrmann, Huber & Kressmann, 2004). Zum anderen ist der Determinationskoeffizient R^2 für die nachgeordnete reflektive LV heranzuziehen. Dieser benennt den Anteil aufgeklärter Varianz der nachgeordneten Kriteriumsvariable und sollte bei mindestens $R^2 \geq 0,3$ liegen (Hermann et al., 2006).

Zur Gütebeurteilung auf Messmodellebene werden die Gewichte der formativen Indikatoren in PLS, wie bereits im Falle des Pfadkoeffizienten β angeführt, mit Hilfe einer Resampling-Prozedur wie Bootstrapping anhand des t-Werts auf Signifikanz überprüft (Herrmann, Huber & Kressmann, 2004). Bei einem t-Wert > 2 erweist sich der Einfluss des jeweiligen Indikators als signifikant ($\alpha = 0,05$). Erweist er sich nicht als signifikant, kann der zugehörige Indikator aus dem Messmodell entfernt werden, da sein Beitrag zur Erklärung der LV zu gering ist. Bei der Entscheidung zur Elimination ist jedoch abzuwägen, ob das Entfernen des Indikators aus theoretischer, inhaltlicher Sicht als gerechtfertigt angesehen werden kann (Christophersen & Grape, 2006).

6.7 Logistische Regression

Die binomiale logistische Regression (LR) wird verwendet, wenn als abhängige Variable eine nominale Variable mit zwei Ausprägungen vorliegt. Das Ziel ist die Schätzung der Regressionsparameter der unabhängigen Variablen sowie die Analyse ihrer Wirkungsbeziehungen auf die Eintrittswahrscheinlichkeit der Ausprägungen der nominal skalierten abhängigen Variablen (Backhaus et al. 2003, S.453ff.). Da die Prädiktorvariablen mit Hilfe von siebenstufigen Lickert-Skalen erhoben wurden, unterscheiden sich die Art der

Daten auf unabhängiger und abhängiger Variablenseite deutlich voneinander. Die Gefahr eines Common Method Bias wird dadurch gemindert (vgl. Podsakoff et al., 2003).

Mit Hilfe der LR soll das nominalskalierte Kriterium Auswahlentscheidung mit den $j = 2$ Ausprägungen *Auswahl Website 1* bzw. *Auswahl Website 2* vorhergesagt werden⁷. Zur Bestimmung der prädiktiven Validität des Modells wird analysiert, inwiefern eine Vorhersage der verhaltensbasierten Variable Auswahlentscheidung auf Basis der verwendeten ufosV2a-Skalen möglich ist (Hypothesen 7a und 7b). Des Weiteren erfolgt die Überprüfung, inwiefern sich der im Modell angenommene Zusammenhang zwischen der Nutzungsintention und der Auswahlentscheidung gemäß Hypothese 19 bestätigt.

6.7.1 Voraussetzungen der LR

Die Anwendung einer LR ist an bestimmte Voraussetzungen gebunden. So sollte die Anzahl der Beobachtungsfälle je Kategorie der abhängigen Variable mindestens $n = 25$ Fälle betragen (Hosmer & Lemeshow, 2000, S. 339 ff.). Außerdem ist zu beachten, dass keine zu hohe Multikollinearitäten zwischen den unabhängigen Variablen vorliegen, um Verzerrungen in den Parameterschätzungen zu vermeiden (Menard, 2001, S. 75 ff.).

6.7.2 Gütekriterien der LR

Zur Beurteilung der Modell-Güte können der *Likelihood-Ratio-Test*, die Pseudo- R^2 -Statistiken von *McFadden*, *Cox und Snell* und *Nagelkerke* sowie die Analyse der *Klassifikationsmatrix* herangezogen werden (Backhaus et al. 2003, S. 437 ff.; Hosmer & Lemeshow 2000, S. 144 ff.).

Beim *Likelihood-Ratio-Test* wird die globale Nullhypothese geschätzt, alle Regressionsparameter im untersuchten Modell seien Null (vgl. Rohrlack, 2006, S. 208). Zur Bestimmung des LR-Werts werden ein vollständiges Modell unter Berücksichtigung aller erklärenden Variablen und ein Nullmodell, das nur die Konstante enthält, miteinander verglichen und deren LogLikelihood-Werte (LL) voneinander abgezogen. Sobald der LR-Wert, den Wert der χ^2 -Verteilung für j Freiheitsgrade entsprechend der Anzahl der unabhängigen Variablen übersteigt, kann von einem bedeutenden Einfluss der unabhängigen Variablen ausgegangen werden.

Das *Pseudo- R^2* nach *McFadden* definiert der Autor als eins minus Quotient aus vollständigem Modell und Nullmodell. Der Wertebereich liegt zwischen null für ein schlechtes Modell und annähernd eins für ein gutes Modell.

⁷ Das nominalskalierte Kriterium Auswahlentscheidung verfügt über $j=3$ Ausprägungen *Auswahl Website 1*, *Auswahl Website 2* sowie *keine Auswahl*. Die Ausprägung *keine Auswahl* wird aufgrund eines zu geringen $N=12$ in der LR nicht berücksichtigt.

$$\text{McFadden } R^2 = 1 - \frac{LL_v}{LL_0} \quad (6)$$

LL_v LogLikelihood-Wert des vollständigen Modells

LL_0 LogLikelihood-Wert des Nullmodells

McFadden- R^2 steigt auf annähernd eins an, wenn das vollständige Modell und das Nullmodell stark auseinander fallen. In diesem Falle geht der Quotient gegen null. Es ist davon auszugehen, dass die aufgenommenen Variablen einen bedeutenden Einfluss haben.

Cox und Snell haben ein weiteres Gütemaß entwickelt, das in Gleichung (7) wiedergegeben ist. Der Wertebereich liegt ebenfalls zwischen null und annähernd eins.

$$\text{Cox und Snell } R^2 = 1 - \left[\frac{L_0}{L_v} \right]^{\frac{2}{n}} \quad (7)$$

L_0 Likelihood-Wert des Nullmodells

L_v Likelihood-Wert des vollständigen Modells

n Stichprobenumfang

Da das Maß praktisch nicht den Wert eins annehmen kann hat *Nagelkerke* es erweitert zu:

$$\text{Nagelkerke } R^2 = \frac{1 - \left[\frac{L_0}{L_v} \right]^{\frac{2}{n}}}{1 - \left(\frac{L_0}{L_v} \right)^{\frac{2}{n}}} \quad (8)$$

Werte für die Pseudo- R^2 Maße über 0,2 lassen auf ein *akzeptables*, Werte über 0,4 auf ein *gutes* Modell schließen. Im Falle des R^2 nach Nagelkerke kann ein Wert über 0,5 als *sehr gut* interpretiert werden. Nach Backhaus et al. ist diesem Gütemaß der Vorzug zu geben (2003, S. 440 f.).

Bei der Analyse der *Klassifikationsmatrix* werden die beobachteten mit den modellierten Klassifikationen verglichen. Der Anteil der korrekt zugeteilten Fälle spiegelt die Güte des Modells wider und gibt an, wie hoch der Anteil der AV-Werte ist, die durch das Modell korrekt vorhergesagt werden. Er sollte höher sein als eine zufällig richtige Einsortierung anhand eines Basismodells ohne Prädiktoren. Zum Vergleich bietet sich das „proportional chance criterion“ (PCC) an (Morrison, 1969). In Tabelle 6-3 werden die Maße zur Beurteilung der Modellgüte und deren Wertebereiche dargestellt.

Tab. 6-3: Gütemaße der logistischen Regression

<i>Gütemaß</i>	<i>Wertebereich</i>
Likelihood-Ratio-Test	LR-Wert größer χ^2 -Wert
McFadden-R ²	größer 0,2 (akzeptabel); größer 0,4 (gut)
Cox und Snell-R ²	größer 0,2 (akzeptabel); größer 0,4 (gut)
Nagelkerke-R ²	größer 0,2 (akzeptabel); größer 0,4 (gut); größer 0,5 (sehr gut)
Klassifikationsmatrix	Anteil korrekter Klassifikationen größer PCC

Quelle: Eigene Darstellung in Anlehnung an Backhaus et al. 2003, S. 447f.

6.8 Weitere Testverfahren

6.8.1 t-Test

Die Berechnung eines t-Tests setzt eine Normalverteilung der Messwerte in der zugrunde liegenden Population voraus. Allerdings reagiert der Test bei größeren Stichproben relativ robust gegenüber einer Verletzung der Normalverteilungsannahme. In diesem Falle kann ein schwacher Effekt mit einer Teststärke von 80% ($1-\beta$ 0,8) und einer Irrtumswahrscheinlichkeit von α 0.05 bei einem Stichprobenumfang von $N > 310$ identifiziert werden (vgl. Bortz & Döring, 2005, S.139).

Um zu prüfen, inwiefern die Reihenfolge der reflektiven bzw. formativen Usability-Skalen sowie ihre Positionierung in der Abfrage einen Einfluss auf die Ausprägung der AV Nutzungsintention hat, wird ein t-Test für unabhängige Stichproben berechnet (vgl. Bortz & Döring, 2005, S. 140). Dabei wird geprüft, ob sich signifikante Mittelwertunterschiede in der AV Nutzungsintention für unterschiedliche Alternativen bzgl. der Abfolge und Positionierung der reflektiven und formativen Usability-Skalen zeigen. Das Signifikanzniveau wird bei α 0.05 festgelegt. Desweiteren ist bei Durchführung eines t-Tests auf Mittelwertunterschiede von Bedeutung, dass die betrachtete Variable in den miteinander zu vergleichenden Teilgruppen die gleiche Varianz aufweist. Allerdings ist Varianzgleichheit keine zwingende Voraussetzung für die Anwendung des t-Tests. Vielmehr stehen für den Fall gleicher bzw. ungleicher Varianzen unterschiedliche Teststatistiken zur Verfügung.

6.8.2 Einfaktorielle Varianzanalyse

Mit Hilfe einer mehrdimensionalen einfaktoriellen Varianzanalyse (ANOVA) soll überprüft werden, inwieweit die eingesetzten reflektiven Skalen geeignet sind, zwischen unterschiedlichen Websites zu differenzieren. Die Berechnung einer ANOVA setzt eine Normalverteilung der Werte sowie Varianzgleichheit in der Grundgesamtheit in allen Gruppen voraus (vgl. Backhaus et al., 2003, S. 150). Allerdings erweist sich das Verfahren bei einer Verletzung dieser Annahmen als robust. Die Testung auf Varianzhomogenität in den verschiedenen Gruppen erfolgt durch die Berechnung von Levene-Tests.

6.8.3 Levene-Test

Der Levene-Test überprüft für verschiedene Wertegruppen die Nullhypothese, die Varianzen seien in der Grundgesamtheit für alle betrachteten Gruppen gleich (vgl. Dayton, 1970, S. 34 f.). Er reagiert relativ robust gegenüber einer Verletzung der Normalverteilungsvoraussetzung (Bortz & Döring, 2005, S. 286). Als statistisches Prüfmaß wird der F-Wert berechnet, mit dessen Hilfe getestet werden kann, mit welcher Wahrscheinlichkeit eine Gleichheit der Varianzen in der Grundgesamtheit vorliegt (Brosius, 2006, S. 402). Im Falle eines signifikanten Testergebnisses ist die Nullhypothese, die Varianzen in den Gruppen seien in der Grundgesamtheit gleich, abzulehnen.

7. Ergebnisse

In den nachfolgenden Abschnitten werden die Ergebnisse der Untersuchung dargestellt. Die Darstellung folgt dabei in der Reihenfolge nicht zwingend dem Ablauf der Analysen, sondern orientiert sich inhaltlich an den in Kapitel drei dargestellten Hypothesen. Die Abschnitte können anhand der Tabelle 6.1 den einzelnen Hypothesen zugeordnet werden.

7.1 Vorbereitende Datenanalysen

Vorbereitend zur Überprüfung der Hypothesen wurden Datenanalysen durchgeführt, die in den Abschnitten 7.1.1 bis 7.1.5 dargestellt werden.

7.1.1 Zusammenfassung der Bewertungen

Für alle $N = 256$ Teilnehmer liegen je Datensatz die Bewertungen von zwei Websites vor. Entsprechend findet sich jeder Indikator zur Erfassung der Modellvariablen in jedem Datensatz zweimal. Für die Auswertungsschritte, in denen die Information der relativen Bewertung der Websites nicht benötigt wird, erscheint es praktikabel, jeden Fall aufzuteilen. Damit verdoppelt sich die Anzahl der Datenfälle auf $N = 512$ mit einer bewerteten Website je Fall.

Nachfolgend wird für jeden Analyseschritt angezeigt, welcher Datensatz bei der Berechnung zugrunde gelegt wurde (*Datensatz mit einer vs. zwei Bewertungen je Fall*).

7.1.2 Analyse und Behandlung fehlender Werte

Nachfolgend werden die Analyse und der Umgang mit fehlenden Werten dargestellt. Im ersten Schritt werden die Fälle identifiziert, die aufgrund eines zu hohen Missing-Anteils von der weiteren Untersuchung ausgeschlossen werden (Abschnitt 7.1.2.1). Daran anschließend wird der zugrunde liegende Fehlendmechanismus untersucht (Abschnitt 7.1.2.2) und das Vorgehen zur Ersetzung fehlender Indikatorwerte dargestellt (Abschnitt 7.1.2.3).

7.1.2.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte

Der Anteil fehlender Werte je Fall wird anhand des Datensatzes mit zwei Bewertungen je Fall bestimmt. In Tabelle 7-1 wird eine Aufteilung der Teilnehmer nach ihrem Anteil an fehlenden Werten dargestellt. Insgesamt hat ein großer Teil von 84,79% der Teilnehmer mehr als 90% aller Items im Verlaufe des Versuchs beantwortet. 28,90% haben mindestens 99% aller Items beantwortet; dies bedeutet im Umkehrschluss bei 263 Items dass maximal zwei Items nicht beantwortet wurden. Lediglich sieben Fälle weisen mehr als 30% fehlende Werte auf und werden entsprechend Abschnitt 6.1.1 von der weitergehenden Analyse ausgenommen. Die Vpn-Codes der Teilnehmer werden in Tabelle E-1 im Anhang E aufgelistet. Es verbleiben

somit $N = 256$ Fälle im Datensatz mit zwei Bewertungen je Fall, bzw. $N = 512$ Fälle im Datensatz mit einer Bewertung je Fall.

Tab. 7-1: Missing Data im Datensatz - Aufteilung der Teilnehmer

<i>Anteil Missing Data pro Fall</i>	<i>n</i>	<i>Kumuliertes n</i>	<i>Gültige Prozente</i>	<i>Kumulierte Prozente</i>
0 bis 1%	76	76	28,90%	28,90%
>1 bis 3%	36	112	13,69%	42,59%
>3 bis 5%	54	166	20,53%	63,12%
>5 bis 10%	57	223	21,67%	84,79%
>10 bis 15%	30	253	11,41%	96,20%
>15 bis 30%	3	256	1,14%	97,34%
>30 bis 52%	7	263	2,66%	100,00%

Anmerkung: Anzahl Teilnehmer als absolute Häufigkeit (n). Lesebeispiel erste Zeile: Bei 76 Teilnehmern (28,90% der Stichprobe) liegt der Anteil fehlender Werte pro Fall zwischen 0 und 1%.

7.1.2.2 Analyse des Fehlendmechanismus

Um den zugrundeliegenden Fehlendmechanismus zu untersuchen, werden zwei methodische Ansätze verfolgt (siehe Abschnitt 6.1.2). Zunächst wird geprüft, inwiefern zwischen der abhängigen Variable Nutzungsintention und dem Fehlendmechanismus ein Zusammenhang besteht. Hierfür werden die bivariaten Korrelationen zwischen dem fallweisen Anteil fehlender Werte und dem Skalenwert der Nutzungsintention anhand des Datensatzes mit zwei Bewertungen je Fall berechnet. Für den Fall der ersten Bewertung wird ein nicht signifikanter Korrelationskoeffizient von $r < 0,01$ berechnet, für den Fall der zweiten Website ein auf dem Niveau $\alpha = 0,05$ signifikanter Korrelationskoeffizient von $r = 0,156$.

In einem zweiten methodischen Ansatz wird eine Berechnung des *MCAR-Test nach Little* durchgeführt. Im Falle der ersten Bewertung erweist sich eine Signifikanz von $\alpha = 0,89$ (t-Wert 0,146; $df = 89$), im Falle der zweiten Bewertung von $\alpha = 0,39$ (t-Wert 0,857; $df = 101$). Nach Tabachnik und Fidell (2006, S. 63) kann entsprechend ein zufälliges Auftreten fehlender Werte unterstellt werden. Auf Basis der rechnerischen Analysen sowie der Ausführungen zu den Ursachen fehlender Werte (siehe Abschnitt 6.1.2) kann geschlossen werden, dass ein MCAR- oder ein MAR-Mechanismus vorliegt.

7.1.2.3 Ersetzen fehlender Werte

Wie in Abschnitt 6.1.3 dargestellt, liegt es bei den reflektiven Konstrukten nahe, fehlende Werte für deren Indikatoren durch den Mittelwert der verfügbaren Werte der übrigen Indikatoren derselben Skala zu ersetzen. Dieses Verfahren der Mittelwertergänzung wird für die reflektiven Indikatoren der Konstrukte Usability (ufosV2ra-Skala), Vertrauen, Nutzervergnügen, Ästhetik, Reputation, Involvement, Nutzerzufriedenheit sowie Nutzungsintention angewandt.

Nach Ersetzen der fehlenden Werte für die reflektiven Indikatoren verbleiben 29 manifeste Variablen im Datensatz, die fehlende Werte aufweisen. Der Anteil an Missing Data liegt mit

578 Zellen fehlender Werte von insgesamt 15.872 Zellen bei 3,64% (Datensatz mit zwei Bewertungen je Fall und $N = 256$). Das oben angewandte Verfahren der Mittelwertergänzung kann zur Ersetzung der fehlenden Werte bei den formativen Indikatoren der ufosV2fa-Skala nicht genutzt werden, da bei formativen Variablen nicht zwingend von einer hohen Korrelation zwischen den Indikatoren ausgegangen werden kann. Die Ersetzung der verbleibenden fehlenden Werte erfolgt deshalb mit Hilfe der multiplen Imputationsmethode unter Anwendung der Software Norm (Schafer, 1997). Da davon auszugehen ist, dass ein MCAR-Ausfall oder ein MAR-Ausfall vorliegt und sich das Verfahren als robust gegenüber Verletzungen der Verteilungsannahmen erweist (Schafer & Graham, 2002), liegen die Voraussetzungen zur Berechnung vor. In Tabelle E-5 im Anhang E sind alle Indikatoren aufgelistet, die in den Imputationsdatensatz aufgenommen werden. Grundlage der Berechnung der multiplen Imputation ist der Datensatz mit einer Bewertung pro Fall ($N = 512$)⁸.

7.1.3 Verteilungen der erhobenen Indikatoren

Für alle erhobenen Indikatoren werden die Verteilungen anhand des Datensatzes mit einer Bewertung je Fall untersucht. Es werden für alle Indikatoren signifikante Ergebnisse im Kolmogorov-Smirnov-Test ausgewiesen ($p < .001$), sodass die Normalverteilungshypothese zurückgewiesen werden kann. Brosius (2004) weist jedoch darauf hin, dass eine Signifikanz des Kolmogorov-Smirnov-Tests alleine nicht zwangsläufig ausreicht, um eine Variable als nicht normalverteilt auszuweisen. Es werden daher zusätzlich die Werte für Schiefe (Sch) und Exzess (Ex) für die Verteilungen sowie die entsprechenden z-Werte berechnet (siehe Tabelle E-7 im Anhang E). Dabei wird deutlich, dass für keinen der Indikatoren die Normalverteilungshypothese aufrechterhalten werden kann. Auf eine Transformation der Daten zur Herstellung von Normalverteilungen wird verzichtet. Zum einen können transformierte Daten häufig schlechter interpretiert werden als nicht transformierte (vgl. Tabachnik & Fidell, 2006, S. 86 f.) und zum anderen setzen die wesentlichen Auswertungen anhand von PLS und LR nicht zwingend normalverteilte Variablen voraus.

7.1.4 Prüfung auf Vorliegen eines Common Method Bias

Zur statistischen Untersuchung inwiefern ein Common Method Bias im Datensatz vorliegt, wird der *Harman's Single-Factor Test* eingesetzt (siehe Abschnitt 6.4). Grundlage ist der Datensatz mit einer Bewertung je Fall ($N = 512$). Es wird über alle Indikatoren der Untersuchung eine Hauptkomponentenanalyse (HKA) ohne Rotation der Faktoren durchgeführt. Dabei werden die Indikatoren zu den Konstrukten Usability (formative und reflektive Skalen ufosV2fa bzw. ufosV2ra), Nutzerzufriedenheit, Nutzungsintention, Nutzervergnügen, Ästhetik, Vertrauen, Involvement und einige weitere Einzelindikatoren in

⁸ Der in der Multiplen Imputation berechnete EM (Expectation Maximization) Algorithmus konvergiert bei einem standardmäßigen Konvergenzkriterium von 0,0001 nach 128 Iterationen.

die Analyse einbezogen. Bei der Berechnung der HKA werden 13 Komponenten mit einem Eigenwert $\lambda > 1$ extrahiert. Die Eigenwerte und Gesamtvarianz der Faktorenlösung werden in Tabelle E-6 in Anhang E dargestellt. Die erste Komponente klärt einen maßgeblichen Anteil an der Gesamtvarianz von 35,95% auf. Die zugehörige Komponentenmatrix wird in Tabelle E-8 im Anhang E dargestellt. 52 von 60 Indikatoren weisen eine Ladung $a_{il} \geq 0.3$ auf dem ersten Faktor auf. Die zweite Komponente trägt weitere 7,52% zur Aufklärung der Modellvarianz bei. Insgesamt tragen die Komponenten 2 bis 13 36,21% zur Erklärung der Modellvarianz bei. Auf Grundlage der Ergebnisse des Harman's Single-Factor Test kann davon ausgegangen werden, dass kein kritischer Common Method Bias im Datensatz vorliegt.

7.1.5 Identifikation problematischer Fälle

Wie in Abschnitt 6.2 dargestellt, werden zwei Analysen vorgenommen, um problematische Fälle zu identifizieren. Einerseits werden fallweise die Varianzen im Antwortverhalten der Teilnehmer untersucht und andererseits die Konsistenz des Antwortverhaltens bei der Beantwortung negativ gepolter Items überprüft. Grundlage ist der Datensatz mit zwei Bewertungen je Fall.

Bei der Analyse der Varianzen im Antwortverhalten werden $N = 25$ Teilnehmer auffällig. Bemerkenswert ist, dass die Anzahl der mit einer niedrigen Varianz als „Durchklicker“ identifizierten Teilnehmer bei der zweiten Website deutlich höher ausfällt. In diesem Fall zeigt sich bei 19 Teilnehmern eine Streuung über alle Items von $SD \leq 1$, während bei der Bewertung der ersten Website lediglich sieben Teilnehmer auffällig werden. Einer der Teilnehmer zeigt sowohl bei der Bewertung der ersten als auch der zweiten Website eine Streuung über alle Items von $SD \leq 1$.

Um die Konsistenz des Antwortverhaltens bei der Beantwortung negativ gepolter Items zu überprüfen, werden die Abweichungen zwischen den Mittelwerten der positiv gepolten Usability-Indikatoren und den Werten für das negativ gepolte Item ur_{2ac} berechnet. Bei 15 Teilnehmern zeigt sich ein inkonsistentes Antwortverhalten. Tabelle E-9 im Anhang E gibt einen Überblick über die insgesamt $N = 40$ Teilnehmer, die in mindestens einer der beiden Analysen ein auffälliges Antwortverhalten zeigten⁹.

7.2 Reihenfolgeeffekte

Wie in Abschnitt 4.7 dargestellt, wurden in der vorliegenden Untersuchung potentielle Störeffekte aufgrund der Reihenfolge in der Itempräsentation unter mehreren Aspekten berücksichtigt. So wurde einerseits die Abfolge der reflektiven und formativen Usability-Skalen systematisch variiert und andererseits die Usability-Skalen in der Abfolge der Konstruktabfrage einmal am Anfang des Fragebogens im Anschluss an die AVs platziert

⁹ Die als auffällig identifizierten Fälle werden in Abschnitt 7.7.1 einer weiteren Überprüfung hinsichtlich ihrer Verwendbarkeit im Rahmend der Logistischen Regression unterworfen.

(Alternative I) und einmal am Ende des Fragebogens vor der Preiseinschätzung (Alternative II). Um zu prüfen, inwieweit die dargestellten Variationen in der Abfolge der Usability-Skalen Einfluss auf die AV Nutzungsintention nehmen, wurde für jede der vier in Tabelle 7-2 dargestellten Varianten ein t-Test für unabhängige Stichproben mit einem Datensatz mit zwei Bewertungen je Fall berechnet (siehe Abschnitt 6.8.1). Somit wird untersucht, ob sich signifikante Unterschiede in den Mittelwerten der AV Nutzungsintention bei Variation der Reihenfolge der Erhebung der Usability-Skalen ergeben. Die Mittelwerte der AV Nutzungsintention für die vier Varianten werden in den Tabellen E-2.1 bis E-2.4 im Anhang E dargestellt. Eine Voraussetzung des t-Tests ist die Normalverteilung der untersuchten Variable Nutzungsintention. Obwohl diese verletzt wird, wird dennoch eine Berechnung durchgeführt, da der t-Test diesbezüglich robust reagiert. Eine weitere Voraussetzung des t-Tests ist Varianzhomogenität. Diese wird für die Varianten drei und vier aufgrund signifikanter Ergebnisse der Levene-Tests ebenfalls verletzt (siehe Tabellen E-3.3 und E-3.4 im Anhang E). Für den Fall ungleicher Varianzen steht jedoch eine alternative Statistik zur Überprüfung der Signifikanz des Mittelwertunterschiedes in SPSS zur Verfügung. Die Durchführung der t-Tests weist für alle vier Varianten insignifikante Ergebnisse aus (siehe Tabellen E-3.1 bis E-3.4 im Anhang E). Somit kann gefolgert werden, dass die Abfolge und Platzierung der Usability-Skalen keinen Einfluss auf die AV Nutzungsintention hat.

Tab. 7-2: Abfolge der reflektiven und formativen Usability-Skalen

<i>Variante</i>	<i>Abfolge der usability-Skalen</i>	<i>Platzierung im Fragebogen</i>	<i>Levene-Test auf Varianzgleichheit</i>	<i>t-Test auf Mittelwertgleichheit</i>
1	Abfolge 1 vs. 2	Alternative I	n.s.	n.s.
2	Abfolge 1 vs. 2	Alternative II	n.s.	n.s.
3	Abfolge 1	Alternative I vs. II	sign.	n.s.
4	Abfolge 2	Alternative I vs. II	sign.	n.s.

Anmerkung: Abfolge 1: reflektiv – formativ, Abfolge 2: formativ – reflektiv; Alternative I: Usability-Skalen werden in der Abfolge der Konstrukte am Anfang des Fragebogens im Anschluss an die AV platziert; Alternative II: Usability-Skalen werden in der Abfolge der Konstrukte am Ende des Fragebogens vor der Preiseinschätzung platziert; Signifikanz auf dem Niveau von 0.05 (2-seitig; Sign.).

7.3 Erfassung der Variablen auf Prädiktorebene

In den folgenden Abschnitten 7.3.1 bis 7.3.3 werden die Ergebnisse der in Abschnitt 6.5 vorgestellten Prozedur zur Gütebeurteilung reflektiver Skalen nach Homburg und Giering (1996) dargestellt. Sämtliche Berechnungen werden auf Grundlage des Datensatzes mit einer Bewertung je Fall ($N = 512$) durchgeführt. Um zu prüfen, inwiefern die reflektiven Skalen geeignet sind, Unterschiede in den Merkmalsausprägungen der betrachteten Websites zu identifizieren, wird eine mehrdimensionale einfaktorielle Varianzanalyse berechnet (siehe Abschnitt 7.3.4).

7.3.1 Prüfung der Skalenstruktur mit Hilfe explorativer Faktorenanalyse

Die Überprüfung der Güte der reflektiven Skalen, die in der Untersuchung auf Prädiktorenebene herangezogen werden, erfolgt mit Hilfe des dreistufigen Vorgehens, das sich an den Empfehlungen von Homburg und Giering (1996) orientiert. Im ersten Schritt der Prozedur wird mit Hilfe der explorativen Faktorenanalyse (EFA) untersucht, inwiefern sich die theoretisch angenommene Struktur auf Prädiktorenebene bestätigen lässt. Hierfür werden mehrere rotierte Hauptachsenanalysen (HAA) unter Verwendung der Rotationstechnik *Direct Oblimin* berechnet, da dieses Verfahren korrelierende Faktoren unterstützt. Da Zusammenhänge zwischen den latenten Variablen aus theoretischer Sicht möglich erscheinen (siehe Abschnitt 2.11), wird der Delta-Wert mit δ^{-2} als Maß für eine mittlere angenommene Korrelation zwischen den Faktoren festgelegt (siehe Abschnitt 6.5.1).

Es werden die reflektiven Skalen zur Erfassung der Konstrukte Usability (ufosV2ra; Items ur1 bis ur8), Nutzungsvergnügen (vg1, vg2, vg3), Vertrauen (v1, v2, v3), Ästhetik (ae1, ae2) und Involvement (i1 bis i5) in die Analyse einbezogen. Die grundsätzlichen Voraussetzungen zur Berechnung einer HAA sind gegeben. Das Verfahren ist nicht an bestimmte Verteilungsannahmen gebunden, angesichts des siebenstufigen Likert-Antwortformates sind univariate Ausreißer auszuschließen und der Stichprobenumfang ist mit $N = 512$ als sehr gut zu bewerten (siehe Abschnitt 6.5.1.1). Kritisch anzumerken ist, dass zur Erhebung des Konstruktes Ästhetik lediglich zwei Items eingesetzt wurden. Nach Bühner (2004, S.157) sollten jedoch mehr Indikatoren pro Konstrukt in eine EFA einbezogen werden.

Im Verlauf der EFA werden mehrere HAAs berechnet, da sich die theoretisch angenommene Struktur zunächst nicht zeigt. Nach einer iterativen Elimination einzelner Items wird die angenommene Skalenstruktur bestätigt und eine angemessene faktorielle Validität der vier extrahierten Skalen erreicht. Tabelle 7-3 gibt einen Überblick über den Verlauf und die Ergebnisse der insgesamt fünf berechneten HAAs. Die Voraussetzungen zur Berechnung der HAAs sind für alle fünf Analysen gegeben. Sowohl die Werte für das KMO- als auch für das MSA-Maß sind ausreichend hoch; die Bartlett-Tests auf Sphärizität zeigen signifikante Ergebnisse.

Tab. 7-3: Übersicht über alle berechneten explorativen Faktorenanalysen zur Überprüfung der angenommenen Skalenstruktur

Nr. der HAA	Prüfung der Voraussetzungen zur Berechnung	Bewertung und vorgenommene Änderung der Skalen	Mustermatrix in Tabelle
1	KMO= 0,914; MSA-Werte mit Werten von 0,823 bis 0,965 „recht gut“ bis „fabelhaft“; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (ungefähres $\chi^2=4824,860$; $df=210$; $p < .001$) → Voraussetzungen zur Berechnung erfüllt*	Extraktion von vier statt fünf erwarteten Faktoren; Alle Kommunalitäten liegen bei $h_i^2 > 0,5$. Indikator ur7 hat hohe Nebenladung auf Faktoren 3 und 4 (0,275; 0,298). → Elimination von ur7	Tab. E-10.1 (Anhang E)
2	KMO= 0,908; MSA-Werte mit Werten von 0,811 bis 0,961 „recht gut“ bis „fabelhaft“; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (ungefähres $\chi^2=4589,20$; $df=190$; $p < .001$) → Voraussetzungen zur Berechnung erfüllt*	Extraktion von vier statt fünf erwarteten Faktoren; Alle Kommunalitäten liegen bei $h_i^2 > 0,5$. Indikatoren ae 1,2 und vg 1,2,3 laden auf den vierten Faktor. Item ae1 sowie Item ae2 weisen Nebenladungen auf den Faktor 1 (0,369 bzw. 0,205) auf. → Berechnung einer dritten HAA, in die nur die Items zum Nutzervergnügen und Ästhetik eingehen, um zu prüfen, inwiefern sich Eindimensionalität von Nutzervergnügen und Ästhetik bestätigt.	Tab. E-10.2 (Anhang E)
3	KMO= 0,804; MSA-Werte mit Werten von 0,764 bis 0,875 „mittelpfänglich“ bis „recht gut“; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (ungefähres $\chi^2=2038,78$; $df=10$; $p < .001$) → Voraussetzungen zur Berechnung erfüllt*	Es wird Eindimensionalität ausgewiesen. Alle Kommunalitäten liegen bei $h_i^2 > 0,5$. → Berechnung einer vierten HAA, in die nur die Items zur Usability und Ästhetik eingehen, um zu prüfen, inwiefern sich Eindimensionalität von Usability und Ästhetik bestätigt.	Tab. E-10.3 (Anhang E)
4	KMO= 0,937; MSA-Werte mit Werten von 0,838 bis 0,973 „recht gut“ bis „fabelhaft“; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (ungefähres $\chi^2=4664,34$; $df=36$; $p < .001$) → Voraussetzungen zur Berechnung erfüllt*	Es wird Eindimensionalität ausgewiesen. Alle Kommunalitäten liegen bei $h_i^2 > 0,5$. → Elimination der Indikatoren ae1 und ae2	Tab. E-10.4 (Anhang E)
5	KMO= 0,900; MSA-Werte mit Werten von 0,811 bis 0,956 „recht gut“ bis „fabelhaft“; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (ungefähres $\chi^2=4005,528$; $df=153$; $p < .001$) → Voraussetzungen zur Berechnung erfüllt	Extraktion von vier Faktoren entsprechend der erwarteten Anzahl; Alle Kommunalitäten liegen bei $h_i^2 > 0,5$.	Tab. 7-4

Anmerkung: * Als Einschränkung der Voraussetzungserfüllung ist die teilweise geringe Anzahl an Indikatoren für das Konstrukt Ästhetik zu sehen. Kaiser-Meyer-Olkin-Maß (KMO), Kommunalität (h_i^2), Eigenwert (λ_j), Measure of Sampling Adequacy (MSA), Freiheitsgrade (df), Signifikanzniveau (p), Chi-Quadrat-Wert (χ^2).

Eine erste HAA wird nicht als zufrieden stellend angesehen. Anstelle der erwarteten fünf Faktoren werden nur vier extrahiert. Alle Kommunalitäten liegen bei $h_i^2 > 0,5$. Allerdings hat der Indikator ur7 Nebenladungen auf die Faktoren 3 und 4 (siehe Mustermatrix in Tabelle E-10.1 im Anhang E). Das Item wird zur Berechnung der zweiten HAA eliminiert. In dieser Analyse laden die beiden Indikatoren zur Ästhetik gemeinsam mit den Indikatoren zum Nutzervergnügen auf den vierten Faktor, der 8,03% der Gesamtvarianz erklärt (siehe Mustermatrix in Tabelle E-10.2 in Anhang E). Die beiden Indikatoren zur Ästhetik weisen außerdem Nebenladungen auf den Faktor eins auf.

Um die Eindimensionalität des vierten Faktors zu prüfen, wird eine dritte HAA durchgeführt, in die lediglich die drei Indikatoren zum Nutzervergnügen sowie die beiden Indikatoren zur Ästhetik eingehen. Es wird Eindimensionalität ausgewiesen. Entgegen der ursprünglichen Erwartung zeigt sich somit ein eigenständiger Faktor „Ästhetik“ nicht. Des Weiteren wird aufgrund der Nebenladungen der Ästhetik-Indikatoren auf den ersten Faktor dieser auf Eindimensionalität geprüft. Hierfür wird eine vierte HAA berechnet, in die lediglich die sieben Usability- sowie die beiden Ästhetikindikatoren eingehen. Auch hier wird als Ergebnis der Analyse Eindimensionalität bestätigt. Es wird daher die Entscheidung getroffen, die beiden Indikatoren zur Ästhetik zu eliminieren. Nach Ausschluss dieser Items werden in der fünften HAA wie erwartet vier Faktoren extrahiert (siehe Tabelle 7-4). Auf Grundlage der Ergebnisse der berechneten Faktorenanalysen erweist sich die Skala uvosV2ra nach Ausschluss der Items ur7 sowie ae1 und ae2 als faktoriell valide. Damit kann die **Hypothese 2**, die Skala ufosV2ra sei gegenüber Skalen zur Erfassung weiterer Erfolgsfaktoren valide, bestätigt werden.

Tab. 7-4: Mustermatrix der fünften berechneten Hauptachsenanalyse

Faktor	1	2	3	4	h_i^2	MSA-Wert
Usability ur5	,827	,033	,043	,127	,775	,936
Usability ur4	,812	-,005	,038	,072	,705	,937
Usability ur3	,793	,027	-,016	,129	,693	,950
Usability ur2	,788	,010	,020	,043	,654	,933
Usability ur6	,775	-,012	,144	,231	,843	,925
Usability ur8	,692	-,058	,201	,186	,721	,953
Usability ur1	,654	-,029	,190	,192	,655	,959
Involvement i4	,064	,906	-,027	-,054	,819	,841
Involvement i1	-,038	,905	,084	,026	,806	,904
Involvement i3	-,006	,897	,102	,054	,825	,864
Involvement i2	-,049	,893	-,021	,031	,797	,870
Involvement i5	,023	,893	,029	-,020	,795	,876
Vertrauen v4	-,022	,084	,806	-,020	,538	,811
Vertrauen v2	,072	,024	,802	,026	,567	,844
Vertrauen v1	,047	-,035	,776	,078	,554	,844
Nutzervergnügen vg2	,076	-,003	,059	,861	,726	,831
Nutzervergnügen vg3	,098	,017	,106	,837	,727	,854
Nutzervergnügen vgl	,040	,008	-,043	,719	,493	,900

Tab. 7-5: Skalen- und Itemkennwerte der reflekt. Skalen auf Prädiktorebene (Fortsetzung)

Konstrukt/ Skalenkennwerte	Item- Code	M	SD	a_{ij}	% Mis- sing	p_m	r_{it}
Vertrauen $\alpha=0.89$ $M=5,01$ $SD=1,24$ Mittlere korr. Trennschärfe= 0,780 Erklärte Varianz: 81,47%	v1	5,07	1,42	0,860	13,88	0,724	0,785
	v2	4,81	1,39	0,872	12,55	0,687	0,791
	v4	5,15	5,15	0,889	7,03	0,736	0,765
Nutzervergnügen $\alpha=0.90$ $M=3,19$ $SD=1,56$ Mittlere korr. Trennschärfe= 0,804 Erklärte Varianz: 83,52%	vg1	3,29	1,56	0,901	4,75	0,470	0,727
	vg2	3,24	1,55	0,895	4,94	0,463	0,851
	vg3	3,03	1,57	0,873	3,61	0,433	0,834
Involvement $\alpha=0.95$ $M=4,59$ $SD=1,66$ Mittlere korr. Trennschärfe= 0,872 Erklärte Varianz: 84,52 %	i1	4,61	1,77	0,916	2,66	0,659	0,886
	i2	4,31	1,51	0,912	3,42	0,616	0,854
	i3	4,42	1,57	0,912	3,04	0,631	0,882
	i4	4,53	1,72	0,925	3,04	0,647	0,877
	i5	5,08	1,75	0,909	3,80	0,726	0,863

Anmerkung: Cronbachs Alpha (α), Mittelwert (M), Standardabweichung (SD), Faktorladung (a_{ij}), Anteil fehlender Werte in Prozent (% Missing), korrigierte Trennschärfe (r_{it}), Itemschwierigkeit (p_m).

Für die Indikatoren ur2 und ur4 ergeben sich vergleichsweise hohe Itemschwierigkeiten von p_m 0,74 und p_m 0,73. Die übrigen Indikatoren der ufosV2ra-Skala liegen im Bereich von p_m 0,63 bis p_m 0,68, sodass dies nicht als problematisch angesehen wird.

Neben den Indikatoren ur2 und ur4 zeigen auch die Indikatoren v1, v4 zum Vertrauen und i5 zum Involvement vergleichsweise hohe Itemschwierigkeiten mit Werten von $p_m > 0,7$. Die Werte der korrigierten Trennschärfe r_{it} fallen für sämtliche Indikatoren ausreichend hoch aus. Ebenso weisen die durchgeführten explorativen Faktorenanalysen positive Ergebnisse aus. Für alle vier reflektiven Skalen wird lediglich ein Faktor extrahiert. Die Varianzaufklärung liegt mit Werten zwischen 78,81% für die Usabilityskala und 84,52% für das Involvement ausreichend hoch. Die Werte der Faktorladungen liegen für die Usability für alle Indikatoren über a_{ij} 0,6, für die Faktoren zum Vertrauen, zum Nutzervergnügen über a_{ij} 0,8 und zum Involvement über a_{ij} 0,9. Sie weisen damit eine angemessene Höhe aus. Insgesamt kann auf Grundlage der Ergebnisse des zweiten Prüfschrittes auf eine gute Messgüte der reflektiven Skalen geschlossen werden.

7.3.3 Prüfung der Skalenstruktur mit Hilfe konfirmatorischer Faktorenanalyse

Im dritten Schritt der Prozedur zur Messgüteprüfung reflektiver Skalen nach Homburg und Giering (1996) werden die reflektiven Skalen zur Usability (ufosV2ra; ur1, ur2, ur3, ur4, ur5, ur6 und ur8), zum Vertrauen (v1, v2 und v4), zum Nutzervergnügen (vg1, vg2 und vg3) sowie zum Involvement (i1, i2, i3, i4 und i5) einer konfirmatorischen Faktorenanalyse (CFA)

unterzogen (siehe Abschnitt 6.5.3). Die Werte der in die CFA einbezogenen Indikatoren für Schiefe und Exzess liegen unterhalb der kritischen Werte von $Sch > 2,0$ bzw. $Ex > 7,0$ (siehe Tabelle E-7 im Anhang E). Zur Berechnung der CFA wird die Maximum Likelihood Methode (ML) ausgewählt. Diese setzt eine multivariate Normalverteilung der in die Analyse einbezogenen Variablen voraus. McDonald und Ho (2002) konnten allerdings zeigen, dass die ML-Methode robust gegenüber Verletzungen der Normalverteilungsannahme ist.

Bei einem Umfang von 18 Indikatoren, die in die CFA eingehen, ist ein $N = 512$ als gut einzuschätzen. Die Schätzung des in Abbildung 7-1 dargestellten CFA-Modells erfolgt mit Hilfe der Software AMOS in der Version 5.0 (Baltès-Götz, 2005).

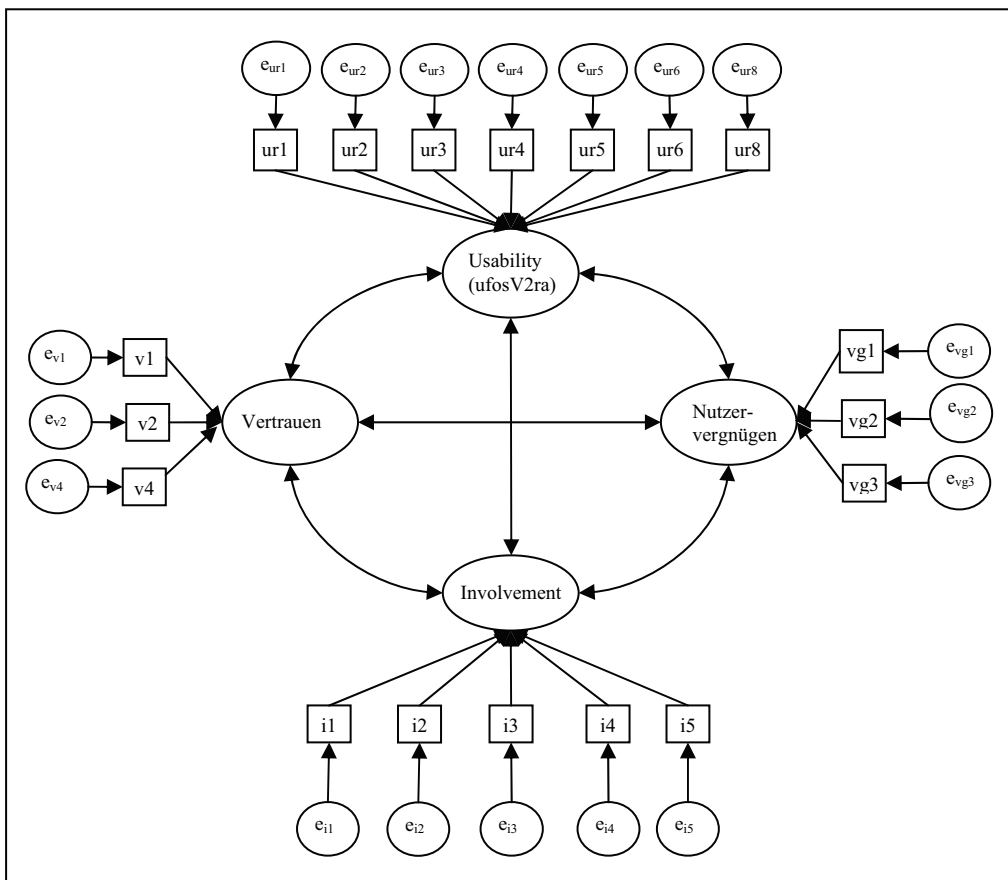


Abb. 7-1: CFA-Modell zur Überprüfung der Skalenstruktur

Anmerkung: Messfehler des Indikators (e).

Die nicht standardisierten partiellen Regressionsgewichte sind für alle einbezogenen Indikatoren auf einem Niveau von $p < .001$ signifikant; In Tabelle 7-6 werden die standardisierten Faktorladungen der Indikatoren ausgewiesen.

Tab. 7-6: Standardisierte Faktorladungen und Messfehlervarianzen der CFA zur Analyse der Skalenstruktur

ξ_j	Indikator	λ_{ij}	θ
Usability (ufosV2r) $\phi_1=2,365$	ur1	0,856	0,872
	ur2	0,799	1,420
	ur3	0,881	0,756
	ur4	0,811	1,051
	ur5	0,909	0,605
	ur6	0,942	0,382
	ur8	0,864	0,824
Vertrauen $\phi_2=1,361$	v1	0,825	0,556
	v2	0,873	0,463
	v4	0,851	0,568
Nutzervergnügen $\phi_3=1,208$	vg1	0,750	1,056
	vg2	0,931	0,321
	vg3	0,921	0,375
Involvement $\phi_4=2,395$	i1	0,916	0,503
	i2	0,892	0,498
	i3	0,906	0,432
	i4	0,898	0,569
	i5	0,895	0,598

Anmerkung: Standardisierte Faktorladung zwischen Indikator x_i und Faktor j (λ_{ij}), Messfehlervarianz (θ), Faktor j (ξ_j), geschätzte Varianz von ξ_j (ϕ_j).

In Tabelle 7-7 werden außerdem die Ausprägungen der Gütemaße zur Beurteilung des CFA-Modells dargestellt. Der Koeffizient aus χ^2 -Wert und Freiheitsgraden (df) liegt mit einem Wert von 2,39 wie gefordert unter einem kritischen Maximalwert von 3 (siehe Abschnitt 6.5.3). Der RMSEA bleibt mit einem Wert von 0,052 deutlich unterhalb des Maximalwertes von 0,08. Die Fit-Indizes CFI und NFI erfüllen beide deutlich den Mindestwert von 0,95.

Bis auf den Indikator vg1 weisen sämtliche Indikatorreliabilitäten Werte über dem Schwellenwert $rel(x_i)$ 0,5 aus; die Werte der Faktorreliabilitäten liegen deutlich über dem geforderten Mindestwert von $rel(\xi_j)$ 0,6. Das Fornell-Larcker-Kriterium, nach dem die durchschnittlich erfasste Varianz eines Faktors $DEV(\xi_1)$ größer sein sollte als die quadrierte Korrelation des Faktors mit den übrigen LVn, wird ebenfalls für jeden Faktor erfüllt.

Tab. 7-7: Evaluation des Messmodells der CFA

1. Globalkriterien:						
χ^2	= 308,65	CFI	= 0,98			
Df	= 129	RMSEA	= 0,052 (LO 90 = 0,45; HI 90 = 0,60)			
χ^2/df	= 2,39	NFI	= 0,96			
2. Detailkriterien:						
Konstrukt ξ_j	Indikator	Indikator- reliabilität $rel(x_i)$	Konvergenzvalidität		Diskriminanzvalidität	
			$rel(\xi_j)$	DEV(ξ_j)	DEV(ξ_j)	φ_{ij}^2
ξ_1 Usability (ufosV2r)	ur1	0,667	0,936	0,680	0,680	> 0,287 (Vertrauen) > 0,442 (Nutzerverg.) > 0,002 (Involvement)
	ur2	0,517				
	ur3	0,710				
	ur4	0,599				
	ur5	0,765				
	ur6	0,847				
	ur8	0,683				
ξ_2 Vertrauen	v1	0,625	0,838	0,650	0,650	> 0,287 (Usability) > 0,199 (Nutzerverg.) > 0,023 (Involvement)
	v2	0,691				
	v4	0,634				
ξ_3 Nutzervergnügen	vg1	0,392	0,843	0,611	0,611	> 0,442 (Usability) > 0,199 (Vertrauen) > 0,004 (Involvement)
	vg2	0,765				
	vg3	0,732				
ξ_4 Involvement	i1	0,800	0,949	0,789	0,789	> 0,004 (Nutzerverg.) > 0,002 (Usability) > 0,023 (Vertrauen)
	i2	0,793				
	i3	0,820				
	i4	0,772				
	i5	0,762				

Anmerkung: Faktor j (ξ_j), Indikatorreliabilität ($rel(x_i)$), Faktorreliabilität ($rel(\xi_j)$), Chi-Quadrat-Wert (χ^2), Freiheitsgrade (df), durchschnittlich erfasste Varianz eines Faktors DEV(ξ_j), quadrierte Korrelation mit LVn (φ_{ij}^2).

In Tabelle 7-8 werden die Korrelationen zwischen den latenten Variablen ausgewiesen. Auffällig sind die hohe Korrelation zwischen den Variablen Usability und Nutzervergnügen ($r = 0,67$) einerseits sowie die verhältnismäßig niedrigen Korrelationen zwischen den Variablen Involvement und Usability bzw. Nutzervergnügen andererseits. Im Ergebnis kann festgestellt werden, dass sowohl die Globalkriterien als auch die Detailkriterien auf eine angemessene Passung des Messmodells an die Daten schließen lassen. Auf Grundlage dieser Ergebnisse findet die **Hypothese 2**, dass sich die reflektive Skala ufosV2ra gegenüber anderen Skalen zur Erfassung weiterer Erfolgsfaktoren als faktoriell valide erweist, eine weitere empirische Bestätigung (siehe Abschnitt 7.3.1).

Tab. 7-8: Konstruktinterkorrelationen auf Basis der CFA

Konstrukt	ξ_1 Usability (ufosV2r)	ξ_2 Vertrauen	ξ_3 Nutzervergnügen	ξ_4 Involvement
ξ_1 Usability (ufosV2ra)	1			
ξ_2 Vertrauen	0,536	1		
ξ_3 Nutzervergnügen	0,665	0,446	1	
ξ_4 Involvement	0,040	0,150	0,064	1

Anmerkung: Faktor j (ξ_j).

7.3.4 Diskriminationsfähigkeit der reflektiven Skalen

Im nächsten Schritt wird geprüft, inwiefern die reflektiven Skalen geeignet sind, Unterschiede in den Merkmalsausprägungen der unterschiedlichen Websites festzustellen. Hierfür wird eine mehrdimensionale einfaktorielle Varianzanalyse (ANOVA) berechnet (siehe Abschnitt 6.8.2).

Die Websites der Krankenkassen stellen die Ausprägungen der unabhängigen Variable dar. Die Mittelwerte der drei reflektiven Maße Usability (ufosV2ra), Vertrauen sowie Nutzervergnügen werden als abhängige Variablen herangezogen. Mit der Durchführung der ANOVA wird die Nullhypothese H_0 geprüft, es bestünden keine Unterschiede zwischen den Mittelwerten der verschiedenen Websites. Entsprechend geht Alternativhypothese H_1 davon aus, das mindestens eine Website sich hinsichtlich des Mittelwerts von den übrigen unterscheidet. Aus den Tabellen E12 bis E14 im Anhang können die Mittelwerte und Standardabweichungen über die sieben Websites hinsichtlich der Skalen Usability (ufosV2ra), Vertrauen und Nutzervergnügen entnommen werden. Eine Voraussetzung zur Berechnung einer ANOVA ist die Normalverteilung der einbezogenen Variablen. Diese wird mit Hilfe von Kolmogorov-Smirnov-Tests überprüft. Für alle drei Variablen zeigen sich signifikante Testergebnisse ($p < .001$), sodass für keine der drei Variablen Normalverteilung anzunehmen ist. Da trotz signifikanter Testergebnisse nach Kolmogorov-Smirnov die Werte einer Variable in der Grundgesamtheit annähernd normalverteilt sein können (Brosius 2006, S. 848), werden zusätzlich Schiefe und Exzess der Skalenmittelwerte sowie deren z-Werte bestimmt (siehe Tabelle 7-9). Die Ergebnisse weisen ebenfalls für alle drei Variablen auf eine Abweichung von der Normalverteilung hin. Somit wird die Voraussetzung der Normalverteilung für die Durchführung einer Varianzanalyse verletzt. Allerdings erweist sich das Verfahren der ANOVA gegen Verletzungen der Normalverteilungsannahme robust (Johnson & Wichern, 2002, S. 179f.), sodass eine Berechnung durchgeführt wird.

Tab. 7-9: Prüfung der Normalverteilungsannahme für die reflektiven Skalenwerte

<i>Variable</i>	<i>Signifikanz im Kolmogorov- Smirnov-Test</i>	<i>Sch</i>	<i>z-Wert Sch</i>	<i>Ex</i>	<i>z-Wert Ex</i>
Usability (ufosV2r)	0,000	-0,492	-4,556	-0,891	-4,144
Vertrauen	0,000	-0,543	-5,028	0,012	0,056
Nutzervergnügen	0,000	0,243	2,250	-0,654	-3,042

Anmerkung: Exzess (Ex), Schiefe (Sch).

Neben der Annahme der Normalverteilung liegt der einfaktoriellen ANOVA die Annahme der Varianzhomogenität, vergleichbare Varianzen der betrachteten abhängigen Variablen in Abhängigkeit der verschiedenen Ausprägungen der unabhängigen Variable Website, zugrunde. Die Überprüfung, inwiefern die Voraussetzung der Varianzhomogenität erfüllt wird, erfolgt anhand der Berechnung von Levene-Tests (siehe Abschnitt 6.8.3). Die Testergebnisse werden in Tabelle 7-10 dargestellt. Im Fall der reflektiven Usability-Skala ufosV2ra ergibt sich ein signifikantes Ergebnis. Die Hypothese, die Varianzen seien in den

sieben Gruppen in der Grundgesamtheit gleich, ist damit abzulehnen, womit die Voraussetzung der Varianzhomogenität als verletzt anzusehen ist.

Tab. 7-10: Ergebnisse der Levene-Tests auf Varianzhomogenität

<i>Abhängige Variable</i>	<i>Levene-Statistik</i>	<i>df1</i>	<i>df2</i>	<i>Signifikanz</i>
Usability (ufosV2ra)	7,734	6	505	0,000
Vertrauen	0,764	6	505	0,599
Nutzervergnügen	0,973	6	505	0,443

Anmerkung: Freiheitsgrade (df).

Die Ergebnisse der ANOVA, die in Tabelle 7-11 dargestellt werden, können deshalb für die Usability-Skala nur mit Einschränkungen interpretiert werden. Für alle drei Skalen zeigt sich ein signifikantes Ergebnis der Varianzanalyse. Es kann damit davon ausgegangen werden, dass sich die Skalenmittelwerte in Abhängigkeit der verschiedenen Websites signifikant unterscheiden. Die **Hypothese H3**, die ufosV2ra-Skala sei geeignet, zwischen mehreren Websites zu differenzieren, kann damit bestätigt werden.

Tab. 7-11: Ergebnisse der mehrdimensionalen einfaktoriellen Varianzanalyse

<i>Reflektives Maß</i>		<i>Quadratsumme</i>	<i>df</i>	<i>Mittel der Quadrate</i>	<i>F</i>	<i>Signifikanz</i>
Usability (ufosV2ra)	Zwischen den Shops	223,776	6	37,296	15,248	0,000
	Innerhalb der Shops	1235,216	505	2,446		
	Gesamt	1458,992	511			
Vertrauen	Zwischen den Shops	48,021	6	8,004	5,150	0,000
	Innerhalb der Shops	748,820	505	1,554		
	Gesamt	832,842	511			
Nutzervergnügen	Zwischen den Shops	61,276	6	10,213	5,101	0,000
	Innerhalb der Shops	1011,003	505	2,002		
	Gesamt	1072,279	511			

Anmerkung: Freiheitsgrade (df), F-Wert (F).

7.3.5 Erfassung der wahrgenommenen Höhe des Beitragssatzes

Der Anteil fehlender Werte liegt für das Single-Item zur Erfassung der wahrgenommenen Höhe des Beitragssatzes (subp) bei 28,3 %. Der Mittelwert des Items beträgt M 4,31 (SD 1,74). In der Tabelle E-18.1 im Anhang E werden die deskriptiven Ergebnisse zur Ausprägung dieses Items für die verschiedenen einbezogenen Websites in Form einer Rangreihe vorgestellt.

7.4 Erfassung der abhängigen Variablen

Nachdem in Abschnitt 7.3 die Messgüte der Variablen auf Prädiktorebene überprüft wurde, wird in den folgenden Abschnitten 7.4.1 bis 7.4.3 die Erfassung der Erfolgsgrößen Nutzerzufriedenheit, Nutzungsintention sowie der Auswahlentscheidung betrachtet.

7.4.1 Reflektive Erfassung der Nutzerzufriedenheit

Die Prüfung der Messgüte der reflektiven Skala Nutzerzufriedenheit erfolgt anhand der in Abschnitt 6.5.2 dargestellten Prüfkriterien der klassischen Testtheorie analog Schritt zwei der Prüfprozedur nach Homburg und Giering (1996).

Bei Einbeziehung der beiden Indikatoren z1 und z2 in die Skala ergibt sich ein Cronbachs Alpha von α 0,94. Weitere Kennwerte der Skala lassen sich Tabelle 7-12 entnehmen. Die Werte der korrigierten Trennschärfe sind mit $r_{it} > 0,8$ wünschenswert hoch; die Itemschwierigkeiten liegen ebenfalls im akzeptablen Wertebereich zwischen $0,2 < p_m < 0,8$. Die Messgüte der reflektiven Skala zur Nutzerzufriedenheit ist damit ebenfalls positiv zu werten.

Wie bereits in Abschnitt 7.1.3 angeführt, kann angesichts der signifikanten Ergebnisse in den Kolmogorov-Smirnov-Tests nicht von einer Normalverteilung der Indikatoren zur Nutzerzufriedenheit ausgegangen werden. Das Gleiche gilt auch für die Skalenmittelwerte aus den beiden Indikatoren zur Nutzerzufriedenheit. Neben einem signifikanten Ergebnis im Kolmogorov-Smirnov-Test weisen auch die Werte für Schiefe und Exzess auf eine Abweichung des Skalenwertes von der Normalverteilung hin (*sch* -0,43; *z-Wert sch* -4,02; *ex* -1,02; *z-Wert ex* -4,75).

Tab. 7-12: Skalen- und Itemkennwerte der reflektiven Skala zur Nutzerzufriedenheit

Skalenkennwerte	Item-Code	M	SD	a_{ij}	% Missing	r_{it}	p_m
$\alpha=0,94$ M=4,55; SD=1,80 Mittlere korr. Trennschärfe=0,896 Homogenität=0,896 Erklärte Varianz: 94,56%	z1	4,52	1,73	0,972	2,15	0,896	0,646
	z2	4,58	1,88	0,972	0,59	0,896	0,654

Anmerkung: Cronbachs Alpha (α), Mittelwert (M), Standardabweichung (SD), Faktorladung (a_{ij}), Anteil fehlender Werte in Prozent (% Missing), korrigierte Trennschärfe (r_{it}), Itemschwierigkeit (p_m).

In der Tabelle E-16 im Anhang E werden die deskriptiven Ergebnisse zur Ausprägung der Skalenmittelwerte zur Nutzerzufriedenheit für die sieben im Quasi-Experiment berücksichtigten Websites anhand einer Rangreihenbildung über die Skalenmittelwerte dargestellt.

7.4.2 Reflektive Erfassung der Nutzungsintention

Die Güteprüfung der Skala zur Nutzungsintention wird analog zu dem in Abschnitt 7.4.1 dargestellten Vorgehen der klassischen Testtheorie vorgenommen. Bei Einbeziehung der vier Indikatoren ni1, ni2, ni3 und ni4 in die Skala Nutzungsintention ergibt sich auf Grundlage eines Datensatzes mit einer Bewertung pro Fall ein Cronbachs Alpha in Höhe von α 0,92. Der Wert kann durch Elimination des Indikators ni4 auf α 0,93 gesteigert werden. Tabelle 7-13 gibt einen Überblick über die Kennwerte der Skala zur Nutzungsintention. In einer explorativen Faktorenanalyse wird erwartungsgemäß ein Faktor bestätigt, der 87,9% der Datenvarianz aufklärt. Die Werte der korrigierten Trennschärfe der drei Items sind mit $r_{it} > 0,8$ wünschenswert hoch; die Itemschwierigkeiten liegen im angemessenen Wertebereich

zwischen $0,2 < p_m < 0,8$ (Fisseni, 1997, S.124). Die Messgüte der reflektiven Skala zur Nutzungsintention ist damit positiv zu bewerten.

Wie bereits in Abschnitt 7.1.3 beschrieben, zeigen sich bei allen reflektiven manifesten Variablen signifikante Ergebnisse im Kolmogorov-Smirnov-Test (sämtlich $p < .001$), so dass auch im Falle der drei Indikatoren zur Nutzungsintention von einer Abweichung von der Normalverteilung ausgegangen werden muss. Das Gleiche gilt auch für den Fall des Skalenmittelwertes aus den drei Indikatoren zur Nutzungsintention, für den ebenfalls ein signifikantes Testergebnis im Kolmogorov-Smirnov-Test vorliegt ($p < .001$). Eine Überprüfung des Ergebnisses anhand der Werte für Schiefe und Exzess bestätigt eine Abweichung des Skalenmittelwertes von der Normalverteilung (*sch* -0,20; *z-Wert sch* -1,85; *ex* -1,13; *z-Wert ex* -5,52).

Tab. 7-13: Skalen- und Itemkennwerte der reflektiven Skala zur Nutzungsintention

Skalenkennwerte	Item-Code	M	SD	a_{ij}	% Missing	r_{it}	p_m
$\alpha = 0,931$ M=4,12; SD=1,78 Mittlere korr. Trennschärfe=0,858 Homogenität=0,819 Erklärte Varianz: 87,91% α vor Elimination von ni4= 0,92	ni1	4,23	1,89	0,944	4,75	0,837	0,604
	ni2	4,16	1,89	0,868	2,85	0,881	0,594
	ni3	3,97	1,92	0,944	3,42	0,858	0,567

Anmerkung: Cronbachs Alpha (α), Mittelwert (M), Standardabweichung (SD), Faktorladung (a_{ij}), Anteil fehlender Werte in Prozent (% Missing), korrigierte Trennschärfe (r_{it}), Itemschwierigkeit (p_m).

In der Tabelle E-17 im Anhang E werden die deskriptiven Ergebnisse zur Ausprägung der Skalenmittelwerte zur Nutzungsintention für die 7 im Quasi-Experiment berücksichtigten Websites anhand einer Rangreihenbildung über die Skalenmittelwerte dargestellt.

7.4.3 Erfassung der Auswahlentscheidung

Die Güte der abhängigen Variable Auswahlentscheidung kann nicht anhand des bei den Variablen Nutzungsintention und Nutzerzufriedenheit gewählten Vorgehens beurteilt werden, da es sich bei der Auswahlentscheidung um eine kategoriale Variable handelt. In diesem Abschnitt werden zunächst die deskriptiven Ergebnisse zur Ausprägung der Variable Auswahlentscheidung dargestellt. Tabelle 7-14 gibt eine Übersicht über die Verteilung der Auswahlentscheidungen auf die sieben einbezogenen Websites. In $N = 244$ Fällen wurde eine Auswahlentscheidung für eine der sieben Websites getroffen. Es haben somit 95,3% der 256 Teilnehmer eine Auswahlentscheidung getroffen. In 4,7% der Fälle wurde keine Auswahl vorgenommen. Der Anteil der Auswahlentscheidungen für eine Website in Bezug zur Anzahl der vorliegenden Bewertungen streut über die Anbieter stark. So entschieden sich lediglich 16 von 71 Teilnehmern (22,5%) die im Verlauf des Quasi-Experiments eine Bewertung bezüglich der Website der GEK - Gmünder Ersatzkasse vornahmen für diese Website, wohingegen im Falle der Website der BARMER 82,5% der beurteilenden Teilnehmer sich für

diese entschieden. Weitere deskriptive Ergebnisse zu den einzelnen Websites werden in Tabelle E-18 im Anhang E in Form einer Rangreihe dargestellt.

Tab. 7-14: Wahlhäufigkeit der Website

Rang	Website	N	Wahlhäufigkeit absolut	Wahlhäufigkeit relativ (%)	Verteilung (%)	Verteilung kumul. (%)
1.	BARMER	80	66	82,5	25,8	25,8
2.	Techniker Krankenkasse	75	41	54,7	16,0	41,8
3.	KKH - Die Kaufmännische	78	39	50,0	15,2	57,0
4.	DAK	79	34	43,0	13,3	70,3
5.	Deutsche BKK	68	23	33,8	9,8	80,1
6.	AOK - Gesundheitskasse	75	25	33,3	9,0	89,1
7.	GEK - Gmünder Ersatzkasse	71	16	22,5	6,3	95,3
Keine Auswahl			12		4,7	100,0
Gesamt		526	256		100,0	

Anmerkung: Anzahl der vorliegenden Bewertungen (N).

7.5 Festlegung der formativen Skala ufosV2fa

Im folgenden Abschnitt wird die formative Usability-Skala ufosV2fa festgelegt. Nachdem im vierten Kapitel die Zusammenstellung des Itempools dargestellt wurde, erfolgt im ersten Schritt eine Festlegung der Items, die in die endgültige ufosV2fa-Skala eingehen (Abschnitt 7.5.1). Im zweiten Schritt werden die Items auf Multikollinearität untersucht. Da die formativen ufosV2fa-Items im weiteren Verlauf in PLS-Modelle einbezogen werden sollen, sind im Falle hoher Multikollinearität Unterindizes zu bilden (Abschnitt 7.5.2). Im dritten Schritt wird die Usability-Skala ufosV2fa mit Hilfe eines Zwei-Konstrukt-Modells in PLS validiert (Abschnitt 7.5.3).

7.5.1 Endgültige Bestimmung der ufosV2fa-Indikatoren

Um endgültig festzulegen, welche der formativen Indikatoren in der Usability-Skala ufosV2fa verbleiben, wird zunächst ein potentieller Ausschluss von Indikatoren aufgrund eines zu hohen Anteils fehlender Werte geprüft (Abschnitt 7.5.1.1). Daran anschließend wird anhand der bivariaten Korrelationen zwischen den Indikatoren geprüft, inwiefern einzelne Indikatoren aufgrund inhaltlicher Redundanzen auszuschließen sind (Abschnitt 7.5.1.2).

7.5.1.1 Elimination von Indikatoren aufgrund eines zu hohen Anteils fehlender Werte

In die endgültige Usability-Skala ufosV2fa sollen nur Indikatoren aufgenommen werden, die einerseits als relevant einzuschätzen sind und andererseits von einem Großteil der Nutzer von *Online-Dienstleistungsangeboten* beantwortet werden können. Aus diesem Grunde wurde festgelegt, dass der Anteil fehlender Werte für einen Indikator 30% nicht überschreiten soll. Dieses Niveau wurde durch keinen der ufosV2fa-Indikator verletzt. In Tabelle 7-15 werden die Indikatoren mit absolut mindestens 10 fehlenden Werten je Indikator im Wortlaut sowie die jeweiligen Missing-Anteile der Items angeführt. Nur zwei Indikatoren überschreiten einen

Missing-Anteil von 10%, sodass insgesamt davon ausgegangen werden kann, dass die verwendeten ufosV2fa-Indikatoren im Rahmen einer Anwendung auf Online-Dienstleistungsangebote von den Nutzern gut beantwortet werden können.

Tab. 7-15: Formative Usability Indikatoren mit mehr als 10 fehlenden Werten (absolut)

Item-Code	Missing Absolut	Missing %	Wortlaut des formativen Indikators
uf1a	10	1,95	Bei der Nutzung dieser Website kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.
uf25ac	11	2,15	Es dauert zu lange, bis die Website auf meine Eingaben reagiert.
uf2ac	12	2,34	Ich muss mir viel merken, wenn ich die Website nutze.
uf24a	38	7,42	Die Website empfiehlt mir interessante Angebote, nach denen ich ursprünglich nicht gesucht habe.
uf31a	40	7,81	Ich erhalte nützliche Informationen zum Unternehmen, das die Website betreibt.
uf23a	41	8,01	Ich finde die verwendeten Bilder und Grafiken hilfreich.
uf5a	51	9,96	Ich kann jede Aktion wieder rückgängig machen.
uf30a	57	11,13	Diese Website ist eine Quelle qualitativ hochwertiger Angebote.
uf20a	62	12,11	Die Suchfunktionen liefern mir sinnvolle Ergebnisse.

Anmerkung: Anteil fehlender Werte (% Missing).

7.5.1.2 Eliminierung inhaltlich redundanter Indikatoren

Im zweiten Schritt ist der Ausschluss inhaltlich redundanter Indikatoren zu prüfen, um den Item-Umfang der formativen Skala ufosV2fa nicht unnötig groß ausfallen zu lassen. Hierfür werden auf Grundlage der bivariaten Korrelationen zwischen den Indikatoren die Indikatorpaarungen mit einem Koeffizienten $r \geq 0,5$ näher betrachtet (siehe Tabelle E-19 im Anhang E).

In Tabelle 7-16 werden sechs Indikatorenpaarungen mit hohen bivariaten Korrelationen aufgeführt, bei denen aufgrund hoher inhaltlicher Redundanzen ein Ausschluss gerechtfertigt erscheint. Als Kriterien zur Beurteilung, welche der Items eliminiert werden sollen, werden neben einer inhaltlichen Abwägung die Itemschwierigkeiten sowie Korrelationen mit anderen Indikatoren herangezogen. Der Anteil fehlender Werte ist hier aufgrund der niedrigen Ausprägungen zur Bewertung der Items nicht geeignet. Im Ergebnis werden die Indikatoren uf8a, uf11a, uf19a sowie uf27a von der weiteren Verwendung in der formativen Skala ufosV2fa ausgeschlossen, um inhaltliche Redundanzen im Fragebogeninventar zu vermeiden und den Itemumfang einzugrenzen. Somit verbleiben 27 Items in der ufosV2fa-Skala.

Tab. 7-16: Indikatorpaarungen mit hoher bivariater Korrelation und inhaltlicher Redundanz

Item-Code	VIF	p_m	% Missing	Item	Elimination	r
uf8a	7,370	0,658	0,0	Diese Website ist gut strukturiert.	X	0,766
uf18a	3,489	0,683	0,0	Es ist schnell erkennbar, welche Inhalte auf der Website angeboten werden.		
uf8a	7,370	0,658	0,0	Diese Website ist gut strukturiert.	X	0,738
uf21a	3,342	0,713	1,4	Die Angebote sind in sinnvolle Kategorien eingeteilt.		
uf9ac	3,170	0,655	0,0	Es sind zu viele Schritte nötig, um das Gesuchte zu finden.		0,749
uf11a	6,642	0,658	0,2	Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.	X	
uf10a	5,175	0,623	0,0	Ich finde mich auf den einzelnen Seiten leicht zurecht.		0,825
uf11a	6,642	0,658	0,2	Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.	X	
uf10a	5,175	0,623	0,0	Ich finde mich auf den einzelnen Seiten leicht zurecht.		0,850
uf19a	6,212	0,664	0,2	Ich finde leicht was ich suche.	X	
uf21a	3,342	0,713	1,4	Die Angebote sind in sinnvolle Kategorien eingeteilt.		0,698
uf27a	5,236	0,655	0,2	Die Themenlisten auf der Website sind übersichtlich.	X	

Anmerkung: Variance Inflation Factor (VIF), Itemschwierigkeit (p_m), Anteil fehlender Werte in Prozent (% Missing), bivariate Pearson-Korrelation (r), Elimination des Items (X).

7.5.2 Behandlung von Multikollinearität zwischen den ufosV2fa-Items

Multikollinearität zwischen den formativen Indikatoren kann bei Einbindung in PLS-Modelle zu Verzerrungen der Schätzparameter und damit zu negativen Auswirkungen auf die Aussagekraft des Modells führen. Um zu prüfen, inwieweit Multikollinearität zwischen den Indikatoren vorliegt, werden die in Abschnitt 6.6.2.3 vorgestellten Ansätze verfolgt.

Zunächst werden die Pearson'schen Korrelationskoeffizienten der Indikatoren herangezogen. Aus Tabelle E-19 im Anhang E ist ersichtlich, dass für die gemäß Vorabschnitt verbleibenden 27 Indikatoren 73 bivariate Korrelationen mit $r \geq 0,5$ identifiziert werden können, von denen 29 Werte bei $r \geq 0,6$ liegen.

Zur weiteren Prüfung werden die Toleranz- und VIF-Werte der verbleibenden 27 formativen Indikatoren einbezogen (siehe Tabelle E-20 im Anhang E). Zehn Items weisen eine Toleranz $< 0,4$ auf; 15 VIF-Werte liegen oberhalb des kritischen Wertes von 2. Angesichts dieser Werte und der hohen Interkorrelationen muss von einer erheblichen Multikollinearität zwischen den formativen Indikatoren ausgegangen werden. Einen weiteren Hinweis auf ein zu hohes Maß an Multikollinearität erbringt die Betrachtung der Konditionsindizes. Für die 27. Dimension ergibt sich ein Konditionsindex von 42,3 (Eigenwert λ 0,014), der deutlich über dem kritischen Wert von 30 liegt.

Angesichts dieser Hinweise auf ein hohes Maß an Multikollinearität soll diese im Zwei-Konstrukt-Modell methodisch angemessen berücksichtigt werden. Hierfür empfiehlt es sich, wie in Abschnitt 2.6.4 dargestellt, untereinander multikollineare Indikatoren der Skala ufosV2a zu Indizes zusammenzufassen (Albers & Hildebrandt, 2006). Methodisch wird dabei dem Prinzip einer Hauptkomponentenregression gefolgt (vgl. Moosmüller, 2004, S. 132; Schneider, 2006, S.199). Um möglichst viel Information der vorhandenen Daten beizubehalten, wird als geeignete Form einer explorativen Faktorenanalyse für die formativen Usability-Indikatoren eine HKA mit orthogonalen Hauptkomponenten berechnet. Die Berechnung einer HKA setzt keine Normalverteilung der Indikatoren voraus, sodass die teilweise schiefen Verteilungen der formativen Usability-Items unkritisch sind (Bühner, 2004, S.159). Um eine Verteilung der erklärten Varianz über mehrere Komponenten zu erreichen, wird eine orthogonale Rotation (Varimax-Methode) der Komponenten vorgenommen. Die extrahierten Komponenten können als eigenständige formative Indikatoren in das PLS-Zwei-Konstrukt-Modell aufgenommen werden (siehe Abschnitt 2.6.4). Eine inhaltliche Interpretation dieser Komponenten ist nicht möglich, da im Falle formativer Indikatoren trotz hoher Korrelationen nicht auf eine inhaltliche Nähe geschlossen werden kann.

Auf Grundlage des *Kaiser-Guttman-Kriteriums*, nach dem die Zahl der zu extrahierenden Faktoren gleich der Zahl der Faktoren mit *Eigenwert* $\lambda > \text{eins}$ ist, werden fünf Komponenten extrahiert (vgl. Heckner, 1998, S. 187f.). Tabelle E-21 im Anhang E stellt die rotierte Komponentenmatrix dar. Der KMO-Koeffizient liegt bei 0,95. Die Indikatoren können daher als sehr gut zur Berechnung einer Faktorenanalyse geeignet angesehen werden. Der Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (*ungefähres* χ^2 7302,823; *df* 325; $p < .001$).

Als Entsprechung zum KMO-Maß werden auf Indikatorebene die *MSA-Werte* betrachtet. Mit Ausnahme des Indikators uf16a liegt kein MSA-Wert unterhalb des kritischen Werts von 0,5. Für drei Indikatoren ergeben sich Kommunalitäten mit Werten von $h^2 < 0,5$ (uf2a, uf20a und uf23a). Von einer Elimination der Indikatoren wird allerdings abgesehen, da damit ein Verlust der inhaltlichen Usability-Aspekte verbunden wäre.

Der durch die 5 Komponenten aufgeklärte Anteil an der Gesamtvarianz beträgt 62,05%, wovon die erste Komponente bereits 42,28% Varianzanteil aufklärt. Eine inhaltliche Interpretation der extrahierten Komponenten kann nicht erfolgen, da eine inhaltliche Dimensionierung eines Sets an formativen Indikatoren mit Hilfe einer HKA nicht möglich scheint. Die formativen Indizes werden gebildet, indem mit Hilfe der HKA die Faktorwerte für die fünf Komponenten berechnet werden (Bühner, 2004, S. 169). Diese werden im Folgenden mit uf_in1 bis uf_in5 benannt.

Zusammenfassend kann festgestellt werden, dass aufgrund eines zu hohen Maßes an Multikollinearität zwischen den einzelnen Usability-Items diese nicht als Einzelindikatoren in eine PLS-Analyse eingehen können. Deshalb werden die Items mit Hilfe einer HKA zu

Indizes zusammengefasst. Diese Indizes können als formative Indikatoren in die nachfolgend berechneten PLS-Analysen aufgenommen werden, da zwischen ihnen keine Multikollinearität vorliegt.

7.5.3 PLS-Zwei-Konstrukt-Modell zur Validierung

Entsprechend dem in Abschnitt 6.6.2.3 vorgestellten Vorgehen wird die formative latente Variable Usability (Skala ufosV2fa) überprüft, indem sie in einem Zwei-Konstrukt-Modell als Prädiktor zu der endogenen reflektiven latenten Variable Usability (Skala ufosV2ra) in Beziehung gesetzt wird. Das Modell wird mit Hilfe des PLS-Algorithmus unter Nutzung des Path Weighting Scheme geschätzt. Als Software wird das Program Smart-PLS in der Version 2.0.M3 eingesetzt (Hansmann & Ringle, 2004). Die Stichprobengröße mit $N = 512$ Fällen erweist sich angesichts der Anzahl der Indikatoren in dem Modell als ausreichend groß. Zur Überprüfung der Modellgüte wird eine Bootstrapping Resampling Prozedur mit 200 Subsamples berechnet.

Tab. 7-17: Kennwerte des PLS-Messmodells zum Zwei-Konstrukt-Modell

<i>Latente Variable</i>	<i>Indikator</i>	<i>Gewicht / Ladung*</i>	<i>Standardfehler</i>	<i>t-Statistik</i>
Usability formativ (ufosV2fa-Skala)	uf in1	0,844	0,058	14,501
	uf in2	0,459	0,068	6,753
	uf in3	0,246	0,046	5,320
	uf in4	0,117	0,050	2,340
	uf in5	0,022	0,046	0,473
Usability reflektiv (ufosV2ra-Skala) <i>Composite Reliability</i> =0,966 <i>AVE</i> =0,806	ur1a	0,885	0,023	38,402
	ur2a	0,845	0,043	19,413
	ur3a	0,902	0,023	39,147
	ur4a	0,889	0,024	36,187
	ur5a	0,922	0,019	48,251
	ur6a	0,935	0,013	69,585
	ur8a	0,900	0,019	46,837

Anmerkung: * Im Fall der formativen Indikatoren werden die Gewichte angegeben, bei den reflektiven Indikatoren werden die Ladungen angegeben.

Im ersten Schritt erfolgt eine Prüfung der Messmodellgüte. In Tabelle 7-17 werden die Kennwerte für die formativen und reflektiven Indikatoren angeführt. Vier der fünf formativen Indizes zur Erfassung der formativen Usability zeigen angemessene Gewichte von $\gamma > 0,1$ sowie signifikante t -Werte über 2,00. Die fünfte in der HKA ausgewiesene Hauptkomponente uf_in5 weist ein Gewicht von lediglich $\gamma < 0,1$ sowie einen insignifikanten t -Wert von $t = 0,47$ aus. Die Ladungen der sieben reflektiven Usability-Indikatoren liegen wie gefordert über $\lambda = 0,707$ bei signifikanten t -Werten über 1,66. Die *Composite Reliability* als Maß für die interne Konsistenz der reflektiven Usability-Skala lässt mit einem Wert von 0,966 auf eine hohe Reliabilität der Skala ufosV2ra schließen. Die **Hypothese 1**, die reflektive Skala ufosV2ra sei reliabel kann damit ein weiteres Mal bestätigt werden (siehe Abschnitt 7.3.2).

Die *durchschnittlich erfasste Varianz liegt* mit einem AVE-Wert von 0,806 deutlich oberhalb des geforderten Wertes von 0,5. In Anbetracht dieser Ergebnisse ist die Messmodellgüte als hervorragend anzusehen. Das niedrige Gewicht von $\gamma=0,022$ für den Index uf_in5 bei einem insignifikanten t -Wert von t 0,47 kann auf den geringen Beitrag der Hauptkomponente zur Varianzaufklärung zurückgeführt werden (4,05% aufgeklärte Varianz; λ 1,05; siehe Tabelle E-21 im Anhang E).

Trotz der nicht zufriedenstellenden Kennwerte des Index uf_in5 wird von seinem Ausschluss abgesehen. Da alle formativen ufosV2fa-Indikatoren jeweils in alle fünf Indizes eingehen, würde der Ausschluss eines Index keine Elimination im herkömmlichen Sinne bewirken. Eine vergleichende inhaltliche Interpretation der Indizes und ihrer Einflussstärken kann nicht erfolgen, da eine inhaltliche Dimensionierung der jeweils aggregierten formativen Indikatoren nicht möglich ist (siehe Abschnitt 7.5.2).

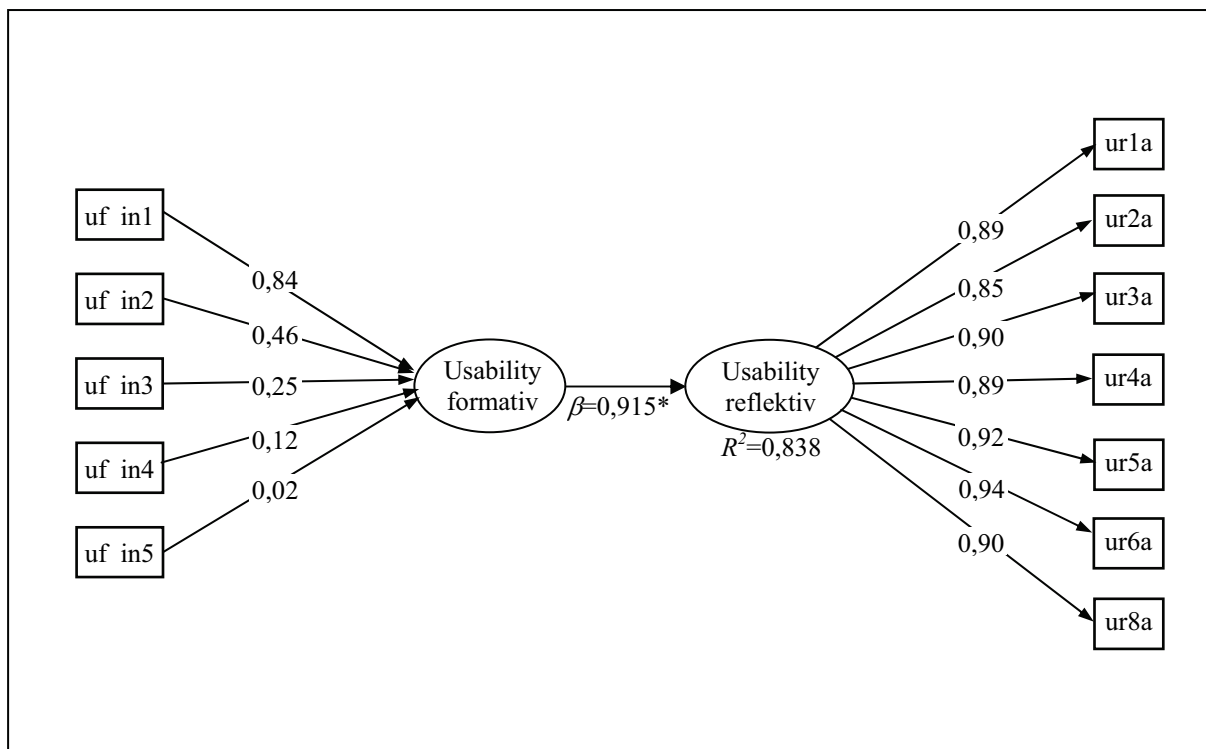


Abb. 7-2: Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala ufosV2fa mit Hilfe der reflektiven Usability-Skala ufosV2ra

Anmerkung: * Pfadkoeffizient ist signifikant ($t > 1,66$). Für die formativen Usability-Indikatoren werden die Gewichte, für die reflektiven Usability-Indikatoren die Ladungen angegeben.

Nachdem die Messmodellgüte als sehr gut bewertet wurde, erfolgt im zweiten Schritt eine Prüfung der Güte des Strukturmodells. In Abbildung 7-2 wird das geschätzte Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala ufosV2fa graphisch dargestellt. Es zeigt sich ein sehr hoher signifikanter Pfadkoeffizient von β 0,915 sowie ein sehr zufriedenstellender Determinationskoeffizient von R^2 0,838. Diese Werte lassen auf einen starken Zusammenhang zwischen der latenten formativen und der reflektiven Usability-Variable schließen. Entsprechend kann davon ausgegangen werden, dass das Set der 27

verwendeten formativen Usability-Indikatoren das Globalurteil auf Basis der reflektiven Usability-Skala sehr gut abdeckt. Die **Hypothese H4**, wonach ein starker Zusammenhang zwischen der formativen ufosV2fa-Skala und der reflektiven ufosV2ra-Skala besteht (siehe Abschnitt 2.6.4), kann somit bestätigt werden.

7.6 Untersuchung des nomologischen Modells mit Hilfe von PLS

In den folgenden Abschnitten werden die Ergebnisse der empirischen Untersuchung des nomologischen Modells (siehe Abschnitt 2.10) dargestellt. Die Hypothesen bezüglich der Zusammenhänge zwischen den Modellelementen werden mit Hilfe des PLS-Strukturgleichungsansatzes überprüft. Hierfür werden zwei PLS-Modelle berechnet. Die beiden Modelle unterscheiden sich dahingehend, dass im einen Modell die reflektive Skala ufosV2ra zur Erfassung der latenten Variable Usability eingebunden wird, während im anderen Modell die formative Usability-Skala ufosV2fa verwendet wird. Im Folgenden werden die Modelle deshalb als ufosV2ra-Modell bzw. ufosV2fa-Modell bezeichnet. Das nomologische Modell beinhaltet Hypothesen bezüglich der Wirkungszusammenhänge zum Erfolgsfaktor Ästhetik. Da sich die Skala zur Erfassung der Variable Ästhetik gegenüber den übrigen Skalen nicht als faktoriell valide erwies (siehe Abschnitt 7.3.1), wird diese nicht weiter im nomologischen Modell berücksichtigt. Die Hypothesen H10, H11, H14 sowie H15 können somit nicht überprüft werden. Zur Schätzung der beiden Modelle wird jeweils der Datensatz mit einer Bewertung von $N = 512$ Fällen je Website herangezogen. Als Schätzalgorithmus wird das Path Weighting Scheme verwendet. Zur Überprüfung der Modellgüte wird eine Bootstrapping Resampling Prozedur mit 200 Subsamples berechnet.

7.6.1 Beurteilung der Messmodelle

Im ersten Schritt wird die Güte der reflektiven Messmodelle beurteilt. In Tabelle 7-18 werden die Kennwerte für beide PLS-Modelle angeführt. Alle Ladungen der reflektiven Indikatoren weisen in beiden PLS-Modellen zufriedenstellende Werte von $\lambda > 0,707$ auf (Barclay et al, 1995). Die zugehörigen t -Werte weisen Signifikanz für sämtliche Ladungen aus. Tabelle 7-19 enthält die Werte der durchschnittlich erfassten Varianz sowie der Composite Reliability für die reflektiven Skalen in beiden PLS-Modellen. Die durchschnittlich erfasste Varianz liegt in allen Fällen bei AVE-Werten deutlich oberhalb des geforderten Wertes von 0,5. Die Ausprägungen der Composite Reliability überschreiten für alle Skalen den kritischen Wert von 0,7 und erweisen sich mit Werten über 0,9 als hoch. Sie lassen damit auf eine hohe Reliabilität der Skalen schließen. Der sehr hohe Wert für die Composite Reliability der reflektiven Usability-Skala ufosV2ra in Höhe von 0,966 wurde bereits im vorangehenden Abschnitt 7.5.3 zum Zwei-Konstrukt-Modell als Bestätigung der **Hypothese 1** gewürdigt.

Tab. 7-18: Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte

Latente Variable	Item	ufosV2ra-Modell (reflektive Usability)			ufosV2fa-Modell (formative Usability)		
		Gewicht bzw. Ladung	SE	t-Wert	Gewicht bzw. Ladung	SE	t-Wert
Usability - ufosV2ra (reflektiv)	ur1a	0,889	0,023	39,380			
	ur2a	0,841	0,039	21,743			
	ur3a	0,900	0,025	36,245			
	ur4a	0,887	0,022	41,237			
	ur5a	0,923	0,019	49,111			
	ur6a	0,938	0,014	65,944			
	ur8a	0,903	0,020	44,282			
Usability - ufosV2fa (formativ)	uf in1				0,634	0,078	8,182
	uf in2				0,699	0,069	10,184
	uf in3				0,276	0,071	3,896
	uf in4				0,181	0,063	2,857
	uf in5				-0,016	0,057	0,280
Vertrauen (reflektiv)	v1	0,922	0,020	47,146	0,921	0,020	46,809
	v2	0,914	0,021	44,318	0,913	0,023	39,843
	v4	0,897	0,031	29,387	0,900	0,026	35,352
Nutzervergnügen (reflektiv)	vg1	0,852	0,038	22,702	0,855	0,039	21,749
	vg2	0,947	0,011	90,142	0,946	0,012	79,980
	vg3	0,943	0,013	71,777	0,941	0,013	74,459
Involvement (reflektiv)	i1	0,918	0,178	5,158	0,918	0,193	4,751
	i2	0,877	0,203	4,326	0,877	0,214	4,092
	i3	0,920	0,176	5,232	0,919	0,198	4,649
	i4	0,934	0,164	5,682	0,934	0,189	4,936
	i5	0,942	0,185	5,092	0,942	0,174	5,429
Nutzungsintention (reflektiv)	ni1	0,929	0,021	44,469	0,929	0,023	41,192
	ni2	0,947	0,014	65,843	0,947	0,015	61,908
	ni3	0,937	0,016	57,021	0,937	0,019	49,717
Nutzerzufriedenheit (reflektiv)	z1	0,974	0,005	180,260	0,975	0,006	178,193
	z2	0,973	0,006	159,417	0,973	0,007	141,649
Wahrgen. Preis (Single-Item)	Subp	1			1		
Bekanntheitsgrad (Single-Item)	b1	1			1		
Kontrollvar.	co1	1			1		
Kontrollvar.	co2	1			1		
Moderatoreffekte		Siehe Tabelle E-22 im Anhang E					
Involv. x Nutzerzuf.							
Involv. x ufosV2ra							
Involv. x ufosV2fa							

Anmerkung: Die Ladungen und Gewichte werden getrennt für die Verwendung der reflektiven Usability-Skala ufosV2ra sowie für die formative Usability-Skala ufosV2fa angegeben. Im Fall des formativen Konstruktes werden in der Spalte „Gewicht bzw. Ladung“ die Gewichte angegeben; im Fall reflektiver Konstrukte handelt es sich um Ladungen. Standardfehler (SE).

Desweiteren wird die faktorielle Validität der reflektiven Skalen geprüft. Hierfür werden in den Tabellen E-25 und E-26 im Anhang E die Korrelationskoeffizienten der latenten Variablen sowie die Werte für die Quadratwurzel des AVE-Wertes dargestellt (Tabelle E-25

zum ufosV2ra-Modell; Tabelle E-26 zum ufosV2fa-Modell). Für alle reflektiven Skalen überschreitet in beiden Modellen die Quadratwurzel des AVE-Wertes die Korrelationskoeffizienten der LV mit den übrigen LVn.

In den Tabellen E-27 und E-28 im Anhang E werden die Kreuzladungen sämtlicher Indikatoren im PLS-Modell aufgeführt. Alle Indikatoren korrelieren mit den ihnen jeweils zugehörigen LVn höher als mit den übrigen Variablen. Auf Grundlage dieser Ergebnisse kann die faktorielle Validität der reflektiven Skalen für beide PLS-Modelle als gegeben betrachtet werden. Die **Hypothese 2**, die reflektive Usability-Skala ufosV2ra sei gegenüber weiteren reflektiven Skalen zur Erfassung anwendernaher Erfolgsfaktoren von Websites faktoriell valide, wird damit bestätigt.

Im nächsten Schritt wird die Messgüte der formativen Skala zur Usability (ufosV2fa) untersucht. Die Gewichte der formativen Usability-Indizes uf_in1 bis uf_in4 weisen hypothesenkonform positive Werte über $\gamma_n > 0,1$ mit signifikanten t -Werten auf. Im Falle des Usability-Indizes uf_in5 liegt das Gewicht mit $\gamma_5 = -0,016$ unter dem geforderten Wert $\gamma_n > 0,1$. Trotz der nicht zufriedenstellenden Kennwerte des Index uf_in5 wird von seinem Ausschluss aus inhaltlicher Sicht abgesehen, da in jeden Index sämtliche Indikatoren der ufosV2fa-Skala eingehen. Insgesamt wird die Messgüte der formativen Skala zur Usability (ufosV2fa) als gut angesehen.

Tab. 7-19: Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability

Latente Variable	ufosV2ra-Modell (reflektive Usability)		ufosV2fa-Modell (formative Usability)	
	AVE	Composite Reliability	AVE	Composite Reliability
Usability - ufosV2ra	0,805	0,966		
Nutzervergnügen	0,837	0,938	0,837	0,939
Vertrauen	0,830	0,936	0,830	0,936
Involvement	0,843	0,964	0,843	0,964
Nutzungsintention	0,879	0,956	0,879	0,956
Nutzerzufriedenheit	0,947	0,973	0,948	0,973

Anmerkung: Average Variance Extracted (AVE). Die Werte für das AVE-Maß sowie für die Composite Reliability werden getrennt für die Verwendung der reflektiven Usability-Skala ufosV2ra sowie für die formative Usability-Skala ufosV2fa angegeben.

7.6.2 Beurteilung der Strukturmodelle

Zur Evaluation des PLS-Strukturmodells werden zunächst Höhe und Signifikanz der Pfadkoeffizienten überprüft. Die Ergebnisse werden in Abbildung 7-3 dargestellt.

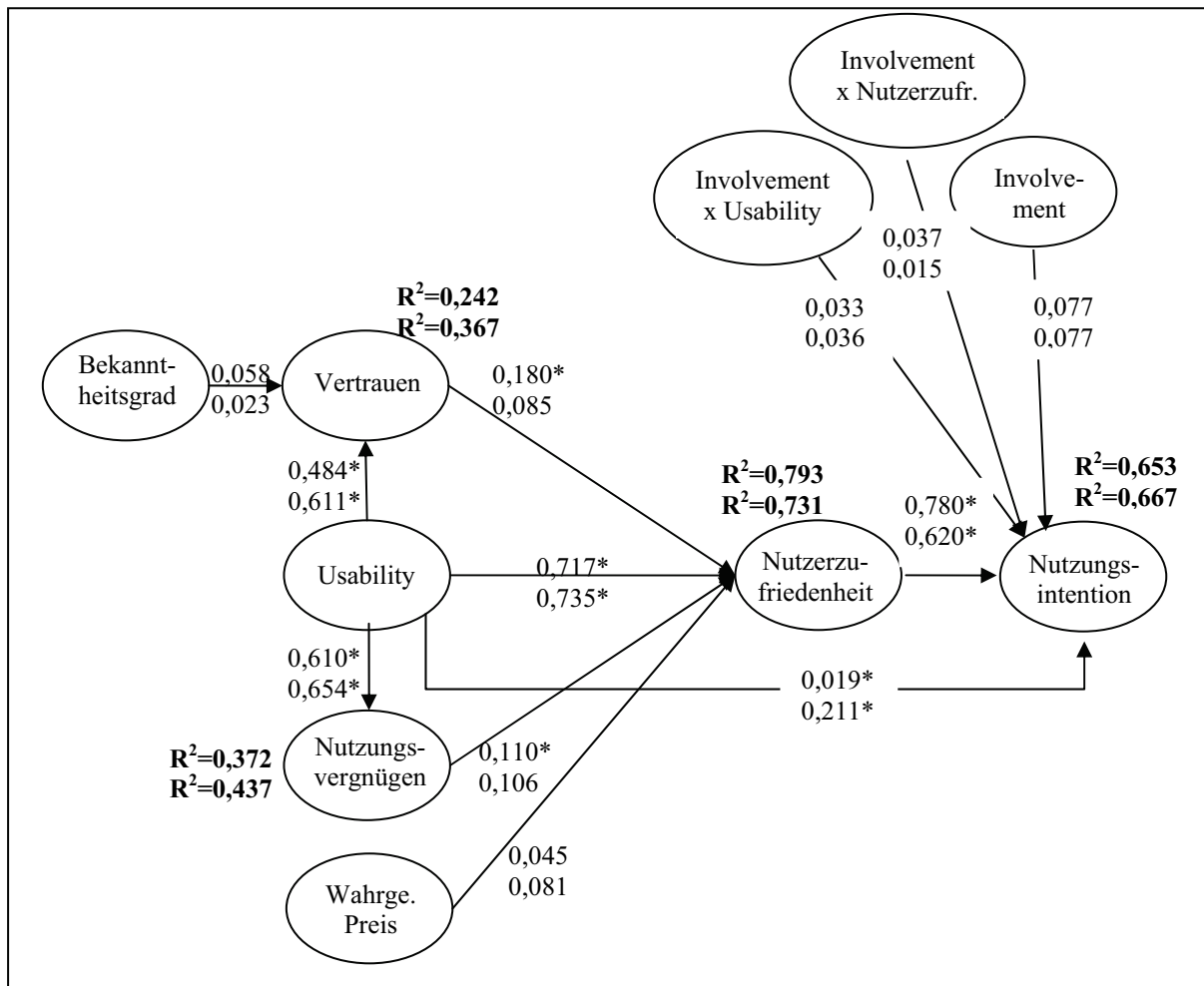


Abb. 7-3: Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen

Anmerkung: * signifikanter Pfadkoeffizient, fettgedruckt Determinationskoeffizient (R^2), Pfadkoeffizienten (β). Der jeweils obere Wert bezieht sich auf das ufosV2ra-Modell (Verwendung der reflektiven Usability-Skala ufosV2ra) während der untere Wert sich auf das ufosV2fa-Modell (Verwendung der formativen Usability-Skala ufosV2fa) bezieht.

In Tabelle 7-20 werden die Determinationskoeffizienten (R^2) der endogenen Variablen für beide PLS-Modelle dargestellt. Die abhängigen Variablen Nutzerzufriedenheit und Nutzungsintention werden durch die Prädiktoren in substantiellem Maße erklärt. Zwischen den beiden PLS-Modellen sind nur geringfügige Unterschiede festzustellen (*Nutzerzufriedenheit*: R^2 0,793 im ufosV2ra-Modell bzw. R^2 0,731 im ufosV2fa-Modell; *Nutzungsintention*: R^2 0,653 im ufosV2ra-Modell bzw. R^2 0,667 im ufosV2fa-Modell). Entsprechend Chin (1998, S. 323) sind Werte mit $R^2 > 0,66$ als substantiell gut zu bewerten. Entsprechend kann gefolgert werden, dass das nomologische Modell zur Bestimmung der Nutzerzufriedenheit und der Nutzungsintention als sehr gut geeignet anzusehen ist.

Der Anteil der aufgeklärten latenten Variable Nutzungsvergnügen ist ebenfalls als ausreichend hoch zu bezeichnen, wobei das Nutzungsvergnügen im ufosV2fa-Modell mit R^2 0,437 leicht besser erklärt wird, als im ufosV2ra-Modell (R^2 0,372). Die LV Vertrauen wird ebenfalls bei Verwendung der formativen Usability-Skala ufosV2fa besser erklärt und weist mit R^2 0,367 nach Chin (1998, S.323) einen „durchschnittlichen“ Wert auf. Im ufosV2ra-Modell hingegen wird mit R^2 0,242 ein eher schwacher Wert für das Bestimmtheitsmaß ausgewiesen.

Die Schätzrelevanz der Modelle in Bezug auf das Kriterium Nutzungsintention wird anhand des Stone-Geisser-Q-Kriteriums auf Grundlage einer Blindfolding-Prozedur berechnet. Die Schätzrelevanz Q^2 nimmt im Falle des ufosV2ra-Modells einen Wert von 0,429 an. Für das ufosV2fa-Modell liegt der Wert bei 0,574. Entsprechend kann in beiden Modellen von einer substantiellen Schätzrelevanz ausgegangen werden.

Tab. 7-20: Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: aufgeklärte Varianz (R^2)

<i>Endogene Variable</i>	<i>Modell mit ufosV2ra (reflektive Usability) R^2</i>	<i>Modell mit ufosV2fa (formative Usability) R^2</i>
Vertrauen	0,242	0,367
Nutzervergnügen	0,372	0,437
Nutzerzufriedenheit	0,793	0,731
Nutzungsintention	0,653	0,667

Anmerkung: Determinationskoeffizient (R^2).

Aus Tabelle 7-21 können die für beide Modelle geschätzten Pfadkoeffizienten unter Angabe der zugehörigen t -Werte entnommen werden. Zusätzlich wurden die Effektstärken f^2 für die exogenen latenten Variablen berechnet, um deren Einfluss auf die endogenen LV zu prüfen. Die Ausprägungen der Effektstärken und ihre Beurteilung werden in Tabelle E-29 im Anhang E dargestellt.

Tab. 7-21: Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Pfadkoeffizienten

<i>Zusammenhang</i>	<i>Modell mit ufosV2ra (reflektive Usability)</i>		<i>Modell mit ufosV2fa (formative Usability)</i>	
	β	t -Wert	β	t -Wert
Usability → Nutzerzufriedenheit	0,717	29,025	0,735	29,141
Usability → Nutzungsintention	0,019	0,168	0,211	16,731
Usability → Vertrauen	0,484	6,006	0,611	8,119
Usability → Nutzungsvergnügen	0,610	10,939	0,654	10,251
Vertrauen → Nutzerzufriedenheit	0,180	3,069	0,085	1,281
Nutzungsvergnügen → Nutzerzufriedenheit	0,110	1,803	0,106	1,436
Nutzerzufriedenheit → Nutzungsintention	0,780	5,959	0,620	5,101
Bekanntheit → Vertrauen	0,058	0,670	0,023	0,293
Wahrgeh. Preis → Nutzerzufriedenheit	0,045	0,845	0,081	1,608

Involvement → Nutzungsintention	0,077	1,021	0,077	1,110
Involvement x Usability → Nutzungsintention	0,033	0,041	0,036	0,305
Involv. x Nutzerzufried. → Nutzungsintention	0,037	0,324	0,015	0,198

Anmerkung: Pfadkoeffizient (β). Der t-Wert der t-Statistik in PLS folgt nur annähernd der t-Verteilung.

Von besonderem Interesse sind die Ergebnisse, die die LVn Usability und Involvement betreffen. In beiden ufosV2a-Modellen kann eine positive Wirkung der Usability sowohl auf die Nutzerzufriedenheit als auch auf die Nutzungsintention bestätigt werden. Im Falle der Nutzerzufriedenheit erweisen sich sowohl die reflektive Skala ufosV2ra als auch die formative Skala ufosV2fa mit signifikanten Pfadkoeffizienten von β 0,717 bzw. β 0,735 als Prädiktoren mit nahezu vergleichbar hohem Einfluss auf die Nutzerzufriedenheit. Die Ausprägungen der Effektstärke f^2 weisen den Einfluss der Usability auf die Nutzerzufriedenheit für beide Skalen als hoch aus (f^2 1,367 im Fall von ufosV2ra bzw. f^2 0,828 ufosV2fa). Im Falle der Nutzungsintention fallen die Werte der Pfadkoeffizienten deutlich niedriger aus. Während für die formative Skala ufosV2fa ein signifikanter Einfluss β 0,211 auf die Nutzungsintention festgestellt werden kann, erweist sich für die reflektive Skala ufosV2ra kein signifikanter Zusammenhang zwischen den LVn Usability und Nutzungsintention. Dies wird auch durch die Werte der Effektstärke bestätigt. Im Falle der formativen Skala ufosV2fa fällt diese mit f^2 0,039 gering aus, während im Falle der reflektiven Skala ufosV2ra ein Einfluss der Usability auf die Nutzungsintention nicht nachgewiesen werden kann. Die **Hypothesen H5** und **H6b** (siehe Tabelle 7-22) können anhand dieser Ergebnisse somit bestätigt werden.

Desweiteren zeigt sich in beiden PLS-Modellen ein hoher Zusammenhang für den Pfad zwischen den LVn Usability und Vertrauen. Dabei fällt der Pfadkoeffizient zur Wirkungsbeziehung der Usability auf das Vertrauen im Falle der ufosV2fa-Variable mit β 0,611 höher aus als im Falle der ufosV2ra-Variable β 0,484. Dies kommt auch in unterschiedlichen Effektstärken zum Ausdruck, welche für die ufosV2fa-Skala hoch (f^2 0,566) und bei der ufosV2ra-Skala mittel hoch (f^2 0,307) ausfällt.

Ebenso kann ein hoher Zusammenhang für den Pfad zwischen den LVn Usability und Nutzervergnügen in beiden Modellen festgestellt werden. Auch hier fällt der Pfadkoeffizient im ufosV2ra-Modell mit β 0,610 etwas niedriger aus als für das ufosV2fa-Modell mit β 0,654. Die Effektstärken weisen sowohl für die ufosV2fa-Skala (f^2 0,774) als auch für die ufosV2ra-Skala (f^2 0,592) einen hohen Einfluss aus. Die **Hypothesen H8** und **H9**, in denen diese Konstruktbeziehungen angenommen werden, können damit bestätigt werden. Die Hypothese **H10** zum Zusammenhang der Ästhetik und der Usability kann aus den genannten Gründen nicht überprüft werden.

Für den Pfad zwischen den LVn Involvement und Nutzungsintention kann in beiden PLS-Modellen bei einem Pfadkoeffizienten von β 0,077 kein signifikanter Zusammenhang festgestellt werden. Dies wird auch durch die niedrigen Ausprägungen der Effektstärke

bestätigt. Im Falle der formativen Skala ufosV2fa fällt diese mit f^2 0,024 ebenso wie auch im Falle der reflektiven Skala ufosV2ra mit f^2 0,017 gering aus. Entsprechend fallen auch die moderierenden Effekte der LV Involvement auf die Beziehungen zwischen Usability und Nutzungsintention bzw. zwischen Nutzerzufriedenheit und Nutzungsintention bei nicht signifikanten t -Werten gering aus.

Um den angenommenen moderierenden Effekt der LV Involvement auf die Wirkungsbeziehung zwischen den LVn Usability und Nutzungsintention weitergehend zu überprüfen, wurde ein Extremgruppenvergleich vorgenommen. Herangezogen wurde das obere und untere Fünftel der Ausprägungen der LV Involvement (N 204), für die separate Schätzungen vorgenommen wurden. Aus Tabelle E-57 im Anhang E können die für beide Modelle geschätzten Pfadkoeffizienten unter Angabe der zugehörigen t -Werte entnommen werden. Bei Verwendung nicht standardisierter Originaldaten erweist sich für das formative Modell ein signifikanter Einfluss der LV Involvement auf die Nutzungsintention mit einem Pfadkoeffizienten β 0,123 ($p < .05$). Ebenso kann eine moderierende Wirkung des Involvement auf die Beziehung zwischen Usability und Nutzungsintention bei einem signifikanten Pfadkoeffizienten von β 0,303 ($p < .05$) festgestellt werden. Dies wird durch den Wert der Effektstärke von f^2 0,030 bestätigt. Allerdings kann bei Verwendung standardisierter Werte für beide PLS-Modelle kein signifikanter moderierender Effekt der LV Involvement festgestellt werden. Da das Vorgehen anhand standardisierter Werte bei der Berechnung moderierender Effekte in PLS-Modellen dem allgemein angewandten Standardverfahren in der Literatur entspricht, wird diesem Ergebnis gefolgt (Ringle, 2004). Die **Hypothesen H 16** und **H17** können somit nicht bestätigt werden.

Bis auf die Hypothesen **H10** zum Zusammenhang der Ästhetik und Usability, **H6a** zum Zusammenhang der Usability und der Nutzungsintention im ufosV2ra-Modell sowie **H16** zur moderierenden Wirkung des Involvement auf die Beziehung zwischen Usability und Nutzungsintention bestätigen sich in beiden PLS-Modellen alle Annahmen, die die latente Variable Usability betreffen. Die Ergebnisse stützen damit sowohl die nomologische Validität als auch die konkurrente Kriteriumsvalidität für die Skalen ufosV2ra und ufosV2fa.

Von nachgeordneter Bedeutung sind die Ergebnisse, die die übrigen Modellzusammenhänge betreffen. Am deutlichsten bestätigt sich der im nomologischen Modell angenommene Zusammenhang zwischen der Nutzerzufriedenheit und der Nutzungsintention. Dabei fällt der Pfadkoeffizient im ufosV2ra-Modell mit β 0,780 bei einem t -Wert von t 5,959 höher aus als im ufosV2fa-Variable mit β 0,620 und einem t -Wert von t 5,101. Die **Hypothese H18** kann somit bestätigt werden. Von den übrigen im Modell angenommenen Zusammenhängen bestätigen sich lediglich die Pfadbeziehungen zwischen den LVn Vertrauen und Nutzerzufriedenheit sowie zwischen den LVn Nutzungsvergnügen und Nutzerzufriedenheit für das ufosV2ra-Modell. Während im Falle des Vertrauens ein Pfadkoeffizient β 0,180 bei einem t -Wert von t 3,069 festgestellt werden kann, zeigt sich im Falle des Nutzervergnügens

nur ein niedriges β 0,110 mit einem t-Wert von t 1,803, der nur knapp über dem kritischen Wert von 1,66 liegt. Die **Hypothesen H12** und **H13** können damit nur für das ufosV2ra-Modell bestätigt werden. Die Hypothese **H15** zum Zusammenhang zwischen Reputation und Vertrauen kann wie auch die Hypothesen zur Ästhetik nicht überprüft werden. Zusammenfassend können die **Hypothesen H6a, H10, H11, H14, H15, H16 und H17** auf Grundlage der PLS-Analysen nicht angenommen werden (siehe Tabelle 7-22).

Tab. 7-22: Übersicht über die mit Hilfe der PLS-Modelle überprüften Hypothesen

Nr.	Annahme	Bestätigung
H5a	ufosV2ra → Nutzerzufriedenheit	+
H5b	ufosV2fa → Nutzerzufriedenheit	+
H6a	ufosV2ra → Nutzungsintention	n.s.
H6b	ufosV2fa → Nutzungsintention	+
H8a	ufosV2ra → Vertrauen	+
H8b	ufosV2fa → Vertrauen	+
H9a	ufosV2ra → Nutzungsvergnügen	+
H9b	ufosV2fa → Nutzungsvergnügen	+
H10a	Ästhetik → ufosV2ra	kann nicht überprüft werden
H10b	Ästhetik → ufosV2fa	kann nicht überprüft werden
H11	Ästhetik → Nutzerzufriedenheit	kann nicht überprüft werden
H12	Vertrauen → Nutzerzufriedenheit	+ (refl.)
H13	Nutzungsvergnügen → Nutzerzufriedenheit	+ (refl.)
H14	Ästhetik → Nutzungsvergnügen	kann nicht überprüft werden
H15	Reputation → Vertrauen	kann nicht überprüft werden
H16	Involvement x Usability → Nutzungsintention	n.s.
H17	Involv. x Nutzerzufriedenh. → Nutzungsintention	n.s.
H18	Nutzerzufriedenheit → Nutzungsintention	+

Anmerkung: Zusammenhang wird durch die berechneten PLS-Modelle bestätigt (+), Zusammenhang wird nur durch das berechnete ufosV2ra-Modell bestätigt (+ refl.), nicht signifikant (n.s.).

7.7 Logistische Regression zur Vorhersage der Auswahlentscheidung

Im letzten Schritt der Datenanalyse soll geprüft werden, inwiefern die beiden ufosV2a-Skalen bzw. die Nutzungsintention einen Einfluss auf die tatsächliche Auswahlentscheidung von Nutzern einer Krankenkassenwebsite haben (Hypothese H7 zum Zusammenhang zwischen den Usability-Skalen und der Auswahlentscheidung bzw. Hypothese H19 zum Zusammenhang zwischen der Nutzungsintention und der Auswahlentscheidung). Diese Hypothesen konnten mit Hilfe des PLS-Strukturgleichungsansatzes nicht überprüft werden, da PLS keine Einbindung kategorialer Daten auf Ebene der abhängigen Variablen zulässt. Aus diesem Grunde erfolgt die Überprüfung der Hypothesen H7 und H19 mit Hilfe einer binomialen logistischen Regression (LR; Abschnitt 6.7).

In der LR wird eine Vorhersage berechnet, ob die Versuchsteilnehmer die erste oder die zweite Website auswählen. Die *Auswahlentscheidung* mit den Ausprägungen *Auswahl Website1* und *Auswahl Website2* stellt dabei die AV dar. Die Berechnung der LR wird anhand eines Datensatzes mit zwei bewerteten Websites je Fall vorgenommen. Dabei wird sequentiell vorgegangen, indem die Prädiktoren blockweise in die Analyse einbezogen werden

(Tabachnick & Fidell, 2006, S. 481ff.). Bei jedem der vier Analyseschritte werden die Veränderungen in den Gütekriterien des Gesamtmodells betrachtet. Hieraus werden Rückschlüsse auf die Stärke des Einflusses der einzelnen Prädiktoren gezogen. Abbildung 7-4 gibt einen Überblick über den Ablauf der sequentiellen logistischen Regressionsanalyse. Dabei werden die nachfolgend dargestellten Schritte durchlaufen.

Schritt 1: Im ersten Schritt der Analyse wird ein Modell zugrunde gelegt, in dem als Prädiktoren Variablen zur Demographie und zu Vorerfahrungen im Umgang mit Internetanwendungen herangezogen werden. Im Einzelnen handelt es sich um die demographischen Variablen *Alter*, *Einkommen* und *Geschlecht* sowie zur Vorerfahrung um die Variablen *Anzahl der Online-Einkäufe* sowie *Anzahl der Internetrecherchen*.

Schritt 2a: In Schritt 2a der sequentiellen LR werden zusätzlich zu den Variablen aus Schritt 1 die Skalenwerte der *ufosV2ra*-Skala als Prädiktoren aufgenommen. Auf Grundlage des Ergebnisses soll die Hypothese H7a zur prädiktiven Validität der reflektiven Usability-Skala überprüft werden. Diese besagt, dass die Skala *ufosV2ra* substantziell zur Auswahlentscheidung der Nutzer einer Krankenkassenwebsite beiträgt.

Schritt 2b: Für die formative Usability-Skala *ufosV2fa* wurde bezüglich ihrer prädiktiven Validität eine analoge Annahme getroffen (Hypothese H7b). Entsprechend werden in Schritt 2b anstelle der Skalenwerte der *ufosV2ra*-Skala die der *ufosV2fa*-Skala einbezogen. Als Skalenwerte kommen die Latenten Variablen Scores aus der PLS-Analyse zum Zwei-Konstrukt-Modell (Abschnitt 7.5.3) zum Einsatz.

Schritt 2c: Im letzten Schritt der Analyse wird untersucht, inwiefern sich die Hypothese H19 zum Zusammenhang zwischen der Nutzungsintention und der Auswahlentscheidung bestätigt. Hierzu werden anstelle der Skalenwerte der *ufosV2a*-Skalen die der Nutzungsintention als Prädiktoren für die Auswahlentscheidung hinzugezogen.

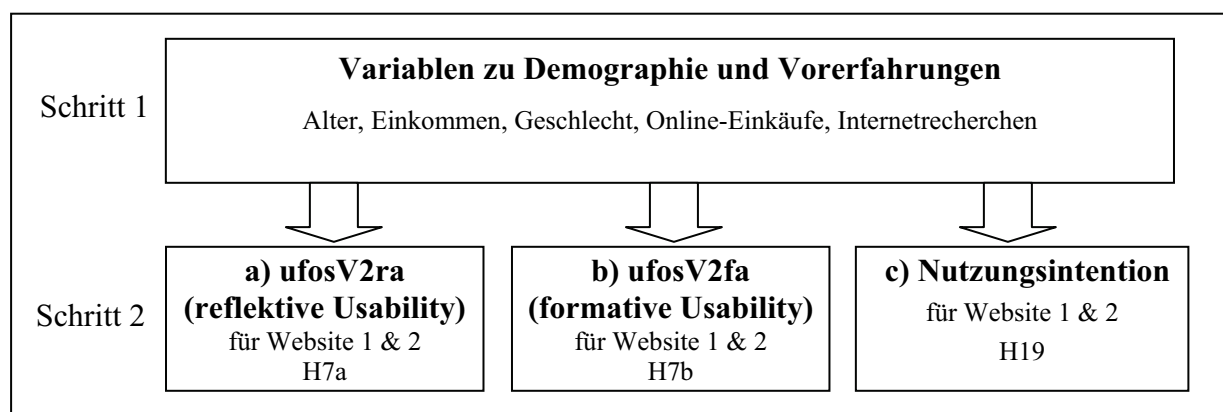


Abb. 7-4: Schritte der sequentiellen logistischen Regressionsanalyse unter Angabe der jeweils eingesetzten Prädiktoren

7.7.1 Prüfung der Voraussetzungen und Ausreißerdiagnostik

Zunächst erfolgt eine Überprüfung der in die LR eingehenden Prädiktorvariablen auf Multikollinearität sowie eine Ausreißerdiagnose. Zur Überprüfung von Multikollinearität zwischen den LR-Prädiktoren werden die in Abschnitt 6.6.2.3 dargestellten Diagnoseverfahren eingesetzt. Zunächst werden die bivariaten Korrelationen zwischen den Prädiktorvariablen untersucht (Tabelle E-30 im Anhang E). Lediglich zwischen dem Alter und dem Einkommen kann erwartungsgemäß eine relativ hohe Korrelation festgestellt werden ($r = 0,44$; $p < .01$). Des Weiteren werden in Tabelle E-31 im Anhang E die Toleranz- und VIF-Werte der Prädiktorvariablen sowie die maximalen Konditionsindizes dargestellt.

Alle Toleranzwerte liegen oberhalb von 0,7, sämtliche VIF-Werte unterhalb des kritischen Wertes von 2. Der maximale Konditionsindex liegt bei 19,015. Somit kann gefolgert werden, dass zwischen den einbezogenen Prädiktorvariablen kein kritisches Maß an Multikollinearität besteht und damit die Voraussetzung zur Berechnung einer LR gegeben ist. Anschließend an die Überprüfung auf Multikollinearität wird eine Ausreißerdiagnose vorgenommen. Es wird untersucht, inwiefern Ausreißer in den Daten den in dem untersuchten Modell unterstellten Zusammenhängen nicht entsprechen und das Untersuchungsergebnis damit verzerren. Hierfür werden für beide Logit-Modelle die vorhergesagten Variablenausprägungen der AV mit den beobachteten Werten verglichen. Es wird eine binomiale logistische Regression berechnet, bei der die Entscheidung *Auswahl Website 1 vs. Website 2* als Kriterium dient (siehe Abschnitt 6.7). In Tabelle E-32 im Anhang E findet sich eine Auflistung der Probanden, die ein standardisiertes Residuum mit einem Wert größer 1 aufweisen. Die Entscheidung einen Ausreißer zu eliminieren, sollte allerdings nur dann getroffen werden, wenn es Hinweise darauf gibt, dass der jeweilige Fall nicht der Population zuzurechnen ist (vgl. Tabachnik & Fidell, 2006, S. 73). Entsprechend werden die anhand der Ausreißerdiagnostik identifizierten Fälle mit den Fällen abgeglichen, die bereits im Verlauf der Analyse auf problematisches Antwortverhalten auffällig geworden sind (Abschnitt 7.1.5). In Tabelle E-33 im Anhang E werden die Codes von drei Versuchsteilnehmern aufgelistet, die in beiden Analysen als auffällig identifiziert wurden. Die drei Fälle werden aus der nachfolgend durchgeführten LR ausgeschlossen, sodass sich N auf 241 Fälle reduziert.

7.7.2 Sequentielle Berechnung der LR

In Tabelle 7-23 werden die Ergebnisse aus der sequentiell durchgeführten LR im Überblick dargestellt. Eine detaillierte Darstellung der Ergebnisse der Schritte 1 – 2c erfolgt anhand der Tabellen E-34 bis E-49 im Anhang E.

Schritt 1: Der Likelihood-Quotiententest zur Prüfung der globalen Anpassungsgüte ergibt ein nicht signifikantes Ergebnis ($\chi^2 = 3,548$; $df = 5$). Das Nagelkerke $R^2 = 0,02$ bringt eine niedrige Gesamtgüte des Ausgangsmodells zum Ausdruck. Der Anteil korrekt klassifizierter Fälle liegt bei 53,80% (siehe Tabelle E-35 im Anhang E). Die Vorhersagegüte ist im Vergleich zur

Zufallswahrscheinlichkeit von 50% somit nicht als hoch zu bewerten. Auf Variablenebene durchgeführte Likelihood-Quotienten-Tests zeigen keine Signifikanz für die Variablen Alter, Einkommen, Geschlecht, Erfahrung mit Online-Einkäufen sowie Erfahrung in der Durchführung von Internetrecherchen (siehe Tabelle E-36 im Anhang E). Dies spricht dafür, dass kein Einfluss der Variablen auf die Auswahlentscheidung besteht.

Tab. 7-23: Kennwerte der sequentiellen LR Schritte 1 bis 2c

Prädiktoren	χ^2 -Wert	Df	Modell χ^2 -Wert	Δ Modell χ^2 -Wert	Nagel- Kerke R^2	Δ Nagel- kerke R^2	Vorher- sage in %
<i>Schritt 1: Demographie etc.</i>							
Alter	0,907	1					
Einkommen	0,599	1					
Geschlecht	0,027	1					
Online-Einkäufe	0,506	1					
Internetrecherchen	2,374	1					
Alle UVn Modell 1			3,548 (df=5)		0,020		53,80
<i>Schritt 2a: ufosV2ra-Skala</i>							
Alter	4,474*	1					
Geschlecht	1,265	1					
Einkommen	2,624	1					
Online-Einkäufe	0,509	1					
Internetrecherchen	5,291*	1					
ufosV2ra Website 1	57,564***	1					
ufosV2ra Website 2	147,520***	1					
Alle UVn Modell 2a im Vergleich zu Schritt 1			211,035*** (df=7)	207,487*** (df=2)	0,781	0,761	89,60
<i>Schritt 2b: ufosV2fa-Skala</i>							
Alter	9,256**	1					
Geschlecht	0,104	1					
Einkommen	2,115	1					
Online-Einkäufe	1,551	1					
Internetrecherchen	6,225*	1					
ufosV2fa Website 1	79,877***	1					
ufosV2fa Website 2	160,220***	1					
Alle UVn Modell 2b im Vergleich zu Schritt 1			239,185*** (df=7)	235,637*** (df=2)	0,843	0,823	93,80
<i>Schritt 2c: Nutzungsintention</i>							
Alter	0,814	1					
Geschlecht	1,281	1					
Einkommen	0,825	1					
Online-Einkäufe	2,855	1					
Internetrecherchen	10,222***	1					
Nutzungsint.Website 1	73,497***	1					
Nutzungsint.Website 2	182,743***	1					
Alle UVn Modell 2c im Vergleich zu Schritt 1			232,063*** (df=7)	228,515*** (df=2)	0,828	0,808	92,90

Anmerkung: * $p < .05$, ** $p < .01$, *** $p < .001$, χ^2 -Wert des Likelihood-Quotienten-Tests auf Variablenebene (χ^2 -Wert); Freiheitsgrade (df); Modell- χ^2 -Wert des Likelihood-Quotienten-Tests zur globalen Modellanpassung, Inkrement des Parameters (Δ); Pseudo-Determinationskoeffizient (Nagelkerke R^2), Inkrement des Parameters (Δ); Prozent korrekt vorhergesagter Fälle (Vorhersage in %).

Schritt 2a: Die Einbeziehung der Skalenwerte der reflektiven Usability-Skala *ufosV2ra* führt zu einer signifikanten Verbesserung der Prognosekraft des Modells. Der Anteil korrekt vorhergesagter Fälle ist mit 89,60% deutlich höher als beim Ausgangsmodell (siehe Tabelle

E-39 im Anhang E). Das Inkrement des Modell- χ^2 -Wertes des Likelihood-Quotiententests zur globalen Modellanpassung ist hoch und signifikant ($\Delta\chi^2$ 207,487; df 2; $p < .001$; siehe Tabelle E-38 im Anhang E). Das Inkrement zur Varianzaufklärung fällt in Schritt 2a mit ΔR^2 0,761 ebenfalls hoch aus. Die Skalenwerte der reflektiven Usability-Skala *ufosV2ra* erweisen sich in den auf Variablenebene durchgeführten Likelihood-Quotiententests für beide Websites als signifikant (χ^2 57,564; df 1; $p < .001$ bzw. χ^2 147,520; df 1; $p < .001$; siehe Tabelle E-40 im Anhang E). Die **Hypothese H7a**, wonach die Skala *ufosV2ra* eine hohe Vorhersagekraft für die Auswahlentscheidung von Nutzern einer Krankenkassenwebsite hat, kann somit als bestätigt betrachtet werden.

Schritt 2b: Auch die Einbeziehung der Latenten Variablen Scores der formativen Usability-Skala *ufosV2fa* führt zu einer signifikanten Verbesserung der Prognosekraft des Modells. Der Anteil korrekt klassifizierter Fälle ist mit 93,80% ebenfalls deutlich höher als im Ausgangsmodell (siehe Tabelle E-43 im Anhang E). Das Inkrement des Modell- χ^2 -Wertes des Likelihood-Quotiententests zur globalen Modellanpassung erweist sich als hoch und signifikant ($\Delta\chi^2$ 235,637; df 2; $p < .001$; siehe Tabelle E-42 im Anhang E). Das Nagelkerke R^2 nimmt einen hohen Wert von 0,843 an, sodass das Inkrement in Schritt 2b im Vergleich zum Ausgangsmodell mit ΔR^2 0,823 hoch ausfällt. Auf Variablenebene durchgeführte Likelihood-Quotiententests weisen die Skalenwerte der formativen Usability-Skala *ufosV2fa* für beide Websites als signifikante Prädiktoren aus (χ^2 79,877; df 1; $p < .001$ bzw. χ^2 160,220; df 1; $p < .001$; siehe Tabelle E-44 im Anhang E). Die **Hypothese H7b**, wonach die Skala *ufosV2fa* eine hohe Vorhersagekraft für die Auswahlentscheidung von Nutzern einer Krankenkassenwebsite hat, kann somit ebenfalls bestätigt werden.

Schritt 2c: Im letzten Schritt der Analyse werden anstelle der Skalenwerte der *ufosV2a*-Skalen die der Nutzungsintention als Prädiktoren für die Auswahlentscheidung hinzugezogen. Die Vorhersagegüte des Modells wird hierdurch gegenüber dem Ausgangsmodell ebenfalls deutlich verbessert. Das Inkrement des Modell- χ^2 -Wertes des Likelihood-Quotiententests zur globalen Modellanpassung ist hoch und signifikant ($\Delta\chi^2$ 228,515; df 2; $p < .001$; siehe Tabelle E-46 im Anhang E). Das Nagelkerke R^2 weist einen hohen Wert von 0,828 aus. Der Anteil korrekt klassifizierter Fälle ist mit 92,90% ebenfalls deutlich höher als im Ausgangsmodell (siehe Tabelle E-47 im Anhang E). Die Skalenwerte der Nutzungsintention erweisen sich in den auf Variablenebene durchgeführten Likelihood-Quotiententests für beide Websites als signifikant (χ^2 73,497; df 1; $p < .001$ bzw. χ^2 182,743; df 1; $p < .001$; siehe Tabelle E-48 im Anhang E). Die **Hypothese H19**, wonach die Nutzungsintention einen signifikanten Prädiktor für die Auswahlentscheidung darstellt, kann somit bestätigt werden.

Es kann als Fazit der LR festgestellt werden, dass beide *ufosV2a*-Skalen eine hohe Vorhersagekraft in Bezug auf die Auswahlentscheidung von Nutzern einer Krankenkassenwebsite besitzen. Die Hypothesen H7a und H7b werden damit bestätigt. Des Weiteren konnte

ein starker Einfluss der Nutzungsintention auf die Auswahlentscheidung der Nutzer nachgewiesen werden, wodurch die Hypothese H19 ebenfalls bestätigt werden konnte.

7.8 Übersicht der Befunde

(1) *Festlegung der Items der ufosV2a-Skalen*: Es wurden acht reflektive und 31 formative Usability-Items in die Datenerhebung einbezogen. Auf Grundlage der Durchführung einer EFA wurde ein reflektiver Indikator eliminiert (siehe Abschnitt 7.3.1). Die endgültige ufosV2ra-Skala setzt sich somit aus sieben Items zusammen, welche in Tabelle 7-24 dargestellt werden.

Tab. 7-24: Items der endgültigen reflektiven Skala ufosV2ra

<i>Code</i>	<i>Reflektives Item</i>
ur1a	Diese Website macht das Suchen nach Krankenversicherungsinformationen im Netz leicht.
ur2ac	Es ist mir zu kompliziert, diese Website zu nutzen (R).
ur3a	Man kann sich schnell einen Überblick über die Website verschaffen
ur4a	Der Umgang mit der Website ist leicht zu erlernen.
ur5a	Auf dieser Website kann eine erfolgreiche Suche zügig erledigt werden.
ur6a	Insgesamt bin ich mit der Benutzerfreundlichkeit dieser Website zufrieden.
ur8a	Ich kann diese Website so benutzen, wie ich es erwarte.

Basierend auf der Analyse inhaltlicher Überschneidungen von Indikatoren wurden im Verlauf der Auswertung vier der 31 formativen Indikatoren eliminiert (siehe Abschnitt 7.5.1). Tabelle 7-25 enthält die 27 Items der endgültigen formativen Skala ufosV2fa.

Tab. 7-25: Items der endgültigen formativen Skala ufosV2fa

<i>Code</i>	<i>Formatives Item</i>
uf1a	Bei der Nutzung der Website kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.
uf2ac	Ich muss mir viel merken, wenn ich auf dieser Website etwas suche.
uf3a	Es ist einfach, zwischen den verschiedenen Seiten hin- und her zu springen.
uf4a	Zentrale Seiten (z.B. die Startseite) sind jederzeit schnell zu erreichen.
uf5a	Ich kann jede Aktion wieder rückgängig machen.
uf6a	Mir ist bei der Nutzung dieser Website stets klar, welche Handlungen ich als nächstes durchführen kann.
uf7ac	Ich weiß manchmal nicht, wo ich mich innerhalb der Website befinde.
uf9ac	Es sind zu viele Schritte nötig, um das Gesuchte zu finden.
uf10a	Ich finde mich auf den einzelnen Seiten gut zurecht.
uf12ac	Es werden auf dieser Website häufig zu viele Informationen dargestellt
uf13ac	Die Texte sind manchmal nur schwer lesbar.
uf14ac	Das optische Design der Website stört mich bei der Nutzung.
uf15a	Wenn ich auf dieser Website auf etwas klicke, passiert genau das, was ich erwarte.
uf16ac	Manchmal finde ich die verwendeten Texte inhaltlich unverständlich.
uf17a	Die Benennungen sind so gewählt, dass ich die Website gut benutzen kann.
uf18a	Es ist schnell erkennbar, welche Inhalte auf der Website angeboten werden.
uf20a	Die Suchfunktionen liefern mir sinnvolle Ergebnisse.
uf21a	Die Angebote sind in sinnvolle Kategorien eingeteilt.

uf22a	Ich bin mit den Beschreibungen der Angebote zufrieden.
uf23a	Ich finde die verwendeten Bilder und Grafiken hilfreich.
uf24a	Die Website empfiehlt mir interessante Angebote, nach denen ich ursprünglich nicht gesucht habe.
uf25ac	Es dauert zu lange, bis die Website auf meine Eingaben reagiert.
uf26a	Die Startseite stellt einen guten Ausgangspunkt für meine Nutzung der Website dar.
uf28a	Insgesamt verfügt die Website über ein reichhaltiges Informationsangebot.
uf29a	Ich bin mit den Informationen, die mir der Anbieter zur Verfügung stellt, sehr zufrieden.
uf30a	Diese Website ist eine Quelle hochwertiger Angebote.
uf31a	Ich erhalte nützliche Informationen zum Unternehmen, das die Website betreibt.

(2) *Ergebnisse zu Güte und zu Wirkungszusammenhängen der Usability-Skalen*: Es wurden diesbezüglich diverse Hypothesen empirisch untersucht. Tabelle 7-26 gibt einen Überblick, inwiefern die in Kapitel drei vorgestellten inhaltlichen Hypothesen H1 bis H10 bestätigt werden konnten.

Tab. 7-26: Hypothesen zu Güte und Wirkungszusammenhängen der ufosV2a-Skalen

Nr.	Hypothese (siehe Abschnitt x.x)	Verfahren Abschnitt	Befund	Empirische Bestätigung
H1	Interne Konsistenz der Skala ufosV2ra	α 7.3.2	$\alpha=0,96$	ja
		Composite Reliability 7.5.3	Compos. Reliability =0,97	ja
H2	Faktorielle Validität der Skala ufosV2ra	EFA 7.3.1	Vier anstelle der fünf erwarteten Faktoren	ja (eingeschr.)
		CFA 7.3.3	Gütemaße	ja (eingeschr.)
		PLS: Komplexes Modell 7.6.1	Kreuzladungen der Indikatoren auf LVn	ja
H3	Differenzierungsfähigkeit der Skala ufosV2ra zwischen Websites	ANOVA 7.3.4	$F=15,25$ ($p < .001$)	ja
H4	ufosV2fa \rightarrow ufosV2ra	PLS: Zwei-Konstrukt-Modell 7.5.3	$\beta=0,915$ (sign.); $R^2=0,838$	ja
H5a	ufosV2ra \rightarrow Nutzerzufriedenheit	PLS: Komplexes Modell 7.6.2	$\beta=0,717$ (sign.)	ja
H5b	ufosV2fa \rightarrow Nutzerzufriedenheit		$\beta=0,735$ (sign.)	ja
H6a	ufosV2ra \rightarrow Nutzungsintention		$\beta=0,019$ (nicht sign.)	nein
H6b	ufosV2fa \rightarrow Nutzungsintention		$\beta=0,211$ (sign.)	ja
H7a	ufosV2ra \rightarrow Auswahlentscheidung	LR 7.7	$\Delta\chi^2=207,487$; $\Delta R^2=0,761$; $AkkF=89,60\%$	ja
H7b	ufosV2fa \rightarrow Auswahlentscheidung		$\Delta\chi^2=235,637$; $\Delta R^2=0,823$; $AkkF=93,80\%$	ja

H8a	ufosV2ra → Vertrauen	PLS: Komplexes Modell 7.6.2	$\beta=0,484$ (sign.)	ja
H8b	ufosV2fa → Vertrauen		$\beta=0,611$ (sign.)	ja
H9a	ufosV2ra → Nutzungsvergnügen		$\beta=0,610$ (sign.)	ja
H9b	ufosV2fa → Nutzungsvergnügen		$\beta=0,654$ (sign.)	ja
H10a	Ästhetik → ufosV2ra	-	-	k.A.
H10b	Ästhetik → ufosV2fa	-	-	k.A.

Anmerkung: Hypothese (H), Partial Least Squares (PLS), binomiale logistische Regression (LR), einfaktorielle Varianzanalyse (ANOVA), explorative Faktorenanalyse (EFA), Cronbachs Alpha (α), konfirmatorische Faktorenanalyse (CFA), Anteil korrekt klassifizierter Fälle (AkkF), keine Aussage möglich (k.A.).

(3) *Ergebnisse zu weiteren Wirkungszusammenhängen innerhalb des nomologischen Modells:* Des Weiteren wurden Hypothesen (H11 bis H19) zu Zusammenhängen zwischen den übrigen Konstrukten innerhalb des nomologischen Modells geprüft. Tabelle 7-27 gibt einen Überblick über die Ergebnisse hierzu.

Tab. 7-27: Weitere Hypothesen zu Zusammenhängen des nomologischen Modells

Nr.	Hypothese	Verfahren Abschnitt	Befund	Empirische Bestätigung
H11	Ästhetik → Nutzerzufriedenheit	-	-	k.A.
H12	Vertrauen → Nutzerzufriedenheit	PLS: Komplexes Modell 7.6.2	$\beta=0,180$ (sign.) $\beta=0,085$ (nicht sign.)	ja n.s.
H13	Nutzungsvergnügen → Nutzerzufriedenheit		$\beta=0,110$ (sign.) $\beta=0,106$ (nicht sign.)	ja n.s.
H14	Ästhetik → Nutzungsvergnügen		-	-
H15	Reputation → Vertrauen	-	-	k.A.
H16	Involvement x Usability → Nutzungsintention	PLS: Komplexes Modell 7.6.2	$\beta=0,033$ (nicht sign.) $\beta=0,036$ (nicht sign.)	n.s.
H17	Involvement x Nutzerzufriedenheit → Nutzungsintention		$\beta=0,037$ (nicht sign.) $\beta=0,015$ (nicht sign.)	n.s.
H18	Nutzerzufriedenheit → Nutzungsintention		$\beta=0,780$ (sign.) $\beta=0,620$ (sign.)	ja
H19	Nutzungsintention → Auswahlentscheidung	LR 7.7	$\Delta\chi^2=228,515$; $\Delta R^2=0,808$; AkkF=92,90%	ja

Anmerkung: Hypothese (H), Partial Least Squares (PLS), binomiale logistische Regression (LR), Befund: oberer Wert/ ufosV2ra-Model, unterer Wert/ ufosV2fa-Model, Anteil korrekt klassifizierter Fälle (AkkF), nicht signifikant und deshalb nicht bestätigt (n.s.), keine Aussage möglich (k.A.).

8. Diskussion

Im Zuge der Diskussion wird zunächst eine Interpretation der Untersuchungsergebnisse vorgenommen. Hierfür werden in Abschnitt 8.1 die Hypothesen im nomologischen Modell aufgegriffen. Daran anschließend erfolgt eine kritische Auseinandersetzung mit dem methodischen Vorgehen. Es werden die interne und externe Validität der Untersuchungsergebnisse sowie das Vorgehen der statistischen Auswertung betrachtet (Abschnitt 8.2). Daran anschließend wird in Abschnitt 8.3 eine zusammenfassende Bewertung der Untersuchungsergebnisse vorgenommen. In Abschnitt 8.4 erfolgt eine Interpretation der Ergebnisse aus der Perspektive der Erfolgsfaktorenforschung. Die Implikationen der Befunde für Forschung und Praxis werden in Abschnitt 8.5 dargestellt. Ihren Abschluss erfährt die Arbeit in Abschnitt 8.6 mit einem Ausblick auf mögliche weiterführende Fragestellungen.

8.1 Diskussion der Befunde

Nachfolgend werden zunächst die Befunde im Zusammenhang mit den ufosV2a-Skalen betrachtet (Abschnitt 8.1.1). Daran anschließend folgt eine kritische Auseinandersetzung mit den Befunden zu weiteren Hypothesen im nomologischen Modell (Abschnitt 8.1.2) sowie abschließend mit den Befunden zu möglichen Effekten aufgrund der Itemreihenfolge (Abschnitt 8.1.3).

8.1.1 Befunde zur Güte der ufosV2a-Skalen

8.1.1.1 Reliabilität der reflektiven Usability-Skala ufosV2ra

Die Hypothese H1 betrifft die Frage nach der Reliabilität der aus sieben Items bestehenden reflektiven Skala. Im Verlaufe der Analyse sprechen zwei Befunde für diese Annahme: So zeigte sich zum einen im Rahmen der Prüfung der Skalengüte mit einem Wert von α 0.95 ein sehr hohes Cronbachs Alpha für die Skala (siehe Abschnitt 7.3.2). Zum anderen wurde in der PLS-Analyse zum Zwei-Konstrukt-Modell eine sehr hohe *Composite Reliability* von 0,97 erreicht, was ebenfalls auf eine hohe Reliabilität der Skala ufosV2ra schließen lässt (siehe Abschnitt 7.5.3). Eine wesentliche Voraussetzung für die Anwendbarkeit der Skala bei der Bewertung von Websites, die Gewährleistung einer hohen Reliabilität, wird damit erfüllt. Des Weiteren kann für den Fall, dass die Anzahl von sieben Items in einem bestimmten Untersuchungskontext als zu hoch angenommen wird, alternativ eine reduzierte Skala mit nur drei ausgewählten Items herangezogen werden. Mit einem α 0.90 wird immer noch ein hoher Wert für Cronbachs Alpha erreicht (siehe Abschnitt 7.3.2). Die Einschränkung der Messgenauigkeit, die damit einhergeht, muss vor dem Hintergrund des jeweiligen Forschungsanliegens bewertet werden. Dabei ist zu beachten, dass die weitergehenden Aussagen zur Skalengüte für die ufosV2ra-Skala mit allen sieben Items abgeleitet wurden.

8.1.1.2 Faktorielle Validität der reflektiven ufosV2ra-Skala

Der Hypothese 2 liegt die Annahme zugrunde, die reflektive Skala ufosV2ra sei gegenüber Skalen zur Erfassung weiterer kundennaher Erfolgsfaktoren faktoriell valide. Zur Überprüfung dieser Annahme wurden drei statistische Verfahren angewandt.

Zunächst wurde eine EFA unter Einbeziehung der Items der Skalen Usability, Vertrauen, Vergnügen, Involvement und Ästhetik durchgeführt. Ähnlich wie bereits in der Untersuchung von Christophersen (2007), wiesen die zwei Items zur Ästhetik hohe Nebenladungen auf denselben Faktor wie auch die Usability-Indikatoren auf (zweite berechnete HHA in Tabelle E-10.2; siehe Abschnitt 7.3.1). Wenn auch zu berücksichtigen ist, dass bei Berechnung einer EFA mehr als zwei Indikatoren pro Konstrukt einbezogen werden sollten (Bühner, 2004, S. 157), spricht das Ergebnis zunächst gegen eine angemessene faktorielle Validität der Skala ufosV2ra. Nach Elimination der beiden Indikatoren zur Ästhetik werden allerdings wie erwartet vier Faktoren extrahiert, was unter Einschränkung für eine Bestätigung der Hypothese 2 zur faktoriellen Validität der Skala ufosV2ra spricht.

In einer CFA wurde ein angemessener Modell-Fit ausgewiesen (siehe Abschnitt 7.3.3). Es könnte allerdings angemerkt werden, dass bei zwei der vier LVn mit jeweils drei Indikatoren die Anforderung an die Mindestanzahl an Indikatoren um eine CFA zu berechnen, gerade erfüllt wurde (vgl. Marsh, Hau, Balla & Grayson, 1998).

Eine eindeutige Bestätigung erfährt die Hypothese 2 zur faktoriellen Validität der reflektiven Usability-Skala ufosV2ra angesichts der Ergebnisse der PLS-Analyse zum nomologischen Modell. Eine Analyse der Kreuzladungen sämtlicher Indikatoren im PLS-Modell zeigte, dass diese, ohne unerwünscht hohe Nebenladungen, mit den ihnen jeweils zugehörigen LVn höher korrelieren als mit den übrigen Variablen.

Zusammenfassend wird die faktorielle Validität der reflektiven Usability-Skala ufosV2ra angesichts der Ergebnisse der EFA mit leichter Einschränkung und durch die Ergebnisse zur CFA und zur PLS-Analyse uneingeschränkt gestützt.

8.1.1.3 Diskriminationsfähigkeit der reflektiven ufosV2ra-Skala

Um zu überprüfen, inwiefern die reflektive Skala ufosV2ra gemäß Hypothese 3 geeignet ist, Unterschiede in den Merkmalsausprägungen der unterschiedlichen Websites festzustellen, wurde eine ANOVA berechnet (siehe Abschnitt 7.3.4). Aufgrund des signifikanten Ergebnisses der ANOVA kann das Gütekriterium der Diskriminationsfähigkeit als erfüllt angesehen werden. Es ergeben sich für verschiedene Websites offensichtlich unterschiedliche Skalenmittelwerte für die reflektive Skala ufosV2ra, was eine Mindestanforderung für die praktische Anwendung einer Evaluationsskala darstellt. Im Rahmen einer *Post-hoc-Power-Analyse* (Cohen, 1988) wurde unter Zugrundelegung eines Signifikanzniveaus von α 0,001 und einer, angesichts der Unterschiedlichkeit der ausgewählten Websites angemessenen, mittleren Effektstärke f 0,25 ein zufriedenstellend hohes Powerniveau von 0,91 berechnet.

Einschränkend muss angemerkt werden, dass die Voraussetzung einer ANOVA, die Annahme der Varianzhomogenität, nicht erfüllt wurde. Die Überprüfung der Annahme anhand eines Levene-Tests ergab ein signifikantes Ergebnis, sodass Varianzhomogenität abzulehnen ist. Allerdings erweist sich das Verfahren bei einer Verletzung dieser Annahme als robust (vgl. Backhaus et al., 2003, S. 150).

8.1.1.4 Überprüfung der formativen Skala ufosV2fa

Entsprechend des in Abschnitt 6.6.2.3 vorgestellten Vorgehens wurde die formative Skala ufosV2fa überprüft, indem sie in einem Zwei-Konstrukt-Modell als Prädiktor zu der reflektiven Skala ufosV2ra in Beziehung gesetzt wurde. Um möglichen Verzerrungen der Schätzparameter aufgrund von Multikollinearität zwischen den formativen Indikatoren zu begegnen, wurden untereinander multikollineare Indikatoren der Skala ufosV2fa zu Indizes zusammengefasst (vgl. Albers & Hildebrand, 2004). Hierfür wurde, um möglichst viel Information der vorhandenen Daten beizubehalten, dem bei Multikollinearität empfohlenen Prinzip der Hauptkomponentenregression folgend, eine Hauptkomponentenanalyse mit orthogonaler Rotation gerechnet (Moosmüller, 2004, S. 132).

Es zeigte sich ein sehr hoher, signifikanter Pfadkoeffizient mit einem Wert von β 0,92 sowie ein sehr zufriedenstellender Determinationskoeffizient hinsichtlich der LV ufosV2ra von R^2 0,84. Diese Werte lassen auf einen hohen Zusammenhang zwischen den beiden LVn schließen, sodass die Hypothese H4, wonach ein Zusammenhang zwischen der formativen ufosV2fa-Skala und der reflektiven ufosV2ra-Skala besteht, bestätigt werden kann. Offensichtlich bildet das Set der 27 verwendeten formativen Usability-Indikatoren das Globalurteil der ufosV2ra-Skala sehr gut ab.

Angesichts der hohen Korrelationen zwischen den formativen Indikatoren könnte vermutet werden, dass bei einer Reduktion des Skalenumfangs der ufosV2fa-Skala durch Ausschluss einzelner Items trotzdem ein hoher Zusammenhang zum Globalurteil der ufosV2ra-Skala bestehen bliebe. Um inhaltliche Redundanzen im Fragebogeninventar zu vermeiden und die Itemanzahl einzugrenzen wurden bereits vier Indikatoren von der weiteren Verwendung in der formativen Skala ufosV2fa ausgeschlossen. Es verbleiben 27 Items in der ufosV2fa-Skala (siehe Abschnitt 7.5.1.2). Der Skalenumfang ist damit im Vergleich zu der von Christophersen (2007) verwendeten Skala ufosV2f mit 46 Items deutlich reduziert.

Dennoch stellt sich angesichts der hohen Itemkorrelationen die Frage, ob es angeraten wäre, noch weitere Items aus der Skala zu eliminieren. Zielsetzung einer Anwendung der formativen Usability-Skala ufosV2fa ist es, auf Grundlage der Ausprägungen von Einzelitems, inhaltlich differenzierende Rückschlüsse auf die Usability der jeweils dahinter stehenden Gestaltungsmerkmale zu erhalten. Weitere Items auszuschließen hätte möglicherweise zur Folge, die Beurteilungsbreite der ufosV2fa-Skala und damit die Inhaltsvalidität einzuschränken, was der angeführten Zielsetzung entgegenstände. Zudem ist zu berücksichtigen, dass die Skalenitems im Vorfeld der Untersuchung auf Grundlage von Experten-

urteilen abgeleitet wurden. Es sollte im Rahmen der vorliegenden Untersuchung deshalb nicht ohne wichtigen Grund auf weitere Items verzichtet werden. Der Skalenumfang von 27 Items wird als sinnvoll erachtet. Einerseits hält sich der Aufwand, der mit der Beantwortung verbunden ist, in angemessenen Grenzen und andererseits kann davon ausgegangen werden, dass alle wesentlichen Einzelaspekte der Usability von Websites erfasst werden.

8.1.1.5 Kriteriumsvalidität der ufosV2a-Skalen

Die konkurrente Kriteriumsvalidität der ufosV2a-Skalen ist Gegenstand der Hypothesen H5 und H6, die jeweils einen Zusammenhang zwischen den ufosV2a-Skalen und den Erfolgsgrößen Nutzerzufriedenheit bzw. Nutzungsintention annehmen. Im Falle der Nutzerzufriedenheit erwiesen sich in beiden berechneten PLS-Modellen die Pfade zwischen der exogenen LV Usability und der endogenen LV Nutzerzufriedenheit als signifikant, wobei der Pfadkoeffizient im Falle der formativen Skala ufosV2fa etwas höher ausfiel. Die unterschiedliche Höhe der Ausprägungen wird am Ende des Abschnitts 8.1.1.6 diskutiert. Im Falle der Nutzungsintention zeigte sich nur für die formative Skala ufosV2fa ein signifikanter Zusammenhang (Hypothese H6b). Die Hypothese H6a bezüglich eines Zusammenhangs zwischen der reflektiven Skala ufosV2ra und der Erfolgsgröße Nutzungsintention konnte nicht bestätigt werden.

Da PLS eine Einbeziehung der kategorialen Variable Auswahlentscheidung nicht zulässt, wurde eine sequentielle LR berechnet, um die Hypothesen H7 und H19 statistisch zu überprüfen (siehe Abschnitt 7.7)¹⁰. Beiden Hypothesen liegt die Annahme eines substanziellen Einflusses der Usability-Skalen (Hypothese H7) bzw. der Nutzungsintention (Hypothese H19) auf die Erfolgsgröße Auswahlentscheidung zugrunde.

Es wurde sequentiell vorgegangen, indem die Prädiktoren blockweise in die Analyse einbezogen wurden. Im ersten Schritt der LR wurden Variablen zur Demographie und zur Vorerfahrung mit Internetanwendungen als Prädiktoren einbezogen. Die Skalenwerte der ufosV2ra- bzw. ufosV2fa-Skalen wurden erst im zweiten Schritt in die Analyse eingebracht. Die Einbeziehung der reflektiven ufosV2ra- (Schritt 2a) als auch der formativen ufosV2fa-Skalenwerte (Schritt 2b) führte zu einer signifikanten Verbesserung der Prognosekraft des Modells. Die Hypothesen H7a und H7b, wonach die Skalen ufosV2ra bzw. ufosV2fa im Sinne einer prädiktiven Kriteriumsvalidität eine hohe Vorhersagekraft für die Auswahlentscheidung besitzen, bestätigen sich damit. Dabei lassen die Unterschiede in den Ausprägungen der Kennwerte der sequentiellen LR darauf schließen, dass die formative ufosV2fa-Skala eine leicht bessere Vorhersage ermöglicht, als die reflektive ufosV2ra-Skala.

¹⁰ Aufgrund eines zu geringen $N=12$ der Ausprägung *keine Auswahl* wurde anstelle einer multinomialen logistischen Regression, wie sie in der vorhergehenden Untersuchung durch Christophersen (2007) zur Anwendung kam, eine binomiale logistische Regression (LR) berechnet.

Die durchgeführten Analysen bestätigen für beide Usability-Skalen sowohl eine konkurrente Validität bezüglich der Erfolgsgröße Nutzerzufriedenheit als auch eine prädiktive Validität in Bezug auf die Auswahlentscheidung. Konkurrente Validität bezüglich der Erfolgsgröße Nutzungsintention konnte lediglich im Falle der formativen Usability-Skala ufosV2fa bestätigt werden. Diese Ergebnisse untermauern die Relevanz der Usability-Maße. Einerseits erweisen sich die Skalen dahingehend relevant, ob ein Anwender einer Website mit dieser zufrieden ist und eine Intention ausbildet, die Website zukünftig wieder zu nutzen. Und andererseits kann beiden Maßen eine Relevanz für die Auswahlentscheidung des Anwenders zugemessen werden, welche der beiden Krankenkassenwebsites er zukünftig bevorzugt aufsuchen würde, um sich zu Krankenversicherungsthemen zu informieren. Für die praktische Anwendung der ufosV2a-Skalen bedeutet dies, dass eine verbesserte Ausprägung der Usability mit einer hohen Wahrscheinlichkeit zu einer höheren Nutzung einer Website führen wird.

Bezüglich der Aussagen zur prädiktiven Validität könnten die Ergebnisse der LR dahingehend infrage gestellt werden, dass die Auswahlentscheidung unmittelbar anschließend an die Erfassung der Prädiktoren erfolgte. Es könnte argumentiert werden, dass damit die Zeitspanne zwischen Erfassung der Prädiktoren und Erfassung der AV zu kurz ausfiel (vgl. Bühner, 2004, S.31). Einer solchen Kritik folgend stellen die Befunde der LR eine weitere Bestätigung der konkurrenten Validität dar. Angesichts des gewählten Versuchsszenarios und unter Berücksichtigung der in Abschnitt 8.2.2.2 angeführten Einschränkungen zur Operationalisierbarkeit der Auswahlentscheidung erscheint es dem Verfasser jedoch angemessen, von einer Vorhersage der AV Auswahlentscheidung und infolgedessen von prädiktiver Validität zu sprechen.

8.1.1.6 Nomologische Validität der ufosV2a-Skalen

Die Hypothesen H8, H9 und H10 beziehen sich auf die Zusammenhänge der ufosV2a-Skalen zu den weiteren nutzernahen Erfolgsgrößen Vertrauen, Nutzungsvergnügen und Ästhetik innerhalb des nomologischen Modells. Die Überprüfung dieser Hypothesen erfolgte mit Hilfe des PLS-Strukturgleichungsansatzes (siehe Abschnitt 7.6.2).

Den Hypothesen H8a und H8b lag die Annahme zugrunde, dass ein Einfluss der Skalen ufosV2ra und ufosV2fa auf das Vertrauen in den Anbieter der Website bestehe. In beiden PLS-Modellen konnte dieser Zusammenhang anhand von hohen signifikanten Pfadkoeffizienten bestätigt werden, wobei der Pfadkoeffizient für das formative Modell höher ausfiel als für das reflektive. Diese Befunde bestätigen die Ergebnisse diverser Studien, in denen die Wirksamkeit der Usability auf das Vertrauen in den Anbieter von Websites und Online-Shops empirisch belegt wurde (siehe Abschnitt 2.11.1). Es ließe sich kritisch anmerken, dass der Analyse lediglich Querschnittsdaten zugrunde liegen und in Folge dessen Zweifel hinsichtlich der unterstellten Kausalrichtung zwischen den beiden Konstrukten geäußert werden können. Allerdings erscheint es nicht als schlüssig, dass sich das Vertrauen in den Anbieter der Website, entgegen der unterstellten Kausalrichtung, auf die wahrgenommene Usability der

Website auswirken sollte. Eine Möglichkeit der Überprüfung wäre durch ein längsschnittliches Design der Untersuchung bei Wiederholung der Erhebung gegeben.

Die Annahme eines positiven Einflusses der Usability auf das Nutzungsvergnügen liegt den Hypothesen H9a und H9b zugrunde. Beide Hypothesen wurden durch hohe signifikante Pfadkoeffizienten zwischen der Usability und dem Nutzungsvergnügen bestätigt. Die hohen Parameterausprägungen weisen auf einen hohen Zusammenhang zwischen den Konstrukten hin. Diese Befunde bestätigen die Ergebnisse verschiedener Studien zum Zusammenhang zwischen der wahrgenommenen Usability einer Website und dem während des Nutzungsvorgangs erlebten Vergnügen (siehe Abschnitt 2.11.3). Die Richtung der Kausalität zwischen den beiden LVn wurde in den PLS-Modellen der vorliegenden Arbeit entsprechend den zitierten Studien dahingehend angenommen, dass die wahrgenommene Usability das Nutzungsvergnügen beeinflusst. Da der Analyse Querschnittsdaten zugrunde liegen, kann die Signifikanz des festgestellten Zusammenhangs nicht als Beleg für die angenommene Kausalrichtung gewertet werden. Wie bereits für den Zusammenhang von Usability und Vertrauen dargestellt, fiel der Pfadkoeffizient für das formative Modell erneut höher aus als für das reflektive Modell.

Die Hypothese H10 zum Zusammenhang zwischen der Usability und der Ästhetik konnte nicht untersucht werden, da sich die Skala zur Erfassung der Variable Ästhetik gegenüber den übrigen Skalen nicht als faktoriell valide erwies (siehe Abschnitt 7.3.1).

Sowohl in der PLS-Analyse als auch in der sequentiellen LR fiel der Zusammenhang zwischen der exogenen LV Usability und der jeweiligen endogenen LV für die formative Skala ufosV2fa etwas höher aus als für die reflektive Skala ufosV2ra. Es könnte geschlossen werden, dass die ufosV2ra-Skala eine leicht geringere Kriteriumsvalidität besitzt als die ufosV2fa-Skala. Eine Überprüfung der Pfadkoeffizienten in den beiden PLS-Modellen anhand der Berechnung der t-Werte für das reflektive bzw. formative Modell weist jedoch keine signifikanten Unterschiede aus.

Der Hypothese H16 liegt die Annahme eines moderierenden Effekts der LV Involvement auf die Wirkungsbeziehung zwischen den LVn Usability und Nutzungsintention zugrunde. Weder für den vollständigen Datensatz, noch nach Reduzierung des Datensatzes auf das obere und untere Fünftel der Ausprägungen der LV Involvement, konnte eine moderierende Wirkung des Involvement auf die Beziehung zwischen Usability und Nutzungsintention festgestellt werden. Diese zeigte sich lediglich bei Verwendung nicht standardisierter Daten. Da dieses Vorgehen bei der Berechnung moderierender Effekte in PLS-Modellen jedoch nicht dem in der Literatur allgemein empfohlenen Standardverfahren entspricht, wurden die Ergebnisse verworfen (vgl. Ringle, 2004). Auch die der Hypothese H17 zu Grunde liegende Annahme einer moderierenden Wirkung des Involvements auf den Zusammenhang von Nutzerzufriedenheit und Nutzungsintention bestätigte sich entgegen der in der Literatur dargestellten Zusammenhänge nicht (Ranaweera et al., 2004, 2005). Dies mag dem Umstand zuzuschreiben

sein, dass das der Untersuchung zugrunde liegende Untersuchungsfeld der Websites von Krankenkassen tendenziell eher von einem niedrigen Involvement geprägt ist. Es kann daher vermutet werden, dass diesbezügliche moderierende Effekte eher schwach ausgeprägt sind.

Für das Vorgehen anhand eines reduzierten Datensatzes wurde im Rahmen einer *Sensitivitäts-Analyse* (Cohen, 1988) unter Zugrundelegung des von Cohen geforderten Richtwerts für die Power von 0,80 eine minimale Effektstärke d 0,17 berechnet. Nach Cohen (1988) ist diese als eher „klein“ zu bezeichnen. Der Umfang der Stichprobe ist also ausreichend, um mit einer Wahrscheinlichkeit von 80% existente Wirkungen auch bei einer Effektstärke d 0,17 als signifikant schätzen zu können.

Zusammenfassend ist festzuhalten, dass abgesehen von den Hypothesen H6a, H10 und H16 alle Hypothesen des nomologischen Modells, die die LV Usability betreffen, empirische Bestätigung finden. Diese Ergebnisse lassen auf eine hohe nomologische Validität der beiden ufosV2a-Maße schließen. Zusammenhänge zwischen der Usability und anderen kundennahen Erfolgsfaktoren, die aufgrund vorangegangener Studien als empirisch abgesichert gelten, erfahren aufgrund dieser Ergebnisse auch bei Verwendung der ufosV2a-Skalen erneut Bestätigung. Folglich erscheinen die beiden Skalen als geeignet, um die Beziehungen zwischen der Usability und weiteren Konstrukten in empirischen Studien zu untersuchen. Besonders hervorzuheben ist dabei, dass, aufbauend auf den Ergebnissen der Untersuchung von Christophersen (2007), die beiden ufosV2a-Skalen sich nun auch zur Anwendung im Falle von Websites von Webdienstleistern als geeignet erwiesen.

8.1.2 Befunde zu den weiteren Hypothesen im nomologischen Modell

Neben den Hypothesen, die die LV Usability betreffen, wurden in den inhaltlichen Hypothesen H11 bis H18 weitere Zusammenhänge zwischen den übrigen LVn des nomologischen Modells betrachtet. Da diese für die Zielsetzungen der Untersuchung nur von nachgeordneter Bedeutung sind, wird die Diskussion der Einzelbefunde lediglich knapp gehalten.

Die Hypothesen H11 und H14, die Zusammenhänge zwischen der Ästhetik und der Nutzerzufriedenheit bzw. dem Nutzungsvergnügen annehmen, konnten aufgrund mangelnder faktorieller Validität der Skala zur Ästhetik nicht anhand der PLS-Analysen überprüft werden. Die Hypothese H15 zur Beziehung zwischen Reputation und Vertrauen konnte ebenfalls nicht überprüft werden, da die Skala zur Reputation aufgrund eines zu hohen Anteils fehlender Werte von der weiteren Analyse ausgeschlossen wurde.

Die Hypothesen H12 und H13 betreffen den Einfluss des Vertrauens bzw. des Nutzervergnügens auf die Nutzerzufriedenheit. Beide Annahmen konnten auf Grundlage niedriger signifikanter Pfadkoeffizienten für das reflektive ufosV2ra-Modell bestätigt werden. Am deutlichsten bestätigte sich der in Hypothese H18 angenommene Zusammenhang

zwischen der Nutzerzufriedenheit und der Nutzungsintention. Die Hypothese H18 konnte somit für beide Modelle bestätigt werden.

Es stellt sich die Frage nach den Erklärungen für die insignifikanten Ergebnisse bei der Überprüfung der Hypothese H6a für das reflektive ufosV2ra-Modell, sowie im Falle der Hypothesen H12 und H13 für das formative ufosV2fa-Modell.

Überraschend ist, dass die Einflüsse der LVn Vertrauen und Nutzervergnügen auf die Nutzerzufriedenheit im Rahmen der PLS-Analyse für das formative Modell überhaupt nicht und für das reflektive Modell lediglich schwach bestätigt wurden (H12 und H13). Diese Zusammenhänge konnten in der Vergangenheit in diversen Studien untermauert werden (siehe Abschnitte 2.10.1 und 2.10.4). Auch weist die Korrelationsmatrix für die LVn im formativen PLS-Modell für beide LVn hohe Korrelationen von $r = 0,57$ zwischen Vertrauen und Nutzerzufriedenheit bzw. $r = 0,61$ zwischen Nutzervergnügen und Nutzerzufriedenheit aus (siehe Tabellen E-25 und E-26 im Anhang E).

Mögliche Erklärungen dafür, dass sich die Pfadkoeffizienten im vorliegenden Modell dennoch nicht als signifikant erweisen, könnten in den Beziehungen zwischen Usability und Vertrauen bzw. Nutzervergnügen angenommen werden. So erwies sich die Usability als signifikanter Prädiktor für die beiden LVn (siehe Abschnitt 7.6.2). Möglicherweise liegen den Beziehungen zwischen den betrachteten LVn Suppressoreffekte zugrunde, die die Effekte zwischen den LVn Vertrauen bzw. Nutzervergnügen und Nutzerzufriedenheit teilweise aufheben (Urban & Mayerl, 2006). Eine weitergehende Analyse der Ursachen für die insignifikanten Ergebnisse liegt jedoch außerhalb des Fokus der vorliegenden Arbeit und soll deshalb nicht vorgenommen werden.

Der Zusammenhang zwischen Usability und Nutzungsintention konnte in der Vergangenheit ebenfalls in diversen Studien zum TAM untermauert werden (vgl. Chen, Gillenson & Sherrell, 2002; Gefen, Karahanna & Straub, 2003; Koufaris, 2002; Pavlou & Fygenson, 2006). In der vorliegenden Untersuchung zeigte sich jedoch lediglich für das formative Modell ein schwacher Effekt (H6b), für das reflektive Modell erwies sich kein signifikanter Zusammenhang (H6a). Als Erklärung können hierfür möglicherweise Mediatoreffekte aufgrund der zusätzlich in das Modell einbezogenen Variable Nutzerzufriedenheit herangezogen werden. So erwiesen sich in beiden PLS-Modellen starke Zusammenhänge sowohl zwischen den LVn Usability und Nutzerzufriedenheit als auch zwischen den LVn Nutzerzufriedenheit und Nutzungsintention. Auch hier soll eine weitergehende Analyse der Ursachen für das insignifikante Ergebnis nicht vorgenommen werden, da diese Fragestellung außerhalb des Fokus der vorliegenden Arbeit liegt.

Der Zusammenhang zwischen den beiden Erfolgsfaktoren Nutzungsintention und Auswahlentscheidung wurde nicht im Rahmen der PLS-Analyse untersucht, sondern mittels einer LR-Analyse. Die Skalenwerte der Nutzungsintention wurden in Schritt 2c der sequentiellen LR als Prädiktoren für die Auswahlentscheidung hinzugezogen (siehe Abschnitt 7.7.2). Der

Anteil korrekt klassifizierter Fälle erwies sich mit 92,90% als hoch, das Inkrement des Modell- χ^2 -Wertes im globalen Likelihood-Quotiententest als signifikant und es ergab sich ein hoher Anteil zusätzlich aufgeklärter Varianz (*Nagelkerke* ΔR^2 0,81). Damit wird deutlich, dass das Maß der Nutzungsintention für die Auswahlentscheidung der Teilnehmer am Quasi-Experiment bedeutsam ist. Die Hypothese H19 kann folglich angenommen werden.

Dieses Ergebnis bestätigt damit nicht nur die der Hypothese H19 zugrunde liegende Kernannahme des nomologischen Modells, sondern es untermauert auch die Güte der Erfolgsgrößen Nutzungsintention und Auswahlentscheidung (siehe Abschnitt 8.2.3).

8.1.3 Befunde zu möglichen Effekten aufgrund der Itemreihenfolge

Eine weitere Zielsetzung der vorliegenden Untersuchung bestand darin, die Auswirkungen einer Veränderung der Abfrageposition der Usability auf die Untersuchungsergebnisse zu untersuchen. Um mögliche Prozesse zu identifizieren, die zu einer Beeinflussung der Bewertung führen können, wurde eine systematische Variation der Itemreihenfolge vorgenommen. Dabei wurde einerseits die Abfolge der reflektiven und formativen Usability-Skalen und andererseits die Positionierung der Usability-Skalen in der Abfolge der Konstruktabfrage systematisch variiert. Die Annahme, die Versuchsteilnehmer zeigten eine unterschiedlich hohe Einschätzung der AV Nutzungsintention, je nachdem, an welcher Stelle die Usability im Rahmen der Befragung abgefragt wurde, bestätigte sich anhand der Ergebnisse eines t-Tests jedoch nicht. Es kann daher davon ausgegangen werden, dass eine Veränderung der Abfrageposition der Usability keinen Einfluss auf die Ergebnisse der Untersuchung nimmt.

8.2 Bewertung des methodischen Vorgehens

Die Bewertung des methodischen Vorgehens erfolgt anhand verschiedener Aspekte. So wird einerseits das Quasi-Experiment hinsichtlich möglicher Einschränkungen der internen und externen Validität betrachtet, sowie andererseits das statistische Vorgehen hinsichtlich seiner Angemessenheit bewertet.

8.2.1 Interne Validität des Quasi-Experiments

Bei der Durchführung eines Experiments ist darauf zu achten, inwiefern die Zusammenhänge zwischen UVn und AV im nomologischen Modell durch systematische Störeinflüsse verzerrt werden und damit die interne Validität des Experiments eingeschränkt wird (vgl. Rack & Christophersen, 2007, S. 31). Um die interne Validität zu sichern, wurden in der durchgeführten quasi-experimentellen Untersuchung verschiedene versuchsplanerische Vorkehrungen getroffen, die positiv hervorzuheben sind.

Einflüsse der Reihenfolge der Websitedarbietung: Wie den Abschnitten 4.7 und 4.8 zur Berücksichtigung von Reihenfolgeeffekten und zum Versuchsplan entnommen werden kann, wurde die Präsentationsreihenfolge der Websites vollständig ausbalanciert. Dieses Vorgehen

sollte dem Umstand Rechnung tragen, dass ein Einfluss der Präsentationsreihenfolge der Websites auf die Bewertung der Variablen nicht auszuschließen ist (vgl. Huber 1995, S.154f.). So kann die Bewertung der zweiten Website unterschiedlich ausfallen, je nachdem wie die zuvor präsentierte Website von der Versuchsperson wahrgenommen wird. Effekte dieser Art dürften sich über den Versuchsablauf dadurch ausbalancieren, dass im Versuchsplan einerseits jede mögliche Paarung der sieben Websites berücksichtigt wurde und andererseits die Reihenfolge der Präsentation für jede Kombination systematisch variiert wurde.

Einflüsse der Reihenfolge der Konstruktabfrage: Die Reihenfolge der Konstruktabfrage orientiert sich an den Wirkungszusammenhängen zwischen den latenten Variablen, wie sie in Abschnitt 2.11 beschrieben wurden. Es wurde darauf geachtet, dass die jeweils endogene LV der exogenen LV in der Abfrage vorangestellt wurde. Dem entsprechend wurde die Nutzungsintention als zentrale AV in der Untersuchung an erster Stelle in der Abfrage präsentiert. So wurden mögliche verzerrende Effekte auf die Messergebnisse aufgrund einer vorhergehenden kognitiven Auseinandersetzung mit der jeweils beeinflussenden Variablen weitestgehend berücksichtigt. Eine Ausnahme wurde lediglich bei der Positionierung der beiden Usability-Skalen gemacht, da diesbezüglich neben den kausalen Wirkungszusammenhängen auch die Auswirkungen einer Veränderung der Abfrageposition der Usability auf die Untersuchungsergebnisse untersucht werden sollten.

Einflüsse der Reihenfolge der Usabilityabfrage reflektiv vs. formativ: Wie bereits in Abschnitt 4.7 dargestellt wurde, konnte in diversen Studien gezeigt werden, dass die Reihenfolge der Itempräsentation einen Einfluss auf die Ergebnisse einer fragebogenbasierten Erhebung von Produktbewertungen haben kann (vgl. Konradt et al., 2006; Schwarz et al., 1991; Sudman et al., 1996). So können verzerrende Effekte daraus entstehen, dass die formativen Einzelaspekte bzw. reflektiven Globalurteile zur Usability einer Website unterschiedlich bewertet werden, je nachdem, ob der Bewertung eine Abfrage mit Hilfe eines reflektiven bzw. formativen Messmodells zur Usability vorausgeht. Um solchen verzerrenden Effekten entgegenzuwirken, wurde die Präsentationsreihenfolge der reflektiven und formativen Usability-Skalen systematisch variiert (Abschnitt 4.7.2). Mögliche Effekte aufgrund der Reihenfolge der Usabilityabfrage wurden folglich über den gesamten Datensatz vollständig ausbalanciert.

Operationalisierung der Auswahlentscheidung: Die Auswahlentscheidung wurde im Quasi-Experiment erst erhoben, nachdem die Versuchsteilnehmer die Bewertung der Websites vorgenommen hatten. Infolgedessen ist eine Einflussnahme der Auseinandersetzung mit den Fragebogenitems auf die Auswahlentscheidung möglich, indem die hinter den Items liegenden Bewertungsaspekte in die Auswahlentscheidung einfließen. Die Gefahr der Einflussnahme auf die Auswahlentscheidung wird jedoch als gering angesehen.

Einflussnahme des aktuellen Versicherungsgebers: Um eine mögliche Beeinflussung der Befragungsergebnisse durch Vorerfahrungen mit dem Versicherungsgeber aufgrund eines

aktuell bestehenden Versicherungsverhältnisses berücksichtigen zu können, wurde der aktuelle Krankenversicherungsgeber erfragt. In 10 % der Fälle bestand eine Übereinstimmung zwischen dem Anbieter der zu bewertenden Website und dem aktuellen Versicherungsgeber der Vpn, sodass die Website des aktuellen Versicherungsgebers in die Auswahlentscheidung einging. In 10 von 26 Fällen, in denen eine Übereinstimmung vorlag, wählten die Vpn die Website der eigenen Krankenkasse aus.

Einflüsse der Versuchsumgebung: Die Hypertextumgebung lief überwiegend stabil. Lediglich in sieben von insgesamt 271 Fällen meldeten sich Versuchspersonen aufgrund von Schwierigkeiten beim Zugang zur Versuchsumgebung. Nachfragen zu den Teilnehmerinstruktionen gab es keine, sodass davon ausgegangen werden kann, dass diese verständlich ausgefallen sind. Versuchsleitereffekte können aufgrund der computergestützten ortsungebundenen Online-Versuchsdurchführung ausgeschlossen werden.

Erhebungszeitraum: Die Erhebung zog sich über eine Dauer von sieben Monaten hin. Während dieses Zeitraums wurden die Websites der sieben herangezogenen Krankenkassen durch die Versuchsleitung regelmäßig daraufhin überprüft, ob die im Rahmen des Quasi-Experiments gestellten Aufgaben bearbeitet werden konnten. Dies war über die gesamte Versuchslaufzeit sichergestellt. Abgesehen von fortlaufenden Aktualisierungen in den redaktionellen Beiträgen konnten keine auffälligen Veränderungen im Aufbau bzw. im Design der Websites festgestellt werden. In Folge dessen konnten wesentliche Veränderungen in den Ausprägungen der Usability oder der weiteren betrachteten kundennahen Erfolgsfaktoren der Websites während des Versuchsablaufs weitestgehend ausgeschlossen werden.

8.2.2 Externe Validität des Quasi-Experiments

Externe Validität bezieht sich auf die Fragestellung, inwieweit die Ergebnisse einer Studie auf andere Situationen und Personen verallgemeinert werden können. Dabei ist zu berücksichtigen, dass Maßnahmen, die der Experimentator zur Erhöhung der internen Validität ergreift, häufig zu einer Einschränkung der externen Validität führen können, so dass die Ergebnisse nur eingeschränkt verallgemeinert werden können. Die interne und externe Validität eines Quasi-Experiments stehen somit in einer wechselseitigen Abhängigkeit (vgl. Rack & Christophersen, 2006, S. 31). Hinsichtlich des durchgeführten Quasi-Experiments kann kritisch hinterfragt werden, inwieweit a) die herbeigeführte Bewertungssituation auf andere Erhebungskontexte übertragbar ist (Abschnitt 8.2.2.1), b) die Operationalisierung der Auswahlentscheidung als realitätsnah betrachtet werden kann (Abschnitt 8.2.2.2), c) die Zusammensetzung der Stichprobe angemessen erscheint (Abschnitt 8.2.2.3) und d) möglicherweise ein Störeffekt aufgrund der getroffenen Auswahl an Websites vorliegt (Abschnitt 8.2.2.4).

8.2.2.1 Realitätsbezug der quasi-experimentellen Nutzungssituation

Um die externe Validität zu beurteilen, soll zunächst der Frage nachgegangen werden, inwieweit die quasi-experimentelle Nutzungssituation auf den alltäglichen Umgang mit Websites zu übertragen ist, sodass eine Verallgemeinerung der Ergebnisse gerechtfertigt ist. So können gegenüber einer normalen Auswahl- und Nutzungssituation sowohl Parallelen als auch Unterschiede angeführt werden.

Angemessenheit der Aufgaben: Die Auswahl der gestellten Aufgaben orientiert sich an den Ergebnissen einer vorangegangenen Kontextanalyse unter Einbeziehung der Untersuchungsergebnisse einer Studie zur Nutzung von Websites der gesetzlichen Krankenversicherungsunternehmen (TNS Infratest, 2005). Insofern konnte sichergestellt werden, dass die Auswahl der Aufgaben an typischen Nutzungsszenarien ausgerichtet wurde und unter besonderer Berücksichtigung der Interessen der Zielgruppe der Studenten vorgenommen wurde. Neben der inhaltlichen Angemessenheit der Aufgabenstellungen kann sicherlich auch die Art der Aufgabenstellung Einfluss auf die Untersuchungsergebnisse nehmen. So können sich im Falle von Aufgaben zu komplexen Benutzungsvorgängen möglicherweise sehr spezifische und hohe Anforderungen an die Ausprägungen der Usability oder auch anderer kundennaher Erfolgsfaktoren einer Website ergeben. Infolgedessen könnten Websites bei komplexen Aufgabenstellungen zum einen tendenziell als weniger benutzerfreundlich eingeschätzt werden und zum anderen bestünde die Gefahr, dass Versuchsteilnehmer die Aufgaben nicht vollständig bearbeiten können. Die getroffene Auswahl typischer Nutzungsszenarien (Allgemeines Stöbern und gezielte Suche nach Themen hoher Relevanz) erscheint für das Untersuchungsobjekt Krankenkassenwebsites generell als angemessen, da die Website einer Krankenkasse diese Abläufe unterstützen muss, damit ein Anwender die Website erfolgreich nutzen kann.

Standardisierung der Interaktion: Der Ablauf des Websitebesuchs wurde durch die vorgegebenen Aufgabenstellungen bis zu einem gewissen Grad standardisiert. Die Versuchspersonen wurden dadurch zur Ausführung definierter für die Nutzung der Website wesentlicher Abläufe motiviert. Dadurch wurde sichergestellt, dass die Websitebewertungen auf einem vergleichbaren Erfahrungsschatz der Versuchsteilnehmer beruhen¹¹. Ergänzend erhielten die Versuchsteilnehmer die Möglichkeit, die Websites nach eigenem Interesse zu erkunden und sich so einen allgemeinen Überblick zu verschaffen.

Dauer und Umfang der Interaktion: Die maximale Dauer der Interaktion mit den Websites wurde den Versuchsteilnehmern freigestellt. Die vorgegebenen Aufgaben machten eine minimale Interaktionsdauer von ca. 15 Minuten je Website erforderlich. Dabei wurde darauf geachtet, dass die Versuchspersonen in der Interaktion mit der Website ausreichende Erfah-

¹¹ Die Standardisierung dient damit ebenso der Unterstützung der internen Validität.

rungen sammeln konnten, um eine umfangliche Bewertung der verschiedenen Bewertungsaspekte in der typischen Nutzung einer Krankenkassenwebsite vornehmen zu können.

Zusammenfassend kann festgestellt werden, dass eine Übertragbarkeit der Ergebnisse auf Websites weiterer Krankenkassen gerechtfertigt scheint. Nach Auffassung des Autors ist die externe Validität der quasi-experimentellen Nutzungssituation für den Bereich der Krankenkassen als hoch anzusehen. Bezüglich der Übertragbarkeit der Ergebnisse auf Websites anderer Anbietersegmente lassen sich jedoch keine Aussagen treffen.

8.2.2.2 Operationalisierung der Auswahlentscheidung

Von einer Auswahlentscheidung kann gesprochen werden, wenn mindestens drei Entscheidungsalternativen bestehen und ein Mindestmaß an kognitiver Beteiligung durch den Auswählenden erbracht wird (Wiswede, 1995). Beide Bedingungen werden durch die Operationalisierung der Auswahlentscheidung erfüllt. Einerseits können sich die Versuchsteilnehmer zwischen einer der beiden Websites und der Ausweichkategorie *keine Auswahl* entscheiden und andererseits kann aufgrund der vorangehenden Bewertung der Websites von einer kognitiven Beteiligung der Auswählenden ausgegangen werden.

Kritisch angemerkt werden könnte, dass eine tatsächliche Auswahl zwischen den beiden Krankenkassen als Versicherungsgeber im Rahmen des Versuchs nicht erfolgt. Alternativ wäre eine fiktive Auswahl unter den Krankenkassen vorstellbar. Allerdings wäre auch in diesem Falle kritisch anzumerken, dass in einer realistischen Auswahlentscheidung voraussichtlich mehr als nur zwei Krankenkassen in den Auswahlprozess einbezogen würden. Eine Sichtung ähnlicher Studien zeigt jedoch, dass nur selten tatsächliches Auswahl- oder Kaufverhalten als abhängige Variable erfasst wird (z.B. Gefen & Straub, 2000; Jarvenpaa & Todd, 1997a; Jarvenpaa, Tractinsky, Saarinen & Vitale, 1999; Jarvenpaa, Tractinsky & Vitale, 2000; Konradt, Wandke, Balazs & Christophersen, 2003; Koufaris & Hampton-Sosa, 2004; Li, Kuo & Russell, 1999; Lohse & Spiller, 1999). Vielmehr wird meist die Kaufintention abgefragt, wobei offen bleibt, inwieweit diese Erfolgsgröße tatsächlich Verhaltensrelevanz besitzt.

Zusammenfassend kann festgestellt werden, dass die im Rahmen des Quasi-Experiments operationalisierte Auswahlentscheidung sich von einer realen Auswahlentscheidung teilweise unterscheidet. Inwieweit dies zu Einschränkungen der externen Validität führt, bleibt unklar.

8.2.2.3 Repräsentativität der Stichprobe

Zur Bewertung der externen Validität des Quasi-Experiments sind auch Effekte aus der Zusammensetzung der Stichprobe zu berücksichtigen (Kaya & Himme, 2006, S. 90; Homburg & Krohmer 2003, S. 226). Wie in Abschnitt 5.1 dargestellt, wurde gezielt vorrangig die Zielgruppe der Studenten und Auszubildenden betrachtet. Für diese besteht bei Aufnahme eines sozialversicherungspflichtigen Beschäftigungsverhältnisses die Anforderung, sich selbst zu

versichern (§§ 5 Abs.1 Nr.9, 10 Abs. 2 SGB V). Damit kann die Notwendigkeit unterstellt werden, sich mit der Thematik Krankenversicherung auseinanderzusetzen. Infolge des gewählten Vorgehens unterscheidet sich die gezogene Stichprobe in ihrer Zusammensetzung gegenüber der Gesamtpopulation in der deutschen Bevölkerung.

Neben diesen Unterschieden ist möglicherweise mit einem Effekt der Selbstselektion der Versuchsteilnehmer zu rechnen. So hat sich sicherlich ein Teil der Versuchspersonen aus eigenem Interesse am Untersuchungsgegenstand oder am Versuch zur Teilnahme angemeldet. Bei anderen ist davon auszugehen, dass die Belohnung in Höhe von 15 Euro bzw. 1,5 Versuchspersonenstunden den Anreiz für die Versuchsteilnahme darstellten.

Zusammenfassend kann festgestellt werden, dass eine Übertragbarkeit der Untersuchungsergebnisse auf die Grundgesamtheit der Gesamtpopulation der deutschen Bevölkerung nicht ohne weiteres gegeben ist. Diese Einschränkung wird dem Umstand geschuldet, dass die Zusammensetzung der Stichprobe bewusst gewählt war, um die praktische Relevanz der Untersuchungsergebnisse zu steigern. Positiv hervorzuheben ist hingegen die vergleichsweise große Stichprobe. Eine direkte Übertragbarkeit auf andere Länder ist aufgrund des Untersuchungsgegenstandes ausgeschlossen.

8.2.2.4 Abhängigkeit der Ergebnisse von der Auswahl der Websites

Neben der Repräsentativität der Stichprobe ist die Auswahl der Websites zu beachten (Monk, 2004). Es stellt sich die Frage, inwieweit möglicherweise ein Störeffekt aufgrund der getroffenen Auswahl an Websites vorliegt. Insgesamt wurden sieben Krankenkassenwebsites ausgewählt. Die Auswahl der Websites erfolgte anhand verschiedener Kriterien (siehe Abschnitt 4.1.2). So wurde darauf geachtet, dass sowohl Websites mit eher hoher als auch geringer Usability-Ausprägung in der Auswahl enthalten waren. Ebenso wurden Websites großer bzw. kleiner, sowie bekannter und weniger bekannter Krankenkassen einbezogen. Die endgültige Auswahl erfolgte auf Grundlage einer Voruntersuchung zur Verwendbarkeit der Websites im Rahmen der Hauptuntersuchung. Angesichts der Unterschiedlichkeit der Websites, die zur Variation der UV ausgewählt wurden, ist es eher unwahrscheinlich, dass eine andere, zufällig gezogene Auswahl an Websites von Krankenkassen zu stark abweichenden Untersuchungsergebnissen geführt hätte.

Kritisch angemerkt werden könnte, dass die Validierungsuntersuchung anhand nur einer Kategorie von Benutzerschnittstellen, Websites von Krankenkassen, vorgenommen wurde. Die Untersuchungsergebnisse könnten insofern als kritisch betrachtet werden, als dass sie lediglich hinsichtlich der ausgewählten Kategorie interpretiert werden können (vgl. Finn & Kayande, 1997).

8.2.3 Vorgehen der statistischen Auswertung

Neben möglichen Einschränkungen aufgrund von Defiziten bezüglich der internen bzw. externen Validität der Untersuchungsergebnisse könnten sich Einschränkungen aufgrund methodischer Probleme bei der statistischen Auswertung ergeben. Entsprechend wird die Angemessenheit des Vorgehens bei der Datenanalyse einer kritischen Betrachtung unterzogen.

Bei der Bewertung des methodischen Vorgehens werden zunächst die Maßnahmen zur Aufbereitung des Datensatzes betrachtet. So wurde mit der Multiplen Imputation ein valides Verfahren zur Ersetzung fehlender Werte, insbesondere im Falle formativer Skalen, verwendet (Tabachnik & Fidell, 2006). Ebenso wurde sorgfältig mit der Frage umgegangen, wie Ausreißer in der berechneten LR zu behandeln sind. Diese wurden ausgeschlossen (siehe Abschnitt 7.7.1), wenn im Abgleich mit den Ergebnissen einer Prozedur zur Identifikation problematischer Fälle begründete Zweifel an der Eignung der Fälle bestanden (siehe Abschnitt 7.1.5).

Um die Angemessenheit der Stichprobe, insbesondere bezüglich nicht signifikanter Ergebnisse, zu validieren, wurden Power-Analysen durchgeführt. Im Rahmen einer *Sensitivitäts-Analyse* (Cohen, 1988) wurde unter Zugrundelegung eines Signifikanzniveaus von α 0,05 und einer geforderten Power von 0,95 eine minimale Effektstärke d 0,14 berechnet. Nach Cohen (1988) ist diese als „klein“ zu bezeichnen. Der Stichprobenumfang ist ausreichend groß, um existente Wirkungen auf einem Signifikanzniveau von α 0,05 auch bei kleiner Effektstärke signifikant zu schätzen.

Von zentraler Bedeutung für die Bewertung des methodischen Vorgehens ist die Frage, inwieweit möglicherweise ein Common Method Bias die Untersuchungsergebnisse beeinflusst haben könnte (Podsakoff et al., 2003; siehe Abschnitt 6.4). Angesichts des gewählten Vorgehens, dass die Datenerhebung sowohl für alle LVn als auch für die AV Nutzungsintention anhand derselben Methode erfolgte, könnte vermutet werden, dass die Ergebnisse der PLS-Analyse zum nomologischen Modell einem Common Method Bias unterliegen. Auf Grundlage der Ergebnisse des Harman's Single-Factor Tests kann jedoch davon ausgegangen werden, dass kein kritischer Common Method Bias im Datensatz vorliegt (siehe Abschnitt 7.1.4). Allerdings kritisieren Podsakoff et al. (2003) den Test als eher insensitiv und messen einem solchen Ergebnis nur eine eingeschränkte Aussagekraft zu. Dennoch kann vor dem Hintergrund der Ergebnisse der sequentiell berechneten LR (siehe Abschnitt 7.7) davon ausgegangen werden, dass die Untersuchungsergebnisse der PLS-Analyse nicht wesentlich durch einen Common Method Bias beeinflusst wurden. Diese Annahme wird durch den hohen Zusammenhang zwischen den Variablen in der LR unterstützt. So konnte in Schritt 2c der LR aufgezeigt werden, dass die Nutzungsintention eine hohe Vorhersagekraft in Bezug auf die Auswahlentscheidung hat (Hypothese 19). Da die Art der Datenerhebung sich auf unabhängiger und abhängiger Variablenseite deutlich

voneinander unterscheidet, ist die Wahrscheinlichkeit eines möglichen Common Method Bias deutlich als gering einzuschätzen (vgl. Podsakoff et al., 2003).

Darüber hinaus ergaben sich im Rahmen der statistischen Auswertung methodische Probleme, die im weiteren Verlauf dargestellt werden. Entgegen der ursprünglichen Erwartung und entgegen den Ergebnissen der Untersuchung von Christophersen (2007) zeigte sich in der EFA kein eigenständiger Faktor „Ästhetik“. Dabei ist allerdings zu berücksichtigen, dass die Anzahl der zugehörigen Indikatoren für die Berechnung einer EFA streng genommen zu gering war (siehe Abschnitt 7.3.1). Erst nach Ausschluss der Items zur LV Ästhetik wurden wie erwartet vier Faktoren extrahiert.

Desweiteren lagen zur Erfassung der Konstrukte wahrgenommener Preis sowie Bekanntheitsgrad jeweils nur ein Item vor (siehe Abschnitt 7.3.5). Zwar zeigen verschiedene Studien, dass Single-Item-Messungen von ausreichender Reliabilität und Validität möglich sind (vgl. Christophersen & Konradt, in press; Nagy, 2002; Wanous & Hudy, 2001). Dies gilt insbesondere dann, wenn sich das Konstrukt wie im Falle des wahrgenommenen Preises auf einen konkreten Bewertungsgegenstand bezieht und nicht über verschiedene Facetten verfügt. Im vorliegenden Falle liegen jedoch keine weiteren Erkenntnisse zur Messgüte der verwendeten Single-Items vor. Zweifel bezüglich ihrer Messgüte müssen deshalb eingeräumt werden.

Ein weiteres methodisches Problem wurde im Zusammenhang mit der Analyse des PLS-Zwei-Konstrukt-Modells deutlich (siehe Abschnitt 7.5.3). Angesichts der Hinweise auf Multikollinearität zwischen den formativen Indikatoren der ufosV2fa-Skala wurden die Indikatoren mit Hilfe einer HKA zu Unterindizes zusammengefasst (vgl. Albers & Hildebrand, 2004). Der Einfluss einzelner formativer Usability-Aspekte auf die Erfolgsgröße Nutzungsintention konnte in der Folge nicht mehr identifiziert werden. Ebenso wenig war eine inhaltliche Interpretation der Komponenten möglich, da im Falle formativer Indikatoren trotz hoher Korrelationen nicht auf eine inhaltliche Nähe geschlossen werden kann. Der Anteil aufgeklärter Varianz lag in der HKA bei 62%. Damit verbleibt ein Anteil von immerhin 38% unaufgeklärt, sodass der damit verbundene Informationsverlust möglicherweise Auswirkungen auf die nachfolgend berechneten Analysen hatte (siehe Abschnitt 7.6.).

Es könnte auch kritisch hinterfragt werden, dass bei der Berechnung der sequentiellen LR keine weiteren kundennahen Erfolgsfaktoren, wie beispielsweise das Nutzervergnügen oder das Vertrauen, einbezogen wurden. Im Vordergrund der Untersuchung stand jedoch die Frage bzgl. des Einflusses der beiden ufosV2a-Skalen bzw. der Nutzungsintention auf die Auswahlentscheidung von Nutzern einer Krankenkassenwebsite. Diese Hypothesen konnten mit Hilfe des PLS-Strukturgleichungsansatzes nicht überprüft werden. Eine Analyse der Einflussstärke der weiteren Konstrukte auf die Auswahlentscheidung stand hingegen nicht im Focus der

Untersuchung. Eine Untersuchung ihrer Zusammenhänge untereinander erfolgte im Rahmen einer PLS-Analyse (siehe Abschnitt 7.6).

8.3 Zusammenfassende Bewertung der Ergebnisse zu den ufosV2a-Skalen

Nachfolgend wird eine zusammenfassende Bewertung der Ergebnisse zu den ufosV2a-Skalen vorgenommen. Insgesamt können die Ziele der Untersuchung im Wesentlichen als erreicht betrachtet werden. Ausgehend vom ufosV2-Fragebogen (Christophersen, 2007; Christophersen & Konradt, 2008) wurden eine reflektive und eine formative Skala zur Erfassung der Usability von Websites für Dienstleistungen entwickelt und am Beispiel der gesetzlichen Krankenkassen überprüft. Die reflektive Skala ufosV2ra umfasst sieben reflektive Items zur Erfassung eines Globalurteils des Konstruktes der Usability. Die formative Skala ufosV2fa besteht aus 27, im Vergleich zu ursprünglich 46 formativen Items, die die wesentlichen Facetten der Usability von Websites abbilden.

Die beiden ufosV2a-Skalen konnten hinsichtlich sämtlicher untersuchter Gütekriterien empirisch betätigt werden. Die reflektive Usability-Skala ufosV2ra erweist sich als hoch reliabel hinsichtlich ihrer internen Konsistenz und als geeignet, um zwischen den Websites verschiedener Anbieter zu differenzieren. Die faktorielle Validität der reflektiven Skala kann als gegeben angesehen werden. Eine leichte Einschränkung dieses Kriteriums angesichts der Ergebnisse einer EFA ist vermutlich auf eine Voraussetzungsverletzung zur Berechnung der EFA zurückzuführen. Zwischen der formativen ufosV2fa-Skala und der reflektiven ufosV2ra-Skala besteht ein hoher Zusammenhang, was als Hinweis für eine hohe Inhaltsvalidität der Skalen gewertet werden kann.

Die Kriteriumsvalidität der beiden Usability-Skalen wird angesichts der Untersuchungsergebnisse in zweifacher Hinsicht bestätigt. So wurden einerseits in einem komplexen PLS-Modell starke Zusammenhänge zwischen den ufosV2a-Skalen und der Erfolgsgröße Nutzerzufriedenheit deutlich. Diese Ergebnisse weisen auf eine hohe konkurrente Kriteriumsvalidität der ufosV2a-Skalen hin. Eine gewisse Schwächung der Ergebnisse ließe sich möglicherweise darin sehen, dass sich in Bezug auf die Erfolgsgröße Nutzungsintention lediglich für die formative ufosV2fa-Skala ein signifikanter Zusammenhang zeigte, was vermutlich auf einen Mediatoreffekt zurückzuführen ist.

Zum anderen konnte mit Hilfe einer binomialen LR gezeigt werden, dass beide ufosV2a-Skalen eine hohe Vorhersagekraft für die Auswahlentscheidung eines Teilnehmers besitzen, welche der beiden Krankenkassenwebsites er zukünftig bevorzugt aufsuchen würde, um sich in Bezug auf Krankenversicherungsthemen zu informieren. In Folge dessen kann auch die prädiktive Kriteriumsvalidität der beiden ufosV2a-Skalen als hoch eingeschätzt werden.

Weiterhin ist festzuhalten, dass auch die angenommenen Zusammenhänge zwischen den beiden ufosV2a-Skalen und den LVn Vertrauen und Nutzervergnügen empirische Bestätigung fanden. Diese Ergebnisse stützen die nomologische Validität der beiden Usability-Skalen.

Eine gewisse Schwächung der Ergebnisse könnte darin gesehen werden, dass Zusammenhangsannahmen im nomologischen Modell zwischen den LVn Vertrauen, Nutzervergnügen und Nutzerzufriedenheit, nur eingeschränkt empirische Unterstützung fanden. Als Erklärung können Suppressoreffekte vermutet werden, die die Effekte zwischen den LVn teilweise aufheben.

Die vorliegende Arbeit unterscheidet sich gegenüber der Untersuchung durch Christophersen (2007) sowie anderen Studien zum Online-Shopping hinsichtlich mehrerer Aspekte. So wählte Christophersen im Falle der zitierten Untersuchung das aufwändigere Vorgehen anhand eines Laborexperiments. In der vorliegenden Untersuchung hingegen kam ein quasiexperimentelles Vorgehen anhand eines onlinegestützten Fragebogens zur Anwendung. Dieses zeigte sich bei Sicherstellung einer ausreichenden Datenqualität als vorteilhaft in Bezug auf die Gewinnung der Versuchsteilnehmer (vgl. Evans & Mathur, 2005). Während Christophersen eine möglichst weitgehende Annäherung der Stichprobe an die Gesamtpopulation der Internetnutzer anstrebte, wurde in der vorliegenden Untersuchung bewusst auf ein Teilsegment der Population fokussiert. Um die praktische Relevanz der Ergebnisse zu steigern, wurde vorrangig die Zielgruppe der Studenten und Auszubildenden betrachtet.

Die Untersuchung von Christophersen zielte auf den Untersuchungsbereich von Online-Shops ab. Hierfür wurden 35 Online-Shops über fünf verschiedene Produktarten einbezogen. Eine Zielsetzung der vorliegenden Untersuchung bestand hingegen darin, das vorliegende Fragebogeninventar ufosV2 hinsichtlich seiner Eignung zur Erfassung des Konstruktes Usability von Websites bei Dienstleistungen weiter zu entwickeln. Die Überprüfung erfolgte am Untersuchungsobjekt der Websites gesetzlicher Krankenkassen, Anbieter von Dienstleistungen mit eher geringem Kundeninvolvement.

Eine weitere Zielsetzung der vorliegenden Arbeit bestand darin, die Wirkung des Involvement auf die Nutzungsprozesse von Websites für Dienstleistungen zu untersuchen. Entgegen der in der Literatur dargestellten Zusammenhänge bestätigte sich eine moderierende Wirkung des Involvement auf die genannten Zusammenhänge jedoch nicht. Dies ist möglicherweise dem Umstand zuzuschreiben, dass Websites von Krankenkassen tendenziell eher von einem niedrigen Involvement geprägt sind. Es kann vermutet werden, dass moderierende Effekte des Involvement eher schwach ausgeprägt sind.

Im Vergleich zur Untersuchung von Christophersen (2007) war es nicht möglich, tatsächliches Kauf- oder Auswahlverhalten zu operationalisieren und als AV in die statistischen Analysen einzubeziehen. Eine tatsächliche Auswahlentscheidung zwischen den Krankenkassen als Versicherungsgeber war im Rahmen des Versuchs nicht praktikabel durchführbar. Als kategoriale Variable unterscheidet sich die Auswahlentscheidung jedoch hinsichtlich ihrer Operationalisierung von den Variablen auf Prädiktorebene, sodass einem potentiellen Common Method Bias entgegen gewirkt wird (vgl. Podsakoff et al., 2003).

Es stellt sich die Frage, inwieweit die dargestellten Untersuchungsergebnisse aus methodischen Gründen möglicherweise in ihrer Validität eingeschränkt sind. Um eine hohe interne Validität sicherzustellen wurden einige versuchsplanerische Vorkehrungen getroffen. Hinweise auf eine Einschränkung der internen Validität aufgrund verzerrender Störeinflüsse ergaben sich nicht.

Hinsichtlich der externen Validität kann festgestellt werden, dass eine Übertragbarkeit der Ergebnisse auf Websites weiterer Krankenkassen als gerechtfertigt scheint. Diesbezüglich ist die externe Validität als hoch anzusehen. Bezüglich der Übertragbarkeit auf Websites anderer Anbietersegmente lassen sich jedoch keine Aussagen begründen.

In Bezug auf die Teilnehmerstichprobe kann kritisch angemerkt werden, dass die Untersuchungsergebnisse nicht ohne weiteres auf die Gesamtpopulation der deutschen Internetnutzer übertragbar sind. Diese Einschränkung ist allerdings dem Umstand geschuldet, dass die Konzentration auf das Segment der Studenten und Auszubildenden bewusst gewählt wurde, um die praktische Relevanz der Untersuchungsergebnisse zu steigern. Eine direkte Übertragbarkeit auf andere Länder ist aufgrund des deutschsprachigen Fragebogeninventars und des Untersuchungsgegenstandes ausgeschlossen.

8.4 Erfolgswirksamkeit der Usability bei Websites

Nachdem eine Bewertung der Untersuchungsergebnisse zu den ufosV2a-Skalen vorgenommen wurde, lassen sich die Befunde auch hinsichtlich der Erfolgswirksamkeit der Usability im Rahmen der Anwendung von Websites für Dienstleistungen interpretieren. Anhand der durchgeführten PLS- und LR-Analysen kann nicht nur die Validität der Skalen bestätigt werden, sondern auch die Bedeutung der Usability für die Akzeptanz und Nutzung von Websites für Dienstleistungen untermauert werden. So zeigen die hohen positiven Pfadkoeffizienten zwischen der Usability und der Nutzerzufriedenheit, dass die Zufriedenheit des Nutzers einer Website wesentlich durch dessen Wahrnehmung der Usability-Ausprägung der Website beeinflusst wird. Ebenso steht das Wahlverhalten des Anwenders zwischen unterschiedlichen Websites gemäß der Ergebnisse der LR-Analyse in einem starken Zusammenhang zu seiner Einschätzung der Usability der Websites. Nachfolgend sollen Implikationen der Skalen und der Befunde für die Forschung und Anwendung in der Praxis betrachtet werden.

8.5 Implikationen für Forschung und Praxis

Die Bedeutung der Usability auf den Erfolg von Websites wurde bereits in verschiedenen Studien empirisch bestätigt (siehe Abschnitt 2.7.5). Die vorliegende Untersuchung untermauert diese Befunde und ergänzt sie, indem sie den Einfluss des Involvement auf die Usability mit einbezieht. Dabei bestätigt sie die Untersuchungsergebnisse von Christophersen (2007) zur Erfassung des Konstruktes Usability bei E-Commerce-Anwendungen für den

Untersuchungsbereich der Online-Dienstleistungen. Entgegen der in der Literatur dargestellten moderierenden Wirkung des Involvement auf den Zusammenhang von Usability bzw. Nutzerzufriedenheit einerseits und Nutzungsintention andererseits bestätigten sich diese Zusammenhänge in der vorliegenden Untersuchung nicht (vgl. Ranaweera et al., 2004, 2005).

Für die praktische Anwendung der ufosV2a-Skalen bedeuten die angeführten Befunde, dass eine verbesserte Ausprägung der Usability einer Website mit einer hohen Wahrscheinlichkeit zu einer höheren Nutzung der Website führen wird. Allerdings ist nicht davon auszugehen, dass dieser Zusammenhang durch eine Intensivierung des Kundeninvolvement, beispielsweise aufgrund des Einsatzes aktivierender Gestaltungselemente, verstärkt werden kann.

Die beiden ufosV2a-Skalen zeichnen sich entsprechend der Spezifikation des Messmodells durch unterschiedliche Eigenschaften aus, die wiederum ihre Eignung für die Anwendung in unterschiedlichen Kontexten bestimmen. So zeichnet sich die reflektive Skala ufosV2ra dadurch aus, dass sie lediglich sieben Items umfasst, die von Seiten des Nutzers Globalurteile zur Usability abfragt. Der Einsatz der Skala ist entsprechend ihrer Kürze mit wenig Aufwand verbunden und eignet sich besonders für Befragungssituationen, in denen auf Seiten der Internetnutzer nur wenig Bereitschaft zur Beantwortung umfangreicher Fragebögen zu erwarten ist. Angesichts der Abfrage summativer Globalurteile eignet sich die Skala nicht, um Rückschlüsse auf konkrete Usability-Defizite in der Ausgestaltung einer Website zu erhalten. Häufig besteht das Ziel der Anwendung von Usability-Skalen jedoch nicht darin, konkrete Gestaltungsmängel zu identifizieren.

Der Einsatz der reflektiven Usability-Skala ufosV2ra bietet sich beispielsweise an, wenn es darum geht, verschiedene Prototypen in der Entwicklungsphase einer Website zu vergleichen, oder eine Website im zeitlichen Verlauf wiederholt zu evaluieren (vgl. Kim, Shaw & Schneider, 2003). Auch im wissenschaftlichen Verwendungsbereich, etwa in Studien zur Erfolgsfaktorenforschung, liegt das Interesse häufig nicht darin, gezielte Erkenntnisse zu einzelnen Usability-Aspekten zu erlangen, sondern in einer reliablen und validen Erfassung der Gesamtausprägung des Konstruktes Usability. Die reflektive Usability-Skala ufosV2ra scheint somit sowohl für den Einsatz im Rahmen praktischer Messungen als auch für Forschungszwecke geeignet.

Die formative Usability-Skala ufosV2fa zeichnet sich hingegen durch eine vergleichsweise hohe Anzahl von 27 Items aus. Diese ermöglichen differenzierte Bewertungen einzelner Usability-Aspekte einer Website. So bietet sich ihre Anwendung beispielsweise an, wenn es darum geht, unbefriedigend ausgestaltete Funktionsbereiche einer Website zu identifizieren (Hamborg & Gediga, 2006). Denkbar ist sowohl der Vergleich verschiedener Versionen derselben Website als auch die vergleichende Evaluation unterschiedlicher Websites.

Wie in Abschnitt 2.8.2 dargestellt ist der Einsatz von Usability-Fragebögen potentiell mit verschiedenen Vor- und Nachteilen verbunden. So sind Fragebögen enge Grenzen gesetzt, wenn es darum geht, im Rahmen ihrer Anwendung konkrete Gestaltungsempfehlungen abzu-

leiten. Auch im vorliegenden Fall ist es weder mit Hilfe der reflektiven noch der formativen Skala möglich, konkrete Empfehlungen zur Verbesserung der Usability einer Website abzuleiten (vgl. Hamborg & Gediga, 2006).

8.6 Weiterführende Fragestellungen

Ausgehend von den dargestellten Ergebnissen können ergänzende Fragestellungen abgeleitet werden, die an die vorliegende Arbeit anknüpfen. So können die Validierungsuntersuchungen zunächst auf weitere Dienstleistungsarten ausgeweitet werden. Dabei bieten sich Websites von Dienstleistungsarten an, die durch ein höheres Involvement gekennzeichnet sind, um mögliche Moderatorwirkungen des Involvement erneut zu untersuchen. So ist beispielsweise bei Seiten, die dem Herunterladen von Musiktiteln oder von Filmen dienen, von einem höheren Involvement auszugehen.

Angesichts des hohen Maßes an Multikollinearität zwischen den formativen ufosV2fa-Indikatoren und der daraus resultierenden Bildung von Unterindizes war es nicht möglich, die individuellen Gewichte der einzelnen formativen Usability-Indikatoren zu bestimmen. So wäre es von Interesse, die Einflüsse der einzelnen formativen Indikatoren zu bestimmen und daraus auf ihre Relevanz in verschiedenen Nutzungskontexten schließen zu können.

Desweiteren könnte eine Reduktion des Skalenumfangs, der mit 27 Items immer noch recht umfangreichen formativen ufosV2fa-Skala, von Interesse sein, um den mit der Beantwortung verbundenen Aufwand zu reduzieren. Eine mögliche Vorgehensweise wäre es, Items von nachgeordneter Relevanz zu eliminieren. Grundlage könnte es sein, für jedes Item eine Bewertung seiner Wichtigkeit vornehmen zu lassen. Die reduzierte Skala wäre entsprechend der in Abschnitt 6.6.2.3 dargestellten Vorgehensweise anhand der Stärke des Zusammenhangs zum Globalurteil der ufosV2ra-Skala zu überprüfen.

Es sollte auch empirisch überprüft werden, inwieweit sich die der Arbeit zugrunde liegende Annahme bestätigt, dass fehlspezifizierte Usability-Skalen gegenüber korrekt spezifizierten Skalen eine geringere Validität aufweisen. So erscheint es untersuchenswert, inwiefern sich ein korrekt spezifiziertes Maß zur Erfassung der nutzerseitig bewerteten Usability gegenüber einer fehlspezifizierten Skala als valider erweist. Dies könnte analog zur in Abschnitt 7.7 dargestellten Vorgehensweise zur Prüfung der prädiktiven Validität der Usability-Skalen erfolgen. So könnten die konkurrierenden Skalen hinsichtlich ihrer prognostischen Güte bezüglich eines geeigneten Kriteriums verglichen werden. Eine höhere Prognosegüte im Falle von korrekt spezifizierten formativen und reflektiven Skalen im Vergleich zu einer fehlspezifizierten Skala würde für die Bedeutung einer klaren Unterscheidung zwischen beiden Messmodellen bei der Entwicklung von Usability-Skalen sprechen.

Zusammenfassung der Untersuchung

Das Internet gewann in den letzten Jahren immer stärker an Bedeutung. Laut der AGOF-Studie Internet Facts 2008 I, an der 110.947 Internetnutzer ab 14 Jahren teilnahmen, liegt die durchschnittliche Reichweite des Internet in den Altersgruppen ab 14 Jahren bei annähernd 63% (AGOF, 2008). Allerdings nimmt ein großer Teil der Webnutzer häufig oder gelegentlich Probleme während der Internetsitzung wahr, so z.B. aufgrund von unübersichtlichen Homepages, aufwändiger Informationssuchen oder aufgrund eines langsamen Seitenaufbaus (vgl. Dzida & Wandke, 2006, S. 483; Van Eimeren, Gerhard & Frees, 2002).

Die Bedeutung der Usability (Gebrauchstauglichkeit) auf den Erfolg von Websites wurde bereits in verschiedenen Studien empirisch bestätigt. Die Usability einer Website beschreibt deren Qualität aus einer benutzerorientierten Perspektive. Eine Website wird dann als gebrauchstauglich bezeichnet, wenn sie in einem bestimmten Nutzungskontext effektiv, effizient und in zufriedenstellender Weise genutzt werden kann (DIN-ISO 9241-11). Die vorliegende Untersuchung untermauert die Befunde für den Untersuchungsbereich der Online-Dienstleistungen am Beispiel der gesetzlichen Krankenversicherung und ergänzt sie, indem sie den Einfluss des Involvement auf die Usability mit einbezieht. Hierbei wird Bezug genommen auf die Arbeit von Christophersen (2007), der mit der Usability-Skala *uvosV2* ein Instrument zur fragebogenbasierten Erfassung des Konstruktes Usability aus Nutzersicht entwickelte.

Ausgehend vom *ufosV2*-Fragebogen wurden eine reflektive Skala (*ufosV2ra*, 7 Items) sowie eine formative Skala (*ufosV2fa*, 27 Items) zur Erfassung der Usability von Websites für Dienstleistungen entwickelt. Das Untersuchungsmodell basiert auf dem Technology Acceptance Model, erweitert um die Variablen Vertrauen, Ästhetik, Nutzungsvergnügen und Involvement. Die Überprüfung der Skalen erfolgte am Untersuchungsobjekt der Websites gesetzlicher Krankenkassen, Anbieter von Dienstleistungen mit tendenziell eher geringem Kundeninvolvement. Der quasi-experimentelle Versuchsablauf erfolgte mit einer computer-gestützten Versuchsumgebung. Jeweils zwei Onlineangebote sollten von den Versuchsteilnehmern anhand von Musteraufgaben genutzt und mit Hilfe eines standardisierten Fragebogens detailliert bewertet werden. Neben der Usability und weiteren Erfolgsfaktoren von Webangeboten wurde eine Auswahlentscheidung zwischen den zwei Onlineangeboten erfasst. Für $N = 256$ Teilnehmer liegen je Datensatz die Bewertungen von zwei Websites vor. Die Anzahl der Datenfälle verdoppelt sich auf $N = 512$ mit einer bewerteten Website je Fall.

Die beiden *ufosV2a*-Skalen konnten hinsichtlich sämtlicher untersuchter Gütekriterien empirisch betätigt werden. Die reflektive Usability-Skala *uvosV2ra* erweist sich mit einem Wert von $\alpha = 0.95$ als hoch reliabel hinsichtlich ihrer internen Konsistenz. Die Skala erwies sich darüber hinaus als faktoriell valide und als geeignet, um zwischen den Websites verschiedener Anbieter zu differenzieren. Die formative latente Variable Usability wurde überprüft, indem

sie gemäß einer Validierungsprozedur nach Diamantopoulos und Winklhofer (2001) in einem Zwei-Konstrukt-Modell als Prädiktor zu der endogenen reflektiven latenten Variable Usability in Beziehung gesetzt wurde. Multikollinearität zwischen den formativen Indikatoren wurde durch Bildung von Indizes berücksichtigt. Den Ergebnissen folgend bildet das Set der 27 verwendeten formativen Usability-Indikatoren das Globalurteil der ufosV2ra-Skala sehr gut ab. Mit der formativen Skala ufosV2fa besteht ein valides Set an formativen Indikatoren zur Erfassung des Konstruktes Usability. Die Überprüfung der Hypothesen zu den Zusammenhängen zwischen den Modellelementen erfolgte mit Hilfe der Berechnung von zwei PLS-Modellen unter jeweiliger Einbindung der formativen bzw. reflektiven Usability-Skala. Dabei erwies sich der Einfluss der Usability auf die Nutzerzufriedenheit als deutlich stärker als bei den übrigen Erfolgsfaktoren.

Entgegen der in der Literatur dargestellten Zusammenhänge bestätigte sich die Annahme einer moderierenden Wirkung des Involvements auf den Zusammenhang von Nutzerzufriedenheit und Nutzungsintention nicht (vgl. Ranaweera et al., 2004, 2005). Dies ist möglicherweise dem Umstand zuzuschreiben, dass das Untersuchungsfeld der Websites von Krankenkassen tendenziell eher von einem niedrigen Involvement geprägt ist. Es kann vermutet werden, dass moderierende Effekte des Involvement eher schwach ausgeprägt sind. Für die Beziehung zwischen Usability und Nutzungsintention erweist sich lediglich für das formative Modell bei Verwendung nicht standardisierter Originaldaten ein signifikanter moderierender Effekt des Involvement.

Die Überprüfung eines Zusammenhangs zwischen Usability bzw. Nutzungsintention und der Auswahlentscheidung erfolgte mit Hilfe einer binomialen logistischen Regression (LR). Sowohl für die beiden ufosV2a-Skalen als auch für die Nutzungsintention konnte eine hohe Vorhersagekraft in Bezug auf die Auswahlentscheidung von Nutzern einer Krankenkassenwebsite festgestellt werden. In Folge dessen kann die prädiktive Kriteriumsvalidität der beiden ufosV2a-Skalen als hoch eingeschätzt werden.

Ein potentieller Einfluss der Abfrageposition der Usability auf die Untersuchungsergebnisse wurde anhand einer systematischen Variation der Itemreihenfolge untersucht, konnte jedoch nicht bestätigt werden. Es kann davon ausgegangen werden, dass eine Veränderung der Abfrageposition der Usability keinen Einfluss auf die Ergebnisse der Untersuchung nimmt.

Hinsichtlich der externen Validität der Ergebnisse kann festgestellt werden, dass eine Übertragbarkeit der Ergebnisse auf Websites weiterer Krankenkassen als gerechtfertigt scheint. Diesbezüglich ist die externe Validität als hoch anzusehen. Bezüglich der Übertragbarkeit auf Websites anderer Anbietersegmente lassen sich jedoch keine Aussagen begründen.

Summary

The importance of the world wide web in every day life was constantly increasing over the last years. According to the AGOF-Study Internet Facts 2008 I, in which 110.947 users of the internet up from the age of 14 participated, the range of internet users covers more than 63% of the population older than the age of 14 (AGOF, 2008). Nevertheless, the mayor part of the internet users notice frequent or occasional problems while using the internet, as there are confusingly build home pages, complex research for information or slow page reproduction (Dzida & Wandke, 2006, S. 483; Van Eimeren, Gerhard & Frees, 2002).

The importance of usability for the success of websites has been empirically proved by several studies. The usability of a website describes the quality of the website from a users' point of view. A website is qualified as usable if it can be used to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (DIN-ISO 9241-11). The study proves the results for the context of online services using the example of compulsory health insurances in Germany. It amends the results by implying the influence of the involvement on the usability. The study refers to the research of Christophersen (2007), who invented with the Usability scale ufosV2 an instrument for the assessment of the construct Usability from a users' point of view based on a questionnaire.

Based on the ufosV2-questionnaire a reflective scale (ufosV2ra, 7 items) and a formative scale (ufosV2fa, 27 items) were developed in order to measure the usability of websites for online services. The research model is based on the Technology Acceptance Model, enlarged by the variables trust, aesthetics, users' enjoyment and involvement. The examination of the scales was made on the analysis subject of websites of compulsory health insurances, hence providers of services with a tendency to low customers' involvement. The quasi-experimental test procedure was carried out within a computer-supported test environment. The participants of the study were supposed to use two online services while applying sample questions and value the services afterwards on the bases of a detailed standardized questionnaire. Apart from the usability and other additional factors for success the choice between two different online services was registered. The evaluation of two websites per dataset exists for $N = 256$ participants. The number of dataset is doubled to $N = 512$ with one evaluated website per case.

Both of the ufosV2a-scales could be empirically confirmed in regard to all of the analyzed effectiveness criterions. The reflective usability-scale ufosV2ra with a value of $\alpha = 0.95$ proves to be highly reliable regarding its internal consistence. Furthermore, the scale proved to be factorial valid and to be suitable for the distinction between the websites of different providers. The formative latent Variable Usability was tested according to the validation procedure by Diamantopoulos and Winklhofer (2001) by putting it as a predictor variable in relation to the endogen reflective latent variable Usability.

Multi-collinearity between the formative indicators was taken into account by formation of indices. Following the results the set of the 27 applied formative Usability indicators displayed the global result of the ufosV2ra-scale very well. The formative scale ufosV2fa forms a valid set of formative indicators to the acquisition of the construct usability. Testing the hypotheses concerning the relation between the model elements was carried out with the calculation of two PLS-models including the formative respectively reflective usability-scale. Usability proved to exert much more influence on the users' contentment than the other factors for success.

Contrary to the literature the assumption of a moderating effect of the involvement on the relationship between users' contentment and user's intention was not confirmed (Ranaweera et al., 2004, 2005). The result can be attributed to the fact that the field of research of websites of health insurances has a tendency to low involvement. It can be assumed that the moderating effects of the involvement are not highly developed. A significant moderating effect of the involvement on the relation between usability and users' intention was only tested for the formative model using not standardized original data.

The coherence between usability and users' intention and the final choice decision was examined by means of a binomial logistic regression (LR). For both of the ufosV2a-scales as well as for the users' intention a high power of prognosis in regard to the final choice decision of the users of a health insurance company website was confirmed. Therefore the predictive criterion validity of the ufosV2a-scales can be estimated as high.

The potential influence of usabilityes' position in the query on the results of the study was examined by a systematic variation of the items order, though could not be confirmed. It can be assumed that the alteration of the position of query has no influence on the results of the study.

In regard to the external validity of the results it can be assessed that the transferability of the results to websites of other health insurance companies can be justified. In this regard the external validity can be considered to be very high.

Literaturverzeichnis:

- ACTA (2004). *Allensbach Computer und Technik-Analyse*. Institut für Demoskopie Allensbach, Allensbach, 2005.
- Adams, D., Nelson, R. & Todd, P. (1992). Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication, *MIS Quarterly*, 16 (2), 227-247.
- AGOF (2007). *Internet facts 2006-III. Berichtsband Teil 1*. Arbeitsgemeinschaft Online-Forschung e.V. (AGOF), 2007.
- Ahlert, D., Becker, J., Kenning, P. & Schütte, R. (2000): *Internet & Co. im Handel: Strategien, Geschäftsmodelle, Erfahrungen*, Springer, Berlin, 2000.
- Aimichai-Hamburger, Y. (2002). Internet and personality, in: *Computers in Human Behavior*, 18 (1), S.1-10.
- Aiken, L.S. & West, S.G. (1991): *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park., CA: Sage.
- Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Eds.), *Action Control: From Cognition to Behavior* (11-39). New York: Springer.
- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Ajzen, I. & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs: Prentice-Hall Inc.
- Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A. & Wood, S. (1997). Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing*, 61, S. 38-53.
- Albers, S. (2007). PLS and success factors studies in marketing. In: V. E. Vinzi, W. W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares - concepts, methods and applications*. Springer, Berlin.
- Albers, S., & Hildebrandt, L. (2006): Methodische Probleme bei der Erfolgsfaktorenforschung: Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells, in: *zfbf, Zeitschrift für betriebswirtschaftliche Forschung*, S. 2-33.
- Albert, H. (1987). *Kritik der reinen Erkenntnislehre*, Mohr Siebeck, Tübingen 1987.
- Aleff, H.J. (2002). *Die Dimension der Zeit im Dienstleistungsmarketing*, Deutscher Universitäts-Verlag, Wiesbaden.

- Al-Gahtani, S.S. & King, M., (1999). Attitudes, satisfaction and usage: factors contributing to each in the acceptance of information technology. *Behaviour and Information Technology* 18, 277-297.
- Alpar, P. (1999). Satisfaction with a website: its measurement, factors and correlations, in: Scheer, A. & Nüttgens, M. (Hrsg.): *Electronic business engineering*. 4. int. Tagung Wirtschaftsinformatik, Heidelberg, S. 271-287.
- Amelang, M. & Bartussek, D. (1970). Untersuchungen zur Validität einer neuen Lügen-Skala. *Diagnostica*, 16, 103-122.
- Anandarajan, M., Simmers, C.A. & Igbaria, M., (2000). An exploratory investigation of the antecedents and impact of internet usage: an individual perspective. *Behaviour and Information Technology* 19, 69-85.
- Anderson, J. C. & Gerbing, D. W. (1991): Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities, in: *Journal of Applied Psychology*, Vol. 76, pp. 732-740.
- Anderson, J. R. (1996): *Kognitive Psychologie*, 2. Aufl., Heidelberg: Spektrum.
- Arbuckle, J.L. & Wothke, W. (1999). *AMOS 4.0 User's Guide*. Chicago, IL: Small Waters.
- Arminger, G. & Muthen, B. (1998). A Bayesian approach to nonlinear latent variable models using Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, 63, 271-300.
- A.T. Kearney (2000). *Satisfying the Experienced Online-Shopper*, New York.
- Ba, S. (2001). Establishing online trust through a community responsibility system. *Decision Support Systems*, 31, S. 323-336.
- Ba, S., Whinston, A.B., & Zhang, H. (2000). Mall companies in the digital economy, in: Brynjolfson, E. & Kahin, B. (Hrsg.): *Understanding the digital economy*, (S. 185-200), Cambridge (MA), London.
- Babin, B. J.; Darden, W. R. & Griffin, M. (1994): Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value, in: *Journal of Consumer Research* Vol. 20, Nr. 3, S. 4-656.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (1996): *Multivariate Analysemethoden*, 8. Aufl., Berlin/Heidelberg/New York.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 10. Aufl., Berlin: Springer.
- Bänsch, A. (1996): *Käuferverhalten*, 7. Aufl., Oldenbourg Verlag, München, 1996
- Bagozzi, R. P. (1975). Marketing as Exchange, *Journal of Marketing*, 39, 32-39.

- Bagozzi, R.P. (1994). Structural equation models in marketing research basic principles. In: *Principles of Marketing Research*, R. Bagozzi (ed), (S. 317-385), Oxford, Blackwell.
- Bagozzi, R.P., Yi, Y. (1988), On the Evaluation of Structural Equation Models, *Journal of the Academy of Marketing Science*, 16, 74-94.
- Bagozzi, R.P., Baumgartner, H. & Yi, Y. (1992). State Versus Action Orientation and the Theory of Reasoned Action: An Application of Coupon Usage, *Journal of Consumer Research*, 18, 505-518.
- Bankhofer, U. (1995): *Unvollständige Daten und Distanzmatrizen in der Multivariaten Datenanalyse*, Bergisch Gladbach, Köln, Verlag Josef Eul.
- Bapna, R., Goes, P. & Gupta, A. (2001). Insights and analyses of online auctions, in: *Communications of the ACM*, 44 (11), S. 42-50.
- Barney, J. (1991): Firm resources and sustained competitive advantage. In: *Journal of Management*, 17. Jg (1991), S. 99 120.
- Baron, M. & Kenny, A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations, In: *Journal of Personality and Social Psychology*, 51 (6), 1173-1183.
- Batiz-Lazo, B. & Wood, D. (2003). A historical appraisal of information technology in commercial banking, in: *Electronic Markets*, 12 (3), S. 192-205.
- Bauer, H.H., Fischer, M. & Sauer, N.E. (2000). Barrieren des elektronischen Einzelhandels Eine empirische Studie zum Kaufverhalten im Internet, in: *Zeitschrift für Betriebswirtschaft, ZfB*, 70 (10), S. 1133-1156.
- Beatty, S.E., Kahle, L.R., & Homer, P. (1988). The involvement commitment model: Theory and implications. In: *Journal of Business Research*, 16, S. 149-167.
- Beck, N., Brüderl, J. & Woywode, M. (2002): *The Causes and Consequences of Organizational Change: How to deal with Unobserved Heterogeneity*, Paper presented at the 20th EGOS annual conference, Barcelona, July 4 6.
- Berekhoven, L., Eckert, W. & Ellenrieder, P. (1996). *Marktforschung Methodische Grundlagen und praktische Anwendungen*, 7. vollst. überarb. und erw. Aufl., Wiesbaden: Gabler, 1996.
- Bettman, J. (1974). Relationship of Information Processing Attitude Structures to Private Brand Purchasing Behavior. In: *Journal of Applied Psychology*, 59 (1), 79-83.
- Bevan, N. & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology*, 13, 132-145.
- BGG (Behindertengleichstellungsgesetz), vgl. <http://bundesrecht.juris.de/bundesrecht/bgg>.

- Bhattacharjee, A. (2001). An empirical analysis of the antecedents of electronic commerce service continuance. *Decision Support Systems*, 32(2), 201-214.
- Bias, R. G. (1994). *The pluralistic usability walkthrough: coordinated empathies. Usability inspection methods*, John Wiley & Sons, Inc., New York, NY, 1994.
- Bieberbach, F. & Hermann, M. (1999). Die Substitution von Dienstleistungen durch Informationsprodukte auf elektronischen Märkten, in: Scheer, A. & Nüttgens, M. (S.67-81), *Electronic business engineering*. 4. int. Tagung Wirtschaftsinformatik, Heidelberg.
- Blackwell, R.D., Miniard, P.W. & Engel, J.F. (2001). *Consumer behavior*. New York, Hartcourt College Publishers.
- Bliemel, F. & Fassott, G. (2000): Produktpolitik im Electronic Business, in: Weiber, R. (Hrsg.): *Handbuch Electronic Business: Informationstechnologien Electronic Commerce Geschäftsprozesse*, (S. 55-522), Wiesbaden.
- Bloch, P. H. & Richins, M. (1983). A Theoretical Model for the Study of Product Importance Perceptions, *Journal of Marketing*, 47, 69-81.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, K.A. & Lennox, R. (1991): Conventional Wisdom on Measurement: A Structural Equation Perspective, *Psychological Bulletin*, 110, 305-314.
- Bollen, K. A. & Stine, R. A. (1993): Bootstrapping goodness-of-fit measures in structural equation models, in: Bollen, K. A. & Long, J. S. (Hrsg.), *Testing structural equation models*, (S. 111 ff.), Newbury Park et al.
- Bopp, C. & Wörmann, C. (2002): Klartext: Auf dem Weg zu Standard-Navigationswordings im deutschsprachigen Internet. In: M. Herczeg, W. Prinz, H. Oberquelle (Hrsg.): *Mensch & Computer 2002: Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*. (S. 405-406), Stuttgart: B. G. Teubner, 2002.
- Borkenau, P. (1981). *Intraindividuelle Variabilität und differentielle Vorhersagbarkeit*. Heidelberg, Univ., Diss., 1982.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. (6. Aufl.).Heidelberg: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Bortz, J., Lienert, G.A. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik* (2. Aufl.). Heidelberg: Springer.
- Bost, E. (1987). *Ladenatmosphäre und Konsumentenverhalten*, Heidelberg: Physica-Verlag.

- Boyer, K., Hallowell, R. & Roth, A. (2002). E-Services: operations strategy, a case study and a method for analysing operational benefits, in: *Journal of Operations Management*, 20 (2), S. 175-188.
- Brady, L. & Phillips, C. (2003). Aesthetics and Usability. A Look at Color and Balance. Software Usability Research Laboratory, Wichita State University, *Usability News* (Nr. 5.1), 2003, revidiert am 8.8.2003,
<http://psychology.wichita.edu/surl/usabilitynews/51/aesthetics.htm>
- Brännback, M. & Puhakainen, J. (1998). Web marketing: has the distinction between products and services become obsolete? In: *Journal of Market Focussed Management*, 3 (1), S. 47-59.
- Braunstein, C. (2001). *Einstellungsforschung und Kundenbindung*. Wiesbaden: Dt. Universitäts-Verlag.
- Breithaupt, H.F. (2002). Dienstleistungsqualität im Internet am Beispiel von Intermediären, in Bruhn, M & Staus, B. (Hrsg.): *Electronic Services Dienstleistungsmanagement Jahrbuch 2002*, (S. 177-207), Wiesbaden, Deutscher Universitäts-Verlag.
- Breithaupt, H.F. (2005). *Dienstleistungen im Internet und ihre Qualität aus Kundensicht*, Deutscher Universitäts-Verlag, Wiesbaden.
- Brielmaier, A. & Diller, H., 1995. Die Organisation internationaler Vertriebsaktivitäten. Problemfelder, Einflussfaktoren und Lösungsansätze aus der Sicht der Transaktionskostentheorie. In: Kaas, K.P. (ed.), *Kontrakte, Geschäftsbeziehungen, Netzwerke Marketing und neue Institutionenökonomik. Zfbf Schmalenbachs Zeitschrift für die betriebswirtschaftliche Forschung*, special edition, 35, 205-222.
- Browne, L., & Arminger, G. (1995). Specification and estimation of mean and covariance-structure models. In G. Arminger, C. Clogg, & M. Sober (Eds.), *Handbook of statistical modeling for the social and behavioral sciences*, 185-249. New York: Plenum Press.
- Bruhn, M. (2000): Qualitätssicherung im Dienstleistungsmarketing eine Einführung in die theoretischen und praktischen Probleme, in: Bruhn, M. & Staus, B. (Hrsg.): *Dienstleistungsqualität: Konzepte Methoden Erfahrungen*, 3. Aufl. (S. 21-48) Wiesbaden, Gabler.
- Bruhn, M. (2001): *Relationship Marketing*, München, Vahlen.
- Bruhn M. (2002): Electronic Services eine Einführung. In: Bruhn, M & Staus, B. (Hrsg.): *Electronic Services Dienstleistungsmanagement Jahrbuch 2002*, (S. 3-41) Wiesbaden, Gabler.
- Bruhn, M. (2003): *Kundenorientierung*, 2. Aufl., München, Vahlen.

- Bruhn, M. & Meffert, H. (2001). *Handbuch Dienstleistungsmanagement*, 2. Aufl., Wiesbaden, Gabler.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Burchert, H. (1998). *Ökonomische Evaluation von Telematik-Anwendungen im Gesundheitswesen und Schlussfolgerungen für ihre Implementierung*, Veröffentlichungen des Lehrstuhls für Allgemeine Betriebswirtschaftslehre und Betriebswirtschaftliche Finanzwirtschaft, insbesondere Untenehmensbewertung, Ernst-Moritz-Arndt-Universität Greifswald, Internet-Verffentlichung Nr. 5, 1998. <http://www.rsf.uni-greifswald.de/bwl/index.html>.
- Burmester, M., Hassenzahl, M. & Koller, F. (2002): Usability ist nicht alles Wege zu attraktiven interaktiven Produkten. *I-Com*, 1 (1), S. 32-40.
- Burnkrant, R. & Sawyer, A. (1983). Effect of involvement and message content on informationprocessing intensity, in: Haris, R. (Hrsg.), *Information processing research in advertising*, Hilldale, S. 3-64.
- Burroughs, R. & Sabherwal, R. (2002). Determination of retail electronic purchasing: a multi-period investigation. *Information systems and operational research*, 40 (1), 35-57.
- Busemeyer J. & Jones, L. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. In: *Psychological Bulletin*, 93, 549-563.
- Buzzell, R. & Gale, B. (1989). *Das PIMS-Programm, Strategien und Unternehmenserfolg*, Wiesbaden: Gabler, 1989.
- Caplan, S.H. (1994). Making Usability a Kodak Produkt Differentiator, in: M. Wiklund (Hrsg.). *Usability in Practice: How Companies Develop User Friendly Products*, Boston: Academic Press, 21-58.
- Carroll, J. M., & Thomas, J. C. (1988). Fun. *SIGCHI Bulletin*, 19(3), 21-24.
- Chau, P.Y., (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems*, 13(2), S.85-104.
- Chain, M.M., Sarasohn-Kahn, J. & Wayne, J.C., (2000). *Health e-people: the online consumer experience*, California Healthcare Foundation, <http://admin.chcf.org/documents/ehealth/healthpeople.pdf>.
- Chatterjee, S. & Price, B. (1977): *Regression Analysis by Example*, New York et al, John Wiley.
- Chen, L., Gillenson, M. & Sherell, D. (2002). Enticing online consumers: an extended technology acceptance perspective. *Information and Management* 39, 705-719.

- Chen, L., Gillenson, M. & Sherell, D. (2004). Consumer acceptance of virtual stores: a theoretical model and critical success factors for virtual stores. *Database for Advances in Information Systems* 35(2), 8-31.
- Chignell, M, 1990, A taxonomy of user interface terminology. *SIGCHI Bull*, 21.4, 27-34.
- Chin, J.P., Diehl, V.A. & Norman, K.L. (1988), Development of an Instrument measuring User satisfaction of the human computer interface, in: *Proceedings of the computer and human interaction conference*, pp. 213-218.
- Chin, W.W. (1998): The Partial Least Squares Approach to Structural Equation Modeling, in: Marcoulides, G.A. (Hrsg.): *Modern Methods for Business Research*, New Jersey, 295-336.
- Chin, W. W. & Newsted, P. R. (1999). Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares. In: *Statistical Strategies for Small Sample Research*, Rick Hoyle (ed.), (pp. 307- 341) Thousand Oaks, CA: Sage Publications.
- Chin, W.W. & P.A. Todd (1995): On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research: A Note of Caution, *MIS Quarterly*, 19, 237-246.
- Chiravuri, A. & Nazareth, D., (2001). Consumer trust in electronic commerce: an alternative framework using technology acceptance. *Proceedings of the seventh Americas Conference on Informations Systems*.
- Choi, S.Y., Stahl, D.O. & Whinston, A.B. (1997). *The economics of electronic commerce*, Indianapolis, Mac-Millan Technical Publishing.
- Christophersen, T. & Konradt, U. (2004). Wettbewerbsvorteile im Online-Shopping durch kundenorientierte Gestaltung der Benutzungsoberfläche. In: Albers, S., Haßmann, V., Somm, F. & Tomczak, T. (Hrsg.). *Verkauf: Kundenmanagement, Vertriebssteuerung, E-Commerce*. Loseblattwerk und Online-Dienst www.verkauf-aktuell.de. Gabler: Wiesbaden.
- Christophersen, T. (2006). *Usability im Online-Shopping: Entwicklung eines Fragebogeninstrumentes (ufosV2) unter Berücksichtigung formativer und reflektiver Messmodelle*. Universität Kiel, Dissertation.
- Christophersen, T. & Grape, C. (2006). Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (115-132). Wiesbaden: Deutscher Universitäts-Verlag.
- Christophersen, T. & Konradt, U. (2008). The development of a formative and a reflective scale for the assessment of on-line store usability. Paper submitted for publication.
- Christophersen, T. & Konradt, U. (in press). Reliability, validity, and sensitivity of a single-item measure of online store usability. *International Journal of Human-Computer Studies*.

- Churchill, G.A., Jr. (1979): A Paradigm for Developing Better Measures of Marketing Constructs, *Journal of Marketing Research*, 16, 64-73.
- Clarke R. (1997). *Regulating Financial Services in the Marketplace: The Public's Interests*, Conference of the Australian Securities Commission Conference on 'Electronic Commerce: Regulating Financial Services in the Marketplace', Sydney, 4-5 February 1997, at <http://www.anu.edu.au/people/Roger.Clarke/EC/ASC97.html>
- Clemons, E. (2007). An Empirical Investigation of Third-Party Seller Rating Systems in E-Commerce: The Case of buySAFE. *Journal of Management Information Systems*, 24(2), 43-71.
- Cloninger C. (2001). *Fresh Styles for Web Designers. Eye Candy from the Underground*. Indianapolis, New Riders.
- Cohen, J. (1988): *Statistical Power Analysis for the Behavioral Sciences*, 2. Aufl., Hillsdale, NJ.
- Conklin, E. J. (1987): Hypertext: An introduction and survey. *IEEE Computer*, 20 (9), 17-41.
- Conner, M., & Armitage, C. J. (1998). Extending the Theory of Planned Behavior: A Review and Avenues for Future Research. *Journal of Applied Social Psychology* (28), pp. 1429-1464.
- Cote, J.A., & Buckley, R. (1988). Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation. *Journal of Consumer Research*, 14, 579-582.
- Coughlan, J., Macredie, R. & Patel, N., (2007). Moving face-to-face communication to Web-based systems, *Interacting with Computers*, 19 (1), 1-6.
- Couper, M., Traugott, M. & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65, S. 230-253.
- Cove, J. F. & Walsh, B. C. (1988). Online text retrieval via browsing. *Information Processing and Management*, 24, 31-37.
- Cyr, D. (2008). Modeling Web Site Design Across Cultures: Relationships to Trust, Satisfaction, and E-Loyalty. *Journal of Management Information Systems*, 24(4), 47-72.
- Dahlberg, T, Mallat, N., & Oorni, A. (2003). Trust enhanced Technology Acceptance Model Consumer acceptance of mobile payment solutions. *Proceedings of the CIC Roundtable 2003*. Elektr. Ressource verfügbar unter http://web.hhs.se/cic/roundtable2003/papers/D_31_dahlberg_et_al.pdf
- Dahui, L., Browne, G., & Wetherbe, J.,(2006). Why Do Internet Users Stick with a Specific Web Site? A Relationship Perspective.Preview. *International Journal of Electronic Commerce*, 10 (4), 105-141.

- Dalbert, C. & Schmitt, M. (1986). Einige Anmerkung zur Formulierung und Überprüfung von Moderatorhypothesen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 7 (1), 29-43.
- Darby, M.R. & Karni, E. (1973). Free competition and the amount of fraud, in: *Journal of Law and Economics*, 16 (1), S. 67-86.
- Davis, F.D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.
- Davis, F.D. (1993). User Acceptance of Information Technology: System Characteristics, User Perceptions and Behavioral Impacts. *International Journal of Man-Machine Studies*, 38, 475-487.
- Dayton, C.M. (1970). *The Design of Educational Experiments*. New York: McGraw-Hill.
- Dellarocas, C. (2002): *The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms*, Arbeitspapier MIT.
- DeLone, W. D. & McLean, E. R. (1992). Information systems success: The quest for the dependent variable; *Inform. Systems Research*, 1, 60-95.
- DeLone, W.D. & McLean, E.R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19 (4), 9-30.
- Deimel, K. (1989). Grundlagen des Involvement und Anwendung im Marketing, *Marketing Zeitschrift für Forschung und Praxis*, 11, S.153-161.
- De Ruyter, K., Wetzels, M. & Kleijnen, M. (2001). Customer adoption of e-services: an experimental study. *International Journal of Service Industry Management* 12(2), 184-207.
- Dess, G. & Davis, P. (1984): Porter's generic strategies as determinants of strategic group membership and organizations performance. In: *Academy of Management Journal* (27), S. 467 488.
- Devaraj, S. & Kohli, R. (2003). Performance Impacts of Information Technology: Is Actual Usage the Missing Link? *Management Science*, Vol. 49, (3), pp. 273 289.
- Dholakia, R.R. & Fortin, D.R (2002). Advertising on the net: what works and why. In: Dholakia, N. Fritz, W., Dholakia, R.R. & Mundorf, N. (Hrsg): *Global E-Commerce and Online marketing*, (S. 1-13) Westport, London.
- Dholakia, R.R. & Dholakia, N. (2001). Bedenken bezüglich des Internet: Ergebnisse einer empirischen Untersuchung. In W. Fritz (Hrsg.), *Internet-Marketing* (2. Aufl.), 423 432. Stuttgart: Schäffer-Poeschel.

- Diamantopoulos, A. (1999). Export performance measurement: reflective versus formative indicators. *International Marketing Review*, 16 (6), S. 444-457.
- Diamantopoulos, A. & Winklhofer, H. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38, S. 269-277.
- Diehl, S. (2002). *Erlebnisorientiertes Internetmarketing: Analyse, Konzeption und Umsetzung von Internetshops aus verhaltenswissenschaftlicher Perspektive*, Deutscher Universitäts-Verlag, Wiesbaden.
- Diekmann, A. (1998). *Empirische Sozialforschung Grundlagen, Methoden, Anwendungen*, 4. Auflage, Reinbeck bei Hamburg: Rowohlt.
- Dietrich, M. (1986). *Konsument und Gewohnheit, eine theoretische und empirische Arbeit zum habituellen Kaufverhalten*, Physica-Verlag, Heidelberg.
- Diezinger, A & Mayr-Kleffel V (1999) *Soziale Ungleichheit*. Freiburg i. Br.: Lambertus.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Dijkstra, Theo (1983), Some Comments on Maximum Likelihood and Partial Least Squares Methods, *Journal of Econometrics*, Vol. 22, S. 67-90.
- Diller, H. (1997): Beziehungsmanagement, in: Die Betriebswirtschaft, Vol. 57, Nr. 4, S. 572-575. Doney, P. M., & Cannon, J. P. (1997) An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, 61, 35-51.
- Diller, H. & Kusterer, M. (1987): Erlebnisbetonte Ladengestaltung im Einzelhandel Eine empirische Studie, in: Trommsdorff, V. (Hrsg.): *Handelsforschung 1986*, Berlin 1987, S. 105 123.
- Dillon, A. & Morris, M.G., 1996. User acceptance of information technology : theories and models. *Annual Review of Information Science and Technology* 31, 3-32.
- DIN EN ISO 9241-11 (1999). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit Leitsätze*. Berlin: Beuth.
- DIN EN ISO 9241-11 (1999). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit Leitsätze*. Berlin: Beuth.
- DIN EN ISO 9241-110 (2006). *Ergonomische Anforderungen der Mensch-System-Interaktion. Teil 110: Grundsätze der Dialoggestaltung*. Berlin: Beuth.
- Doney, P.M. & Cannon, J.P., 1997. An examination of the nature of trust in buyer seller relationships. *Journal of Marketing* 61, 35 51.

- Donovan, R.J., Rossiter, J.R., 1982. Store atmosphere: an experimental psychology approach. *Journal of Retailing*, 58, 34-57.
- Dzida, W. (1994). Qualitätssicherung durch software-ergonomische Normen. In: E. Eberleh, H. Oberquelle und R. Oppermann (Hrsg.): *Einführung in die Software-Ergonomie*. (373-406) de Gruyter, Berlin, 1994.
- Dzida, W. & Wandke, H. (2006). Software-Ergonomie: Gestalten und Bewerten interaktiver Systeme. In: B. Zimolong & U. Konradt (Hrsg.), *Ingenieurpsychologie. Enzyklopädie der Psychologie (Bd. D-III-2)*, S. 462-494, Göttingen: Hogrefe.
- Eberl, M. (2004). Formative und reflektive Indikatoren im Forschungsprozess: Entscheidungsregeln und die Dominanz des reflektiven Modells. *Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung*, 19. Ludwig-Maximilians-Universität, München.
- Eberl, M. & Schwaiger, M. (2005). Corporate reputation: disentangling the effects on financial performance. *European Journal of Marketing*, 39 (7/8), 838-854.
- Edwards, J. & Bagozzi, R. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5 (2), S. 155-174.
- Efron, B. & Gong, G. (1983): A Leisurely Look at the Bootstrap, the Jackknife, and Cross Validation, *The American Statistician*, 37, 1, 36-48.
- Egger, F.N. (2001). Affective Design of E-Commerce User Interfaces: How to Maximise Perceived Trustworthiness. In M.G. Helander, H.M. Khalid & M.P. Tham (Eds.), *Proceedings of the International Conference on Affective Human Factors Design* (317-327). London: ASEAN Academic Press.
- Eggert, A. & Fassott, G. (2003): Zur Verwendung formativer und reflektiver Indikatoren in Strukturgleichungsmodellen, in: VHB (Hrsg.): *Ökonomik, Management und Corporate Governance*, 65. Wissenschaftliche Jahrestagung, 10.-13. Juni 2003, Universität Zürich, 112-115.
- Eighmey, J., (1997). Profiling user responses to commercial websites. *Journal of Advertising Research*, 37(3), S. 59-66.
- Eimeren, B. v.; Gerhard, H. & Frees, B. (2002): ARD/ZDF Online-Studie 2002. Entwicklung der Online-Nutzung in Deutschland: mehr Routine, weniger Entdeckerfreude, in: *Media Perspektiven*, Vol. 7, S. 346-362.
- Eimeren, B. v.; Gerhard, H. & Frees, B. (2003): Internetverbreitung in Deutschland: Unerwartet hoher Zuwachs - ARD/ZDF Online-Studie 2003, in: *Media Perspektiven*, Vol. 8, S. 338-358.

- Einwiller, S. & Will, M. (2001). The role of reputation of engender trust in electronic markets. In: *Proceedings of the fifth international conference on corporate reputation, identity and competitiveness*, Paris. Electronic publication, URL:
<http://www.communicationsmgt.org/modules/pub/view.php/communicationsmgt-11>
- Emmons, R. & Diener, E. (1986) Influence of impulsivity and sociability on subjective well-being. *Journal of Personality and Social Psychology*, 50, 1211-1215.
- Emnid, (2000). Veröffentlicht unter <http://www.wuv-studien.de/wuv/studien/index.htm>
- Enders, A. (1997). *Informationsintegration bei der Produktbeurteilung*, Physica-Verlag, Heidelberg.
- Engel, J.F. , Blackwell, R.D. & Miniard, P.W. (1995). *Consumer Behavior*, 8. Auflage, New York.
- Engelhardt, W. H. & Freiling, J. (1995). Die integrative Gestaltung von Leistungspotentialen, in: *Zeitschrift für Betriebswirtschaft ZfB*, 47 (10), S. 899-910.
- Ericsson, K. A., & Simon, H. A. (1983). *Protocol Analysis*. Cambridge, MA: MIT Press.
- Ernst, H. (2001): *Erfolgsfaktoren neuer Produkte. Grundlagen für eine valide empirische Forschung*, Deutscher Universitäts-Verlag, Wiesbaden.
- Esch, F.-R., Hardiman, M. & Langner, T. (2000) Wirksame Gestaltung von Markenauftritten im Internet, *Thesis: Fachzeitschrift für Marketing*, (3), 10-16.
- Esch, F.-R., Langner, T. & Jungen, P. (1998): Kundenorientierte Gestaltung von Verkaufsauftritten im Internet, in: *Der Markt*, 37. Jg., Nr. 146/147, 1998, S. 129-145.
- Evans, J.R. & Mathur, A. (2005). The Value of Online Surveys. *Internet Research: Electronic Networking Applications and Policy*, 15(2), 195-219.
- Evans, R.G. & Wurster, T.S. (1999). Getting real about virtual commerce, in: *Harvard Business Review*, 77(6), 85-94.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating these of exploratory factor analysis in psychological research. *Psychological Methods*, 4 (3), 272-299.
- Fang, Chan, Brzezinski & Xu, (2005). Moderating Effects of Task Type on Wireless Technology Acceptance. *Journal of Management Information Systems*, 22 (3), 123-157.
- Featherman, M. & Pavlou, P. (2002). *Predicting e-services adoption. A perceived risk facets perspective*. Eighth Americas Conference on Informations Systems, 2002. <Http://www.hp.com/solutions1/e-services/>
- Finn, A. & Kayande, U. (1997). Reliability Assessment and Optimization of Marketing Measurement. *Journal of Marketing Research*, 34, 262-275.

- Fishbein, M. & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Fisseni, H. J. (1997). *Lehrbuch der psychologischen Diagnostik* (2. Aufl.). Göttingen: Hogrefe.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik* (3. Aufl.). Göttingen: Hogrefe.
- Fitzpatrick, R. & Higgins, C. (1998). Usable software and its attributes: A synthesis of software quality, European Community law and human-computer interaction, In: *People and Computers XIII. Proceedings of HCI'98 Conference*, Springer, London, UK
- Fitzsimmons, J.A. & Fitzsimmons, M.J. (2000). *Service management operations, strategy and information technology*, 3 Aufl., Mc Graw-Hill/Irvin, New York.
- Flemming, J. (1998). *Web navigation: designing the user experience*, Beijing, O'Reilly.
- Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What Makes A Web Site Credible? A Report on a Large Quantitative Study. *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems*, 2001. ACM Press
- Fornell, C. (2007). *Improvement in Customer Satisfaction Slows*. Internetressource: http://www.theacsi.org/index.php?option=com_content&task=view&id=169&Itemid=168 (20.08.2007)
- Fornell, C. & Cha, J. (1994): Partial Least Squares, in: Bagozzi, R.P. (Hrsg.), *Advanced Methods of Marketing Research*, Cambridge, 52-78.
- Fornell, C. & Larcker, D.F. (1981): Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 18, 39-50.
- Forrester research inc., (2006). *German online insurance sales forecast: 2006 - 2011*. <http://www.forrester.com/Research/Document/Excerpt/0,7211,40092,00.html>
- Forsa, (2006). *TK-Trendmonitor Gesundheit 2006*. Gesellschaft für Sozialforschung und statistische Analyse mbH, Berlin 2006.
- Foscht, T. (1999): Konsumentenverhalten im Kontext der neuen Medien - Analyse des verhaltenswissenschaftlichen und des systemtheoretischen Erklärungsansatzes, in: *Der Markt*, 38. Jg., Nr. 150/151, S. 139-154.
- Francis, J. & White, L. (2002). PIRQUAL: a scale for measuring customers expectations and perceptions of quality in internet retailing: *Proceedings of the Winter Educator's Conference*, American Marketing Association, Chicago, IL S. 438-443.
- Frese, E. (1985). Exzellente Unternehmungen - Konfuse Theorien. Kritisches zur Studie von Peters und Waterman, *Die Betriebswirtschaft*, 45 (5), S. 604 - 606.

- Friedewald, M. & Kolo, C. (2000): Nutzeranforderungen und zukünftige Endgeräte für Online-Dienste, in: *Information Management & Consulting*, 15, 2000, S. 57-62.
- Fritz, W. (1990). Marketing – ein Schlüsselfaktor des Unternehmenserfolges? Eine kritische Analyse vor dem Hintergrund der empirischen Erfolgsfaktorenforschung, *Marketing Zeitschrift für Forschung und Praxis*, 12 (1990) 2, S. 91–110.
- Fritz W. (2001). *Internet-Marketing und Electronic Commerce. Grundlagen, Rahmenbedingungen, Instrumente*, 2. erheblich erweiterte Auflage, Wiesbaden 2001: Gabler.
- Frokjaer, E., Hertzum, M., & Hornbæk, K. (2000). Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? *Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems*, 345-352. ACM Press: New York.
- Gabott, M. & Hogg, G. (1999). Consumer involvement in services: a replication and extension. *Journal of Business Research* 46, 159-166.
- Ganesan, S. (1994). Determinants of long-term orientation in buyer-seller relationships. *Journal of Marketing*, 58, 1-19.
- Gediga, G. & Hamborg, K.-C. (2002): Ergonomische Evaluation von Software: Methoden und Modelle im Software-Entwicklungsprozess. In: *Zeitschrift für Psychologie*, 210 (1), 40-57.
- Gediga, G., Hamborg, K.-C. & Dünisch, I. (1999). The IsoMetrics usability inventory: An operationalisation of ISO 9241-10. *Behaviour and Information Technology*, 18, 151–164.
- Gefen, D. (1997). *Building users' trust in freeware providers and the effects of this trust on users' perceptions of usefulness, ease of use and intended use of freeware*. Ph.D. dissertation, Georgia State University.
- Gefen, D. (2000). E-Commerce: The role of familiarity and trust. *Omega: The International Journal of Management Science*, 28, S. 725-737.
- Gefen, D. & DeVine, P. (2001). Customer loyalty to an online store: the meaning of online service quality. *Proceedings of the Twenty-second International Conference on Information Systems*, S. 613-617.
- Gefen, D.; Karahanna, E. & Straub, D.W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27 (1), 51–90.
- Gefen, D. & Straub, D. (2000). Managing user trust in B2C e-services. *E-Services Quarterly*, 1,1. Electr. Publication, URL:
<http://www.lebow.drexel.edu/gefen/eservicejournal2001.pdf>.

- Gefen, D., Straub, D. & Boudreau, M. (2000). Structural Equation Models and Regression: Guidelines for Research Practice, in: *Communications of the Association for Information Systems*, Vol. 4, Article 7, S. 1-79.
- Geisser, S. (1974): A predictive approach to the random effect model, *Biometrika*, 61 (1), S. 101 ff.
- Ghani, J.A., Supnick, R. & Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. DeGross, J., Benbasat, I., DeSanctis, G., Eds. *Proceedings of 12th International Conference of Information Systems*, (229-237) New York.
- Gierl, H. & Bambauer, S. (2001): Werbewirkung von Unternehmens-Websites am Beispiel von Banken, in: *Planung und Analyse*, 6, S. 56-61.
- Gillenson, M.L., Sherrell, D.L. & Chen, L.D. (2000). A taxonomy of web site traversal patterns and structures. *Communications of AIS*, 3 (17).
- GKV-WSG (2007). *GKV-Wettbewerbsstärkungsgesetz* Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung, Berlin.
- Gogan, J.L. (1997). The web's impact on selling techniques: historical perspective and early observations. *International Journal of Electronic Commerce* 1(2), 89-108.
- Göthlich, S.E. (2006). Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (133-150). Wiesbaden: Deutscher Universitäts-Verlag.
- Göttgens, O. (1996). *Erfolgsfaktoren in stagnierenden und schrumpfenden Märkten: Instrumente einer erfolgreichen Unternehmenspolitik*, Wiesbaden: Gabler, 1996.
- Götz, O. & Liehr-Gobbers K. (2004): Analyse von Strukturgleichungsmodellen mit Hilfe der Partial-Least-Squares(PLS)-Methode, *Die Betriebswirtschaft*, 64, 714-738.
- Grabner-Kräuter, S. (1993). Diskussionsansätze zur Erforschung von Erfolgsfaktoren, *Journal für Betriebswirtschaft*, 43 (1993) 6, S. 278 300.
- Grabner-Kräuter, S. & Kaluscha, E.A. (2003). Empirical research in on-line trust: a review and critical assessment. *International Journal of Human-Computer Studies* 58, 783-812.
- Gräf, H. (1999). *Online-Marketing Endkundenbearbeitung auf elektronischen Märkten*, Wiesbaden: Deutscher Universitäts-Verlag.
- Grapentine, T. (1997). Managing multicollinearity. *Marketing Research*, 9, 3, 10-21.
- Gray, A. & Bromwich, D. (2001). *Personality and learning: Increasing self-awareness to maximise development*. Paper presented at the BPS Occupational Psychology Conference.

- Gregg, D. & Scott, J. (2006). The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce*, 10(3), 95-120.
- Griffith, D.A., Krampf, R.F. & Palmer, J.W. (2001). The role of interface in electronic commerce: consumer involvement with print vs. online catalogs, in: *International Journal of Electronic Commerce*, 5 (4), S. 135-153.
- Grönroos, C. (1982). An applied service marketing theory. *European Journal of Marketing*, (7), S. 30-41.
- Grönroos, C., Heinonen, F. Isoniemi, K. & Lindholm, M. (1999). *The net offer model: developing internet offerings for the virtual marketplace*, working paper, Swedish School of Economics and Business Administration, Stockholm.
- Gröppel, A. (1988): Erfassung erlebnisorientierter Kunden im Einzelhandel, in: *Marktforschung und Management*, (1), 7-14.
- Grofmann, H.D., Schäfers, V. & Viktorin, S. (1999). Die Beraterbank im Internet: Verstärkung der Kundenbindung durch individuelle Finanzdienstleistungen, in: Scheer, A. & Nüttgens, M. (Hrsg.): *Electronic business engineering*. 4. int. Tagung Wirtschaftsinformatik, Heidelberg.
- Grünig, R., Heckner, F. & Zeus, A. (1996). Methoden zur Identifikation strategischer Erfolgsfaktoren, *Die Unternehmung*, 50 (1), S. 3 12.
- Gujarati, D. N. (2003), *Basic Econometrics*, 4th ed., Burr Ridge, IL.
- Gupta, S., & Kim, H. (2007). The Moderating Effect of Transaction Experience on the Decision Calculus in On-Line Repurchase. *International Journal of Electronic Commerce*, 12(1), 127-158.
- Gwizdka, J. & Spence, I., (2007). Implicit measures of lostness and success in web navigation. *Interacting with Computers*, 19 (3), 357-369.
- Haenecke, H. (2002). Methodenorientierte Systematisierung der Kritik an der Erfolgsfaktorenforschung. *ZfB* 72. (2), 165 183
- Hall, R. (1992): The Strategic Analysis of Intangible Resources, in: *Strategic Management Journal*, Vol. 13, S. 135 144.
- Hall, R. & Hanna, P., (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & Information Technology*, 23 (3), pp. 183 195.
- Hamborg, K.-C. (2007). *Using the thinking aloud method in usability testing - what kind of data do you really get and what they are good for?* 10th Congress of the Swiss Society of Psychology. Differences, Diversity, and Change. September 13-14, 2007, Zuerich.

- Hamborg, K.-C. & Gediga, G. (2006). Methoden und Modelle für die Gestaltung gebrauchstauglicher Software. In: B. Zimolong & U. Konradt (Hrsg.), *Ingenieurpsychologie. Enzyklopädie der Psychologie (Bd. D-III-2)*, S. 495-530, Göttingen: Hogrefe.
- Hamborg, K.-C., Gediga, G. & Hassenzahl, M. (2003). Fragebogen zur Evaluation. In: S. Heinsen & P. Vogt (Hrsg.), *Usability praktisch umsetzen, Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte* (S. 172-187). München: Hanser.
- Hamborg, K.-C. & Greif, S. (2003). New Technologies and Stress. In M. J. Schabracq, J.A. Winnubst & C.L. Cooper & (Eds.), *Handbook of Work and Health Psychology* (2nd ed.), (S. 209-235). Chichester: Wiley.
- Hampton-Sosa, W. & Koufaris, M. (2005). The Effect of Web Site Perceptions on Initial Trust in the Owner Company. *International Journal of Electronic Commerce*, 10 (1), 55-81.
- Hansmann, K.-W. & Ringle, C.M. (2004). *SmartPLS Benutzerhandbuch*. Förderverein Industrielles Management an der Universität Hamburg e.V.. Online im Internet: URL: <http://www.smartpls.de/forum/downloads/handbook/handbook.zip>
- Harms, I. & Schweibenz, W. (2000): Usability Engineering Methods for the Web. Results From a Usability Study. In: *Informationskompetenz - Basiskompetenz in der Informationsgesellschaft*. Proceedings des 7. Internationalen Symposiums für Informationswissenschaft (ISI 2000). Konstanz: UKV. 17-30.
- Harper, B., Slaughter, L., & Norman, K. (1997). *Questionnaire administration via the WWW: A validation and reliability study for a user satisfaction questionnaire*. Paper presented at WebNet 97, Association for the Advancement of Computing in Education, Toronto, Canada.
- Hartwig, S. (2001). Binnenmarkt im Versicherungsbereich versus Verbraucherschutz, in: *Versicherungswirtschaft*, S. 636.
- Harvey, R. J., Billings, R. S., & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology*, 70, 461-468.
- Hassenzahl, M., (2000). "Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal," *Proc. CHI 2000 Conf. Human Factors in Computing*, ACM Press, Addison-Wesley, New York, 2000, pp. 201-208.
- Hassenzahl, M. (2001). The Effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, 13, 481-499.
- Hassenzahl, M. & Hartwig, R. (2005). Certified Fun – Stehen hedonische Qualitätsaspekte und Qualitätssicherung im Widerspruch? In: Hassenzahl M., Peissner M. (Hrsg.): *Usability Professionals 2005*, Darmstadt.

- Hassenzahl, M.; Burmester, M. & Koller, F. (2003): AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J. ; Szwillus, G. (Hrsg.): *Mensch & Computer 2003, Interaktion in Bewegung*. (S. 187-196) Stuttgart, Leipzig: B.G. Teubner.
- Hassenzahl, M.; Kekez, R. & Burmester, M. (2002): The importance of a software's pragmatic quality depends on usage modes. In: Luczak, H.; Cakir A. E.; Cakir, G. (Eds.): *Proceedings of the 6th international conference on Work With Display Units (WWDU 2002)*. Berlin: ERGONOMIC Institut für Arbeits- und Sozialforschung. S. 275-276.
- Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000): Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. In: Turner, T.; Szwillus, G. (Hrsg.), *Proceedings of the CHI 2000 Conference on Human Factors in Computing*. (S. 201-208) New York: ACM, Addison-Wesley.
- Hassenzahl, M., Schöbel, M. & Trautmann, T. (2008). How motivational orientation influences the evaluation and choice of hedonic and pragmatic interactive products: The role of regulatory focus. *Interacting with Computers*, 20 (4-5), 473-479.
- Hassenzahl, M. & Seewald, F. (2004). Vom kritischen Ereignis zum Nutzungsproblem: Die qualitative Analyse in diagnostischen Usability Tests. In: M. Hassenzahl & M. Peissner (Hrsg.), *Usability Professionals 2004*. Stuttgart: German Chapter der Usability Professionals' Association.
- Hauser, R. (1973). Disaggregating a social-psychological model of educational attainment. In: Goldberger, A. Duncan, O. (Eds.), *Structural equation models in the social sciences*. New York, S. 255-284.
- Heckner F. (1998). *Identifikation marktspezifischer Erfolgsfaktoren: Ein heuristisches Verfahren angewendet am Beispiel eines pharmazeutischen Teilmarktes*, Bern u.a.O.: Lang, 1998.
- Heise, D.R. (1986): Estimating Non-linear Models: Correcting for Measurement Error, *Sociological Methods of Research*, 14, 447-472.
- Henderson, R. D., Rickwood, D. & Roberts, P. (1998). The beta test of an electronic supermarket. *Interacting with Computers*, 10 (4), pp. 385-399.
- Henderson, R. D., Smith, M. C., Podd, J. & Varela-Alvarez, H. (1995). A Comparison of the four prominent user-based methods for evaluating the usability of computer software. *Ergonomics*, 38 (10), 2030-2044.
- Hennig-Thurau, T. (1998): *Konsum-Kompetenz: eine neue Zielgröße für das Management von Geschäftsbeziehungen*. Theoretische Begründung und empirische Überprüfung der Relevanz für das Konsumgütermarketing, Dissertation, Frankfurt a.M.

- Hentschel, B. (1992). *Dienstleistungsqualität aus Kundensicht: vom merkmals- zum ergebnisorientierten Ansatz*, Wiesbaden: Deutscher Universitäts-Verlag.
- Henseler, J. (2005). Einführung in die PLS-Pfadmodellierung. *Wirtschaftswissenschaftliches Studium*, 34 (2), 70-75.
- Herrmann, A., Huber, F. & Kressmann, F. (2004): *Partial Least Squares - Ein Leitfaden zur Spezifikation, Schätzung und Beurteilung varianzbasierter Strukturgleichungsmodelle*, Arbeitspapier.
- Herrmann, A., Huber, F. & Kressmann, F. (2006), Varianz- und kovarianzbasierte Strukturgleichungsmodelle Ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung, in: *zfbf*, 58, 34-66.
- Hermanns, A. & Gampenrieder, A. (2002): Wesen und Eigenschaften des E- Commerce, in: Schlögel, M.; Tomczak, T.; Belz, C. (Hrsg.): *Roadmap to E-Business. Wie Unternehmen das Internet erfolgreich nutzen*, St. Gallen 2002, S. 70-91.
- Hewlett-Packard (2000). *E-services*, elektronisch veröffentlicht unter <http://e-services.hp.com>
- Hildebrandt, L. (1986). A Facet Theoretical Approach for Testing Measurement and Structural Theories. An Application of Confirmatory MDS, in: Lutz, Richard J. (ed.), *Advances in Consumer Research*, Vol. XIII, S. 523-528.
- Hildebrandt, L. (1999). Hypothesenbildung und empirische Überprüfung, in Herrmann, A. & Homburg, C. (Hrsg.): *Marktforschung: Methoden, Anwendungen, Praxisbeispiele*, Wiesbaden: Gabler, S. 33 57.
- Hilke, W. (1989). Grundprobleme und Entwicklungstendenzen des Dienstleistungs-Marketing, in: *Schriften zur Unternehmensführung*, SzU, (35), S. 5-44.
- Hiltz, S.R. & Johnson, K. (1990). User satisfaction with computer-mediated communication systems. *Management Science*, 36 (6), 739-764.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21, 967 988.
- Hirshleifer, J. & Riley, J.G., 1979. The analytics of uncertainty and information: an expository survey. *Journal of Economic Literature* 17, 1374-1421.
- Hirshleifer, J. & Riley, J.G., 1979. The analytics of uncertainty and information: an expository survey. *Journal of Economic Literature* 17, 1374-1421.
- Hoffman, D.L. & Novak, T.P. (1996). Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing*, 60(3), S. 50-68.
- Hoffman, D.L., Novak, T.P. & Chatterjee, P. (1995). Comercial scenarios for the web: opportunities and challenges, in: *Journal of Computer Mediated Communications*, 1 (3), elektronisch veröffentlicht unter <http://www.Jcmc.huji.ac.il/vol1/issue3/hoffman.html>.

- Hoffman, D., Novak, T., & Peralta, M. (1999). Building Consumer Trust Online. *Communications of the ACM*, 42 (4), pp. 80-85.
- Homburg, C. & Baumgartner, H. (1995): Beurteilung von Kausalmodellen. Bestandsaufnahme und Anwendungsempfehlungen, *Marketing: Zeitschrift für Forschung und Praxis*, 17, 162-176.
- Homburg, C. & Baumgartner H. (1998): Die Kausalanalyse als Instrument zur Messung der Kundenzufriedenheit im Industriegütermarketing, in: Hildebrandt, L. und C. Homburg (Hrsg.): *Die Kausalanalyse: Instrument der empirischen betriebswirtschaftlichen Forschung*, Stuttgart, 237-264.
- Homburg, C., Becker, A. & Hentschel, F. (2003): Der Zusammenhang zwischen Kundenzufriedenheit und Kundenbindung, in: Bruhn, M.; Homburg, C. (Hrsg.): *Handbuch Kundenbindungsmanagement*, 4. Aufl., (S. 91-121) Wiesbaden 2003, Gabler-Verlag.
- Homburg, C. & Bruhn, M. (2000). Kundenbindungsmanagement - Eine Einführung in die theoretischen und praktischen Problemstellungen, in: Bruhn, M; Homburg, C. (Hrsg.): *Handbuch Kundenbindungsmanagement*, 3. Aufl., (S. 3-36) Wiesbaden 2000, Gabler-Verlag.
- Homburg, C. & Dobratz, A. (1998). Iterative Modellselektion in der Kausalanalyse. In L. Hildebrandt & C. Homburg (Hrsg.), *Die Kausalanalyse: Instrument der empirischen betriebswirtschaftlichen Forschung* (447-474). Stuttgart: Schäffer-Poeschel.
- Homburg, C. & Giering, A. (1996): Konzeptualisierung und Operationalisierung komplexer Konstrukte Ein Leitfaden für die Marketingforschung, *Marketing: Zeitschrift für Forschung und Praxis*, 18, 5-24.
- Homburg, C. & Kebbe, P. (2001). Komplexität als Determinante der Wahrnehmung von Dienstleistungsqualität, *Zeitschrift für betriebswirtschaftliche Forschung*, 53, 8, 478-499.
- Homburg, C. & Krohmer, H. (2004). Die Fliegenpatsche als Instrument des wissenschaftlichen Dialogs. *Die Betriebswirtschaft (DBW)*, 64(5), 626-631.
- Homburg, C. & Pflesser, C. (1999). Strukturgleichungsmodelle mit latenten Variablen: Kausalanalyse, in: Herrmann, A. & Homburg, C. (Hrsg.): *Marktforschung: Methoden, Anwendungen, Praxisbeispiele*, Wiesbaden: Gabler, S. 633 659.
- Homburg, C. & Pflesser, C. (2000). Konfirmatorische Faktorenanalyse. In: A. Herrmann & C. Homburg (Hrsg.), *Marktforschung* (2. Aufl., 415-436). Wiesbaden: Gabler.
- Homburg, C. & Rudolph, B. (1998). Theoretische Perspektiven zur Kundenzufriedenheit. In: H. Simon & C. Homburg (Hrsg.). *Kundenzufriedenheit* (3., aktual. und erw. Aufl.), (S. 33 55). Wiesbaden: Gabler.

- Homburg, C. & Stock, R. (2001). Der Zusammenhang zwischen Mitarbeiter- und Kundenzufriedenheit: Eine dyadische Analyse. *Zeitschrift für Betriebswirtschaft*, 71 (7), 789-806.
- Horton, R.P., Buck, T., Waterson, P.E. & Clegg, C.W. (2001). Explaining intranet use with the technology acceptance model. *Journal of Information Technology*, 16, 237-249.
- Hosmer, D.W. & Lemeshow, S. (2000): *Applied Logistic Regression*, 2. Aufl., New York, Wiley Verlag.
- Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification, *Psychological Methods*, 3(4), 424-453.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling*, 6(1), 1-55.
- Hu, P., Chau, P., Sheng, O. & Tam, K. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems*, 16(2), S.91-112.
- Huber, O. (1995): *Das psychologische Experiment: Eine Einführung*, 2. Aufl., Bern, Verlag Hans Huber.
- Hudlicka, E. (2003): To feel or not to feel: the role of affect in humancomputer interaction. In: *International Journal of HumanComputer Studies*, 59 (1) pp. 1-32.
- Hühnerberg, R. (1997). Was Online-Kommunikation für das Marketing bedeutet, in: *Thesis*, 1, 1997.
- Hulland, J. (1999). Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies, *Strategic Management Journal*, 20, 195-204.
- IBM (2001). *Ease of Use Guidelines*, URL:
http://www-3.ibm.com/ibm/easy/eou_ext.nsf/publish/572.
- Idaszak, J., & Drasgow, F. (1987). A revision of the job diagnostic survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, 72, 69-74.
- Igbaria, M., Schiffman, S. J. & Wieckowski, T. J. (1994): The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology*, 13(6), 349-361.
- Igbaria, M., Zinatelli, N., Cragg, P. & Cavaye, A.L.M., (1997). Personal computing acceptance factors in small firms: a structural equation model. *MIS Quarterly* 21, 279-305.
- Isoniemi, K. & Snellman, K. (2000). *What is a virtual service? Broadening the perspective technology based services*, Working Paper Nr. 431, Swedish School of Business Administration, Stockholm.

- Jaccard, J., Wan, C.K. & Turrisi R. (1990): The Detection and Interpretation of Interaction Effects between Continuous Variables in Multiple Regression, *Multivariate Behavioral Research*, 25, 467-478.
- Jäpel, W. (1985): *Die Qualität alternativer Rating-Formen und ihre Einflußgrößen*, Diss. Univ. Regensburg.
- Jackson, C. M., Chow S. & Leitch, R. A. (1997). Toward an understanding of the behavioral intention to use an information system, in: *Decision Sciences*, 28 (2), 357-389.
- Jaworski, B. J. & Kohli A. K. (1993): Market Orientation: Antecedents and Consequences, in: *Journal of Marketing*, Vol. 57, Nr. 7, S. 53-70.
- Jarvenpaa, S. & Todd, P. (1997a). Consumer reactions to electronic shopping on the world wide web. *International Journal of Electronic Commerce*, 1 (2), S. 59-88.
- Jarvenpaa, S. & Todd, P. (1997b). Is there a future for retailing on the internet? Petersen, R., ed. *Electronic Marketing and the Consumer*. Sage, Thousand Oaks, CA.
- Jarvenpaa, S., Tractinsky, N., Saarinen, L. & Vitale, M. (1999). Consumer trust in an internet store: a crosscultural validation. *Journal of Computer-Mediated Communications*, 5,2. Electronic publication, URL: <http://www.ascusc.org/jcmc/vol15/issue2/jarvenpaa.html>
- Jarvenpaa, S., Tractinsky, N. & Vitale, M. (2000). Consumer trust in an internet store. *Information Technology and Management* 1 (1-2), 45-71.
- Jarvis, C.B., MacKenzie, S.B. & Podsakoff, P.M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification. *Marketing and Consumer Research*, 30, 199-218.
- Jennings, M. (2000). *Theory and Models for Creating Engaging and Immersive Ecommerce Websites*, in: Proceedings of the 2000 ACM Special Interest group on Computer Personnel Research, Chicago, 77-85.
- Jöreskog, K. & Sörbom, D. (1996), *LISREL 8: User's Reference Guide*, Scientific Software International Inc.
- Johnson, R.A. & Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall, New York.
- Jones, S. & Hewitt, J. (1997). Heuristic Evaluation of Web Site Usability: Experience from Two Case Studies. In *HCI'97 Conference Companion*. Bristol, UK: Springer. p. 23 25.
- Jordan, P. (1998). Human factors for pleasure in product use. *Applied Ergonomics* 29 (1), pp. 25-33.
- Jordan, P., Thomas, B. & McClelland, I. (1996). Issues for usability evaluation in industry: seminar discussions. In: Jordan, P., Thomas, B. & McClelland, I. (Eds.), *Usability Evaluation in Industry*, Taylor & Francis, 227-234.

- Kaiser, H.F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34, 111-117.
- Karahanna, E., Straub, D. W. & Chervany, N. L. "Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs," *MIS Quarterly* (23:2), 1999, pp. 183-213.
- Kassim, N., & Abdullah, N. (2008, September). Customer Loyalty in e-Commerce Settings: An Empirical Study. *Electronic Markets*, 18(3), 275-290.
- Katona, G. (1960): *Das Verhalten der Verbraucher und Unternehmer*, Tübingen, 1960.
- Kaya, M. & Himme, A. (2006). Möglichkeiten der Stichprobenbildung. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (89-99). Wiesbaden: Deutscher Universitäts-Verlag.
- Keat, T. & Mohan, A. (2004). Integration of TAM based electronic commerce models for trust. *The Journal of American Academy of Business*, Cambridge, 2004.
- Keeney, R.L. (1999): The Value of Internet Commerce to the Customer, in: *Management Science*, Vol. 45, April, 2000, S. 533-542.
- Kent, N. & Panizzi, E. (2006). Levels of automation and user participation in usability testing. *Interacting with Computers*, 18 (2), 246-264.
- Kim, D. (2008). Self-Perception-Based Versus Transference-Based Trust Determinants in Computer-Mediated Transactions: A Cross-Cultural Comparison Study. *Journal of Management Information Systems*, 24(4), 13-45.
- Kim, J., Lee, J. & Choi, D. (2003). Designing emotionally evocative homepages: An empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies*, 59, pp. 899 940.
- Kim, K.K. & Prabhakar, B. (2004). Initial Trust and the Adoption of B2C E-Commerce: The Case of Internet Banking. *Database for Advances in Information Systems*, 35(2), 50-64.
- Kim, S.-E., Shaw, T. & Schneider, H. (2003). Web Site Design Benchmarking within Industry Groups. *Internet Research: Electronic Networking Applications and Policy*, 13, 17-26.
- Kirakowski, J. (1997). *The use of questionnaire methods for usability assessment*, URL: <http://www.ucc.ie/hfrg/questionnaires/sumi/sumipapp.html>
- Kirakowski, J. & Claridge, N. (2001). *A Professional Tool for Evaluating Web Sites*. Working paper, presented at the IHM-HCI Conference Lille, 2001.
- Kirakowski, J., Claridge, N. & Whitehand, R. (1998). *Human centred measures of success in web site design*, URL, <http://research.att.com/conf/fhweb/conferences/basking4.zip>

- Klein, A. (2000). Moderatormodelle. Verfahren zur Analyse von Moderatoreffekten in Strukturgleichungsmodellen. *Studienreihe Psychologische Forschungsergebnisse*, Bd. 71, Verlag Dr. Kovac, Hamburg.
- Klietmann, M. (2001): Der Online-Verbraucher - auf der Suche nach dem Yeti, in: Klietmann, M (Hrsg.): *Kunden im E-Commerce*, Düsseldorf 2001, S. 11- 27.
- Kline, R.B. (1998). *Principles and practice of structural equation modelling*. New York: The Guilford Press.
- Knospe, J. (2000). Direktversicherer kommen nur langsam voran. In: *Versicherungswirtschaft*, 2000, S. 1190 ff.
- Köszegi, S. (1999): *Vertrauen, Kontrolle und Risiko in virtuellen Organisationen*, Working Paper, Nr. 9903 OP, Wien, 1999.
- Kohli, A.K., Jaworski, B. & Kumar, A. (1993). MARKOR: A Measure of Market Orientation. *Journal of Marketing Research*, 30, 467-477.
- Konradt, U., Christophersen, T. & Schaeffer-Kuelz, U. (2006). Predicting user satisfaction, strain and system usage of employee self-services. *International Journal of Human-Computer Studies*, 64, 1141-1153.
- Konradt, U., Wandke, H., Balazs, B. & Christophersen, T. (2003). Usability in Online Shops: Scale Construction and the Influence on the Buyers' Intention and Decision. *Behaviour & Information Technology*, 22(3), 165-174.
- Kortzfleisch, H. & Winand, U. (2000): Trust in Electronic Learning and Teaching Environments: The case of "WINFOLine", in: *Proceedings of the 8th European Conference on Information Systems, ECIS 2000, A Cyberspace Odyssey*, Vienna, Austria, July 3-5, 2000, Hrsg. von Hansen, R. et al., Wien, 2000, S. 1348-1354.
- Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behaviour. *Information Systems Research* 13, 205-223.
- Koufaris, M. & Hampton-Sosa, W. (2002) : *Customer trust online: examining the role of the experience with the web-site*. CIS Working Paper Series, Zicklin School of Business, Baruch College, New York, NY. Electronic publication, URL:
<http://cisnet.baruch.cuny.edu/papers/cis200205.pdf>.
- Koufaris, M. & Hampton-Sosa, W. (2004). The Development of Initial Trust in an Online Company by new Customers. *Information & Management*, 41, 377-397.
- Kroeber-Riel, W. (1992). *Konsumentenverhalten*, 5. Auflage, München, Vahlen.
- Kroeber-Riel, W. (1996). *Konsumentenverhalten*, 6. Auflage, München, Vahlen.
- Kroeber-Riel, W. & Weinberg, P. (1999). *Konsumentenverhalten*, 7. Auflage, München, Vahlen.

- Kube, C. (1991). *Erfolgsfaktoren in Filialsystemen: Diagnose und Umsetzung im strategischen Controlling*, Wiesbaden: Gabler, 1991.
- Kurosu, M. & Kashimura, K. (1995). Apparent Usability vs. Inherent Usability: Experimental Analysis on the Determinants of the Apparent Usability. In I. Katz, R. Mack & L. Marks (Eds.), *Conference Companion on Human Factors in Computing Systems, CHI '95, Denver, Colorado, United States, May 07-11 (292-293)*. New York, NY: ACM Press.
- Kuß, A. & Tomczak, T. (2000). *Käuferverhalten (2. Aufl.)*. Stuttgart: Lucius & Lucius.
- Kühn, R. (1991): Methodische Überlegungen zum Umgang mit der Kundenorientierung im Marketing-Management, in: *Marketing ZFP*, Vol. 13, Nr. 2. S. 97-101.
- Lamberti, H.J. (2000). Transformation des Bankwesens unter dem Einfluss der Informationstechnik am Beispiel der Deutsche Bank AG in: Weiber, R. (Hrsg.): *Handbuch Electronic Business: Informationstechnologien Electronic Commerce Geschäftsprozesse*, Wiesbaden, S. 623-642.
- Landon, S. & Smith, C. E. (1997). The use of quality and reputation indicators by consumers: The case of Bordeaux wine. *Journal of Consumer Policy*, 20, 289-323.
- Law, K. & Wong, C. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, 25 (2), 143-160.
- Lederer, A., Maupin, D., Sena, M. & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems*, 29 (3), 269-282.
- Lee, S., Katerattanakul, P. & Hong, D. (2005). Framework for user perception of effective e-tail web sites. *Journal of Electronic Commerce in Organizations* 3(1), 13-34.
- Lee, M., K. & Turban, E., 2001. A trust model for consumer Internet shopping. *International Journal of Electronic Commerce* 6 (1), 75-91.
- Legris, P., Ingham, J. & Collerete, P., (2003). Why do People use information technology? A critical review of technology acceptance model. *Information and Management* 40, 191-204.
- Lewis, C. & Norman, D. (1995). Designing for Error. In R. Baecker & W. Buxton, (Eds.), *Human-Computer Interaction: toward the year 2000*, Morgan Kaufmann Publishers, Inc., Los Altos, CA, 686 - 697.
- Lewis, C. & Wharton, C. (1997). Cognitive Walkthroughs. In M. Helander, T.K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction (2nd ed.)*, 717-732.
- Lewis, J. R. (1993). Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction*, 5, 383-392.

- Lewis, J.R. (1995). IBM computer usability satisfaction questionnaires, in : *International Journal of Human Computer Interaction*, 7 (1), 57-78.
- Li, H., Kuo, C. & Russell, M.G. (1999). The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior. *Journal of Computer-Mediated Communications*, 5, 2. Elektronisch unter:
www.ascusc.org/jcmc/vol5/issue2/hairong.html [2.5.2002].
- Liao, Z. & Cheung, M. (2001). Internetbased e-shopping and customer attitudes: an empirical study. *Information & Management*, 38(5), 299-306.
- Licharz, E.-M., 2002. *Vertrauen im Business-to-Consumer Electronic Commerce*, Universität Gesamthochschule Kassel.
- Liechtenstein, D.R., Netemeyer, R.D. & Burton, S. (1999). Distinguishing coupon Proneness from value consciousness: an acquisition transaction utility theory perspective. In: *Journal of Marketing*, 54, S. 54-67.
- Lin, J. (2007). Online stickiness: its antecedents and effect on purchasing intention. *Behaviour & Information Technology*, 26 (6), 507-516.
- Lin, H.X., Choong, Y.-Y. & Salvendy, G. (1997). A Proposed Index of Usability: a Method for Comparing the Relative Usability of Different Software Systems. *Behaviour & Information Technology*, 16(4/5), 267-278.
- Lingenfelder, M. (2001). Die Identifikation und Bearbeitung von Online-Käufersegmenten Ergebnisse einer empirischen Untersuchung. In W. Fritz (Hrsg.), *Internet-Marketing* (2., überarb. u. erw. Aufl.), (S. 373 - 397). Stuttgart: Schäffer-Poeschel.
- Linn, N., Krotzsch, S. & Stein, J. (2001). Finanzdienstleistungen und Online-Banking, in: Hermann, A., Sauter, M. (Hrsg.): *Management-Handbuch Electronic-Commerce: Grundlagen, Strategien, Praxisbeispiele*, 2. Aufl. (S. 557-566) München, Vahlen Verlag.
- Little, R.J.A (1988): A Test of Missing Completely at Random for Multivariate Data with Missing Values, *Journal of the American Statistical Association*, 83, 1198-1202.
- Little, R.J.A. & Rubin d.B. (2002): *Statistical Analysis With Missing Data*, 2. Aufl., Thousand Oaks.
- Logan, J. (1994). Behavioral and emotional usability: Thomson Consumer Electronics. In M. Wiklund (Ed.), *Usability in practice*, Cambridge, Academic Press, 59-82.
- Lohmöller, J.-B. (1989): *Latent variable path modeling with partial least squares*; Heidelberg, Physica-Verlag.
- Lohse, G.L., Bellman, S., Johnson, E.J., (2000). Consumer buying behavior on the Internet: Findings from panel data. *Journal of Interactive Marketing*, 14, 15-29.

- Lohse, G.L., & Spiller, P. (1998). Electronic shopping. *Communications of the ACM*, 41(7), 81-87.
- Lohse, G.L. & Spiller, P. (1999). Internet Retail Store Design: How the User Interface Influences Traffic and Sales. *Journal of Computer Mediated Communication*, 5(2). Online im Internet: URL: <http://www.ascusc.org/jcmc/vol5/issue2/lohse.htm>
- Loos, C. (1998): *Online-Vertrieb von Konsumgütern*, Deutscher Universitäts Verlag, Gabler, Wiesbaden, 1998.
- Lovelock, C.H. (2001). *Service marketing: people, technology, strategy*, 4. Aufl., Upper Saddle river.
- Lowengart, O. & Tractinsky, N. (2001). Differential Effects of Product Category on Shoppers' Selection of Web-based Stores: A Probabilistic Modeling Approach. *Journal of Electronic Commerce Research*, 2 (4), 142-165.
- Luhmann, N., (1989). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*, 3rd Edition. Enke, Stuttgart.
- Lührig, T. & Dholakia, N. (2002). Online auctions: the emerging new electronic agora. In: N. Dholakia, W. Fritz, & N. Mundorf (Hrsg.): *Global E-Commerce and Online-Marketing*, Westport, London, 111-121.
- Lürzer, R. (2000). Eintauchen in die Welt des E-Commerce: Ein Sprung ins kalte Wasser?, in: *Versicherungswirtschaft*, 2000, S. 1024 ff.
- Luxem, R. (2000). *Digital Commerce*, Lohmar, Köln.
- Lynch, J. & Ariely, D. (2000). Wine online: Search cost and competition on price, quality, and distribution. *Marketing Science* 2000, 19 (1), pp. 83 103.
- MacCallum, R.C.& Browne, M.W. (1993): The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues, in: *Psychological Bulletin*, 114 (3), 533-541.
- MacCallum, R. C., Wideman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factoranalysis. *Psychological Methods*, 4 (1), 87 99.
- MacDonald, P. & Paunonen, S.V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62 (6), 921-943.
- MacInnes, I., Li, Y., & Yurcik, W. (2005). Reputation and Dispute in eBay Transactions. *International Journal of Electronic Commerce*, 10(1), 27-54.
- MacLeod, Bowden, Bevan & Curson, (1997). The MUSiC Performance Measurement Methode. *Behaviour & Information Technology*, 16 (4/5), 279-293.

- Mahmood, M.A., Hall, L. & Swanberg, D.L., (2001). Factors affecting information technology usage: a meta-analysis of the empirical literature. *Journal of Organizational Computing and Electronic Commerce* 11, 107-130.
- Maleri, R. (1997). *Grundlagen der Dienstleistungsproduktion*, 4. Aufl., Berlin, Springer.
- Maleri, R. (2001). Grundlagen der Dienstleistungsproduktion. In: Bruhn, M. & Meffert, H. (Hrsg.) *Handbuch Dienstleistungsmanagement*, (125-148) Wiesbaden, 2. Aufl., Gabler-Verlag.
- Mangold, R., Reese, F., Klauck, T. & Stanulla, S. (2000). Designed for Emotions? Mimik-basierte Mikroevaluation von Onlineangeboten. *Planung & Analyse*, 27 (5), 58-61.
- Marcoulides, George A.; Saunders, Carol (2006). PLS: A Silver Bullet? *MIS Quarterly*, 30 (2), 201-211.
- Marsh, H.W., Hau, K.F., Balla, J.R. & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioural Research*, 33 (2), 181-220.
- Massad, N., Heckman, R., & Crowston, K. (2006). Customer Satisfaction with Electronic Service Encounters. *International Journal of Electronic Commerce*, 10(4), 73-104.
- Mathwick, C. (2002). Understanding the online consumer: a typology of online relational norms and behavior, in: *Journal of Interactive Marketing*, 16 (1), S.40-45.
- Mathwick, C., Malhotra, N. & Rigdon, E. (2001). Experimental value: conceptualisation, measurement and application in the catalog and internet shopping environment. *Journal of Retailing*, 77(1), 39-56.
- Matthews, K.K. (1999). *Aesthetics and Usability, Advanced Topics in User Interface Design*, University of Colorado at Boulder, URL:
<http://home.att.net/~kiana.matthews/independent-study/Aesthetics-Paper.htm>
- Mayer, R.C., Davis, J.H. & Schoorman, F.D., (1995). An integrative model of organizational trust. *Academy of Management Review* 20 (3), 709-734.
- McColl-Kennedy, J.R. & Fetter, R.E. (2001). An empirical examination of the involvement to external search relationship in services marketing, in: *Journal of Services Marketing*, 15 (2), S.82-98.
- McDonald, R.P. & Ho, M.H. (2002). Principles and practice in reporting structural equation analysis. *Psychological Methods*, 7(1), 64-82.
- McKinney, V., Yoon, K. & Zahedi, F. (2002). The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. *Information Systems Research*, 13 (3), 296-315.

- McKnight, D. H., Cummings, L. L. & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, 23, 473-490.
- McQuarrie, E.F. & Munson, J.M. (1992). A revised product involvement inventory : Improved usability and validity. *Advances in Consumer Research* 19, 108-115.
- Meffert, H. (1992). *Marketingforschung und Käuferverhalten*, Wiesbaden, Gabler.
- Meffert, H. & Bruhn, M. (1997). *Dienstleistungsmarketing: Grundlagen, Konzepte, Methoden*. 2. Aufl., Wiesbaden, Gabler.
- Meffert, H. & Bruhn, M. (2000). *Dienstleistungsmarketing: Grundlagen, Konzepte, Methoden*. 3. Aufl., Wiesbaden, Gabler.
- Meffert, H. & Bruhn, M. (2003). *Dienstleistungsmarketing*, 4. Aufl., Wiesbaden, Gabler.
- Meffert, H. & Bruhn, M. (2006). *Dienstleistungsmarketing*, 5. Aufl., Wiesbaden, Gabler.
- Menard, S. (2001): *Applied Logistic Regression Analysis*, 2. Aufl., Thousand Oaks.
- Mendoza, J. L., Stafford, K. L. & Stauffer, J.M. (2000). Large sample confidence intervals for validity and reliability coefficients. *Psychological Methods*, 5, 356-369.
- Merz, M. (2002): *E-Commerce und E-Business*, 2. Aufl., Heidelberg, dpunkt-Verlag.
- Meyer, A. & Pfeiffer, M. (1998). Der Einsatz interaktiver Medien: Nutzen und Konsequenzen für Dienstleistungsanbieter, in: A. Meyer (Hrsg.): *Handbuch Dienstleistungs-Marketing*, Bd. 1, (297-318) Stuttgart, Schäffer-Poeschel.
- Mittal, V. & Kamakura, W. (2001). Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics. In: *Journal of Marketing Research*, 38 (1), 131-142.
- Mittal, V. & Sawhney, M. (2001). Learning and using electronic information products and services: a field study, in: *Journal of Interactive Marketing*, 15 (1), S. 2-12.
- Moll, T. (1987). Über Methoden zur Analyse und Evaluation interaktiver Computersysteme. In K.-P. Fähnrich (Hrsg.), *Software-Ergonomie* (S. 179-190). München: Oldenbourg.
- Moon J.W. & Kim, Y.G. (2001). Extending the TAM for a World Wide Web context, in: *Information & Management*, 38 (4), 217-230.
- Moosbrugger, H. (1983). Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. Strukturierung und Reduzierung von Daten. In: J. Bredenkamp & H. Feger (Hrsg.): *Enzyklopädie der Psychologie. Themenbereich Methodologie und Methoden*, Göttingen, Hogrefe.
- Moosbrugger, H., Klein, A., Frank, D. & Schermelleh-Engel, K. (1996). Zum Problem der Schätzung von latenten Moderatoreffekten (On the problem of estimating latent moderator effects). In R. Brandmaier & Ch. Rietz (Hrsg.), *Methodische Grundlagen und*

- Anwendungen von Strukturgleichungsmodellen* (Band 2, S. 5-35). Mannheim: Forschung Raum und Gesellschaft e.V.
- Moosmüller, G. (2004). *Methoden der empirischen Wirtschaftsforschung*. München: Pearson.
- Morrison, D.G. (1969): On the Interpretation of Discriminant Analysis, *Journal of Marketing Research*, 6, 156-163.
- Mukherjee, A., & Nath, P. (2007). Role of electronic trust in online retailing. *European Journal of Marketing*, 41(9/10), 1173-1202
- Muller, M., McClard, A., Bell, B., Dooley, S., Meiskey, L., Meskill, A., Sparks, R., & Tellam, D. (1995). Validating an Extension to Participatory Heuristic Evaluation: Quality of Work and Quality of Work Life. In: *CHI'95 Mosaic of Creativity*, pp 115-116.
- Mummendey, A. (1979). Zum gegenwärtigen Stand der Erforschung der Einstellungs-Verhaltens-Konsistenz. In: H. D. Mummendey (Hrsg.) *Einstellung und Verhalten. Psychologische Untersuchungen in natürlicher Umgebung*. Bern: Huber, 13-30.
- Mundorf, N., (1993). "Effects of Hedonic Components and User's Gender on the Acceptance of Screen-Based Information Services," *Behaviour & Information Technology*, 12 (5), S. 293 303.
- Nagy, M.S. (2002). Using a Single-Item Approach to Measure Facet Job Satisfaction. *Journal of Occupational and Organizational Psychology*, 75, 77-86.
- Neibecker, B. (1990). *Werbewirkungsanalyse mit Expertensystemen*, Heidelberg, Physica-Verlag..
- Nelson, P. (1974). Advertising as information. In: *Journal of Political Economy*, 82 (2), 729-754.
- Nerdinger, F. (2001). *Psychologie des persönlichen Verkaufs*, München: Oldenbourg Wissenschaftsverlag.
- Nicolaou, A. & McKnight, H. (2006). Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use. *Information Systems Research*, 17 (4), 332-351.
- Nielsen, J. (1993). Usability Engineering, San Diego: Academic Press Nielsen, J. (1994a). Heuristic Evaluation, in: J. Nielsen, & R.L. Mack (Eds.). *Usability Inspection Methods*, New York: John Wiley, pp.25-62
- Nielsen, J. (1994). Heuristic Evaluation, in: J. Nielsen & R.L. Mack (Eds.). *Usability Inspection Methods*, New York: John Wiley, pp.25-62 .
- Nielsen, J. (1998). *Using Link Titles to Help Users Predict Where they are Going*, URL: <http://www.useit.com/alertbox/980111.html>.
- Nielsen, J. (2001). *Did Poor Usability Kill E-Commerce?*, URL

- <http://www.useit.com/alertbox/20010819.html>.
- Nielsen, J. (2002). *User Empowerment and the Fun Factor*, URL:
<http://www.useit.com/alertbox/20020707.html>.
- Nielsen, J. & Molich, R. (1990). Heuristic Evaluation of User Interfaces. In J.C. Chew & J. Whiteside (Eds.), *Proceedings of the CHI Conference '90: Human Factors in Computing Systems* (249-256). New York: ACM.
- Nunnally, J.C. (1978). *Psychometric Theory*, (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. & Bernstein, I. (1994). *Psychometric Theory*, 3. Ed, New York, McGraw-Hill.
- O'Connor, B. P. (2006). Programs for Problems Created by Continuous Variable Distributions in Moderated Multiple Regression. In: *Organizational Research Methods*, 9, 554-569.
- Oldenburger, K., Lehto, X., Feinberg, R., Lehto, M., & Salvendy, G. (2008, January). Critical purchasing incidents in e-business. *Behaviour & Information Technology*, 27(1), 63-77.
- Oppermann, R. & Reiterer, H. (1997). Software Evaluation Using the 9241 Evaluator. *Behaviour & Information Technology*, 16(4/5), 232-245.
- Otter, M. & Johnson, H. (2000). Lost in Hyperspace: Metrics and Mental Models. *Interacting with Computers*, 13, 1-40.
- Parasuraman, A. & Grewal, D. (2000). The impact of technology on the quality-value-loyalty chain: A research agenda. In: *Journal of Academy of Marketing Science*, 28 (1), 168-174.
- Parasuraman, A.; Zeithaml, V.A. & Berry, L.L. (2001): Alternative scales for Measuring service quality: a comparative assessment based on psychometric and diagnostic criteria, in: Bruhn, M.; Meffert, H. (Hrsg.): *Handbuch Dienstleistungsmanagement*, (S. 465-498) Wiesbaden 2001, Gabler-Verlag.
- Parasuraman, A., Zeithaml, V. & Malhotra, A. (2005). E-S-QUAL. A Multiple-Item Scale for Assessing Electronic Service Quality. *Journal of Service Research*, 7(3), 213-233.
- Parente, S.T. (2000). Beyond the hype: a taxonomy of e-health business models, in: *Health Affairs* 19 (6), 89-102.
- Park, D., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 11(4), 125-148.
- Pavlou, P.A. (2003). Consumer acceptance of electronic commerce integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce* 7(3), 69-103.
- Pavlou, P.A. & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. *MIS Quarterly* 30(1), 115-143.

- Pavlou, P.A. & Gefen, D. (2005). Psychological Contract Violation in Online Marketplaces: Antecedents, Consequences, and Moderating Role. *Information Systems Research*, 16 (4), 372-399.
- Pereira, R.E. (1999): Consumer Information Search and Decision Making in the Electronic Commerce Environment, in: Sudweeks, F. & Romm, C.T. (1999): *Doing Business on the Internet*, Springer, 1999.
- Peter, J.P. (1979): Reliability: A Review of Psychometric Basics and Recent Marketing Practices, *Journal of Marketing Research*, 26, 6-17.
- Peter, J. & Olson, J. (1987). *Consumer behavior: marketing strategy perspectives*, Irvin, Homewood, IL.
- Peters, T. & Waterman, R. (2000). *Auf der Suche nach Spitzenleistungen: Was man von den bestgeführten US-Unternehmen lernen kann*, 8. Aufl., Landsberg am Lech: Moderne Verlagsges. MVG, 2000.
- Picinelli, G. & Stammers, E. (2000). *From e-processes to e-networks: an e-service-oriented approach*, Working Paper, Hewlett-Packard Laboratories, Filton Park, Bristol.
- Picot, R. & Neuburger, R. (2002). Prinzipien der Internet-Ökonomie, in: M. Schlögel, T. Tomczak & C. Belz (Hrsg.): *Roadm@p to E-Business Wie Unternehmen das Internet erfolgreich nutzen*, St. Gallen, 92-107.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y. & Podsakoff, N.P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88, 879-903.
- Poensgen, A., & Larson, S. (2001). *Patienten, Ärzte und Internet: Mythos Realität und Implikationen*, The Boston Consulting Group, München 2001.
- Popper, K. (1935). *Logik der Forschung*. Springer, Wien 1935.
- Porter, M. E. (1996): What is strategy? In: *Harvard Business Review* (1996), November-December, S. 61-78.
- Preece, J. (1994). *Human-Computer Interaction*. Harlow: Addison-Wesley.
- Preuß, K.J. (1997). Informations- Kommunikations-Technologien und Vernetzung im Gesundheitssektor als Basis für Managed-Care-Konzepte, in: Arnold, M., Lauterbach, K.W., Preuß, K.J. (Hrsg) *Managed Care*, Schriftenreihe der Robert Bosch Stiftung, Beiträge zur Gesundheitsökonomie 31, (S. 259-300) Stuttgart, New York 1997.
- Pristl, K. & Reuther, R., (2005). Leistungsausgaben in der gesetzlichen Krankenversicherung. *Statistisches Monatsheft Baden-Württemberg* 7, 18-20.
- Prümper, J. & Anft, M. (1993). Die Evaluation von Software auf der Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen

- Systemgestaltung ein Fallbeispiel. In K.-H. Rüdiger (Hrsg.), *Software Ergonomie '93* (145-156). Stuttgart: Teubner.
- Quelch, J. A. & Klein, L. R. (1996). The Internet and international marketing. *Sloan Management Review*, 37, Vol. 3, 60-75.
- Rack, O. & Christophersen, T. (2006). Experimente. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (19-37). Wiesbaden: Deutscher Universitäts-Verlag.
- Raffee, H. & Jacobs, S. (1986). The new media: effects on consumer behavior and trade. *Markenartikel*, 48 (12), 568-574.
- Raffee, H. & Silberer, G. (1981). *Informationsverhalten des Konsumenten*, Wiesbaden.
- Rafiei, D. & Mendelzon, A., (2000). *What is this page known for?* Computing webpage reputation. Proceedings of the www9 Conference, Amsterdam, 2000.
- Ranaweera, C., McDougall, G., & Bansal, H. (2004). *A model of online customer behaviour: moderating effects of customer characteristics* (Working Paper No. 2004-01 MKTG). Waterloo: Wilfrid Laurier University, School of Business & Economics.
- Ranaweera, C., McDougall, G., & Bansal, H. (2005). A model of online customer behavior during the initial transaction: Moderating effects of customer characteristics. *Marketing Theory Articles*, 5 (1), 51-74.
- Rapp, R. (1995). *Kundenzufriedenheit durch Servicequalität*. Wiesbaden: Deutscher Universitäts-Verlag.
- Reichheld, F.F. (1997). *Der Loyalitäts-Effekt*. Frankfurt a.M.: Campus.
- Reichheld, F. F. & Sasser, E. (1990): Zero-defections: quality comes to services, in: *Harvard Business Review*, Vol. 68, Nr. 5, S. 105-111.
- Reichwald, R. Piller, F. & Meyer, R (2002). E-Service Customiction Strategien zur effizienteren Individualisierung von Dienstleistungen, in: Bruhn, M. & Stauss, B. (Hrsg.): *Electronic Services Dienstleistungsmanagement Jahrbuch 2002*, Wiesbaden, S. 225-241.
- Rengelshausen, O. (1999). *Online-Marketing in deutschen Unternehmen*. Einsatz, Akzeptanz und Wirkungen, Wiesbaden, Gabler.
- Richter, M. (1999). Online-Befragung als neues Instrument zur Beurteilung der Benutzerfreundlichkeit interaktiver Software am Beispiel einer Internet-Anwendung. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, M. K. & A. Werner (Hrsg.), *Aktuelle Online-Forschung Trends, Techniken, Ergebnisse*. Zürich: Online Press.
- Ring, S. & Winand, U. (2000): Haushaltsgerechte Oberflächen für E-Commerce-Systeme, in: *HMD*, Nr. 215, 2000.

- Ringle, C. (2004). Gütemaße für den PLS-Ansatz zur Bestimmung von Kausalmodellen, *Industrielles Management* 16, 27-30.
- Robertson, T. (1976). Low commitment consumer behaviour, in: *Journal of Advertising Research*, 16, pp. 19-24.
- Rockart, J. F. (1979): Chief executives define their own data needs. In: *Harvard Business Review* 57 (1979), March-April, S. 81-92.
- Rohrlack, C. (2006). Logitische und Ordinale Regression. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (205-222). Wiesbaden: Deutscher Universitäts-Verlag.
- Romita, T. (2001). Familiarity breeds net sales. *Business 2.0* (February 6, 2001), p. 98.
- Roßmanith, T. (2001). *Informationsverhalten und Involvement im Internet*. Universität Karlsruhe.
- Roselli, T. (1991). Control of User Disorientation in Hypertext Systems, in: *Educational Technology*, 31 (12), 42-46.
- Rossiter, J.R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing* 19, 305-335.
- Rossiter, J.R. (2005). Reminder: a Horse is a Horse. *International Journal of Research in Marketing*, 22, 23-25.
- Roth, S. & Pastowski, S. (2002). Ökonomische Analyse von Internet-Auktionen, in: Bruhn, M./Stauss, B. (Hrsg.): *Electronic Services, Jahrbuch Dienstleistungsmanagement 2002*, Wiesbaden, 693-722.
- Rubin, D.B. (1977): Formalizing Subjective Notion About the Effect of Nonrespondents in Sample Surveys, *Journal of the American Statistical Association*, 72, 538-543.
- Rubin, D.B. (1987): *Multiple Imputation for Nonresponse in Surveys*, New York et al., Academic Press.
- Rubin, J. (1994). *Handbook of Usability Testing*. New York: John Wiley & Sons, Inc.
- Rück, H. (2000). *Dienstleistungen in der ökonomischen Theorie*, Wiesbaden.
- Sarris, V. (1992): *Methodologische Grundlagen der Experimentalpsychologie 2: Versuchsplanung und Stadien*, München, Reinhardt.
- Scansaroli, J.A. & Eng, V. (1997). Interactive retailing: Marketing Products. *Chain Store Age*, 73(1), 9A-10A.
- Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London.
- Schafer, J.L. (1999). Multiple Imputation: a Primer, *Statistical Methods in Medical Research*, 8, 3-15.

- Scheer, A. & Nüttgens, M. (1999). *Electronic business engineering*. 4. int. Tagung Wirtschaftsinformatik, Heidelberg.
- Schenkman, B.N. & Jönsson, F.U. (2000). Aesthetics and preferences of web pages. *Behaviour & Information Technology* 19, 367-377.
- Schiller W., 1988. Vom sinnvollen Aufwand in der Faktorenanalyse. *Archiv für Psychologie*, 140, 73-95.)
- Schillewaert, N., Ahearne, M.J., Frambach, R.T. & Moeckart, R.K. (2005). The adoption of information technology in the sales force. *Industrial Marketing Management* 34, 323-336.
- Schmid, B.F. (2001): What is New in the Digital Economy?, in: *Electronic Markets - International Journal of Electronic Commerce & Business Media*, Vol. 11,Nr. 1, S. 1-17.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait multimethod matrices. *Applied Psychological Measurement*, 10, 1 22.
- Schneider, H. (2006). Nachweis und Behandlung von Multikollinearität. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (187-204). Wiesbaden: Deutscher Universitäts-Verlag.
- Schneider, K.C. & Johnson, J.C. (1994). Link between Response-inducing Strategies and Uninformed Response, *Marketing Intelligence & Planning*, 12 (1), 29 36.
- Schneider, D. & Gerbert, P. (1999). *E-Shopping: Erfolgsstrategien im Electronic Commerce Marken schaffen, Shops gestalten, Kunden binden*, Wiesbaden, Gabler.
- Schnell, R., Hill, P. & Esser, E. (1999). *Methoden der empirischen Sozialforschung*, 6. Aufl., München, Oldenburg.
- Schneller, J. (2003). *ACTA 2003- Die Entwicklung des Internet als Vertriebskanal*, Studie von Allensbacher Computer- und Technikanalyse, Institut für Demoskopie Allensbach, Allensbach.
- Schrepp, Held & Laugwitz (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. *Interacting with Computers*, 18 (5), 1055-1069.
- Schröder, H. (1994). Erfolgsfaktorenforschung im Handel – Stand der Forschung und kritische Würdigung der Ergebnisse, *Marketing Zeitschrift für Forschung und Praxis*, 16 (2), S. 89 104.
- Schubert, C. (2000). *Cybermediaries als neue Geschäftsform im Internet: Grundlagen, Erscheinungsformen und strategische Handlungsalternativen*, Wiesbaden, Gabler.

- Schulenburg, J.M. Graf v. d. & Greiner, W. (2000). *Gesundheitsökonomik*, Tübingen, Mohr-Siebeck.
- Schuman, H. (1992). Context effects: state of the past / state of the art. In: Schwarz, N., Sudman, S., (Eds.), *Context Effects in Social and Psychological Research* (pp. 4-20). New York: Springer.
- Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys*. San Diego: Academic Press.
- Schumann, S. (1997): *Repräsentative Umfrage*, 2. Aufl., München et al, Oldenbourg.
- Schwalbach, J. (2004). Reputation. In: Schreyögg, G., v. Werder, A. (Hrsg.), *Handbuch Unternehmensführung und Organisation*, 4. Auflage, S.1266-1272. Stuttgart: Schäffer Pöschel Verlag, 2004.
- Schwarz, N., Groves, R. & Schuman, H. (1998). Survey methods. In: Gilbert, D., Fiske, S., Lindzey, G., (Eds.), *Handbook of Social Psychology*, (4th ed., vol. 1, pp 143-179). New York: McGraw-Hill.
- Schwarz, N. & Hippler, H.J. (1995). Subsequent questions may influence answers to predicting questions in mail surveys. *Public Opinion Quarterly* 59, 93-97.
- Schwarz, N., Strack, F. & Mai, H.P. (1991). Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. *Public Opinion Quarterly* 55, 199-207.
- Schwarz, U. & Müller, S. (2007). Erfolgsfaktoren der Gestaltung von Webseiten. *WiSt Wirtschaftswissenschaftliches Studium*, 12, S. 562-569.
- Sellen, A. & Nicol, A. (1995). Building User-Centred On-Line Help, in: R. Baecker, J. Grudin, W. Buxton & S. Greenberg (Eds.). *Readings in Human-Computer Interaction: Toward the Year 2000*, 2nd Ed., San Francisco, Morgan Kaufmann, 718-723.
- Shackel B. (1991). Usability-Context, Framework, Definition, Design & Evaluation, in: B. Shackel & S. Richardson (Eds.), *Human Factors of Informatics Usability*, (21-37) Cambridge: Cambridge University Press.
- Shankar, V., Smith, A. & Rangaswamy, A. (2003). Customersatisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20, S. 153-175.
- Shih, Y. (2006). The effect of computer self-efficacy on enterprise resource planning usage. *Behaviour & Information Technology*, 25 (5), 407-411.
- Shin, D. (2007). User acceptance of mobile Internet: Implication for convergence technologies, *Interacting with Computers*, 19 (4), 472-483.

- Silberer, G. (1985). Wertewandel und Marketing, in: *WiSt (Wirtschaftswissenschaftliches Studium)*, 14. (3), 119-124.
- Silberer, G. & Yom, M. (2001). Das Verhalten und Erleben von Webnovizen beim Online-Shopping. In W. Fritz (Hrsg.), *Internet-Marketing*, 2. Aufl., 433-452. Stuttgart: Schäffer-Poeschel.
- Sillence, E., Briggs, P., Harris, P. & Fishwick, L. (2007). Going online for health advice: Changes in usage and trust practices over the last five years. *Interacting with Computers*, 19 (3), 397-406.
- Silverstein, M.; Stanger, P. & Abdelmessih, N. (2001): *Winning the online consumer 2.0 - converting traffic into Profitable Relationships*, Studie von The Boston Consulting Group, 2001.
- Simon, H. (2003): E-Business - Quo Adis? Analyses and Perspectives, in: Albach, H.; Hummel, I. (Hrsg.): Die Zukunft des Electronic Business, *ZfB*, Ergänzungsheft 1/2003, Wiesbaden 2003, S. 1-16.
- Simon, H. & Homburg, C. (Hrsg.). (1998). *Kundenzufriedenheit* (3., aktual. und erw. Aufl.). Wiesbaden: Gabler.
- Slaughter, L. A., Harper, B. D. & Norman, K. L. (1994). Assessing the Equivalence of Paper and On-line versions of the QUIS 5.5. *Proceedings of the 2nd Annual Mid-Atlantic Human Factors Conference*, (pp. 87-91), Washington, D.C.
- Smith, J. B., & Barclay, D. W. (1997). The effects of organizational differences and trust on the effectiveness of selling partner relationships. *Journal of Marketing*, 61, 3-21.
- Spence, R. (1999). A Framework for Navigation, in: *International Journal of Human-Computer-Studies*, 51 (5), S. 919-945.
- Spitzenverbände der gesetzlichen Krankenversicherung (2006). *Untersuchung der Auswirkungen befundbezogener Festzuschüsse*, Planungsgruppe M + MAG, AOK-Bundesverband, Bonn 2006.
- Sporer, M. & Blackert, S. (2001). E-Commerce die Konsumentenperspektive, in: Hermann, A. & Sauter, M. (Hrsg.): *Management-Handbuch Electronic Commerce: Grundlagen, Strategien, Praxisbeispiele*, 2.Aufl., München, S 75-83.
- Stahl, B. (1987). Testing for usability can head off disaster. *Computerworld* 21(49), 83-92.
- Starr, P. (1997). Smart technology, stunted policy: developing health information networks, in: *Health Affairs*, (5/6), 91-105.
- Stauss, B. (1991). Dienstleistungen in die vierte Dimension, in: *Harvard Business Manager*, 13 (2), 81-89.

- Stauss, B. (1996). Dienstleistungen als Faktoren, in: Kern, W., Schröder, H.H. & Weber, J. (Hrsg.): *Handwörterbuch der Produktionswirtschaft*, 2. Aufl., Stuttgart, Schäffer-Poeschel, S. 317-327.
- Stauss, B. (1999). Kundenzufriedenheit. *Marketing ZFP*, 21 (1), 5 – 24.
- Stauss, B. (2001a). Technologie in der Dienstleistungsmarketing-Forschung, in: *Die Unternehmung*, 55 (4/5), S. 307-326.
- Stauss, B. (2001b). „Augenblicke der Wahrheit“ in der Dienstleistungserstellung – ihre Relevanz und ihre Messung mithilfe der Kontaktpunkt-Analyse, in: Bruhn, M. & Stauss, B. (Hrsg.): *Dienstleistungsqualität: Konzepte Methoden Erfahrungen*, 3. Aufl., S. (321-340) Wiesbaden, Gabler.
- Steckel, J.H. & Vanhonacker, W.R. (1993): Cross-Validating Regression Models in Marketing Research, *Marketing Science*, 12, 415-427.
- Steffen, A. (1994). *Das Problem der Multikollinearität in Regressionsanalysen*, Dissertation, Frankfurt am Main.
- Stelzer, D. (2000). Digitale Güter und ihre Bedeutung in der Internet-Ökonomie. In: *WISU - Das Wirtschaftsstudium*, 6, S. 835-842.
- Stewart, K. (2006). How Hypertext Links Influence Consumer Perceptions to Build and Degrade Trust Online. *Journal of Management Information Systems*, 23(1), 183-210.
- Stier, W. (1999): *Empirische Forschungsmethoden*, 2. Aufl., Berlin, Springer.
- Straub, D., Limayen, M. & Karahanna-Evaristo, E., 1995. Measuring system usage: implications for IS theory testing. *Management Science* 41, 1328 – 1342.
- Straub, D.W., Keil, M. & Brenner, W. (1997). Testing the technology acceptance model across cultures: a three country study. *Information & Management*, 31 (1), pp. 1-11.
- Su, B. (2008). Characteristics of Consumer Search On-Line: How Much Do We Search?. *International Journal of Electronic Commerce*, 13(1), 109-129.
- Sudman, S., Bradburn, N. & Schwarz, N. (1996). *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Suh, B. & Han, I. (2003). The impact of customer trust and perception of security control on the acceptance of electronic commerce. *International Journal of Electronic Commerce*, 7 (3), 135-161.
- Suh, J., & Yi, Y. (2006). When Brand Attitudes Affect the Customer Satisfaction-Loyalty Relation: The Moderating Role of Product Involvement. *Journal of Consumer Psychology*, 16(2), 145-155.
- Sweeney, M., Maguire, M. & Shackel, B. (1993). "Evaluating usercomputer interaction: a framework", *International Journal for Man-Machine Studies*, 38, 689-711.

- SwissRe (Hrsg.): *Sigma 5/2000*, E-Business in der Versicherungswirtschaft, Zwang zur Anpassung Chance zur Erneuerung, Zürich 2000.
- Swoboda, B. (1999). Ausprägungen und Determinanten der zunehmenden Convenience-orientierung von Konsumenten. *Marketing ZFP*, 21 (2), 95 – 104.
- Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management Science* 42(1) 85 – 92.
- Tabachnik, B. & Fidell, L. (2006). *Using Multivariate Statistics* (5th ed.). Boston: Pearson.
- Thissen, F. (2001). *Screen-Design Handbuch – Effektiv informieren und kommunizieren mit Multimedia*, 2. Aufl., Berlin, Springer.
- Thorelli, H. & Thorelli, S. (1977). *Customer information systems and consumer policy*, Cambridge, Ballinger Publishing.
- TNS-EMNID, (2003). *Optimierung des TK-Geschäftsstellennetzes*. Untersuchungsteil Versicherte, Bielefeld, unveröffentlichte Untersuchung, 2003.
- TNS Infratest, (2005). *TK-Online-Monitor, 2005*. Bielefeld, unveröffentlichte Untersuchung, 2005.
- Tourangeau, R. & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin* 103(3), 299-314.
- Townsend, A. M., Demarie, S.M. & Henrickson, A.R. (2001). Desktop video conferencing in virtual workgroups: anticipation, system evaluation and performance. *Information Systems Journal* 11, 213-227.
- Tractinsky, N., (1997). Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: Sheva, B., *Association for Computing Machinery, CHI 97* Electronic Publications, revidiert am 23.7.1997
<http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm>
- Tractinsky, N., Katz, A. S. & Ikar, D. (2000): What is beautiful is usable. *Interacting with Computers*, 13, 127 – 145.
- Tractinsky, N. & Rao, S. (2001). Incorporating social dimensions in Web-store Design. *Human Systems Management*, 20, 105 – 121.
- Trommsdorff, V. (1990): *Erfolgsfaktorenforschung, Produktinnovation und Schnittstelle Marketing F&E, Diskussionspapier*, TU Berlin 143, Berlin 1990.
- Trommsdorff, V. (1995). Involvement, in Tietz, B. (Hrsg.), *Handwörterbuch des Marketing*, Stuttgart, S. 1067-1078.
- Trommsdorff, V. (1998): *Konsumentenverhalten*, 3. Aufl., Kohlhammer, Stuttgart.

- Urban, D. & Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung*, 2. Auflage, Wiesbaden, VS-Verlag
- Van der Heijden, H. & Verhagen, T. (2004). Online Store Image: Conceptual Foundations and Empirical Measurements. *Information & Management*, 41, 609-617.
- Varianini, V. & Vaturi, D. (2000): Marketing lessons from e-failures, in: *The McKinsey Quarterly*, No. 4, 2000, S. 86-97.
- Venkatesh, A. (1998). Cybermarketscapes and consumer freedoms and identities, in: *European Journal of Marketing*, 32 (7/8), 664-676.
- Venkatesh, V. (1999). Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation, *MIS Quarterly* (23:2), 1999, 239-260.
- Venkatesh, V., Brown, S., Maruping, L., & Bala, H. (2008). Predicting different Conceptualizations of System Use: The competing roles of behavioral intention, facilitating conditions, and behavioral expectations. *MIS Quarterly*, 32(3), 483-502.
- Venkatesh, V., Cheri, S. & Morris, M.G. (2002). User acceptance enablers in individual decision making about technology: towards an integrated model. *Decision Sciences* 33, 297-316.
- Venkatesh, V. & Davis, F.D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science* 46, 186-204.
- Virzi, R.A. (1997). Usability Inspection Methods. In M. Helander, T.K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., 705-715). Amsterdam: Elsevier.
- WAI (Web Accessibility Initiative), vgl:<http://www.w3.org/WAI/Resources>
- Wang, G.C.S. (1996): How to handle multicollinearity in regression modeling, *Journal of Business Forecasting Methods & Systems*, 15, 1, 23-27.
- Ward, M.R & Lee, M.J. (1999). *Internet shopping, consumer search, and product branding*. Working paper. University of Pennsylvania.
- Weber, T. & Zhiqiang (2007). A Model of Search Intermediaries and Paid Referrals. *Information Systems Research*, 18 (4), 414-436.
- Webster, J. & Martocchio, J. (1992). Microcomputer playfulness: Development of a measure with workplace implications, *MIS Quarterly* 16(2), pp. 201 226.
- Wegmann, C. (2002). Der E-Services Marketingmix, in: M. Bruhn, B. Stauss (Hrsg.): *Electronic Services Dienstleistungsmanagement, Jahrbuch 2002*, Wiesbaden, Gabler Verlag, S. 243-262.

- Weiber, R. & Adler, J. (1995). Informationsökonomisch begründete Typologisierung von Kaufprozessen. *ZfbF – Schmalenbachs Zeitschrift für die betriebswirtschaftliche Forschung*, special edition, 47 (1), 43-65.
- Weiber, R. & Weber, M. R. (2002): Customer Relationship Marketing und Customer Lifetime Value im Electronic Business, in: Weiber, R. (Hrsg.): *Handbuch Electronic Business*, 2. Aufl., Wiesbaden, S. 609-643.
- Weinberg, P. (1981). *Das Entscheidungsverhalten der Konsumenten*, Paderborn, Schöningh, 1981.
- Weinberg, P. & Besemer, S. (1999). Shopping-Center in der Zukunft, in: *Marketing ZFP*, 21(3), 237-247.
- Weise, G. (1975). *Psychologische Leistungstests*, Göttingen: Hogrefe.
- Werts, C.E., Linn, R.L. & Jöreskog, K.G. (1974): Interclass Reliability Estimates: Testing Structural Assumptions, *Educational and Psychological Measurement*, 34, 24-33.
- West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and implications*, 56-75. Thousand Oaks: Sage Publications.
- Wiedemann, K.-P.; Frenzel, T. & Buxel, H. (2001): Strategisches E-Commerce-Marketing, in: Eggers, B., Hoppen, G. (Hrsg.): *Strategisches E-Commerce-Management*, (S. 417-443) Wiesbaden, Gabler.
- Wiesner, V. (2004). Usability versus Design – ein Widerspruch? Theorie und Praxis der Gestaltung von Websites. In: *Alles Buch – Studien der Erlanger Buchwissenschaft VI*. Hrsg.: Rautenberg, U. und Titel, V., Universität Erlangen-Nürnberg.
- Wilhelm, T., Yom, M. & Wohlfahrt, J. (2002). Web-Gestaltung: So fühlt sich der Surfer auf Ihrer Homepage heimisch, in: *acquisa – Zeitschrift für Führungskräfte in Verkauf und Marketing*, 50 (10), 50-51.
- Wilke, K. (2000): Die Eignung des Internets für die Reduktion von Qualitätsrisiken im Kaufentscheidungsprozess des Konsumenten, in: *IfH*, 5, 2000, S. 117-133
- Winand, U. & Pohl, W. (2000). Die Vertrauensproblematik in elektronischen Netzwerken. In: J. Link (Hrsg.), Wettbewerbsvorteile durch Online-Marketing. *Die strategischen Perspektiven elektronischer Märkte* (2. Auflage, S. 261-277). Berlin: Springer.
- Wirtz, B.W. (2000). *Electronic Business*, Wiesbaden, Gabler.
- Wirtz, B.W. (2001). *Electronic Business*, 2. Aufl., Wiesbaden, Gabler.
- Wirtz, B. & Olderog, T. (2002). Kundenbindungsmanagement für elektronische Dienstleistungen. In: Bruhn, M.; Stauss, B. (Hrsg.): *Dienstleistungsmanagement Jahrbuch 2002 – Electronic Services*. Gabler, Wiesbaden, 513-562.

- Wirtz, B.W. & Vogt, P. (2001). Kundenbeziehungsmanagement im Electronic Business. *Jahrbuch der Absatz- und Verbrauchsforschung*, 2/2001, 116 - 135.
- Wiswede, G. (1995). *Einführung in die Wirtschaftspsychologie*. München: Reinhardt.
- Wittenberg, J. & Arminger, G. (1997). Bayesian Non-Linear Latent Variable Models - Specification and Estimation with the Program System BALAM, in: Bandilla, W. & Faulbaum, F. (Hrsg.), *SOFTSTAT 97, Advances in Statistical Software 6*, Lucius & Lucius, Stuttgart 1997, 487 - 496.
- Wixom, B.H. & Todd, P.A. (2005). A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*, 16 (1), 85 - 102.
- Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In P.R. Krishnaiah (Ed.), *Multivariate Analysis*, 391-420. New York: Academic Press.
- Wold, H. (1982). Soft modeling : the basic design and some extensions, in: Jöreskog, K. G. & Wold, H. (Hrsg.), *Systems under indirect observations: causality, structure, prediction ; part 2*, Amsterdam, S. 1 ff.
- Wolfenbarger, M. & Gilly, M.C. (2001): Shopping online for freedom, control and fun in: *California Management Review*, Vol. 43, (2), S. 34-55.
- Wolfenbarger, M. & Gilly, M.C. (2003). eTailQ: dimensionalizing, measuring and predicting etail quality. *Journal of Retailing*, 79, 183-198.
- Woywode, M. (2002): *Wege aus der Erfolglosigkeit der Erfolgsfaktorenforschung*, Arbeitspapier, Lehr- und Forschungsgebiet Internationales Management, RWTH Aachen.
- Yom, M. (2001). Wirkungen von Orientierungshilfen auf das Shoppingverhalten und Shopperleben bei Webnovizen, in: Jonas, K.J., Breuer, P., Schauenburg, B. & Boos, M. (Hrsg.). *Perspectives on Internet Research: Concepts and Methods* (Online-Document). URL: <http://server3.uni-psych.gwdg.de/gor/contrib/yom-miriam/article.html>
- Yom, M. (2003). *Web-Usability von Online-Shops*. Göttingen: Better Solutions.
- Yom, M. & Wilhelm, T. (2004). WOOS Ein Messinstrument für die wahrgenommene Orientierung in Online-Shops. In: Keil-Slawik, R., Selke, H. & Szwillus, G. (Hrsg.): *Mensch & Computer 2004: Allgegenwärtige Interaktion*. (S. 43 - 53) München: Oldenbourg Verlag.
- Zahedi, F., & Song, J. (2008). Dynamics of Trust Revision: Using Health Infomediaries. *Journal of Management Information Systems*, 24(4), 225-248.
- Zaichkowsky, J.L. (1985). Measuring the involvement construct. *Journal of Consumer Research* 12, 341-352.

- Zaichkowsky, J.L. (1994). The Personal Involvement Inventory: Reduction, Revision, and Application to Advertising. *Journal of Advertising*, 23(4), 59-70.
- Zeithaml, V.A. (1981). How consumer evaluations differ between goods and services. In Donnelly, H.J. & George, W.R. (Ed.): *Marketing of Services*. Chicago, S. 186-190.
- Zeithaml, V.A. & Bitner, M.J. (2003). *Services marketing: integrating customer focus across the firm*. 3. Aufl., Boston, McGraw-Hill.
-

Anhang A

Ergänzende Darstellungen zum Hintergrund

Screenshots zentraler Seiten der Hypertext-Versuchsumgebung des ufosV2-Fragebogen

Abb. A-1.1: Hypertext-Versuchsumgebung Seite 1 - Begrüßung.....	291
Abb. A-1.2: Hypertext-Versuchsumgebung Seite 2 - Festlegung auf Produkt	291
Abb. A-1.3: Hypertext-Versuchsumgebung Seite 3 - Überblick über den Versuchsablauf.	292
Abb. A-1.4: Hypertext-Versuchsumgebung Seite 6 - erster Shop Aufgabe 1 (Teil 1)	292
Abb. A-1.5: Hypertext-Versuchsumgebung Seite 7 - erster Shop Aufgabe 1 (Teil 2)	293
Abb. A-1.6: Hypertext-Versuchsumgebung Seite 12 - Bewertung erster Shop (Teil 1).....	293
Abb. A-1.7: Hypertext-Versuchsumgebung Seite 24 - Kaufentscheidung	294
Abb. A-1.8: Hypertext-Versuchsumgebung Seite 25 - Bestellung	294
Abb. A-1.9: Hypertext-Versuchsumgebung Seite 27 - Wahl des Gutscheins.....	295

Item-Pools der Skala ufosV2

Tab. A-1.1: Item-Pool der Skala ufosV2r	296
Tab. A-1.2: Item-Pool der Skala ufosV2f	297

Screenshots zentraler Seiten der Hypertext-Versuchsumgebung



Abb. A-1.1: Hypertext-Versuchsumgebung Seite 1 Begrüßung

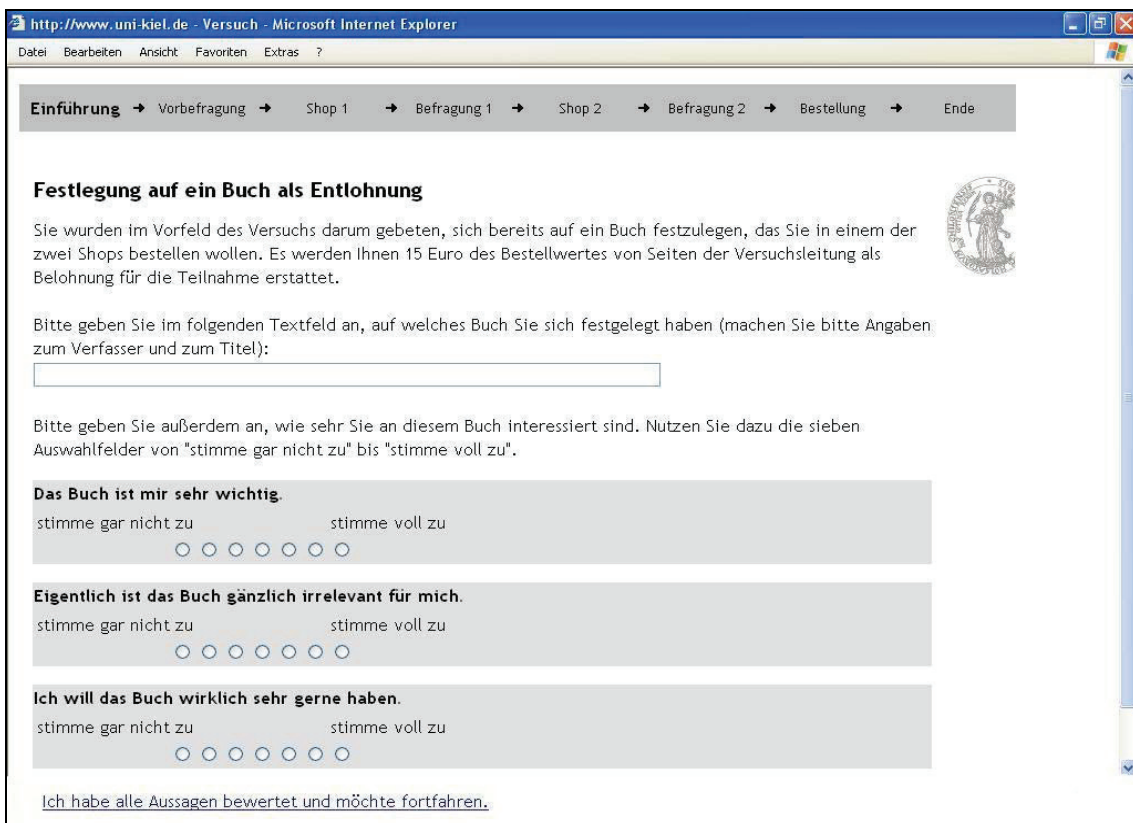


Abb. A-1.2: Hypertext-Versuchsumgebung Seite 2 Festlegung auf Produkt

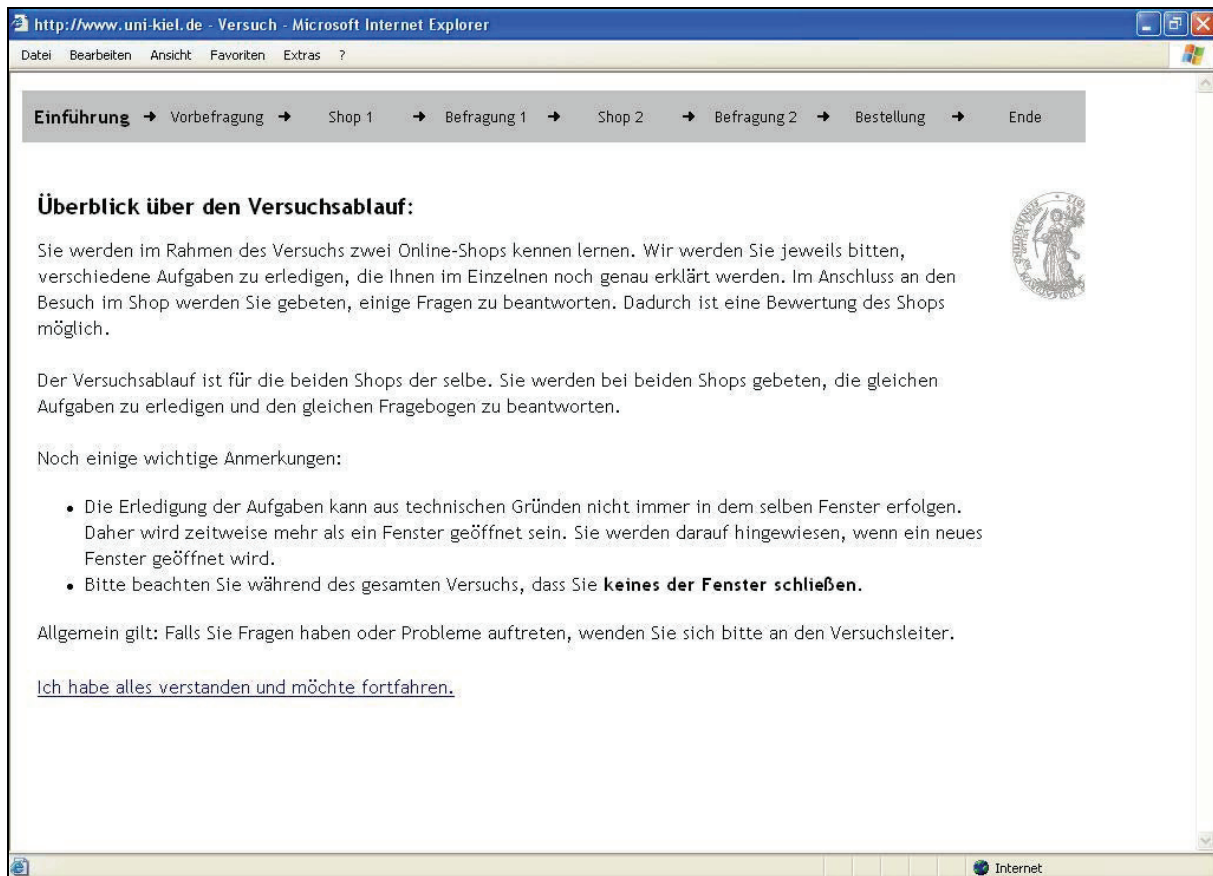


Abb. A-1.3: Hypertext-Versuchsumgebung Seite 3 Überblick über den Versuchsablauf



Abb. A-1.4: Hypertext-Versuchsumgebung Seite 6 erster Shop Aufgabe 1 (Teil 1)



Abb. A-1.5: Hypertext-Versuchsumgebung Seite 7 erster Shop Aufgabe 1 (Teil 2)



Abb. A-1.6 Hypertext-Versuchsumgebung Seite 12 Bewertung erster Shop (Teil 1)



Abb. A-1.7: Hypertext-Versuchsumgebung Seite 24 Kaufentscheidung



Abb. A-1.8: Hypertext-Versuchsumgebung Seite 25 Bestellung



Abb. A-1.9: Hypertext-Versuchsumgebung Seite 27 Wahl des Gutscheins

Tab. A-1.1: Item-Pool der Skala ufosV2r

<i>Code</i>	<i>Reflektives Item</i>
ur1	Dieser Shop macht das Einkaufen im Netz leicht.
ur2c	Es ist mir zu kompliziert, diesen Shop zu benutzen.
ur3	Man kann sich schnell einen Überblick über den Shop verschaffen.
ur4	Der Umgang mit dem Shop ist leicht zu erlernen.
ur6	In diesem Shop kann ein Einkauf zügig erledigt werden.
ur7	Insgesamt bin ich mit der Benutzerfreundlichkeit dieses Shops zufrieden.
ur8	Der Shop bietet alle Möglichkeiten, die ich mir von einem Online-Shop wünsche.
ur9	Ich kann den Shop so benutzen, wie ich es erwarte.

Tab. A-1.2: Item-Pool der formativen Skala ufosV2f

<i>Code</i>	<i>Formatives Item</i>
uf1	Beim Einkauf in diesem Shop kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.
uf5c	Ich muss mir viel merken, wenn ich in dem Shop einkaufe.
uf7	Es ist einfach, zwischen den verschiedenen Seiten des Shops hin- und her zu springen.
uf8	Zentrale Seiten des Shops (z.B. die Startseite) sind jederzeit schnell zu erreichen.
uf9	Ich kann jede Aktion wieder rückgängig machen.
uf10	Mir ist beim Einkauf in dem Shop stets klar, welche Handlungen ich als nächstes durchführen kann.
uf11c	Ich weiß manchmal nicht, wo ich mich innerhalb des Shops befinde.
uf12	Dieser Shop ist gut strukturiert.
uf13c	Es sind zu viele Schritte nötig, um zu einem Produkt zu gelangen.
uf14	Ich finde mich auf den einzelnen Seiten des Shops gut zurecht.
uf15	Ich finde in diesem Shop schnell die Informationen, die für meinen Einkauf wichtig sind.
uf16c	Es werden in diesem Online-Shop häufig zu viele Informationen dargestellt.
uf18c	Die Texte sind manchmal nur schwer lesbar.
uf19c	Das optische Design des Shops stört mich bei meinem Einkauf.
uf20	Wenn ich in diesem Shop auf etwas klicke, passiert genau das, was ich erwarte.
uf21c	Manchmal finde ich die im Shop verwendeten Texte inhaltlich unverständlich.
uf22	Die Benennungen sind so gewählt, dass ich den Shop gut benutzen kann.
uf23	Es ist schnell erkennbar, welche Produkte ich in dem Online-Shop kaufen kann und welche nicht.
uf24	Es ist leicht, gesuchte Produkte zu finden.
uf25	Ich kann sehr gut auf die Suchfunktionen des Shops zugreifen.
uf27	Die Suchfunktionen liefern mir sinnvolle Ergebnisse.
uf28	Die Produktlisten im Shop sind übersichtlich.
uf29	Die Produkte sind in sinnvolle Kategorien eingeteilt.
uf31	Ich bin mit den Beschreibungen der Produkte zufrieden.
uf32	Ich finde die verwendeten Bilder und Grafiken hilfreich.
uf33	Der Shop empfiehlt mir interessante Produkte, nach denen ich ursprünglich nicht gesucht habe.
uf34	Die Preise der einzelnen Produkte sind sofort ersichtlich.
<i>Code</i>	<i>Formatives Item</i>
uf35	Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.
uf36	Der Shop bietet mir nützliche Informationen zu den Lieferbedingungen.
uf38	Ich werde gut über die verschiedenen Zahlungsmöglichkeiten informiert.
uf39c	Man muss lange nach Angaben zu den Lieferzeiten der Produkte suchen.
uf40	Es sind nützliche Informationen zu den allgemeinen Geschäftsbedingungen verfügbar.
uf41	Ich werde ausreichend darüber informiert, wie mit meinen persönlichen Daten umgegangen wird.
uf42	Der Warenkorb ist übersichtlich.
uf43	Ich weiß in diesem Shop stets, welche Produkte sich gerade im Warenkorb befinden.
uf44	Es ist in diesem Shop ganz einfach, Produkte in den Warenkorb zu legen.
uf45	Der Inhalt des Warenkorbes kann ganz einfach verändert werden.
uf46	Der Bestellvorgang kann problemlos durchlaufen werden.
uf47	Die Eingabe meiner persönlichen Daten (z.B. Anschrift, Kreditkartennummer) während des Bestellvorgangs ist unkompliziert.
uf48	Während des Bestellvorgangs weiß ich bei einigen persönlichen Angaben nicht, wofür sie benötigt werden.
uf49c	Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.
uf50	Der Ablauf des Bestellvorgangs gefällt mir.
uf54c	Die Werbung auf den Seiten des Shops lenkt mich vom Einkaufen ab.
uf55	Die Startseite dieses Shops stellt einen guten Ausgangspunkt für meinen Einkauf dar.
uf56c	Bei der Benutzung des Shops treten störende technische Probleme auf.
uf57c	Ich vermisse wichtige Informationen, die ich für meine Kaufentscheidung benötige.

Anhang B

Ergänzende Darstellungen zur ufosV2a-Entwicklung

Tab. B-1.1: Modifikationen und Eliminationen der ufosV2-Items (ur1 bis ur9)	301
Tab. B-1.2: Modifikationen und Eliminationen der ufosV2-Items (uf1 bis uf13)	302
Tab. B-1.3: Modifikationen und Eliminationen der ufosV2-Items (uf14 bis uf30)	303
Tab. B-1.4: Modifikationen und Eliminationen der ufosV2-Items (uf31 bis uf48)	304
Tab. B-1.5: Modifikationen und Eliminationen der ufosV2-Items (uf49 bis uf58)	305
Tab. B-2: Quellen für neue formative Usability-Items der ufosV2fa-Skala	305

Tab. B-1.1: Modifikationen und Eliminationen der ufosV2-Items (Items u1 bis u17)

<i>Item-code</i>	<i>Item</i>
ur1	<i>Dieser Shop macht das Einkaufen im Netz leicht.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur1a: <i>Diese Website macht das Suchen nach Krankenversicherungsinformationen im Netz leicht.</i>
ur2	<i>Es ist mir zu kompliziert, diesen Shop zu benutzen.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur2a <i>Es ist mir zu kompliziert, diese Website zu nutzen (R).</i>
ur3	<i>Man kann sich schnell einen Überblick über den Shop verschaffen.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur3a <i>Man kann sich schnell einen Überblick über die Website verschaffen</i>
ur4	<i>Der Umgang mit dem Shop ist leicht zu erlernen.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur4a: <i>Der Umgang mit der Website ist leicht zu erlernen.</i>
ur6	<i>In diesem Shop kann ein Einkauf zügig erledigt werden.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur5a: <i>Auf dieser Website kann eine erfolgreiche Suche zügig erledigt werden.</i>
ur7	<i>Insgesamt bin ich mit der Benutzerfreundlichkeit dieses Shops zufrieden.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur4a: <i>Insgesamt bin ich mit der Benutzerfreundlichkeit dieser Website zufrieden.</i>
ur8	<i>Der Shop bietet alle Möglichkeiten, die ich mir von einem Online-Shop wünsche..</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur7a <i>Die Website bietet alle Möglichkeiten, die ich mir von einer Krankenversicherungswebsite wünsche.</i>
ur9	<i>Ich kann den Shop so benutzen, wie ich es erwarte.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → ur8a <i>Ich kann diese Website so benutzen, wie ich es erwarte.</i>

Tab. B-1.2: Modifikationen und Eliminationen der ufosV2-Items (Items uf1 bis uf13)

Item-code	Item
uf1	<p><i>Beim Einkauf in diesem Shop kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf1a: <i>Bei der Nutzung der Website kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.</i></p>
uf2	<p><i>Es sind nützliche Informationen zur Benutzung des Shops abrufbar.</i></p> <p>Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.</p>
uf3	<p><i>Bei auftretenden Problemen bietet der Shop von sich aus nützliche Hilfe an.</i></p> <p>Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.</p>
uf4	<p><i>Es ist in diesem Shop leicht, versehentlich durchgeführte Aktionen zu korrigieren.</i></p> <p>Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.</p>
uf5	<p><i>Ich muss mir viel merken, wenn ich in dem Shop einkaufe.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf2a: <i>Ich muss mir viel merken, wenn ich auf dieser Website etwas suche.</i></p>
uf6	<p><i>Es ist leicht, sich innerhalb des Shops zu bewegen.</i></p> <p>Elimination aufgrund inhaltlicher Redundanz mit Item uf7.</p>
uf7	<p><i>Es ist einfach, zwischen den verschiedenen Seiten des Shops hin- und herzuspringen.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf3a: <i>Es ist einfach, zwischen den verschiedenen Seiten hin und her zu springen.</i></p>
uf8	<p><i>Zentrale Seiten des Shops (z.B. die Startseite) sind jederzeit schnell zu erreichen.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf4a: <i>Zentrale Seiten (z.B. die Startseite) sind jederzeit schnell zu erreichen.</i></p>
uf9	<p><i>Ich kann jede Aktion wieder rückgängig machen.</i></p> <p>Unveränderte Übernahme des Items → uf5a: <i>Ich kann jede Aktion wieder rückgängig machen.</i></p>
uf10	<p><i>Mir ist beim Einkauf in dem Shop stets klar, welche Handlungen ich als nächstes durchführen kann.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf6a: <i>Mir ist bei der Nutzung dieser Website stets klar, welche Handlungen ich als nächstes durchführen kann.</i></p>
uf11	<p><i>Ich weiß manchmal nicht, wo ich mich innerhalb des Shops befinde.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf7a: <i>Ich weiß manchmal nicht, wo ich mich innerhalb der Website befinde.</i></p>
uf12	<p><i>Dieser Shop ist gut strukturiert.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf8a: <i>Diese Website ist gut strukturiert.</i></p>
uf13	<p><i>Es sind zu viele Schritte nötig, um zu einem Produkt zu gelangen.</i></p> <p>Konkrete Bezugnahme auf den Bewertungsgegenstand → uf9a: <i>Es sind zu viele Schritte nötig, um das Gesuchte zu finden.</i></p>

Tab. B-1.3: Modifikationen und Eliminationen der ufosV2-Items (Items uf14 bis uf30)

Item-Code	Item
uf14	<i>Ich finde mich auf den einzelnen Seiten des Shops gut zurecht.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf10a: <i>Ich finde mich auf den einzelnen Seiten gut zurecht.</i>
uf15	<i>Ich finde in diesem Shop schnell die Informationen, die für meinen Einkauf wichtig sind.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf11a: <i>Ich finde auf dieser Webseite schnell die Informationen, die für mich wichtig sind.</i>
uf16	<i>Es werden in diesem Online-Shop häufig zu viele Informationen dargestellt.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf12a: <i>Es werden auf dieser Website häufig zu viele Informationen dargestellt.</i>
uf17	<i>Ich kann selbst bestimmen, welche Informationen vom Shop angezeigt werden und welche nicht.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf18	<i>Die Texte sind manchmal nur schwer lesbar.</i> Unveränderte Übernahme des Items → uf13a: <i>Die Texte sind manchmal nur schwer lesbar.</i>
uf19	<i>Das optische Design des Shops stört mich bei meinem Einkauf.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf14a: <i>Das optische Design der Website stört mich bei der Nutzung.</i>
uf20	<i>Wenn ich in diesem Shop auf etwas klicke, passiert genau das, was ich erwarte.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf15a: <i>Wenn ich auf dieser Website auf etwas klicke, passiert genau das, was ich erwarte.</i>
uf21	<i>Manchmal finde ich die im Shop verwendeten Texte inhaltlich unverständlich.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf16a: <i>Manchmal finde ich die verwendeten Texte inhaltlich unverständlich.</i>
uf22	<i>Die Benennungen sind so gewählt, dass ich den Shop gut benutzen kann.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf17a: <i>Die Benennungen sind so gewählt, dass ich die Website gut benutzen kann.</i>
uf23	<i>Es ist schnell erkennbar, welche Produkte ich in dem Online-Shop kaufen kann und welche nicht.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf18a: <i>Es ist schnell erkennbar, welche Inhalte auf der Website angeboten werden.</i>
uf24	<i>Es ist leicht, gesuchte Produkte zu finden.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf19a: <i>Ich finde leicht, was ich suche.</i>
uf25	<i>Ich kann sehr gut auf die Suchfunktionen des Shops zugreifen.</i> Elimination, da das Item durch Experten überprüft werden kann.
uf26	<i>Die Suchfunktionen ermöglichen mir, schnell bestimmte Produkte zu finden.</i> Elimination aufgrund inhaltlicher Redundanz mit Item u25.
uf27	<i>Die Suchfunktionen liefern mir sinnvolle Ergebnisse.</i> Unveränderte Übernahme des Items → uf20a: <i>Die Suchfunktionen liefern mir sinnvolle Ergebnisse.</i>
uf28	<i>Die Produktlisten im Shop sind übersichtlich.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf29	<i>Die Produkte sind in sinnvoll gewählte Kategorien eingeteilt.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf21a: <i>Die Angebote sind in sinnvolle Kategorien eingeteilt.</i>
uf30	<i>Ein Vergleich verschiedener Produkte ist leicht.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.

Tab. B-1.4: Modifikationen und Eliminationen der ufosV2-Items (Items uf31 bis uf48)

Item-Code	Item
uf31	<i>Ich bin mit den Beschreibungen der Produkte zufrieden.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf22a <i>Ich bin mit den Beschreibungen der Angebote zufrieden.</i>
uf32	<i>Ich finde die verwendeten Bilder und Graphiken hilfreich.</i> Unveränderte Weiterverwendung → uf23a <i>Ich finde die verwendeten Bilder und Graphiken hilfreich.</i>
uf33	<i>Der Shop empfiehlt mir interessante Produkte, nach denen ich ursprünglich nicht gesucht habe.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf24a: <i>Die Website empfiehlt mir interessante Angebote, nach denen ich ursprünglich nicht gesucht habe.</i>
uf34	<i>Die Preise der einzelnen Produkte sind sofort ersichtlich.</i> Elimination aufgrund Ergebnisse in der Voruntersuchung.
uf35	<i>Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf36	<i>Der Shop bietet mir nützliche Informationen zu den Lieferbedingungen.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf37	<i>Die Informationen zur Rückgabe von Produkten sind verständlich.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf38	<i>Ich werde gut über die verschiedenen Zahlungsmöglichkeiten informiert.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf39	<i>Man muss lange nach Angaben zu den Lieferzeiten der Produkte suchen.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf40	<i>Es sind nützliche Informationen zu den allgemeinen Geschäftsbedingungen verfügbar.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf41	<i>Ich werde ausreichend darüber informiert, wie mit meinen persönlichen Daten umgegangen wird.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf42	<i>Der Warenkorb ist übersichtlich.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf43	<i>Ich weiß in diesem Shop stets, welche Produkte sich gerade im Warenkorb befinden.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf44	<i>Es ist in diesem Shop ganz einfach, Produkte in den Warenkorb zu legen.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf45	<i>Der Inhalt des Warenkorbes kann ganz einfach verändert werden.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf46	<i>Der Bestellvorgang kann problemlos durchlaufen werden.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf47	<i>Die Eingabe meiner persönlichen Daten (z.B. Anschrift, Kreditkartennummer) während des Bestellvorgangs ist unkompliziert.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf48	<i>Während des Bestellvorgangs weiß ich bei einigen persönlichen Angaben nicht, wofür sie benötigt werden.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.

Tab. B-1.5: Modifikationen und Eliminationen der ufosV2-Items (Items uf49 bis uf58)

<i>Item-Code</i>	<i>Item</i>
uf49	<i>Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf25a: <i>Es dauert zu lange, bis die Website auf meine Eingaben reagiert.</i>
uf50	<i>Der Ablauf des Bestellvorgangs gefällt mir.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf51	<i>Die Auflistungen der Produkte lassen sich bequem nach meinen Bedürfnissen ordnen.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf52	<i>Ich erhalte nützliche Informationen zum Unternehmen, das den Shop betreibt.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf31a: <i>Ich erhalte nützliche Informationen zum Unternehmen, das die Website betreibt.</i>
uf53	<i>Ich bin mit den Informationen zur Sicherheit des Datentransfers zufrieden.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf54c	<i>Die Werbung auf den Seiten des Shops lenkt mich vom Einkaufen ab.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf55	<i>Die Startseite dieses Shops stellt einen guten Ausgangspunkt für meinen Einkauf dar.</i> Konkrete Bezugnahme auf den Bewertungsgegenstand → uf26a: <i>Die Startseite stellt einen guten Ausgangspunkt für meine Nutzung der Website dar.</i>
uf56c	<i>Bei der Benutzung des Shops treten störende technische Probleme auf.</i> Elimination, da das Item durch Experten überprüft werden kann.
uf57c	<i>Ich vermisse wichtige Informationen, die ich für meine Kaufentscheidung benötige.</i> Elimination aufgrund mangelnder inhaltlicher Verwendbarkeit.
uf58c	<i>Die Ladezeiten der Seiten sind bei diesem Online-Shop zu lang.</i> Elimination, da das Item durch Experten überprüft werden kann.

Tab. B-2: Quellen für neue formative Usability-Items der ufosV2fa-Skala

<i>Neuer Item Code</i>	<i>ufos-Item (Quelle)</i>
uf27a	Die Themenlisten auf der Website sind übersichtlich. (vgl. Konradt et al., 2003)
uf28a	Insgesamt verfügt die Website über ein reichhaltiges Informationsangebot. (vgl. Chen, Gillenson & Sherell, 2004)
uf29a	Ich bin mit den Informationen, die mir der Anbieter zur Verfügung stellt, sehr zufrieden. (vgl. Lowengart & Tractinsky, 2002)
uf30a	Diese Website ist eine Quelle hochwertiger Angebote. (vgl. Chen, Gillenson & Sherell, 2004)

Anhang C

Ergänzende Darstellungen zur Untersuchungsplanung

Tab. C-1: Items zur reflektiven Erfassung des Konstruktes Vertrauen.....	309
Tab. C-2: Items zur reflektiven Erfassung des Konstruktes Nutzungsvergnügen.....	309
Tab. C-3: Items zur reflektiven Erfassung des Konstruktes Ästhetik	309
Tab. C-4: Items zur Erfassung des Konstruktes Involvement (Skala „Importance“ von McQuarrie & Munson, 1992).....	310
Tab. C-5: Items zur reflektiven Erfassung des Konstruktes Reputation	310
Tab. C-6: Item zur Erfassung des Konstruktes Bekanntheit	310
Tab. C-7: Item zur Erfassung des Konstruktes wahrgenommener Preis.....	310
Tab. C-8: Items zur reflektiven Erfassung des Konstruktes Nutzerzufriedenheit.....	310
Tab. C-9: Items zur reflektiven Erfassung des Konstruktes Nutzungsintention	311

Tab. C-1: Items zur reflektiven Erfassung des Konstruktes Vertrauen

<i>Neuer Item Code</i>	<i>Verwendetes Item zum Vertrauen Originalitem (Quelle)</i>
v1	Dieser Anbieter ist vertrauenswürdig. (Aktuelle Untersuchung) This store is trustworthy. (Jarvenpaa et al., 2000) Dieser Online-Shop ist vertrauenswürdig. (Christophersen, 2007)
v2	Ich vertraue darauf, dass dieser Anbieter meine Interessen berücksichtigt. (Aktuelle Untersuchung) I trust this store keeps my best interests in mind. (Jarvenpaa et al., 2000) Ich vertraue darauf, dass dieser Shop meine Interessen berücksichtigt. (Christophersen, 2007)
v3	Dieser Anbieter hält seine Versprechen und Verpflichtungen ein. (Aktuelle Untersuchung) This store wants to be known as one who keeps promises and commitments. (Jarvenpaa et al., 2000) Ich bin mir sicher, dass der Online-Shop seine Versprechen und Verpflichtungen einhalten wird. (Christophersen, 2007)
v4	Ich vertraue in die Informationen, die dieser Anbieter mir zur Verfügung stellt. (Aktuelle Untersuchung) I trust in the information of this seller. (Hampton-Sosa, 2002)

Tab. C-2: Items zur reflektiven Erfassung des Konstruktes Nutzungsvergnügen

<i>Neuer Item Code</i>	<i>Verwendetes Item zum Nutzungsvergnügen Originalitem (Quelle)</i>
vg1	Die Website ist unterhaltsam. (Aktuelle Untersuchung) I found my visit to this web site enjoyable. (Koufaris und Hampton-Sosa, 2002)
vg2	Die Website zu nutzen ist reizvoll. (Aktuelle Untersuchung) I found my visit to this web site exciting. (Koufaris und Hampton-Sosa, 2002)
vg3	Es macht wirklich Spaß, diese Website zu nutzen. (Aktuelle Untersuchung) It's really fun to shop at this website. (Wolfenbarger & Gilly, 2003) Es macht wirklich Spaß, bei diesem Online-Shop einzukaufen. (Christophersen, 2007)

Tab. C-3: Items zur reflektiven Erfassung des Konstruktes Ästhetik

<i>Neuer Item Code</i>	<i>Verwendetes Item zur Ästhetik Originalitem (Quelle)</i>
ae1	Ich finde es ansprechend, wie diese Website ihre Informationen darbietet. (Aktuelle Untersuchung) XYZ's Internet site is aesthetically appealing. (Mathwick, Malhotra & Rigdon, 2001) Ich finde es optisch ansprechend, wie dieser Online-Shop seine Produkte anbietet. (Christophersen, 2007)
ae2	Die Website ist ästhetisch ansprechend gestaltet. (Aktuelle Untersuchung) The Way XYZ displays its products is attractive. (Mathwick, Malhotra & Rigdon, 2001) Der Internetauftritt dieses Online-Shops ist ästhetisch gestaltet. (Christophersen, 2007)

Tab. C-4: Items zur Erfassung des Produkt Involvement
(Skala „Importance“ von McQuarrie & Munson, 1992)

Important	Unimportant
Irrelevant	Relevant
Matters to me	Doesn't matter
Means a lot to me	Means nothing to me
Of no concern to me	Of concern to me

Tab. C-5: Items zur reflektiven Erfassung des Konstruktes Reputation

<i>Neuer Item Code</i>	<i>Verwendetes Item zur Reputation Originalitem (Quelle)</i>
r1	Dieser Anbieter hat einen guten Ruf bei den Kunden. (Aktuelle Untersuchung) This Web retailer has a bad reputation in the market. (Pavlou, 2003)
r2	Dieser Anbieter ist dafür bekannt, zuverlässig zu sein. (Aktuelle Untersuchung) This Web retailer is known to be dependable. (Pavlou, 2003)

Tab. C-6: Item zur Erfassung des Konstruktes Bekanntheit

<i>Neuer Item Code</i>	<i>Verwendetes Item zur Bekanntheit Originalitem (Quelle)</i>
b1	Dieser Anbieter ist allgemein sehr bekannt. (Aktuelle Untersuchung) This store ist wellknown. (Jarvenpaa et al., 2000)

Tab. C-7: Item zur Erfassung des Konstruktes Wahrgenommener Preis

<i>Neuer Item Code</i>	<i>Verwendetes Item zum Wahrgenommenen Preis Originalitem (Quelle)</i>
subpc	Der Beitragssatz dieser Krankenversicherung ist mir zu hoch. (Aktuelle Untersuchung) The prices of the product(s) I purchased from XYZ's Internet Site are too high, given the quality of the merchandise. (Mathwick, Malhotra & Rigdon, 2001) Die Preise in diesem Shop sind mir zu hoch. (Christophersen, 2007)

Tab. C-8: Items zur reflektiven Erfassung des Konstruktes Nutzerzufriedenheit

<i>Neuer Item Code</i>	<i>Verwendetes Item zur Nutzerzufriedenheit Originalitem (Quelle)</i>
z1	Insgesamt: Wie zufrieden sind Sie mit dieser Website? (Aktuelle Untersuchung) Overall, how satisfied are you with ... (Wixom & Todd, 2005)
z2	Mit der Nutzung dieser Website bin ich zufrieden. (Aktuelle Untersuchung) All things considered, I am very satisfied with ... (Wixom & Todd, 2005)

Tab. C-9: Items zur reflektiven Erfassung des Konstruktes Nutzungsintention

<i>Neuer Item Code</i>	<i>Verwendetes Item zur Nutzungsintention Originalitem (Quelle)</i>
ni1	Wenn ich eine neue Krankenkasse suche, werde ich diese Website wieder aufsuchen. (Aktuelle Untersuchung) How likely is it that you would return to this store's web site? (Jarvenpaa et al., 2000) Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen. (Christophersen, 2007)
ni2	Wenn ich mich über Krankenkassen informieren möchte, werde ich diese Website nutzen. (Aktuelle Untersuchung) How likely is it that you would return to this store's web site? (Jarvenpaa et al., 2000) Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen. (Christophersen, 2007)
ni3	Wenn ich wieder einmal Informationen zum Thema Krankenversicherung benötige, werde ich diese Website nutzen. (Aktuelle Untersuchung). How likely is it that you would return to this store's web site? (Jarvenpaa et al., 2000) Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen (Christophersen, 2007)
ni4	Ich werde diese Website nach dem Versuch vermutlich wieder aufsuchen. (Aktuelle Untersuchung) How likely is it that you would return to this store's web site? (Jarvenpaa et al., 2000) Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen. (Christophersen, 2007)

Anhang D

Ergänzende Darstellungen zur Durchführung

Verwendete Werbemittel

Abb. D-1: Aushang zur Untersuchung 315

Screenshots zentraler Seiten der Hypertext-Versuchsumgebung

Abb. D-2.1: Screenshot einer Seite der Hypertext Versuchsumgebung Begrüssung..... 316

Abb. D-2.2: Hypertext-Versuchsumgebung Fragen zur Person 316

Abb. D-2.3: Hypertext-Versuchsumgebung Fragen zur Computererfahrung 317

Abb. D-2.4: Hypertext-Versuchsumgebung Fragen zur Bekanntheit des Anbieters 317

Abb. D-2.5: Hypertext-Versuchsumgebung 1. Website, Aufgabe 1 318

Abb. D-2.6: Hypertext-Versuchsumgebung 1. Website, Aufgabe 2..... 318

Abb. D-2.7: Hypertext-Versuchsumgebung 1. Website, Aufgabe 3 319

Abb. D-2.8: Hypertext-Versuchsumgebung 1. Website, Aufgabe 4..... 319

Abb. D-2.9: Hypertext-Versuchsumgebung 1. Website, Aufgabe 5..... 320

Abb. D-2.10: Hypertext-Versuchsumgebung Bewertung der 1. Website (Teil 1)..... 320

Abb. D-2.11: Hypertext-Versuchsumgebung Abfrage des Involvement..... 321

Abb. D-2.12: Hypertext-Versuchsumgebung Abschließende Auswahlentscheidung 321

Informationen zum Versuchsablauf

Teilnehmerinstruktionen der Hauptuntersuchung..... 322

Christian-Albrechts-Universität zu Kiel



15,- Euro

1,5 Versuchspersonenstunden

Studie zum Einkaufen im Internet !!

- Im Rahmen der ca. einstündigen Online-Untersuchung wirst Du zwei Websites von Unternehmen aus dem Bereich der Krankenversicherung kennen lernen und anhand eines Fragebogens bewerten.
- ✓ **Als Belohnung für die Teilnahme an der Untersuchung erhältst Du von uns 15 Euro.**
 - ✓ Teilnehmen können Studenten aus allen Fachrichtungen.
 - ✓ Anstelle der 15,- Euro können Studenten i. R. ihrer Studienordnung **1,5 Versuchspersonenstunden** angerechnet bekommen.
 - ✓ Die Untersuchung erfolgt in einer onlinegestützten Versuchsumgebung und kann bequem am PC online durchgeführt werden.
- **Wie geht es weiter? – Ganz einfach!**

Bitte unter folgender Emailadresse melden, damit wir Dir einen Zugangscode für die Online-Studie zukommen lassen können.

guntherheld@hamburg.de (Betreff: Onlinestudie)

Forschungsgruppe „Einkaufen im Internet“ an der Christian-Albrechts-Universität zu Kiel, Dipl. Wirtschaftsingenieur Gunther Held

P.s.: Die Studie verfolgt rein wissenschaftliche Zwecke und wird im Rahmen einer Promotion durchgeführt. Es besteht kein Kontakt zu den untersuchten Unternehmen. Doktorvater ist Prof. Dr. Konradt, Lehrstuhlinhaber für Arbeits , Organisations & Marktpsychologie an der Christian Albrechts Universität zu Kiel. Die Teilnehmerdaten werden absolut vertraulich behandelt, es erfolgt keine Weitergabe persönlicher Daten an Dritte.

Abb. D-1: Aushang zur Untersuchung

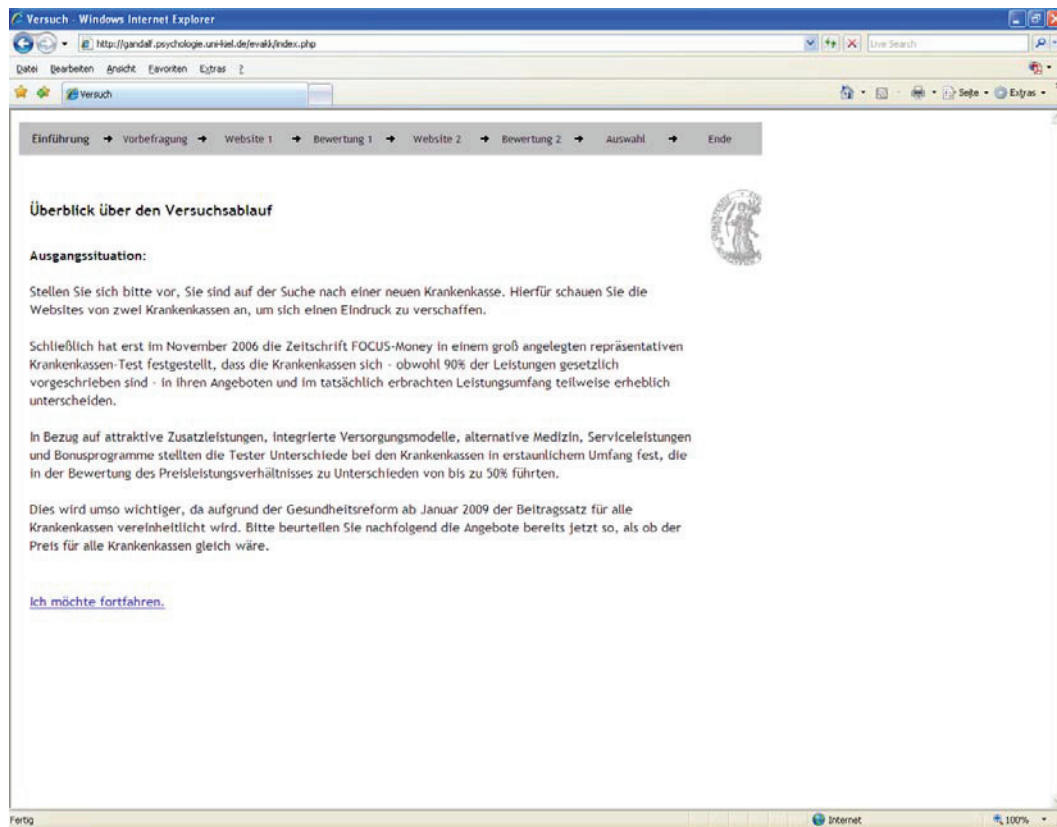


Abb. D-2.1: Screenshot einer Seite der Hypertext Versuchsumgebung

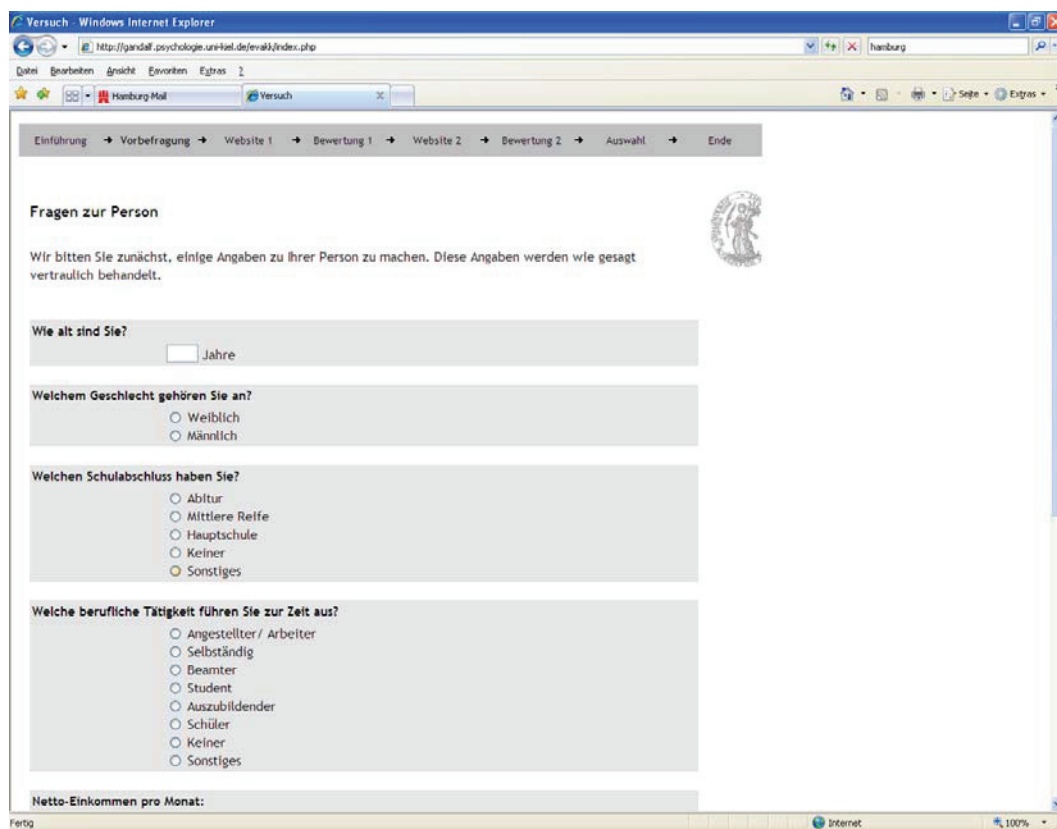


Abb. D-2.2: Hypertext-Versuchsumgebung Fragen zur Person

The screenshot shows a web browser window titled 'Versuch - Windows Internet Explorer'. The address bar contains 'http://gandalf.psychologie.uni-hiel.de/evak3/index.php'. The browser has two tabs: 'Hamburg-Mail' and 'Versuch'. The page content includes a breadcrumb trail: 'Einführung → Vorbefragung → Website 1 → Bewertung 1 → Website 2 → Bewertung 2 → Auswahl → Ende'. The main heading is 'Fragen zu Ihrer Computererfahrung'. Below the heading is a paragraph: 'Bitte machen Sie nun einige Angaben zu Ihrer Erfahrung mit Computern und dem Internet.' To the right of the text is a small circular logo. The survey consists of several questions, each with a text input field and a radio button:

- Seit wie vielen Jahren nutzen Sie einen PC?
ungefähr Jahre
- Wie häufig nutzen Sie einen PC momentan?
ungefähr Tag(e) im Monat
und ungefähr Stunde(n) pro Tag
- Seit wie vielen Jahren nutzen Sie das Internet?
ungefähr Jahre
- Wie häufig nutzen Sie das Internet momentan?
ungefähr Tag(e) im Monat
und ungefähr Stunde(n) pro Tag
- Wie oft haben Sie während des letzten Jahres im Internet eingekauft?
(Bitte beziehen Sie Online-Auktionen wie Ebay mit ein.)
ungefähr mal
- Wie oft nutzen Sie das Internet für Informationsrecherchen?
ungefähr mal im Monat

At the bottom of the form, there is a link: 'Ich möchte jetzt mit dem Versuch fortfahren.'

Abb. D-2.3: Hypertext-Versuchsumgebung Fragen zur Computererfahrung

The screenshot shows a web browser window titled 'Versuch - Windows Internet Explorer'. The address bar contains 'http://gandalf.psychologie.uni-hiel.de/evak3/index.php'. The browser has two tabs: 'Versuch' and 'Live Search'. The page content includes a breadcrumb trail: 'Einführung → Vorbefragung → Website 1 → Bewertung 1 → Website 2 → Bewertung 2 → Auswahl → Ende'. The main heading is 'Bekanntheit des Anbieters'. Below the heading is a paragraph: 'Die erste Website, mit der Sie sich beschäftigen werden, wird die Website der "DAK" sein. Bevor Sie mit der ersten Aufgabe starten machen Sie bitte noch 3 Angaben zum Bekanntheitsgrad des Anbieters der Website. Nutzen Sie zur Bewertung die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu". Falls eine Aussage Ihrer Meinung nach nicht bewertet werden kann, so wählen Sie die Antwortmöglichkeit "keine Angabe".' To the right of the text is a small circular logo. The survey consists of three questions, each with a five-point Likert scale and a 'Keine Angabe' radio button:

- Dieser Anbieter ist allgemein sehr bekannt.
Stimme gar nicht zu Stimme voll zu Keine Angabe
- Dieser Anbieter ist dafür bekannt, zuverlässig zu sein.
Stimme gar nicht zu Stimme voll zu Keine Angabe
- Dieser Anbieter hat einen guten Ruf bei den Kunden.
Stimme gar nicht zu Stimme voll zu Keine Angabe

At the bottom of the form, there is a link: 'Ich möchte jetzt mit dem Versuch fortfahren.'

Abb. D-2.4: Hypertext-Versuchsumgebung Fragen zur Bekanntheit des Anbieters

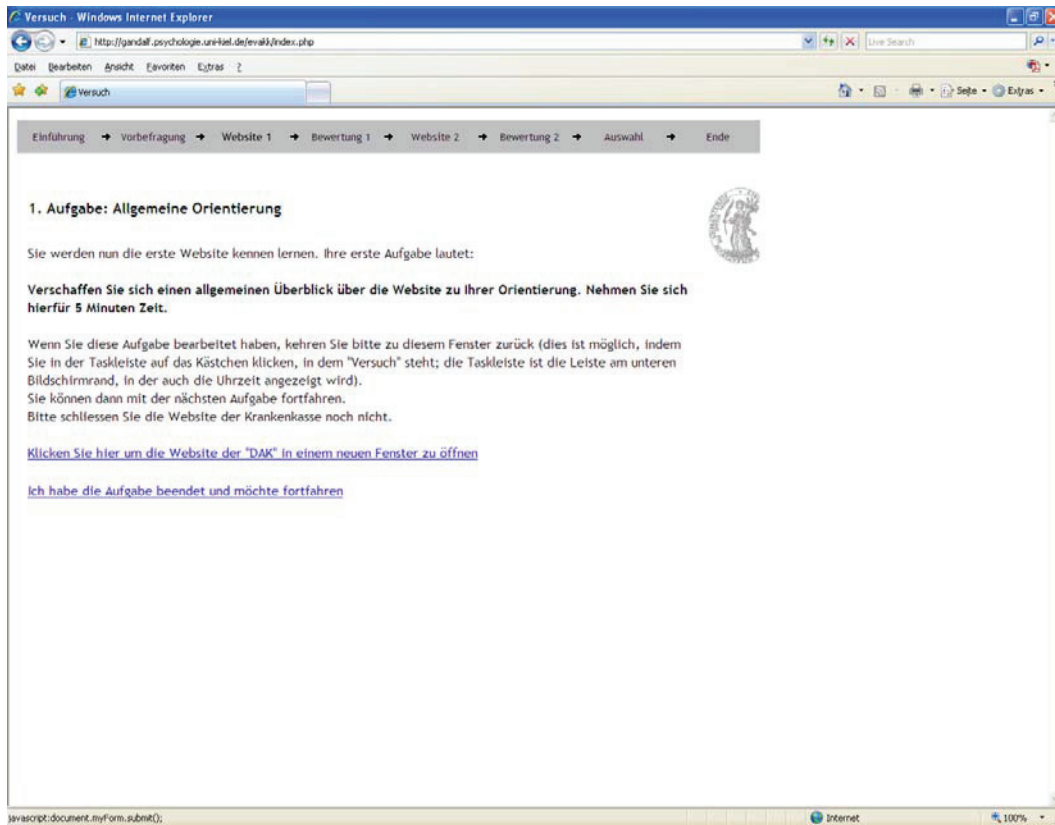


Abb. D-2.5: Hypertext-Versuchsumgebung 1. Website, Aufgabe 1



Abb. D-2.6 Hypertext-Versuchsumgebung 1. Website, Aufgabe 2

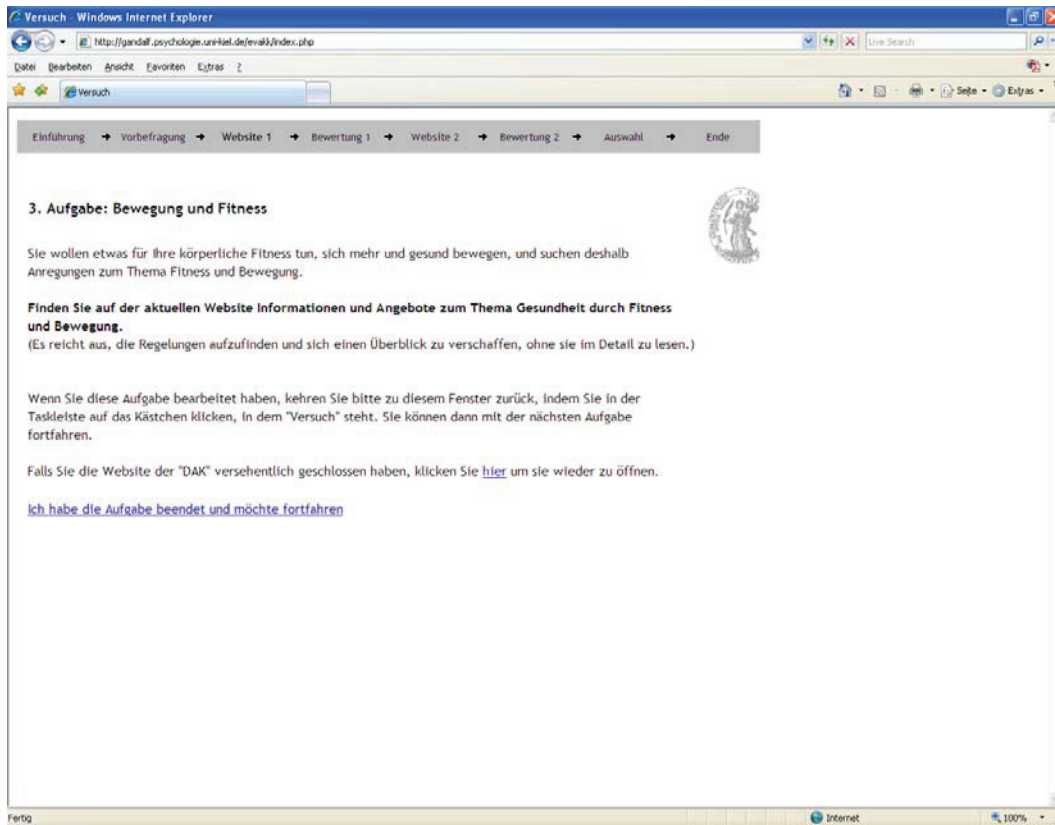


Abb. D-2.7: Hypertext-Versuchsumgebung 1. Website, Aufgabe 3

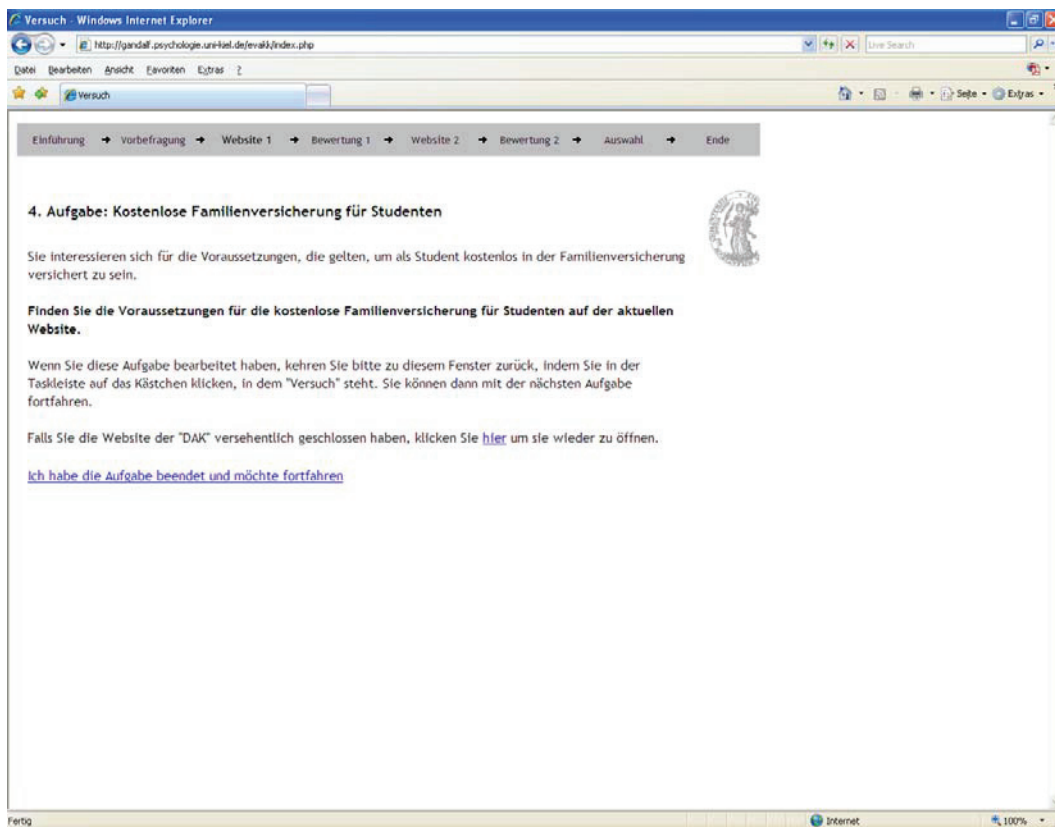


Abb. D-2.8: Hypertext-Versuchsumgebung 1. Website, Aufgabe 4

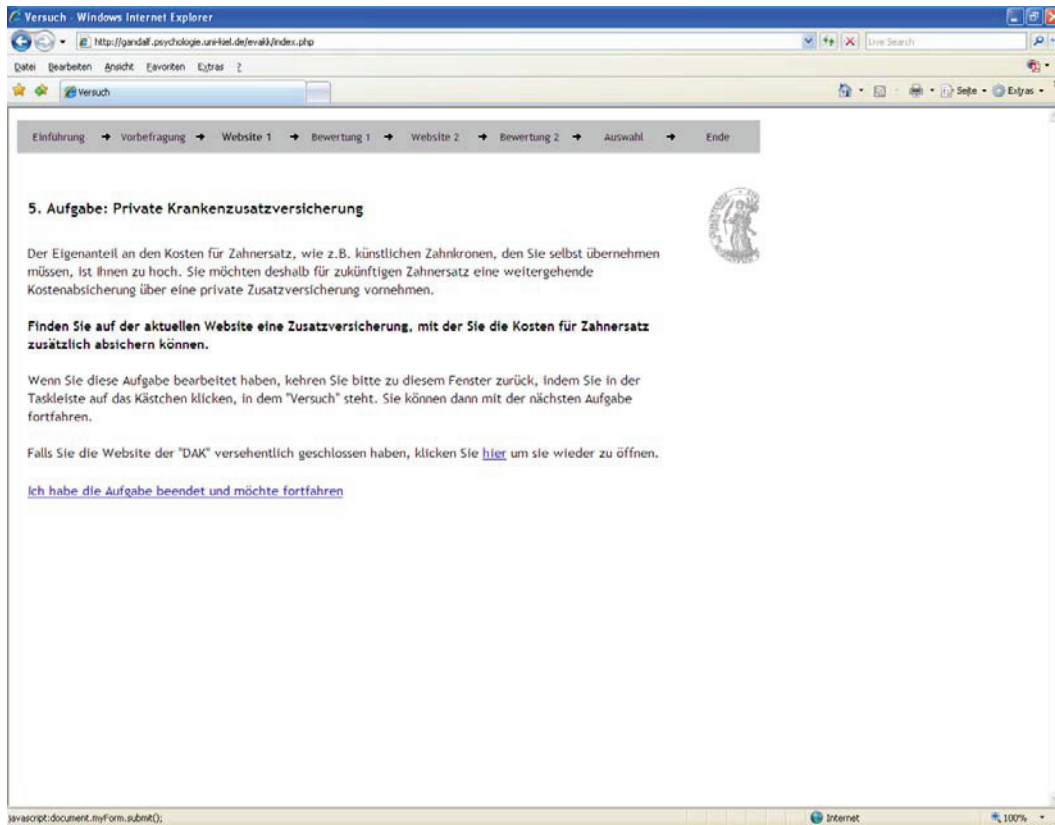


Abb. D-2.9: Hypertext-Versuchsumgebung 1. Website, Aufgabe 5

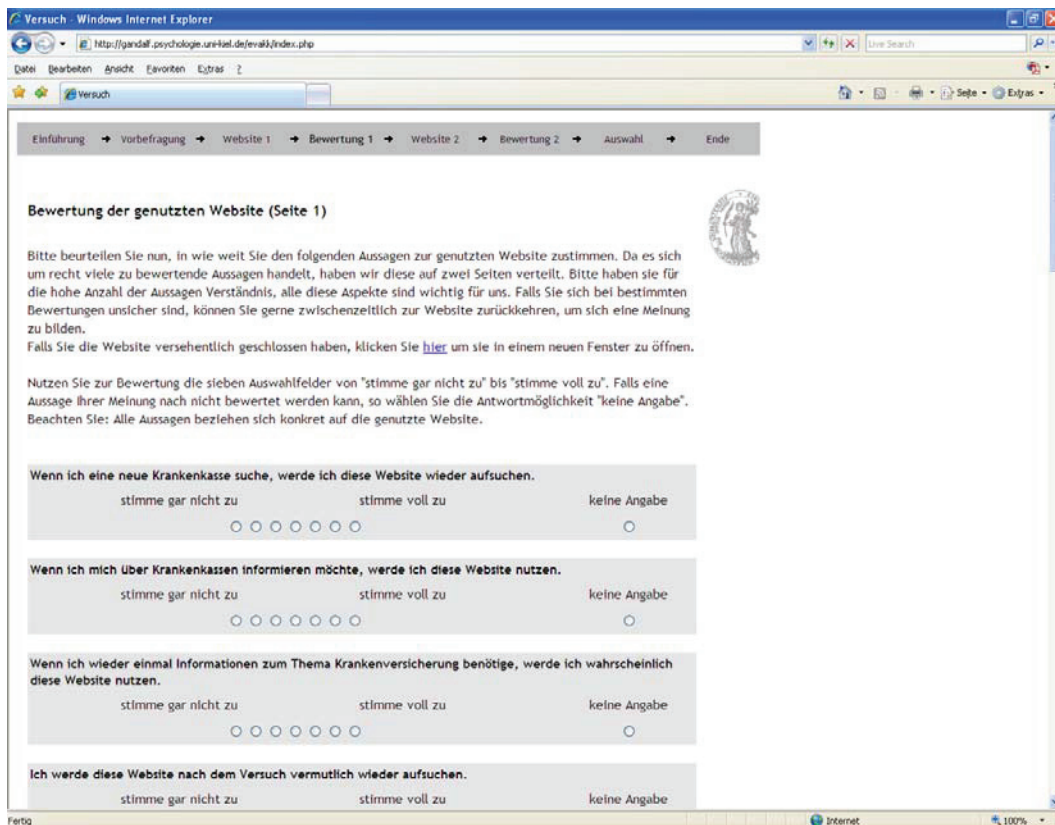


Abb. D-2.10: Hypertext-Versuchsumgebung Bewertung der 1. Website (Teil 1)

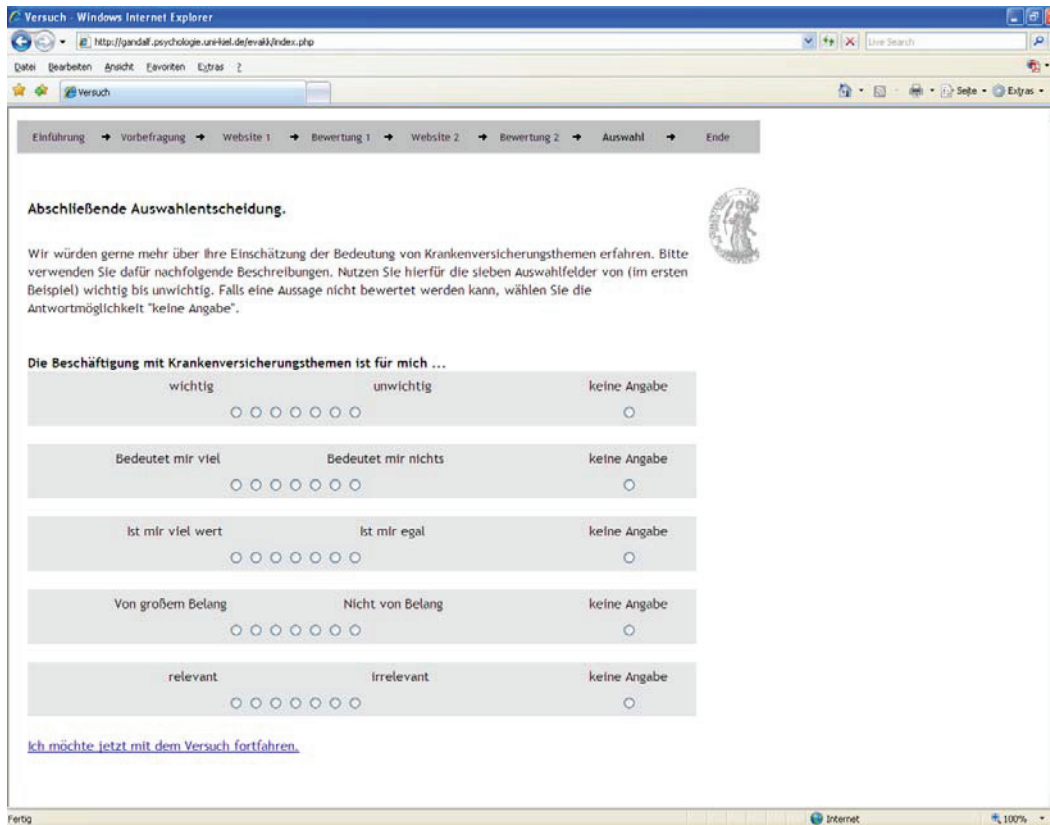


Abb. D-2.11: Hypertext-Versuchsumgebung Abfrage des Involvement

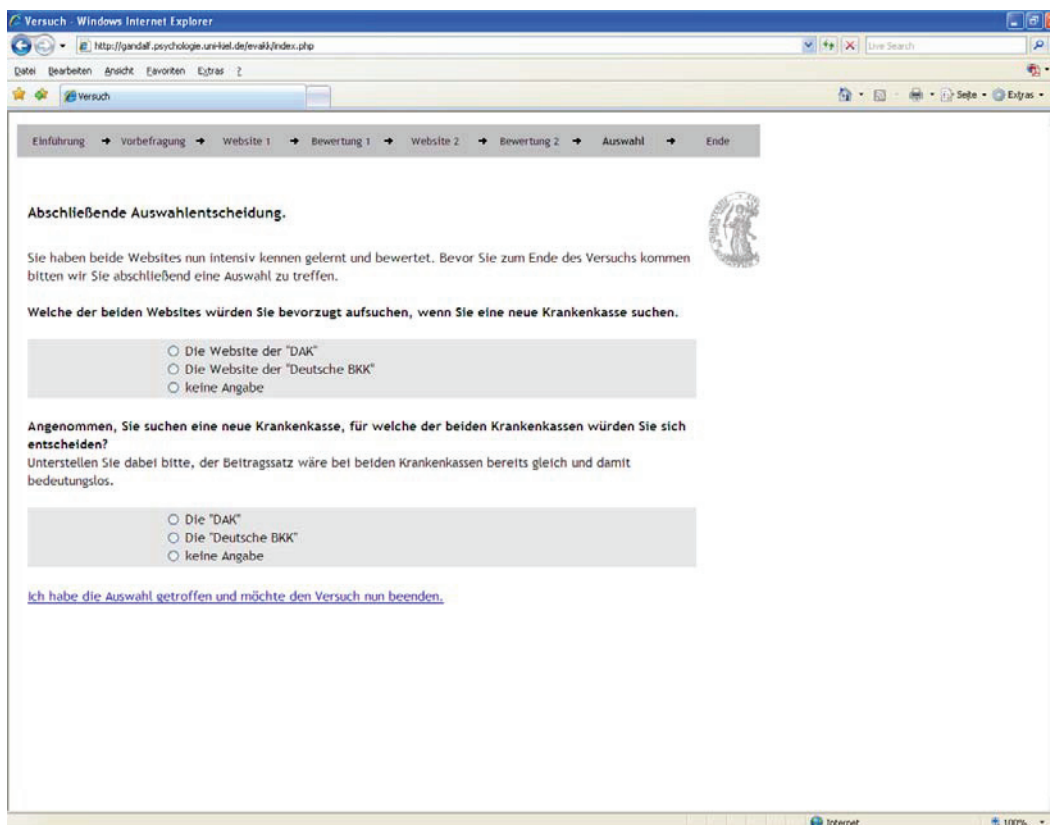


Abb. D-2.12: Hypertext-Versuchsumgebung Abschließende Auswahlentscheidung

Teilnehmerinstruktionen der Hauptuntersuchung

Seite 1 (Phase Einführung)

Überblick über den Versuchsablauf (1):

Liebe(r) Versuchsteilnehmer(in),

vielen Dank, dass Sie an diesem Versuch zum Einkaufen im Internet teilnehmen.

Sie werden sich während dieser Untersuchung die Websites von zwei Krankenkassen anschauen. Zu jeder Website erhalten Sie fünf Aufgaben. Dadurch werden Sie die Website genauer kennen lernen. Im Anschluss an die Bearbeitung der Aufgaben bitten wir Sie, jede Website durch die Beantwortung eines Fragebogens zu bewerten. Insgesamt wird der Versuch ca. 60 min. dauern.

Im Rahmen des Versuchs werden nicht Sie, sondern die Websites getestet. Sie können also nichts "falsch machen". Außerdem werden natürlich alle Angaben vertraulich behandelt.

[Link zu Seite 2:] [Ich möchte fortfahren](#)

Seite 2 (Phase Einführung)

Überblick über den Versuchsablauf (2):

Ausgangssituation:

Stellen Sie sich bitte vor, Sie sind auf der Suche nach einer neuen Krankenkasse. Hierfür schauen Sie sich die Websites von zwei Krankenkassen an, um sich einen Eindruck zu verschaffen. Da aufgrund der Gesundheitsreform ab Januar 2009 der Beitragssatz für alle Krankenkassen vereinheitlicht wird, beurteilen Sie die Angebote bitte bereits jetzt so, als ob der Preis für alle Krankenkassen gleich wäre.

[Link zu Seite 3:] [Ich möchte fortfahren.](#)

Seite 3 (Phase Einführung)

Überblick über den Versuchsablauf (3):

Der Versuchsablauf ist für die beiden Websites derselbe. Sie werden bei beiden Websites gebeten, die gleichen Aufgaben zu erledigen und den gleichen Fragebogen zu beantworten.

Noch einige wichtige Anmerkungen:

- Die Erledigung der Aufgaben kann aus technischen Gründen nicht immer in dem selben Fenster erfolgen. Daher wird zeitweise mehr als ein Fenster geöffnet sein. Sie werden darauf hingewiesen, wenn ein neues Fenster geöffnet wird.

- Bitte beachten Sie während des gesamten Versuchs, dass Sie keines der Fenster schließen.

Zu Klären: Hinweis wie man sich verhält, wenn man eine Aufgabe vorzeitig abbricht und zur Aufgabe zurückkehren möchte:

[Link zu Seite 4:] [Ich möchte fortfahren.](#)

Seite 4 (Phase Vorbefragung)

Fragen zu Ihrer Person

Wir bitten Sie zunächst, einige Angaben zu Ihrer Person zu machen. Diese Angaben werden wie gesagt vertraulich behandelt.

Alter: [freies Eingabefeld] Jahre

Geschlecht: weiblich, männlich

Schulabschluss: Abitur, mittlere Reife, Hauptschule, keiner, Sonstiges

Berufliche Tätigkeit: Angestellter/ Arbeiter, Selbständig, Beamter, Student, Auszubildender, Schüler, Keiner, Sonstiges

Netto-Einkommen pro Monat: bis 300 Euro, 300 bis 750 Euro, 750 bis 1500 Euro, 1500 bis 2000 Euro, 2000 bis 2500 Euro, mehr als 2500 Euro

Haushaltsgröße: [freies Eingabefeld] Personen

Aktuell bin ich bei folgender Krankenversicherung versichert: [freies Eingabefeld]

Ist Deutsch Ihre Muttersprache? ja (in diesem Fall können Sie die nächste Frage überspringen); nein

Wie gut können Sie Deutsch verstehen? (siebenstufige Likert-Skala von „sehr schlecht“ bis „sehr gut“)

[Link zu Seite 5:] Ich habe alle Angaben gemacht und möchte mit dem Versuch fortfahren.

Seite 5 (Phase Vorbefragung)

Fragen zu Ihrer Computererfahrung

Bitte machen Sie nun einige Angaben zu Ihrer Erfahrung mit Computern und dem Internet.

Seit wie vielen Jahren nutzen Sie einen PC? ungefähr [freies Eingabefeld] Jahre

Wie häufig nutzen Sie einen PC momentan? ungefähr [freies Eingabefeld] Tag(e) im Monat und ungefähr [freies Eingabefeld] Stunde(n) pro Tag

Seit wie vielen Jahren nutzen Sie das Internet? ungefähr [freies Eingabefeld] Jahre

Wie häufig nutzen Sie das Internet momentan? ungefähr [freies Eingabefeld] Tag(e) im Monat und ungefähr [freies Eingabefeld] Stunde(n) pro Tag

Wie oft haben Sie während des letzten Jahres im Internet eingekauft? (Bitte beziehen Sie Online-Auktionen wie Ebay mit ein.) ungefähr [freies Eingabefeld] mal

Wie oft nutzen Sie das Internet für Informationsrecherchen? ungefähr [freies Eingabefeld] mal im Monat.

[Link zu Seite 6:] Ich habe alle Fragen beantwortet und möchte mit dem Versuch fortfahren.

Seite 6 (Phase Website 1)

Bekanntheit des Anbieters

Die erste Website, mit der Sie sich beschäftigen werden, wird die Website der [Name der 1. Krankenkasse wird angeführt] sein. Bevor Sie mit der ersten Aufgabe starten machen Sie bitte noch 3 Angaben zum Bekanntheitsgrad des Anbieters der Website.

Nutzen Sie zur Bewertung die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu". Falls eine Aussage Ihrer Meinung nach nicht bewertet werden kann, so wählen Sie die Antwortmöglichkeit "keine Angabe".

Dieser Anbieter ist allgemein sehr bekannt.

stimme gar nicht zu						stimme voll zu	keine Angabe

Dieser Anbieter ist dafür bekannt, zuverlässig zu sein.

stimme gar nicht zu						stimme voll zu	keine Angabe

Dieser Anbieter hat einen guten Ruf bei den Kunden.

stimme gar nicht zu						stimme voll zu	keine Angabe

[Link zu Seite 7:] Ich habe die Angaben gemacht und möchte mit dem Versuch fortfahren.

Seite 7 (Phase Website 1)

1. Aufgabe: Allgemeine Orientierung

Sie werden nun die erste Website kennen lernen. Ihre erste Aufgabe lautet:

Verschaffen Sie sich einen allgemeinen Überblick über die Website zu Ihrer Orientierung. Nehmen Sie sich hierfür 5 Minuten Zeit.

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte zum aktuellen Fenster zurück (dies ist möglich, indem Sie in der Taskleiste auf das Kästchen klicken, in dem "Versuch" steht; die Taskleiste ist die Leiste am unteren Bildschirmrand, in der auch die Uhrzeit angezeigt wird). Sie können dann mit der nächsten Aufgabe fortfahren.

[Link zu Website 1:] Ich möchte jetzt mit dem Versuch fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 8 (Phase Website 1)

2. Aufgabe: Zahnersatz

In der folgenden Aufgabe geht es um Zahnersatz. Ihr Zahnarzt weist Sie darauf hin, dass Sie eine künstliche Zahnkrone benötigen. Sie möchten sich darüber informieren, inwieweit die Krankenkasse die Kosten für die kostspielige Behandlung übernimmt.

Suchen Sie die Regelungen über die Zuschüsse der Krankenkasse zum Zahnersatz auf. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.)

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte zum aktuellen Fenster zurück, indem Sie in der Taskleiste auf das Kästchen klicken, in dem "Versuch" steht. Sie können dann mit der nächsten Aufgabe fortfahren.

[Link zu Website 1:] Ich möchte jetzt mit dem Versuch fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 9 (Phase Website 1)

3. Aufgabe: Bewegung und Fitness

Sie wollen etwas für Ihre körperliche Fitness tun, sich mehr und gesund bewegen, und suchen deshalb Anregungen zum Thema Fitness und Bewegung.

Finden Sie auf der aktuellen Website Informationen und Angebote zum Thema Gesundheit durch Fitness und Bewegung. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.) Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück.

[Link zu Website 1:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 10 (Phase Website 1)

4. Aufgabe: Kostenlose Familienversicherung für Studenten

Sie interessieren sich für die Voraussetzungen, die gelten, um als Student kostenlos in der Familienversicherung versichert zu sein.

Finden Sie die Voraussetzungen für die kostenlose Familienversicherung für Studenten auf der aktuellen Website.

Wenn Sie die Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück.

[Link zu Website 1:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 11 (Phase Website 1)

5. Aufgabe: Private Krankenzusatzversicherung

Der Eigenanteil an den Kosten für Zahnersatz, wie z.B. künstlichen Zahnkronen, den Sie selbst übernehmen müssen, ist Ihnen zu hoch. Sie möchten deshalb für zukünftigen **Zahnersatz** Zahnersatz eine weitergehende Kostenabsicherung über eine private Zusatzversicherung vornehmen.

Finden Sie auf der aktuellen Website eine Zusatzversicherung, mit der Sie die Kosten für Zahnersatz zusätzlich absichern können.

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen **"Versuch"** „Versuch“ in der Taskleiste zum aktuellen Fenster zurück. Sie werden dann mit der Bewertung der Website beginnen.

[Link zu Website 1:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 12 (Phase Bewertung 1)

Bewertung der genutzten Website (Seite 1)

Bitte beurteilen Sie nun, in wie weit Sie den folgenden Aussagen zur genutzten Website zustimmen. Da es sich um recht viele zu bewertende Aussagen handelt, haben wir diese auf zwei Seiten verteilt. Bitte haben sie für die hohe Anzahl der Aussagen Verständnis, alle diese Aspekte sind wichtig für uns. Falls Sie sich bei bestimmten Bewertungen unsicher sind, können Sie gerne zwischenzeitlich zur Website zurückkehren, um sich eine Meinung zu bilden.

Nutzen Sie zur Bewertung die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu". Falls eine Aussage Ihrer Meinung nach nicht bewertet werden kann, so wählen Sie die Antwortmöglichkeit "keine Angabe". Beachten Sie: Alle Aussagen beziehen sich konkret auf die genutzte Website.

[Link zu Seite 13:] Ich möchte meine Bewertung fortsetzen.

Seite 13 (Phase Bewertung 1)

Bewertung der genutzten Website (Seite 2)

Bitte setzen Sie Ihre Bewertung fort. Nutzen Sie dazu erneut die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu" oder das Feld "keine Angabe". Beachten Sie wiederum: Alle Aussagen beziehen sich konkret auf die genutzte Website.

[Link zu Seite 14:] Ich habe die Bewertung abgeschlossen und möchte den Versuch fortsetzen.

Seite 14 (Phase Vorbefragung)

Bekanntheit des Anbieters

Die zweite Website, mit der Sie sich beschäftigen werden, wird die Website der [Name der 2. Krankenkasse anführen] sein. Bevor Sie mit der ersten Aufgabe starten machen Sie bitte wieder 3 Angaben zum Bekanntheitsgrad des Anbieters der Website. Geben Sie dazu an, in wie weit Sie den folgenden Aussagen zustimmen.

[Link zu Seite 15:] Ich habe die Angaben gemacht und möchte mit dem Versuch fortfahren.

Seite 15 (Phase Website 2)

1. Aufgabe: Allgemeine Orientierung

Sie werden nun die zweite Website kennen lernen. Wie zuvor lautet Ihre erste Aufgabe:

Verschaffen Sie sich einen allgemeinen Überblick über die Website zu Ihrer Orientierung. Nehmen Sie sich hierfür 5 Minuten Zeit.

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte zum aktuellen Fenster zurück (dies ist möglich, indem Sie in der Taskleiste auf das Kästchen klicken, in dem "Versuch" steht; die Taskleiste ist die Leiste am unteren Bildschirmrand, in der auch die Uhrzeit angezeigt wird). Sie können dann mit der nächsten Aufgabe fortfahren.

[Link zu Website 2:] Ich möchte jetzt mit dem Versuch fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 16 (Phase Website 2)

2. Aufgabe: Zahnersatz

In der folgenden Aufgabe geht es um Zahnersatz. Ihr Zahnarzt weist Sie darauf hin, dass Sie eine künstliche Zahnkrone benötigen. Sie möchten sich darüber informieren, inwieweit die Krankenkasse die Kosten für die kostspielige Behandlung übernimmt.

Suchen Sie die Regelungen über die Zuschüsse der Krankenkasse zum Zahnersatz auf. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.)

Wenn Sie die Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück.

[Link zu Website 2:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 17 (Phase Website 2)

3. Aufgabe: Bewegung und Fitness

Sie wollen etwas für Ihre körperliche Fitness tun, sich mehr und gesund bewegen, und suchen deshalb Anregungen zum Thema Fitness und Bewegung.

Finden Sie auf der aktuellen Website Informationen und Angebote zum Thema Gesundheit durch Fitness und Bewegung.

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück. (Es reicht aus, die Regelungen aufzufinden und sich einen Überblick zu verschaffen, ohne sie im Detail zu lesen.)

[Link zu Website2:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 18 (Phase Website 2)

4. Aufgabe: Kostenlose Familienversicherung für Studenten

Sie interessieren sich für die Voraussetzungen, die gelten, um als Student kostenlos in der Familienversicherung versichert zu sein.

Finden Sie die Voraussetzungen für die kostenlose Familienversicherung für Studenten auf der aktuellen Website.

Wenn Sie die Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück.

[Link zu Website 2:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 19 (Phase Website 2)

5. Aufgabe: Private Krankenzusatzversicherung

Der Eigenanteil an den Kosten für Zahnersatz, wie z.B. künstlichen Zahnkronen, den Sie selbst übernehmen müssen, ist Ihnen zu hoch. Sie möchten deshalb für zukünftigen **Zahnersatz** Zahnersatz eine weitergehende Kostenabsicherung über eine private Zusatzversicherung vornehmen.

Finden Sie auf der aktuellen Website eine Zusatzversicherung, mit der Sie die Kosten für Zahnersatz zusätzlich absichern können.

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte über das Kästchen „Versuch“ in der Taskleiste zum aktuellen Fenster zurück. Sie werden dann mit der Bewertung der Website beginnen.

[Link zu Website2:] Ich möchte mit der Aufgabe fortfahren. (Die Website wird in einem neuen Fenster geöffnet.)

Seite 20 (Phase Bewertung 2)

Bewertung der genutzten Website (Seite 1)

Bitte beurteilen Sie nun, in wie weit Sie den folgenden Aussagen zur genutzten Website zustimmen. Da es sich um recht viele zu bewertende Aussagen handelt, haben wir diese auf zwei Seiten verteilt. Falls Sie sich bei bestimmten Bewertungen unsicher sind, können Sie gerne zwischenzeitlich zur Website zurückkehren, um sich eine Meinung zu bilden.

Nutzen Sie zur Bewertung die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu". Falls eine Aussage Ihrer Meinung nach nicht bewertet werden kann, so wählen Sie die Antwortmöglichkeit "keine Angabe". Beachten Sie: Alle Aussagen beziehen sich konkret auf die genutzte Website.

[Link zu Seite 21:] Ich möchte meine Bewertung fortsetzen.

Seite 21 (Phase Bewertung 2)

Bewertung der genutzten Website (Seite 2)

Bitte setzen Sie Ihre Bewertung fort. Nutzen Sie dazu erneut die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu" oder das Feld "keine Angabe". Beachten Sie wiederum: Alle Aussagen beziehen sich konkret auf die genutzte Website.

[Link zu Seite 22:] Ich habe die Bewertung abgeschlossen und möchte mit dem Versuch fortfahren.

Seite 22 (Phase Auswahl)

Wir würden gerne mehr über Ihre Einschätzung der Bedeutung von Krankenversicherungsthemen erfahren. Bitte verwenden Sie dafür nachfolgende Beschreibungen. Nutzen Sie hierfür die sieben Auswahlfelder von (im ersten Beispiel) wichtig bis unwichtig. Falls eine Aussage nicht bewertet werden kann, wählen Sie die Antwortmöglichkeit "keine Angabe".

[Link zu Seite 23:] Ich habe alle Fragen beantwortet und möchte mit dem Versuch fortfahren.

Seite 23 (Phase Auswahl)

Abschließende Auswahlentscheidung.

Sie haben beide Websites nun intensiv kennen gelernt und bewertet. Bevor Sie zum Ende des Versuchs kommen bitten wir Sie abschließend eine Auswahl zu treffen.

Welche der beiden Websites würden Sie bevorzugt aufsuchen, wenn Sie eine neue Krankenkasse suchen.

Angenommen, Sie suchen eine neue Krankenkasse, für welche der beiden Krankenkassen würden Sie sich entscheiden? Unterstellen Sie dabei bitte, der Beitragssatz wäre bei beiden Krankenkassen bereits gleich und damit bedeutungslos.

[Link zu Seite 24:] Ich habe die Auswahl getroffen und möchte den Versuch nun beenden.

Seite 24 (Phase Ende)

Vielen Dank für Ihre Teilnahme am Versuch!

Falls Sie noch Fragen zum Versuch haben sollten, wenden Sie sich bitte an den Versuchsleiter unter der folgenden e-mailadresse: guntherheld@hamburg.de

Anhang E

Ergänzende Ergebnisdarstellungen

- Tab. E-1: Ausschluss von Fällen aufgrund eines zu hohen Anteils fehlender Werte
- Tab. E-2.1: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 1 Usability (reflektiv formativ)
- Tab. E-2.2: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 2 Usability (reflektiv formativ)
- Tab. E-2.3: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 3 Usability (reflektiv formativ)
- Tab. E-2.4: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 4 Usability (reflektiv formativ)
- Tab. E-3.1: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Nutzungsintention bei Reihenfolge 1 Usability (reflektiv formativ)
- Tab. E-3.2: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Nutzungsintention bei Reihenfolge 2 Usability (reflektiv formativ)
- Tab. E-3.3: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Nutzungsintention bei Reihenfolge 3 Usability (reflektiv formativ)
- Tab. E-3.4: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Nutzungsintention bei Reihenfolge 4 Usability (reflektiv formativ)
- Tab. E-4: Anteil fehlender Werte je Indikator
- Tab. E-5: Indikatoren, die in die Multiple Imputation eingehen
- Tab. E-6: Harman's Single Factor-Test Eigenwerte und Gesamtvarianz der Faktorenlösung
- Tab. E-7: Prüfung der Normalverteilungsannahme für alle erhobenen Indikatoren
- Tab. E-8: Harman's Single Factor-Test Komponentenmatrix
- Tab. E-9: Auflistung aller Probanden mit potenziell problematischen Antwortverhalten
- Tab. E-10.1: Mustermatrix der ersten berechneten Hauptachsenanalyse
- Tab. E-10.2: Mustermatrix der zweiten berechneten Hauptachsenanalyse
- Tab. E-10.3: Mustermatrix der dritten berechneten Hauptachsenanalyse
- Tab. E-10.4: Mustermatrix der vierten berechneten Hauptachsenanalyse
- Tab. E-11: Cronbachs-Alpha der reflektiven Usability-Skala ufosV2ra für die einzelnen Websites
- Tab. E-12: Rangreihe der Website-Mittelwerte auf der reflektiven Usability-Skala ufosV2ra
- Tab. E-13: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zum Vertrauen
- Tab. E-14: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zum Nutzervergnügen
- Tab. E-15: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Ästhetik
- Tab. E-16: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Nutzerzufriedenheit
- Tab. E-17: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Nutzungsintention
- Tab. E-18: Rangreihe über die Anzahl der im Versuch getätigten Auswahlentscheidungen

- Tab. E-18.1: Rangreihe der Websites auf dem Item *Höhe des Beitragssatzes*
- Tab. E-19: Pearson-Korrelationen zwischen den formativen Usability Indikatoren
- Tab. E-20: Toleranz- und VIF-Werte der formativen ufosV2fa-Indikatoren
- Tab. E-21: Rotierte Komponentenmatrix zur Erstellung der Multikollinearitätsindizes zur formativen Usability-Skala ufosV2fa
- Tab. E-22: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte
- Tab. E-23: Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Pfadkoeffizienten
- Tab. E-24: Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability
- Tab. E-25: Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability-Skala ufosV2ra mit Quadratwurzel des AVE-Maßes
- Tab. E-26: Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability-Skala ufosV2fa mit Quadratwurzel des AVE-Maßes
- Tab. E-27: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra
- Tab. E-28: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2f
- Tab. E-29: Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum nomologischen Modell
- Tab. E-30: Korrelationen der LR-Prädiktoren
- Tab. E-31: Toleranz und Varance-Inflation-Factors (VIF) der LR-Prädiktoren
- Tab. E-32: Teilnehmer mit standardisierten Residuen $z_{Redsid} > 1$ in der LR
- Tab. E-33: Ausreißer in der LR
- Schritt 1 der LR: Hinzufügen der Variablen zur Demographie, Vorerfahrung und Kontrolle
- Tab. E-34: Statistiken zur globalen Modellanpassung
- Tab. E-35: Klassifikationsmatrix sequentielle LR Schritt 1
- Tab. E-36: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 1
- Tab. E-37 Parameterschätzer sequentielle LR Schritt 1
- Schritt 2a der LR: Hinzufügen der ufosV2ra-Scores (Usability reflektiv) für beide Websites
- Tab. E-38: Statistiken zur globalen Modellanpassung
- Tab. E-39: Klassifikationsmatrix sequentielle LR Schritt 2a
- Tab. E-40: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2a
- Tab. E-41: Parameterschätzer sequentielle LR Schritt 2a
- Schritt 2b der LR: Hinzufügen der ufosV2fa-Scores (Usability formativ) für beide Websites
- Tab. E-42: Statistiken zur globalen Modellanpassung sequentielle LR Schritt 2b

- Tab. E-43: Klassifikationsmatrix sequentielle LR Schritt 2b
- Tab. E-44: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2b
- Tab. E-45: Parameterschätzer sequentielle LR Schritt 2b
- Schritt 2c der LR: Hinzufügen der Variable Kaufintention für beide Websites
- Tab. E-46: Statistiken zur globalen Modellanpassung sequentielle LR Schritt 2c
- Tab. E-47: Klassifikationsmatrix sequentielle LR Schritt 2c
- Tab. E-48: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2c
- Tab. E-49: Parameterschätzer sequentielle LR Schritt 2c
- Tab. E-50: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei reduziertem Datensatz: Ladungen und Gewichte
- Tab. E-51: Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability-Skala ufosV2ra mit Quadratwurzel des AVE-Maßes (eingeschränkter Datensatz)
- Tab. E-52: Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability-Skala ufosV2fa mit Quadratwurzel des AVE-Maßes (eingeschränkter Datensatz)
- Tab. E-53: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra bei eingeschränktem Datensatz
- Tab. E-54: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2fa bei eingeschränktem Datensatz
- Tab. E-55: Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: aufgeklärte Varianz (R^2) bei eingeschränktem Datensatz
- Tab. E-56: Kennwerte der PLS-Messmodelle zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability bei eingeschränktem Datensatz
- Tab. E-57: Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei eingeschränktem Datensatz: Pfadkoeffizienten
- Tab. E-58: Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum nomologischen Modell bei eingeschränktem Datensatz

Tab. E-1: Ausschluss von Fällen aufgrund eines zu hohen Anteils fehlender Werte

<i>Teilnehmer-code</i>	<i>Alter</i>	<i>Geschlecht</i>	<i>Anteil fehlender Daten</i>
234-4-5	23	w	91,7%
84-7-5	26	m	91,0%
98-3-1	24	m	91,0%
71-5-4	26	w	90,2%
211-7-5	30	w	90,2%
444-7-4	20	w	90,2%
673-7-4	21	m	48,1%

Anmerkung: weiblich (w), männlich (m).

Tab. E-2.1: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 1 Usability (reflektiv formativ)

<i>Variable</i>	<i>Reihenfolge Usability-Skalen</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE des Mittelwertes</i>
Nutzungsintention	Usab. 1 (Reflektiv – formativ)	152	4,21	1,642	,133
	Usab. 2 (Reflektiv – formativ)	110	4,02	1,771	,169

Anmerkung: Anzahl Fälle (N), Mittelwert (M), SD (Standardabweichung), Standardfehler (SE).

Tab. E-2.2: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 2 Usability (reflektiv formativ)

<i>Variable</i>	<i>Reihenfolge Usability-Skalen</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE des Mittelwertes</i>
Nutzungsintention	Usab. 3 (Reflektiv – formativ)	114	4,24	1,920	,180
	Usab. 4 (Reflektiv – formativ)	136	4,03	1,996	,171

Anmerkung: Anzahl Fälle (N), Mittelwert (M), SD (Standardabweichung), Standardfehler (SE).

Tab. E-2.3: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 3 Usability (reflektiv formativ)

<i>Variable</i>	<i>Reihenfolge Usability-Skalen</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE des Mittelwertes</i>
Nutzungsintention	Usab. 1 (Reflektiv – formativ)	152	4,21	1,642	,133
	Usab. 3 (Reflektiv – formativ)	114	4,24	1,920	,180

Anmerkung: Anzahl Fälle (N), Mittelwert (M), SD (Standardabweichung), Standardfehler (SE).

Tab. E-2.4: Mittelwerte für die AV Nutzungsintention bei Reihenfolge 4 Usability (reflektiv formativ)

<i>Variable</i>	<i>Reihenfolge Usability-Skalen</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE des Mittelwertes</i>
Nutzungsintention	Usab. 2 (Reflektiv – formativ)	110	4,02	1,771	,169
	Usab. 4 (Reflektiv – formativ)	136	4,03	1,996	,171

Anmerkung: Anzahl Fälle (N), Mittelwert (M), SD (Standardabweichung), Standardfehler (SE).

Tab. E-3.1: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV
Nutzungsintention bei Reihenfolge 1 Usability (reflektiv formativ)

	Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit						
							95% Konfidenzintervall der Differenz		
	F	Sign.	t	df	Sign. (2-seitig)	Mittlere Differenz	SE der Differenz	Untere	Obere
Varianzen sind Gleich	,701	,403	,905	260	,366	,19	,212	-,226	,611
Varianzen sind nicht gleich			,894	224,181	,372	,19	,215	-,232	,616

Anmerkung: F-Wert (F), t-Wert (t), Freiheitsgrade (df), Standardfehler (SE).

Tab. E-3.2: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV
Nutzungsintention bei Reihenfolge 2 Usability (reflektiv formativ)

	Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit						
							95% Konfidenzintervall der Differenz		
	F	Sign.	t	df	Sign. (2-seitig)	Mittlere Differenz	SE der Differenz	Untere	Obere
Varianzen sind Gleich	,961	,328	,833	248	,406	,21	,249	-,283	,698
Varianzen sind nicht gleich			,836	243,323	,404	,21	,248	-,282	,696

Anmerkung: F-Wert (F), t-Wert (t), Freiheitsgrade (df), Standardfehler (SE).

Tab. E-3.3: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV
Nutzungsintention bei Reihenfolge 3 Usability (reflektiv formativ)

	Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit						
							95% Konfidenzintervall der Differenz		
	F	Sign.	t	df	Sign. (2-seitig)	Mittlere Differenz	SE der Differenz	Untere	Obere
Varianzen sind Gleich	6,556	,011	-,120	264	,904	-,03	,219	-,457	,405
Varianzen sind nicht gleich			-,118	221,188	,907	-,03	,224	-,467	,415

Anmerkung: F-Wert (F), t-Wert (t), Freiheitsgrade (df), Standardfehler (SE).

Tab. E-3.4: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV
Nutzungsintention bei Reihenfolge 4 Usability (reflektiv formativ)

	Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit						
							95% Konfidenzintervall der Differenz		
	F	Sign.	t	df	Sign. (2-seitig)	Mittlere Differenz	SE der Differenz	Untere	Obere
Varianzen sind Gleich	6,614	,011	-,046	244	,963	-,01	,244	-,491	,468
Varianzen sind nicht gleich			-,047	241,878	,963	-,01	,240	-,485	,462

Anmerkung: F-Wert (F), t-Wert (t), Freiheitsgrade (df), Standardfehler (SE).

Tab. E-4: Anteil fehlender Werte je Indikator

	N	Mittelwert	Standardabweichung	Fehlend	
				Anzahl	Prozent (%)
Verver	154	4,60	1,21	109	41,44
Ruf	155	4,39	1,35	108	41,06
Zuver	156	4,19	1,27	107	40,68
Bezuho	190	4,47	1,77	73	27,76
Verdig	221	5,00	1,40	42	15,97
Sucerg	224	4,90	1,85	39	14,83
Anbint	224	4,82	1,44	39	14,83
Quahoc	228	4,48	1,49	35	13,31
Infnue	230	4,15	1,72	33	12,55
Aktrue	232	6,00	1,18	31	11,79
Bilgra	233	3,75	1,58	30	11,41
Welkasse	235	3,75	1,58	28	10,65
Intang	237	3,97	1,68	26	9,89
Vinver	244	5,09	1,40	19	7,22
Zahvor	244	4,63	1,71	19	7,22
Webunt	248	3,22	1,57	15	5,70
Texunv	249	2,94	1,59	14	5,32
Lanein	249	5,77	1,47	14	5,32
Aufsuc	250	4,51	1,77	13	4,94
Nutrei	250	3,38	1,47	13	4,94
Zufrie	250	4,34	1,86	13	4,94
Reisin	250	4,79	1,80	13	4,94
Wieder	251	3,33	1,85	12	4,56
Angstr	251	4,86	1,61	12	4,56
Moewun	251	4,25	1,77	12	4,56
Paserw	252	5,12	1,52	11	4,18
Bekan	252	4,83	2,17	11	4,18
Nuterw	252	4,28	1,85	11	4,18
Welwebsi	252	4,62	1,71	11	4,18
Viemer	253	4,72	1,80	10	3,80
Spaweb	253	3,05	1,44	10	3,80
involv 5	253	2,94	1,74	10	3,80
Zuvinf	254	4,00	1,86	9	3,42
Angkat	254	5,07	1,48	9	3,42
Benoet	254	3,80	2,03	9	3,42
Sucnet	254	4,20	1,91	9	3,42
Bengut	254	4,59	1,70	9	3,42
Spaweb	254	3,01	1,69	9	3,42
involv 2	254	3,71	1,57	9	3,42
Hinher	255	5,46	1,45	8	3,04
Zenerr	255	6,02	1,35	8	3,04
Schles	255	5,44	1,57	8	3,04
Aesges	255	4,32	1,78	8	3,04
Inform	255	3,95	2,02	8	3,04
Kompli	255	4,85	2,09	8	3,04
Nichwo	255	4,75	1,93	8	3,04
Desnut	255	4,72	2,01	8	3,04

Tab. E-4: Anteil fehlender Werte je Indikator (Fortsetzung)

	N	Mittelwert	Standardabweichung	Fehlend	
				Anzahl	Prozent (%)
Besang	255	4,75	1,54	8	3,04
Ansdar	255	4,06	2,03	8	3,04
involv 3	255	3,58	1,57	8	3,04
involv 4	255	3,47	1,72	8	3,04
Suczue	256	4,87	1,69	7	2,66
Finlei	256	4,79	1,71	7	2,66
Theueb	256	4,78	1,72	7	2,66
Schwic	256	4,85	1,63	7	2,66
Reiinf	256	5,72	1,28	7	2,66
Inanve	256	5,08	1,39	7	2,66
Nutzuf	256	4,27	2,01	7	2,66
Uebbli	256	4,53	1,98	7	2,66
Benzuf	256	4,39	1,96	7	2,66
Zuvinf	256	4,24	1,86	7	2,66
involv 1	256	3,38	1,77	7	2,66
Lernen	257	5,37	1,53	6	2,28
Zurech	257	4,96	1,55	6	2,28
Gutstr	257	4,67	1,83	6	2,28
Zuvsch	257	4,63	1,81	6	2,28
Inhang	257	4,75	1,69	6	2,28
Staaus	257	5,06	1,60	6	2,28

Tab. E-5: Indikatoren, die in die Multiple Imputation eingehen

<i>Code</i>	<i>Item</i>	<i>% Missing</i>	<i>Code</i>	<i>Item</i>	<i>% Missing</i>
Verdig	Vertrauen1	15,97	involv 5	Involvement5	3,80
Sucerg	Usab form 20	14,83	benoet	Nutzungsintention3	3,42
Anbint	Vertrauen2	14,83	nutzufe	Zufriedenheit2	3,42
Sucerg	Usab form 20	13,69	zuvinf	Usab form 12	3,42
Quahoc	Usab form 30	13,31	bengut	Usab form 17	3,42
Quahoc	Usab form 30	13,31	besang	Usab form 22	3,42
Infhue	Usab form 31	12,55	angkat	Usab form 21	3,42
aktrue	Usab form 5	12,55	benoet	Nutzungsintention3	3,42
aktrue	Usab form 5	11,79	sucnet	Usab ref 1	3,42
verdig	Vertrauen1	11,79	suczue	Usab ref 5	3,42
bilgra	Usab form 23	11,41	zenerr	Usab form 4	3,42
intang	Usab form 24	9,89	bengut	Usab form 17	3,42
intang	Usab form 24	9,51	spaweb	Nutzervergnügen3	3,42
bilgra	Usab form 23	9,13	reiinf	Usab form 28	3,42
infhue	Usab form 31	7,60	inanve	Usab form 29	3,42
moewun	Usab ref 7	7,22	angstr	InfoBewegungsang.	3,42
vinver	Vertrauen4	7,22	involv 2	Involvement2	3,42
zahvor	InfoZahnersatz	7,22	hinher	Usab form 3	3,04
vinver	Vertrauen4	6,84	zenerr	Usab form 4	3,04
zahvor	InfoZahnersatz	6,08	schles	Usab form 13	3,04
viemer	Usab form 2	5,70	texunv	Usab form 16	3,04
webunt	Nutzervergnügen1	5,70	aesges	Aesthetik2	3,04
texunv	Usab form 16	5,32	inform	Nutzungsintention2	3,04
lanein	Usab form 25	5,32	kompli	Usab ref 2	3,04
aufsuc	Nutzungsintention1	4,94	nichwo	Usab form 7	3,04
nutrei	Nutzervergnügen2	4,94	schles	Usab form 13	3,04
zufrie	Zufriedenheit1	4,94	desnut	Usab form 14	3,04
reisin	Usab form 1	4,94	besang	Usab form 22	3,04
nutrei	Nutzervergnügen2	4,94	ansdar	Aesthetik1	3,04
wieder	Nutzungsintention4	4,56	aesges	Aesthetik2	3,04
angstr	InfoBewegungsangebote	4,56	involv 3	Involvement3	3,04
aufsuc	Nutzungsintention1	4,56	involv 4	Involvement4	3,04
moewun	Usab ref 7	4,56	inform	Nutzungsintention2	2,66
zufrie	Zufriedenheit1	4,18	suczue	Usab ref 5	2,66
paserw	Usab form 15	4,18	nuterw	Usab ref 8	2,66
bekan	Bekanntheit	4,18	finlei	Usab form 19	2,66
nuterw	Usab ref 8	4,18	theueb	Usab form 27	2,66
paserw	Usab form 15	4,18	schwic	Usab form 11	2,66
hinher	Usab form 3	4,18	nichwo	Usab form 7	2,66
angkat	Usab form 21	4,18	reiinf	Usab form 28	2,66
sucnet	Usab ref 1	3,80	inanve	Usab form 29	2,66
reisin	Usab form 1	3,80	nutzufe	Zufriedenheit2	2,66
viemer	Usab form 2	3,80	uebbli	Usab ref 3	2,66
desnut	Usab form 14	3,80	benzufe	Usab ref 6	2,66
lanein	Usab form 25	3,80	zurech	Usab form 10	2,66
webunt	Nutzervergnügen1	3,80	finlei	Usab form 19	2,66
spaweb	Nutzervergnügen3	3,80	theueb	Usab form 27	2,66
wieder	Nutzungsintention4	3,80	gutstr	Usab form 8	2,66
lernen	Usab ref 4	3,80	schwic	Usab form 11	2,66

Anmerkung: Anteil fehlender Werte in Prozent (% Missing).

Tab. E-5: Indikatoren, die in die Multiple Imputation eingehen (Fortsetzung)

Code	Item	% Missing
zuvsch	Usab form 9	2,66
zuvinf	Usab form 12	2,66
inhang	Usab form 18	2,66
staaus	Usab form 26	2,66
involv 1	Involvement1	2,66
lernen	Usab ref 4	2,28
kompli	Usab ref 2	2,28
uebbli	Usab ref 3	2,28
benzuf	Usab ref 6	2,28
zurech	Usab form 10	2,28
gutstr	Usab form 8	2,28
zuvsch	Usab form 9	2,28
inhang	Usab form 18	2,28
staaus	Usab form 26	2,28
ansdar	Aesthetik1	2,28
bekan	Bekanntheit	1,14
pc oft	PC-Nutzung	-
pc ofth	PC-Nutzung	-
pc seit	PC-Nutzung	-
www oft	PC-Nutzung	-
www ofth	PC-Nutzung	-
recherch	Internetrecherche	-
altervp	Alter	-
sex	Geschlecht	-
schule	Schulbildung	-
beruf	Beruf	-
einkom	Einkommen	-
haushgr	Haushaltsgröße	-
www seit	Web-Nutzung	-
shoperf	Interneteneinkäufe	-
deufae	Fähigk. deut. Sprache	-

Anmerkung: Anteil fehlender Werte in Prozent (% Missing).

Tab. E-6: Harman's Single Factor-Test Eigenwerte und Gesamtvarianz der Faktorenlösung

Komponente	Anfänglicher Eigenwert λ	Erklärte Gesamtvarianz	
		% der Varianz	Kumulierte %
1	21,93	35,95	35,95
2	4,58	7,52	43,47
3	3,13	5,13	48,60
4	2,18	3,57	52,17
5	1,99	3,25	55,42
6	1,71	2,81	58,23
7	1,47	2,42	60,65
8	1,42	2,33	62,98
9	1,26	2,07	65,05
10	1,16	1,90	66,96
11	1,11	1,82	68,78
12	1,05	1,72	70,49
13	1,01	1,66	72,16

Tab. E-7: Prüfung der Normalverteilungsannahme für alle erhobenen Indikatoren

		<i>Sch</i>	<i>z-Wert Sch</i>	<i>Ex</i>	<i>z-Wert Ex</i>
Altervp	Alter VP	2,255	4,775	6,776	7,383
Sex	Geschlecht	0,551	1,168	- 1,859	- 2,025
Schule	Schulabschluss	4,498	9,524	20,859	22,727
Beruf	Aktuelle berufl. Tätigkeit	- 1,796	- 3,804	2,729	2,974
Einkom	Netto-Einkommen pro Monat	1,310	2,774	1,902	2,072
Haushgr	Haushaltsgröße	0,680	1,441	- 0,258	- 0,281
Deumu	Muttersprache Deutsch	0,979	2,073	- 1,145	- 1,247
pc seit	PC-Nutzung Jahre	0,924	1,957	0,281	0,306
pc oft	PC-Nutzung Tage pro Monat	- 1,212	- 2,566	- 0,089	- 0,097
pc ofth	PC-Nutzung Stunden pro Tag	0,422	0,893	- 1,199	- 1,306
www seit	Internetnutzung Jahre	0,660	1,397	- 0,263	- 0,287
www oft	Internetnutzung Tage pro Monat	- 0,759	- 1,607	- 1,214	- 1,323
www ofth	Internetnutzung Stunden pro Tag	2,184	4,625	4,726	5,149
Shoperf	Interneteinkäufe	1,230	2,605	0,751	0,819
Recherch	Internetrecherche	1,066	2,257	0,392	0,427
Bekan	Bekanntheit des Anbieters	- 0,421	- 0,891	- 1,507	- 1,642
Aufsuc	Nutzungsintention1	- 0,130	- 0,276	- 1,069	- 1,165
Aufsuc	Nutzungsintention2	- 0,036	- 0,076	- 1,164	- 1,268
Inform	Nutzungsintention3	-	-	- 1,290	- 1,406
Wieder	Nutzungsintention4	0,419	0,887	- 0,691	- 0,753
Zufrie	Zufriedenheit1	- 0,497	- 1,052	- 0,737	- 0,803
Nutzuf	Zufriedenheit2	- 0,145	- 0,306	- 1,277	- 1,391
Sucnet	Usab ref 1	- 0,380	- 0,805	- 0,901	- 0,982
Kompli	Usab ref 2	- 0,580	- 1,228	- 1,219	- 1,329
Uebbli	Usab ref 3	- 0,826	- 1,750	0,542	0,591
Lernen	Usab ref 4	- 0,893	- 1,891	- 0,403	- 0,439
Suczue	Usab ref 5	- 0,733	- 1,553	- 0,249	- 0,271
Benzuf	Usab ref 6	- 0,325	- 0,688	- 1,508	- 1,643
Nuterw	Usab ref 8	- 0,492	- 1,042	- 1,126	- 1,227
Moewun	Usab ref 7	- 0,346	- 0,733	- 0,056	- 0,061
Reisin	Usab form 1	- 0,554	- 1,174	- 0,774	- 0,843
Viemer	Usab form 2	- 0,499	- 1,057	- 1,063	- 1,158
Zurech	Usab form 10	- 1,088	- 2,304	- 0,077	- 0,084
Finlei	Usab form 19	- 0,465	- 0,984	- 0,681	- 0,742
Infneue	Usab form 31	- 0,607	- 1,285	- 0,505	- 0,550
Theueb	Usab form 27	- 0,694	- 1,469	- 1,027	- 1,118
Gutstr	Usab form 8	- 0,425	- 0,900	- 1,445	- 1,575
Schwic	Usab form 11	- 0,838	- 1,775	- 0,479	- 0,521
Paserw	Usab form 15	0,091	0,193	- 0,368	- 0,401
Hinher	Usab form 3	- 0,558	- 1,182	- 0,924	- 1,007
Zener	Usab form 4	- 1,017	- 2,154	- 0,116	- 0,127
Aktrue	Usab form 5	- 0,855	- 1,811	- 0,041	- 0,045
Nichwo	Usab form 7	- 1,100	- 2,330	0,337	0,368

Tab. E-7: Prüfung der Normalverteilungsannahme für alle erhobenen Indikatoren
(Fortsetzung)

Zuvsch	Usab form 9	- 0,300	- 0,636	- 1,182	- 1,288
Zuvinf	Usab form 12	0,113	0,239	- 1,377	- 1,500
Schles	Usab form 13	- 0,529	- 1,119	- 0,927	- 1,010
Desnut	Usab form 14	- 0,485	- 1,027	- 1,247	- 1,359
Texunv	Usab form 16	0,294	0,622	- 1,071	- 1,167
Bengut	Usab form 17	- 0,780	- 1,652	- 0,195	- 0,212
Inhang	Usab form 18	- 0,609	- 1,289	- 0,561	- 0,611
Sucerg	Usab form 20	- 0,410	- 0,869	- 0,121	- 0,131
Besang	Usab form 22	- 0,550	- 1,164	- 0,485	- 0,528
Angkat	Usab form 21	- 0,860	- 1,821	- 0,426	- 0,464
Bilgra	Usab form 23	- 0,314	- 0,665	- 0,640	- 0,698
Intang	Usab form 24	-	-	- 0,423	- 0,461
Lanein	Usab form 25	- 1,902	- 4,028	5,092	5,548
Staaus	Usab form 26	- 0,643	- 1,362	- 0,840	- 0,915
Reiinf	Usab form 28	- 1,205	- 2,552	1,290	1,406
Inanve	Usab form 29	- 0,463	- 0,980	- 0,810	- 0,883
Quahoc	Usab form 30	0,067	0,141	- 0,745	- 0,812
Zahvor	InfoZahnersatz	- 0,420	- 0,890	- 0,922	- 1,005
Angstr	InfoBewegungsangebote	- 0,268	- 0,567	- 1,225	- 1,335
Verdig	Vertrauen1	- 1,052	- 2,227	0,687	0,749
Anbint	Vertrauen2	-	-	- 0,401	- 0,437
Vinver	Vertrauen4	- 0,466	- 0,987	- 0,341	- 0,371
Inform	Involvement1	- 0,537	- 1,137	- 0,221	- 0,240
involv 2	Involvement2	- 0,400	- 0,847	- 1,133	- 1,234
involv 3	Involvement3	- 0,402	- 0,852	- 0,596	- 0,650
involv 4	Involvement4	- 0,614	- 1,300	0,200	0,218
involv 5	Involvement5	- 1,146	- 2,427	1,592	1,735
Webunt	Nutzervergnügen1	0,193	1,786	- 0,810	- 3,757
Nutrei	Nutzervergnügen2	0,266	2,462	- 0,780	- 3,619
Spaweb	Nutzervergnügen3	0,378	3,505	- 0,748	- 3,472
Ansdar	Aesthetik1	- 0,390	- 0,825	- 1,151	- 1,255
Aesges	Aesthetik2	- 0,241	- 0,511	- 1,207	- 1,315

Anmerkung: Schiefe (Sch), Exzess (Ex). Der Kolmogorov-Smirnov-Test auf Normalverteiltheit wird für alle Indikatoren signifikant ($p < 0,001$).

Tab. E-8: Harman's Single Factor-Test Komponentenmatrix

Item	Komponente												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Bekanntheit					-0,4	0,3				-0,4	0,4		
Nutzungsint.1	0,6		0,4		-0,4								
Nutzungsint.2	0,6		0,3		-0,3		0,3						
Nutzungsint.3	0,7		0,3		-0,3								
Nutzungsint.4	0,5		0,3	-0,3	-0,3								
Zufriedenheit1	0,8												
Zufriedenheit2	0,8				-0,3								
Usab ref 1	0,8												
Usab ref 2	0,7		-0,3										
Usab ref 3	0,8		-0,3										
Usab ref 4	0,8		-0,3										
Usab ref 5	0,8		-0,3										
Usab ref 6	0,9												
Usab ref 8	0,8												
Usab ref 7	0,8												
Usab form 1	0,6												
Usab form 2	0,5							0,5					
Usab form 10	0,8		-0,3										
Usab form 19	0,8		-0,3										
Usab form 31	0,4		0,3										0,6
Usab form 27	0,7												
Usab form 8	0,8												
Usab form 11	0,9												
Usab form 15	0,6					0,3							
Usab form 3	0,6						0,3						
Usab form 4	0,4			0,3		0,3	0,3						0,3
Usab form 5						0,6					-0,3		
Usab form 7	0,6		-0,3										
Usab form 9	0,7		-0,3										
Usab form 12	0,6		-0,3		0,4								
Usab form 13	0,4				0,3		0,3				0,4	0,3	
Usab form 14	0,6				0,4							0,4	
Usab form 16	-0,4					0,3		-0,4	0,3				
Usab form 17	0,8												
Usab form 18	0,7												
Usab form 20	0,3									0,7			
Usab form 22	0,7												
Usab form 21	0,8												
Usab form 23	0,4							0,3	-0,4				
Usab form 24	0,4		0,3									-0,3	
Usab form 25	0,3			0,3		0,3				0,3	0,3		
Usab form 26	0,7												
Usab form 28	0,5		0,4	0,3							0,3		
Usab form 29	0,7		0,3	0,3									
Usab form 30	0,4						-0,3	0,4	0,3				
InfoZahnersatz	0,5						-0,5						

Tab. E-9: Auflistung aller Probanden mit potenziell problematischen Antwortverhalten

<i>Teilnehmer- code</i>	<i>Usability reflektiv Website1</i>	<i>Usability reflektiv Website2</i>	<i>SD Website1</i>	<i>SD Website2</i>
18-6-3	1,00	4,00	1,34	1,82
31-6-4	0,29	- 1,14	1,13	0,88
43-1-2	3,29	0,14	1,95	1,44
61-4-1	4,29	2,71	1,13	1,42
86-1-2	1,00	0,14	1,22	0,97
91-1-7	-0,29	0,14	1,12	0,99
107-4-5	2,57	3,86	2,10	2,06
112-5-3	1,43	- 0,14	1,43	0,97
114-5-6	1,43	-	1,33	0,90
218-1-7	-	0,86	0,85	1,09
219-2-1	-1,86	- 0,71	1,49	0,82
223-2-6	-0,57	0,57	0,87	0,91
227-3-2	-	1,71	1,22	0,80
228-3-5	1,00	5,00	1,85	1,57
230-3-7	1,57	1,43	0,93	1,76
268-3-4	0,29	- 0,14	0,99	1,56
273-4-2	1,86	0,14	1,25	0,98
279-5-2	1,86	-	1,44	0,81
283-5-7	1,43	0,86	1,37	0,87
406-6-1	- 3,43	1,00	1,87	1,63
407-5-2	1,29	1,00	1,43	0,83
426-4-5	3,14	2,71	1,86	1,81
437-6-3	0,43	- 4,00	2,08	1,86
451-1-6	- 0,71	0,57	1,47	0,97
452-1-7	- 0,14	0,43	1,90	0,77
463-3-6	3,86	0,14	2,43	1,67
469-4-6	2,43	-0,71	1,22	0,83
475-7-6	2,29	0,86	1,61	0,95
496-5-3	4,00	-0,43	1,82	1,48
623-7-6	-	0,29	1,04	0,82
626-1-5	0,14	2,00	0,87	1,03
628-1-7	1,71	3,43	1,31	1,16
642-7-2	1,86	3,14	2,35	2,35
659-3-1	1,86	1,71	1,06	0,74
663-3-6	0,86	-1,43	0,92	1,25
665-4-1	3,14	1,00	1,72	1,02
678-1-3	- 0,57	- 3,57	1,39	0,97
680-2-6	0,57	0,43	0,98	1,46
684-2-1	1,00	4,43	1,82	1,92
687-3-2	3,14	0,71	2,24	1,18

Anmerkung: Die grau hinterlegten Zellen weisen auf ein auffälliges Antwortverhalten der jeweiligen Kategorie hin. Weiblich (w), männlich (m), Standardabweichung (SD).

Tab. E-10.1: Mustermatrix der ersten berechneten Hauptachsenanalyse

<i>Faktor</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	h_i^2	<i>MSA-Wert</i>
Usability ur5	,833	,040	,033	,096	,780	,941
Usability ur4	,816	-,001	,033	,054	,708	,947
Usability ur3	,794	,031	-,022	,122	,696	,959
Usability ur2	,794	,013	,015	,024	,654	,946
Usability ur6	,771	-,004	,138	,220	,855	,942
Usability ur8	,697	-,051	,202	,164	,752	,952
Usability ur1	,652	-,018	,182	,165	,658	,965
Usability ur7	,459	-,039	,275	,298	,624	,956
Involvement1	-,036	,907	,085	,022	,808	,898
Involvement4	,071	,904	-,025	-,054	,819	,843
Involvement5	,030	,894	,028	-,029	,799	,871
Involvement2	-,043	,892	-,018	,034	,798	,872
Involvement3	-,005	,899	,104	,053	,827	,858
Vertrauen4	-,043	,093	,814	-,050	,550	,823
Vertrauen2	,048	,035	,806	-,003	,571	,874
Vertrauen1	,017	-,026	,789	,059	,566	,876
Ästhetik ae2	,200	-,110	,190	,581	,720	,878
Ästhetik ae1	,365	-,075	,255	,524	,804	,914
Nutzervergnügen2	,078	,026	,040	,824	,737	,876
Nutzervergnügen3	,101	,046	,083	,787	,732	,889
Nutzervergnügen1	-,003	,029	-,067	,783	,544	,908
$\Sigma=78,77\%$	40,45%	20,64%	9,15%	7,65%		
$\lambda_j \text{ }_5=0,577$	8,496	4,335	1,920	1,606		

Anmerkung: Kommunalität (h_i^2), Eigenwert (λ_j), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % (Σ).

Tab. E-10.2: Mustermatrix der zweiten berechneten Hauptachsenanalyse

<i>Faktor</i>	1	2	3	4	h_i^2	<i>MSA-Wert</i>
Usability ur5	,831	,038	,042	,104	,778	,940
Usability ur4	,813	-,003	,041	,062	,708	,943
Usability ur3	,791	,029	-,014	,129	,698	,957
Usability ur2	,791	,011	,023	,032	,654	,942
Usability ur6	,768	-,006	,145	,227	,851	,938
Usability ur8	,689	-,051	,202	,171	,726	,959
Usability ur1	,651	-,020	,188	,172	,658	,961
Involvement i1	-,035	,907	,085	,021	,807	,902
Involvement i4	,069	,903	-,024	-,054	,819	,842
Involvement i3	-,005	,899	,103	,052	,825	,864
Involvement i2	-,044	,893	-,020	,031	,797	,875
Involvement i5	,030	,893	,031	-,029	,796	,877
Vertrauen v4	-,037	,090	,811	-,044	,549	,811
Vertrauen v2	,055	,032	,805	,004	,569	,864
Vertrauen v1	,024	-,028	,786	,066	,564	,865
Ästhetik ae2	,205	-,112	,194	,583	,720	,869
Ästhetik ae1	,369	-,077	,260	,528	,804	,906
Nutzervergnügen vg2	,083	,025	,043	,819	,732	,870
Nutzervergnügen vg3	,106	,045	,088	,789	,731	,885
Nutzervergnügen vg1	,002	,028	-,063	,780	,542	,903
$\Sigma=78,77\%$	39,50%	21,67%	9,57%	8,03%		
λ_j $s=0,574$	7,900	4,333	1,919	1,606		

Anmerkung: Kommunalität (h_i^2), Eigenwert (λ_j), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % (Σ).

Tab. E-10.3: Mustermatrix der dritten berechneten Hauptachsenanalyse

<i>Faktor</i>	1	h_i^2	<i>MSA-Wert</i>
Nutzervergnügen vg2	,891	,773	,812
Nutzervergnügen vg3	,874	,759	,816
Nutzervergnügen vg1	,748	,566	,875
Ästhetik ae1	,827	,744	,769
Ästhetik ae2	,789	,707	,764
$\Sigma=74,63$	74,63		
λ_j $s=3,73$	3,73		

Tab. E-10.4: Mustermatrix der vierten berechneten Hauptachsenanalyse

<i>Faktor</i>	1	h_i^2	<i>MSA-Wert</i>
Usability ur6	,943	,858	,934
Usability ur5	,892	,797	,945
Usability ur3	,885	,761	,961
Usability ur8	,858	,727	,961
Usability ur1	,844	,698	,973
Usability ur4	,811	,662	,960
Usability ur2	,797	,643	,963
Ästhetik ae1	,827	,805	,877
Ästhetik ae2	,654	,686	,838
$\Sigma=73,32$	73,32		
λ_j $s=6,59$	6,59		

Anmerkung: Kommunalität (h_i^2), Eigenwert (λ_j), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % (Σ).

Tab. E-11: Cronbachs-Alpha der reflektiven Usability-Skala ufosV2ra für die einzelnen Websites

<i>Lfd.Nr</i>	<i>Krankenkasse</i>	<i>Cronbachs Alpha</i>	<i>N</i>
1	AOK Die Gesundheitskasse	0,95	67
2	BARMER	0,95	80
3	DAK	0,95	78
4	Deutsche BKK	0,92	71
5	GEK Gmünder Ersatzkasse	0,97	67
6	KKH Die Kaufmännische	0,96	78
7	Techniker Krankenkasse	0,94	71

Anmerkung: Cronbachs Alpha (α), Anzahl der vorliegenden Bewertungen (N).

Tab. E-12: Rangreihe der Website-Mittelwerte auf der reflektiven Usability-Skala ufosV2ra

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	BARMER	5,49	1,28	80
2.	KKH Die Kaufmännische	4,95	1,75	78
3.	Techniker Krankenkasse	4,89	1,68	71
4.	Deutsche BKK	4,86	1,64	71
5.	DAK	4,52	1,85	78
6.	AOK Die Gesundheitskasse (Hamburg)	4,15	1,85	67
7.	GEK Gmünder Ersatzkasse	3,70	1,87	67
<i>Gesamt</i>				512

Anmerkung: Die reflektive Usability-Skala ufosV2ra umfasst die sieben Indikatoren ur1, ur2, ur3, ur4, ur5, ur6 und ur8. Cronbachs Alpha $\alpha=0,955$. Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-13: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zum Vertrauen

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	BARMER	5,44	1,42	80
2.	Techniker Krankenkasse	5,24	1,26	71
3.	KKH Die Kaufmännische	5,18	1,27	78
4.	DAK	5,12	1,22	78
5.	Deutsche BKK	4,91	1,32	71
6.	GEK Gmünder Ersatzkasse	4,61	1,40	67
7.	AOK Die Gesundheitskasse (Hamburg)	4,57	1,41	67
<i>Gesamt</i>				512

Anmerkung: Die reflektive Vertrauens-Skala umfasst die vier Indikatoren v1, v2 und v4. Cronbachs Alpha $\alpha=0,886$. Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-14: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zum Nutzervergnügen

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	BARMER	3,65	1,56	80
2.	Techniker Krankenkasse	3,50	1,50	71
3.	KKH Die Kaufmännische	3,32	1,50	78
4.	Deutsche BKK	3,22	1,43	71
5.	DAK	3,12	1,65	78
6.	AOK Die Gesundheitskasse (Hamburg)	2,92	1,69	67
7.	GEK Gmünder Ersatzkasse	2,50	1,40	67
<i>Gesamt</i>				512

Anmerkung: Die reflektive Skala zum Nutzervergnügen umfasst die drei Indikatoren vg1, vg2 und vg3. Cronbachs Alpha $\alpha=0,900$. Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-15: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Ästhetik

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	BARMER	5,17	1,68	80,00
2.	KKH Die Kaufmännische	4,70	1,71	78,00
3.	Deutsche BKK	4,44	1,74	71,00
4.	Techniker Krankenkasse	4,06	1,85	71,00
5.	DAK	3,97	2,02	78,00
6.	AOK Die Gesundheitskasse (Hamburg)	3,57	1,95	67,00
7.	GEK Gmünder Ersatzkasse	3,51	1,80	67,00
<i>Gesamt</i>				512

Anmerkung: Die reflektive Ästhetik-Skala umfasst die zwei Indikatoren ae1 und ae2. Cronbachs Alpha $\alpha=0,895$, Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-16: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Nutzerzufriedenheit

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	Barmer	6,09	1,255	80
2.	KKH	4,97	1,736	78
3.	Techniker Krankenkasse	4,93	1,580	71
4.	DAK	4,67	1,665	78
5.	Deutsche BKK	4,52	1,681	71
6.	AOK	4,93	1,580	67
7.	GEK	4,01	1,754	67
<i>Gesamt</i>		4,72	1,777	512

Anmerkung: Die reflektive Skala zur Nutzerzufriedenheit umfasst die zwei Indikatoren z1 und z2. Cronbachs Alpha $\alpha=0,935$, Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-17: Rangreihe der Website-Mittelwerte auf der reflektiven Skala zur Nutzungsintention

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	Barmer	5,23	,171	80
2.	Techniker Krankenkasse	4,65	,172	71
3.	KKH	4,31	,197	78
4.	DAK	4,21	,212	78
5.	Deutsche BKK	3,79	,206	71
6.	AOK	3,58	,209	67
7.	GEK	2,87	,222	67
<i>Gesamt</i>		4,09	,201	512

Anmerkung: Die reflektive Skala zur Nutzungsintention umfasst die drei Indikatoren ni1, ni2 und ni3. Cronbachs Alpha $\alpha=0,93$, Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-18: Rangreihe über die Anzahl der im Versuch getätigten Auswahlentscheidungen

<i>Rang</i>	<i>Website</i>	<i>N</i>	<i>Absolute Häufigkeit</i>	<i>Relative Häufigkeit(%)</i>
1.	BARMER	80	66	82,5
2.	Techniker Krankenkasse	75	41	54,7
3.	KKH - Die Kaufmännische	78	39	50,0
4.	DAK	79	34	43,0
5.	Deutsche BKK	68	23	33,8
6.	AOK - Die Gesundheitskasse	75	25	33,3
7.	GEK - Gmünder Ersatzkasse	71	16	22,5
<i>Gesamt</i>		526	244	46,4

Anmerkung: Anzahl der vorliegenden Bewertungen (N).

Tab. E-18.1: Rangreihe der Websites auf dem Item *Höhe des Beitragssatzes*

<i>Rang</i>	<i>Website</i>	<i>M</i>	<i>SD</i>	<i>N</i>
1.	GEK - Gmünder Ersatzkasse	3,33	1,32	46,00
2.	Techniker Krankenkasse	3,72	1,59	43,00
3.	Deutsche BKK	3,72	1,58	47,00
4.	KKH - Die Kaufmännische	4,25	1,61	51,00
5.	AOK - Die Gesundheitskasse (Hamburg)	4,76	1,96	51,00
6.	BARMER	5,06	1,70	50,00
7.	DAK	5,12	1,52	51,00
<i>Gesamt</i>		4,31	1,74	339

Anmerkung: Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-19.2: Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung)

uf31	uf30	uf29	uf28	uf27	uf26	uf25	uf24	uf23	uf22	uf21	uf20	uf19	uf18	uf17	uf1
,335	,527	,539	,376	,569	,550	,261	,234	,268	,472	,593	,365	,627	,537	,593	uf1
,147	,345	,391	,194	,396	,353	,205	,111	,264	,349	,398	,308	,448	,412	,360	uf2
,287	,486	,523	,388	,525	,530	,231	,250	,295	,493	,519	,385	,566	,486	,514	uf3
,272	,375	,403	,353	,338	,449	,201	,204	,182	,373	,371	,349	,335	,349	,398	uf4
,266	,369	,345	,247	,268	,310	,143	,159	,134	,355	,325	,247	,288	,237	,305	uf5
,263	,405	,446	,329	,539	,501	,259	,153	,268	,403	,501	,305	,539	,514	,514	uf7
,449	,558	,606	,419	,883	,677	,230	,320	,426	,588	,738	,414	,759	,766	,712	uf8
,321	,507	,569	,373	,576	,532	,236	,325	,310	,540	,633	,491	,742	,618	,686	uf9
,390	,544	,600	,392	,740	,609	,255	,278	,371	,570	,694	,420	,850	,702	,681	uf10
,444	,586	,643	,439	,733	,635	,229	,366	,420	,625	,744	,496	,867	,709	,732	uf11
,293	,314	,354	,108	,465	,423	,161	,213	,319	,345	,366	,340	,389	,501	,388	uf12
,196	,326	,315	,288	,334	,330	,349	,237	,202	,382	,372	,231	,326	,336	,317	uf13
,334	,429	,414	,302	,524	,470	,229	,340	,407	,422	,408	,341	,428	,513	,46	uf14
,358	,529	,569	,429	,575	,609	,233	,295	,327	,510	,593	,413	,635	,571	,582	uf15

Tab. E-20: Toleranz- und VIF-Werte der formativen ufosV2fa-Indikatoren

Item-Code	Toleranz	VIF
uf16ac	0,929	1,077
uf25ac	0,798	1,253
uf13ac	0,675	1,482
uf24a	0,669	1,495
uf23a	0,666	1,502
uf2ac	0,635	1,575
uf5a	0,62	1,612
uf3 1a	0,613	1,632
uf4a	0,591	1,692
uf20a	0,573	1,745
uf12ac	0,509	1,965
uf7ac	0,491	2,038
uf14ac	0,486	2,059
uf28a	0,47	2,126
uf1a	0,446	2,24
uf3a	0,402	2,49
uf26a	0,396	2,525
uf15a	0,393	2,543
uf22a	0,36	2,778
uf30a	0,334	2,991
uf9ac	0,315	3,17
uf21a	0,299	3,342
uf18a	0,287	3,489
uf17a	0,285	3,508
uf29a	0,26	3,852
uf10a	0,193	5,175
uf27a	0,191	5,236
uf19a	0,161	6,212
uf1 1a	0,151	6,642
uf8a	0,136	7,37

Anmerkung: Variance Inflation Factor (VIF).

Tab. E-21: Rotierte Komponentenmatrix zur Erstellung der Multikollinearitätsindizes zur formativen Usability-Skala ufosV2fa

Komponente	1	2	3	4	5	h2	MSA-Wert
Uf10	0,773	0,3	0,204	0,109	0,028	0,741	970
Uf18	0,732	0,4	0,036	0,084	0,065	0,709	960
Uf7	0,725	0,058	0,184	0,187	0,045	0,599	964
Uf9	0,725	0,347	-0,001	0,121	-0,034	0,661	947
Uf17	0,684	0,457	0,159	0,072	-0,08	0,713	959
Uf1	0,652	0,19	0,365	0,101	-0,087	0,612	973
Uf15	0,642	0,263	0,417	0,033	-0,044	0,659	971
Uf21	0,64	0,455	0,208	0,148	-0,087	0,688	976
Uf2	0,627	-0,015	0,131	0,17	0,148	0,461*	950
Uf26	0,589	0,408	0,249	0,178	-0,006	0,607	977
Uf3	0,577	0,177	0,497	0,158	0,04	0,638	953
Uf2	0,533	0,273	-0,161	0,154	0,517	0,676	909
Uf24	0,034	0,732	0,044	0,128	-0,017	0,556	948
Uf31	0,194	0,666	0,181	-0,079	0,094	0,529	955
Uf22	0,418	0,615	0,275	0,175	-0,031	0,660	966
Uf30	0,377	0,603	0,336	0,199	-0,115	0,671	953
Uf23	0,275	0,599	-0,074	-0,008	0,175	0,470*	954
Uf29	0,453	0,596	0,325	0,188	-0,17	0,730	941
Uf28	0,163	0,564	0,352	0,302	-0,304	0,652	931
Uf20	0,358	0,434	0,23	0,019	0,216	0,416*	947
Uf14	0,427	0,456	-0,107	0,294	0,333	0,599	936
Uf5	0,194	0,107	0,769	0,016	0,06	0,643	910
Uf25	0,177	0,036	0,092	0,792	-0,072	0,673	939
Uf13	0,231	0,197	0,111	0,696	0,203	0,630	933
Uf16	-0,093	-0,045	0,215	0,019	0,696	0,541	399
Uf4	0,22	0,207	0,65	0,177	0,208	0,588	934
$\Sigma=62,05$	42,28	5,81	5,44	4,46	4,05		
$\lambda_j \sigma=0,932$	10,995	1,510	1,414	1,161	1,053		

Anmerkung: (*) Kommunalität liegt unterhalb des kritischen Wertes $h^2 < 0,5$. Kommunalität (h_i^2), Eigenwert (λ_j), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % (Σ).

Tab. E-22: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte

<i>Latente Variable</i>	<i>Item</i>	<i>ufosV2ra-Modell (reflektive Usability)</i>			<i>ufosV2fa-Modell (formative Usability)</i>		
		<i>Gewicht bzw. Ladung</i>	<i>SE</i>	<i>t-Wert</i>	<i>Gewicht bzw. Ladung</i>	<i>SE</i>	<i>t-Wert</i>
Usability - ufosV2ra (reflektiv)	ur1a	0,889	0,023	39,380			
	ur2a	0,841	0,039	21,743			
	ur3a	0,900	0,025	36,245			
	ur4a	0,887	0,022	41,237			
	ur5a	0,923	0,019	49,111			
	ur6a	0,938	0,014	65,944			
	ur8a	0,903	0,020	44,282			
Usability - ufosV2fa (formativ)	uf in1				0,634	0,078	8,182
	uf in2				0,699	0,069	10,184
	uf in3				0,276	0,071	3,896
	uf in4				0,181	0,063	2,857
	uf in5				-0,016	0,057	0,280
Vertrauen (reflektiv)	v1	0,922	0,020	47,146	0,921	0,020	46,809
	v2	0,914	0,021	44,318	0,913	0,023	39,843
	v4	0,897	0,031	29,387	0,900	0,026	35,352
Nutzervergnügen (reflektiv)	vg1	0,852	0,038	22,702	0,855	0,039	21,749
	vg2	0,947	0,011	90,142	0,946	0,012	79,980
	vg3	0,943	0,013	71,777	0,941	0,013	74,459
Involvement (reflektiv)	i1	0,918	0,178	5,158	0,918	0,193	4,751
	i2	0,877	0,203	4,326	0,877	0,214	4,092
	i3	0,920	0,176	5,232	0,919	0,198	4,649
	i4	0,934	0,164	5,682	0,934	0,189	4,936
	i5	0,942	0,185	5,092	0,942	0,174	5,429
Nutzungsintention (reflektiv)	ni1	0,929	0,021	44,469	0,929	0,023	41,192
	ni2	0,947	0,014	65,843	0,947	0,015	61,908
	ni3	0,937	0,016	57,021	0,937	0,019	49,717
Nutzerzufriedenheit (reflektiv)	z1	0,974	0,005	180,260	0,975	0,006	178,193
	z2	0,973	0,006	159,417	0,973	0,007	141,649
Wahrgen. Preis (Single-Item)	Subp	1			1		
Bekanntheitsgrad (Single-Item)	b1	1			1		
Kontrollvar.	co1	1			1		
Kontrollvar.	co2	1			1		

Tab. E-22: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte (Fortsetzung)

Moderatoreffekte							
Involv. x Nutzerzufr.	z1*i1	0,901	0,169	5,325	0,901	0,162	5,555
	z1*i2	0,975	0,190	5,122	0,975	0,164	5,958
	z1*i3	0,926	0,161	5,740	0,926	0,149	6,226
	z1*i4	0,807	0,170	4,764	0,807	0,136	5,950
	z1*i5	0,803	0,160	5,030	0,803	0,137	5,873
	z2*i1	0,828	0,149	5,538	0,828	0,187	4,415
	z2*i2	0,906	0,171	5,317	0,906	0,178	5,081
	z2*i3	0,865	0,149	5,800	0,865	0,161	5,382
	z2*i4	0,739	0,159	4,634	0,739	0,155	4,763
	z2*i5	0,743	0,150	4,963	0,743	0,160	4,644
Involv. x ufosV2ra	ur1a*i1	0,806	0,181	4,460			
	ur1a*i2	0,815	0,194	4,197			
	ur1a*i3	0,811	0,190	4,261			
	ur1a*i4	0,756	0,190	3,975			
	ur1a*i5	0,775	0,181	4,272			
	ur2a*i1	0,836	0,168	4,973			
	ur2a*i2	0,852	0,188	4,531			
	ur2a*i3	0,835	0,182	4,587			
	ur2a*i4	0,792	0,165	4,790			
	ur2a*i5	0,792	0,173	4,569			
	ur3a*i1	0,842	0,167	5,037			
	ur3a*i2	0,861	0,178	4,832			
	ur3a*i3	0,857	0,174	4,918			
	ur3a*i4	0,798	0,178	4,495			
	ur3a*i5	0,792	0,178	4,460			
	ur4a*i1	0,782	0,162	4,818			
	ur4a*i2	0,809	0,178	4,547			
	ur4a*i3	0,796	0,172	4,619			
	ur4a*i4	0,743	0,168	4,416			
	ur4a*i5	0,741	0,169	4,397			
	ur5a*i1	0,839	0,164	5,116			
	ur5a*i2	0,868	0,180	4,813			
	ur5a*i3	0,863	0,177	4,880			
	ur5a*i4	0,793	0,170	4,672			
	ur5a*i5	0,806	0,168	4,792			
	ur6a*i1	0,851	0,169	5,034			
	ur6a*i2	0,870	0,186	4,692			
	ur6a*i3	0,862	0,179	4,832			
	ur6a*i4	0,796	0,180	4,416			
	ur6a*i5	0,809	0,172	4,701			
	ur8a*i1	0,816	0,165	4,943			
	ur8a*i2	0,841	0,172	4,897			
	ur8a*i3	0,838	0,171	4,913			
	ur8a*i4	0,767	0,171	4,476			
	ur8a*i5	0,771	0,163	4,736			

Tab. E-22: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: Ladungen und Gewichte (Fortsetzung)

Involv. x ufosV2fa	uf in1*i1	0,252	0,110	2,283
	uf in1*i2	0,269	0,118	2,278
	uf in1*i3	0,215	0,100	2,156
	uf in1*i4	0,139	0,095	1,472
	uf in1*i5	0,152	0,100	1,525
	uf in2*i1	0,002	0,101	0,020
	uf in2*i2	0,079	0,098	0,809
	uf in2*i3	0,027	0,093	0,286
	uf in2*i4	0,055	0,092	0,601
	uf in2*i5	0,045	0,090	0,501
	uf in3*i1	0,056	0,092	0,608
	uf in3*i2	0,090	0,102	0,882
	uf in3*i3	0,058	0,101	0,577
	uf in3*i4	0,013	0,100	0,126
	uf in3*i5	0,126	0,100	1,260
	uf in4*i1	-0,142	0,105	1,358
	uf in4*i2	-0,083	0,101	0,819
	uf in4*i3	0,005	0,089	0,057
	uf in4*i4	-0,060	0,091	0,661
	uf in4*i5	-0,101	0,099	1,019
	uf in5*i1	-0,013	0,092	0,145
	uf in5*i2	0,047	0,094	0,497
	uf in5*i3	0,011	0,096	0,110
	uf in5*i4	0,003	0,090	0,030
	uf in5*i5	0,021	0,091	0,234

Tab. E-23: Ergebnisse der PLS-Strukturmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: Pfadkoeffizienten

<i>Zusammenhang</i>	<i>Modell mit ufosV2ra (reflektive Usability)</i>		<i>Modell mit ufosV2fa (formative Usability)</i>	
	β	<i>t-Wert</i>	β	<i>t-Wert</i>
Usability → Nutzerzufriedenheit	0,717	29,025	0,735	29,141
Usability → Nutzungsintention	0,019	0,168	0,211	16,731
Usability → Vertrauen	0,484	6,006	0,611	8,119
Usability → Nutzungsvergnügen	0,610	10,939	0,654	10,251
Vertrauen → Nutzerzufriedenheit	0,180	3,069	0,085	1,281
Nutzungsvergnügen → Nutzerzufriedenheit	0,110	1,803	0,106	1,436
Nutzerzufriedenheit → Nutzungsintention	0,780	5,959	0,620	5,101
Bekanntheit → Vertrauen	0,058	0,670	0,023	0,293
Wahrgen. Preis → Nutzerzufriedenheit	0,045	0,845	0,081	1,608
Involvement → Nutzungsintention	0,077	1,021	0,077	1,110
Involvement x Usability → Nutzungsintention.(alt)	0,033	0,041	0,036	0,305
Involv. x Nutzerzufried. → Nutzungsintention.(alt)	0,037	0,324	0,015	0,198
Involvement x Usability → Nutzungsintention.(standardisierte Werte)	0,005	0,038	0,036	0,321
Involv. x Nutzerzufried. → Nutzungsintention. (standardisierte Werte)	0,037	0,308	0,015	0,197
Involvement x Usability → Nutzungsintention. (originale werte)	0,040	0,089	0,294	1,324
Involv. x Nutzerzufried. → Nutzungsintention. (originale werte)	0,048	0,111	-0,206	0,652

Anmerkung: Pfadkoeffizient (β). Der t-Wert der t-Statistik in PLS folgt nur annähernd der t-Verteilung

Tab. E-24: Kennwerte der PLS-Messmodelle für das nomologische Modell zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability

<i>Latente Variable</i>	<i>ufosV2ra-Modell (reflektive Usability)</i>		<i>ufosV2fa-Modell (formative Usability)</i>	
	<i>AVE</i>	<i>Composite Reliability</i>	<i>AVE</i>	<i>Composite Reliability</i>
Usability - ufosV2ra	0,805	0,966		
Nutzervergnügen	0,837	0,938	0,837	0,939
Vertrauen	0,830	0,936	0,830	0,936
Involvement	0,843	0,964	0,843	0,964
Nutzungsintention	0,879	0,956	0,879	0,956
Nutzerzufriedenheit	0,947	0,973	0,948	0,973

Anmerkung: Average Variance Extracted (AVE). Die Werte für das AVE-Maß sowie für die Composite Reliability werden getrennt für die Verwendung der reflektiven Usability-Skala ufosV2ra sowie für die formative Usability-Skala ufosV2fa angegeben.

Tab. E-25: Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability-Skala ufosV2ra mit Quadratwurzel des AVE-Maßes

	<i>Wahrg. Preis</i>	<i>Vertrauen</i>	<i>Usability * Involv.</i>	<i>Usability reflektiv</i>	<i>Nutzungsintention</i>	<i>Nutzerzufriedenh.</i>	<i>Nutzer- vergnügen</i>	<i>Involvement</i>	<i>Bekanntheit</i>
<i>Bekanntheit</i>	0,143	0,093	-0,058	0,073	0,197	0,141	0,091	0,114	
<i>Involvement</i>	0,083	0,096	-0,069	-0,016	0,089	0,023	0,063	0,918	
<i>Nutzer- vergnügen</i>	-0,076	0,388	0,039	0,609	0,582	0,614	0,915		
<i>Nutzerzufriedenheit</i>	-0,017	0,568	0,053	0,870	0,804	0,974			
<i>Nutzungsintention</i>	-0,032	0,584	0,070	0,706	0,938				
<i>Usability reflektiv</i>	-0,046	0,488	0,036	0,898					
<i>Usability refl. * Involvement</i>	0,051	0,062	0,815						
<i>Vertrauen</i>	-0,108	0,911							
<i>Wahrg. Preis</i>									

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-26: Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability-Skala ufos V2fa mit Quadratwurzel des AVE-Maßes

	<i>Wahrgen. Preis</i>	<i>Vertrauen</i>	<i>Usability * Invol.</i>	<i>Usability formativ</i>	<i>Nutzungsintention</i>	<i>Nutzerzufriedenheit</i>	<i>Nutzer- vergnügen</i>	<i>Involvement</i>	<i>Bekanntheit</i>
<i>Bekanntheit</i>	0,143	0,093	-0,081	0,105	0,197	0,141	0,091	0,114	
<i>Involvement</i>	0,083	0,097	0,014	-0,000	0,089	0,023	0,063	0,918	
<i>Nutzer- vergnügen</i>	-0,075	0,387	0,037	0,661	0,581	0,613	0,915		
<i>Nutzerzu- friedenheit</i>	-0,017	0,568	0,105	0,845	0,804	0,974			
<i>Nutzungs- intention</i>	-0,032	0,584	0,116	0,734	0,938				
<i>Usability formativ</i>	-0,115	0,605	0,044						
<i>Usability * Involvement</i>	0,037	0,048	0,383						
<i>Vertrauen</i>	-0,108	0,911							
<i>Wahrgen. Preis</i>									

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-27: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra

	<i>Bekanntheit</i>	<i>Involvement</i>	<i>Nutzervergnügen</i>	<i>Nutzerzufriedenheit</i>	<i>Nutzungsintention</i>	<i>Usability reflektiv</i>
<i>b1</i>	1,000	0,114	0,092	0,141	0,197	0,074
<i>inv1</i>	0,098	0,917	0,052	0,005	0,061	-0,049
<i>inv2</i>	0,108	0,877	0,048	-0,037	0,022	-0,067
<i>inv3</i>	0,103	0,919	0,089	0,016	0,083	-0,016
<i>inv4</i>	0,103	0,934	0,020	0,031	0,084	-0,003
<i>inv5</i>	0,113	0,942	0,072	0,039	0,108	0,006
<i>z1</i>	0,161	0,019	0,615	0,974	0,801	0,841
<i>z2</i>	0,113	0,026	0,581	0,973	0,765	0,854
<i>ni1</i>	0,205	0,072	0,533	0,768	0,929	0,657
<i>ni2</i>	0,164	0,089	0,538	0,743	0,947	0,649
<i>ni3</i>	0,186	0,092	0,567	0,751	0,937	0,669
<i>ur1a</i>	0,064	0,028	0,585	0,795	0,687	0,888
<i>ur2a</i>	0,030	-0,047	0,450	0,684	0,546	0,842
<i>ur3a</i>	0,048	-0,027	0,548	0,753	0,586	0,901
<i>ur4a</i>	0,061	-0,019	0,501	0,748	0,596	0,887
<i>ur5a</i>	0,075	-0,002	0,563	0,798	0,628	0,923
<i>ur6a</i>	0,066	0,004	0,611	0,852	0,695	0,938
<i>ur8a</i>	0,112	-0,048	0,553	0,819	0,657	0,902
<i>v1</i>	0,082	0,033	0,368	0,547	0,554	0,466
<i>v2</i>	0,103	0,103	0,368	0,524	0,545	0,463
<i>v4</i>	0,070	0,135	0,323	0,480	0,496	0,400
<i>vg1</i>	0,041	0,043	0,852	0,433	0,417	0,433
<i>vg2</i>	0,087	0,060	0,947	0,610	0,557	0,603
<i>vg3</i>	0,112	0,068	0,943	0,615	0,600	0,611
<i>subp1</i>	0,144	0,084	-0,076	-0,017	-0,033	-0,047

Anmerkung: Fett gedruckt dargestellt sind die Ladungen der Indikatoren auf ihren jeweiligen latenten Variablen.

Tab. E-27: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra (Fortsetzung)

	<i>Usability * Involvement</i>	<i>Vertrauen</i>	<i>Wahrg. Preis</i>
<i>b1</i>	-0,058	0,094	0,144
<i>inv1</i>	-0,044	0,095	0,105
<i>inv2</i>	-0,036	0,030	0,052
<i>inv3</i>	-0,056	0,110	0,056
<i>inv4</i>	-0,075	0,054	0,115
<i>inv5</i>	-0,080	0,111	0,054
<i>z1</i>	0,046	0,568	-0,020
<i>z2</i>	0,058	0,539	-0,013
<i>ni1</i>	0,062	0,557	-0,039
<i>ni2</i>	0,068	0,548	-0,027
<i>ni3</i>	0,069	0,540	-0,026
<i>ur1a</i>	0,055	0,495	-0,030
<i>ur2a</i>	0,060	0,339	-0,058
<i>ur3a</i>	0,028	0,375	-0,032
<i>ur4a</i>	0,017	0,405	-0,040
<i>ur5a</i>	-0,009	0,438	-0,057
<i>ur6a</i>	0,038	0,485	-0,052
<i>ur8a</i>	0,043	0,505	-0,029
<i>v1</i>	0,100	0,922	-0,084
<i>v2</i>	0,022	0,914	-0,128
<i>v4</i>	0,046	0,897	-0,084
<i>vg1</i>	0,009	0,248	-0,011
<i>vg2</i>	0,037	0,381	-0,081
<i>vg3</i>	0,056	0,412	-0,102
<i>subp1</i>	0,052	-0,109	1,000

Tab. E-28: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2fa

	<i>Bekanntheit</i>	<i>Involvement</i>	<i>Nutzer- vergnügen</i>	<i>Nutzerzufriedenheit</i>
<i>b1</i>	1,000	0,114	0,091	0,141
<i>inv1</i>	0,098	0,917	0,052	0,005
<i>inv2</i>	0,108	0,877	0,048	- 0,037
<i>inv3</i>	0,103	0,919	0,089	0,016
<i>inv4</i>	0,103	0,934	0,020	0,031
<i>inv5</i>	0,113	0,942	0,072	0,039
<i>vg1</i>	0,041	0,043	0,855	0,434
<i>vg2</i>	0,087	0,060	0,946	0,610
<i>vg3</i>	0,112	0,068	0,941	0,615
<i>z1</i>	0,161	0,019	0,614	0,974
<i>z2</i>	0,113	0,026	0,580	0,973
<i>ni1</i>	0,205	0,072	0,532	0,768
<i>ni2</i>	0,164	0,089	0,537	0,743
<i>ni3</i>	0,186	0,092	0,566	0,751
<i>uf ind1</i>	0,013	0,011	0,414	0,614
<i>uf ind2</i>	0,107	0,011	0,499	0,531
<i>uf ind3</i>	0,120	- 0,086	0,098	0,231
<i>uf ind4</i>	- 0,015	0,047	0,093	0,124
<i>uf ind5</i>	- 0,148	- 0,052	0,105	- 0,047
<i>v1</i>	0,082	0,033	0,367	0,547
<i>v2</i>	0,103	0,103	0,367	0,524
<i>v4</i>	0,070	0,135	0,322	0,480
<i>subp1</i>	0,144	0,084	- 0,076	- 0,017

Tab. E-28: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2fa (Fortsetzung)

	<i>Nutzungs- intention</i>	<i>Usability formativ</i>	<i>Vertrauen</i>	<i>Wahrgen. Preis</i>
ni1	0,929	0,690	0,556	- 0,039
ni2	0,947	0,687	0,548	- 0,027
ni3	0,937	0,693	0,539	- 0,026
z1	0,801	0,831	0,568	- 0,020
z2	0,765	0,815	0,538	- 0,013
uf ind1	0,464	0,644	0,289	- 0,028
uf ind2	0,532	0,704	0,450	- 0,130
uf ind3	0,180	0,243	0,291	0,009
uf ind4	0,120	0,170	0,193	- 0,027
uf ind5	- 0,059	0,034	- 0,030	- 0,097
v1	0,554	0,569	0,921	- 0,084
v2	0,545	0,560	0,913	- 0,128
v4	0,496	0,525	0,900	- 0,084
vg1	0,417	0,498	0,248	- 0,011
vg2	0,557	0,651	0,380	- 0,081
vg3	0,600	0,648	0,412	- 0,102
aes1	0,538	0,669	0,407	- 0,100
aes2	0,647	0,811	0,508	- 0,147
b1	0,197	0,105	0,094	0,144
inv1	0,061	- 0,027	0,095	0,105
inv2	0,022	- 0,046	0,030	0,052
inv3	0,083	0,014	0,111	0,056
inv4	0,084	- 0,010	0,054	0,115
inv5	0,108	0,022	0,112	0,054
subp1	- 0,033	- 0,115	- 0,109	1,000

Tab. E-29: Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum nomologischen Modell

Latente exogene Variable	Latente endogene Variable	Modell mit ufosV2ra (reflektive Usability)		Modell mit ufosV2fa (formative Usability)	
		f^2	Beurteilung	f^2	Beurteilung
Usability	Nutzungsintention	0,009	keiner	0,039	gering
	Nutzerzufriedenheit	1,367	hoch	0,828	hoch
	Nutzungsvergnügen	0,592	hoch	0,774	hoch
	Vertrauen	0,307	mittel	0,566	hoch
Vertrauen	Nutzerzufriedenheit	0,116	gering	0,026	gering
Nutzungsvergnügen	Nutzerzufriedenheit	0,024	gering	0,007	keiner
Nutzerzufriedenheit	Nutzungsintention	0,421	hoch	0,341	hoch
Involvement	Nutzungsintention	0,017	gering	0,024	gering
Bekanntheit	Vertrauen	0,319	mittel	0,012	Keiner
Wahrgen. Preis	Nutzerzufriedenheit	0,010	keiner	0,026	gering

Anmerkung: Effektgröße (f^2). Die Effektgrößen werden getrennt für das ufosV2r-Modell (Verwendung der reflektiven Usability-Skala ufosV2r) sowie für das ufosV2f-Modell (Verwendung der formativen Usability-Skala ufosV2f) angegeben.

Tab. E-30: Korrelationen der LR-Prädiktoren

Nutzungsint. Webs.2	Nutzungsint. Webs.1	ufosV2fa Webs.2	ufosV2fa Webs.1	ufosV2ra Webs.2	ufosV2ra Webs.1	Internetrecherche	Onlineeinkäufe	Einkommen	Geschlecht	Alter
,025	-,021	,092	-,035	-,007	,018	-,032	,002	,439(**)	,103	1
-,006	-,148(*)	,014	,104	,080	-,017	,046	,117	,172(**)	1	
-,063	-,143(*)	,051	,023	,003	-,041	-,003	,219(**)	1		
-,045	-,084	-,014	,051	-,025	-,111	,184(**)	1			
,086	,028	-,038	,087	,048	,058	1				
-,180(**)	,551(**)	,102	,001	-,189(**)	1					
,767(**)	-,168(**)	,011	,030	1						
-,026	,050	-,190(**)	1							
-,005	,024	1								
-,076	1									

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant

Tab. E-31: Toleranz und Varance-Inflation-Factors (VIF) der LR-Prädiktoren

Prädiktor	Schritt 1		Schritt 2a		Schritt 2b		Schritt 2c	
	Tol.	VIF	Tol.	VIF	Tol.	VIF	Tol.	VIF
Alter	,796	1,257	,795	1,258	,791	1,264	,791	1,264
Einkommen	,748	1,337	,747	1,338	,747	1,338	,733	1,364
Geschlecht	,960	1,041	,953	1,049	,951	1,052	,944	1,059
Online-Einkäufe	,902	1,109	,886	1,129	,901	1,110	,897	1,115
Internetrecherchen	,963	1,038	,952	1,050	,957	1,045	,952	1,051
ufosV2raWebsite1			,942	1,061				
ufosV2raWebsite2			,951	1,052				
ufosV2faWebsite1					,947	1,056		
ufosV2faWebsite2					,957	1,045		
Nutzungsintention Website1							,947	1,056
Nutzungsintention Website2							,973	1,027
Maximaler Konditionsindex	13,437		18,536		13,572		16,879	

Anmerkung: Toleranz (Tol.), Variance Inflation Factor (VIF).

Tab. E-32: Teilnehmer mit standardisierten Residuen $zRedsid > 1$ in der LR

Teilnehmer	$zResiduum$
38-7-2	-3,68
76-6-3	1,168
79-6-7	-1,177
81-7-2	-11,668
87-1-3	-2,748
225-3-1	1,112
283-5-7	-2,636
627-1-6	-2,683
628-1-7	-1,223
636-6-2	2,169
642-7-2	1,673
692-4-1	1,388

Anmerkung: standardisiertes Residuum der logistischen Regression mit dem Kriterium *Kauf in Shop 1 vs. Gutschein* ($zRedsid1$), für das Kriterium *Kauf in Shop 1 vs. Kauf in Shop 2* ($zRedsid2$), für das Kriterium *Kauf in Shop 2 vs. Gutschein* ($zRedsid3$).

Tab. E-33: Ausreißer in der LR

Teilnehmer-code	Usaility reflektiv Website1	Usaility reflektiv Website2	SD Website1	SD Website2	$zResiduum$
283-5-7	1,43	0,86	1,37	0,87	-2,636
628-1-7	1,71	3,43	1,31	1,16	-1,223
642-7-2	1,86	3,14	2,35	2,35	1,673

Anmerkung: standardisiertes Residuum der logistischen Regression mit dem Kriterium *Auswahl Website1 vs. Auswahl Website2*.

Schritt 1 der LR: Hinzufügen der Variablen zur Demographie, Vorerfahrung und Kontrolle

Tab. E-34: Statistiken zur globalen Modellanpassung

1. Likelihood-Quotiententest:				
<i>Modell</i>	<i>-2 Log-Likelihood</i>	χ^2 -Wert	<i>df</i>	<i>Sign.</i>
<i>Nur konstanter Term</i>	327,776			
<i>Endgültig</i>	324,228	3,548	5	,616
2. χ^2 - bzw. Devianz-Statistik:				
	χ^2 -Wert	<i>df</i>	<i>Sign.</i>	
<i>Pearson</i>	234,608	225	,316	
<i>Abweichung</i>	321,220	225	,000	
3. Pseudo-R ² -Statistiken:				
<i>Cox und Snell</i>	<i>Nagelkerke</i>	<i>McFadden</i>		
0,015	0,020	0,011		

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R²).

Tab. E-35: Klassifikationsmatrix sequentielle LR Schritt 1

Bebachtet	Vorhergesagt		
	<i>Auswahl Website 1</i>	<i>Auswahl Website 2</i>	<i>% richtig</i>
<i>Auswahl Website 1</i>	104	25	80,6
<i>Auswahl Website 2</i>	86	25	22,5
<i>% insgesamt</i>			53,8

Tab. E-36: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 1

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	χ^2 -Wert	<i>df</i>	<i>Sign.</i>
Konstanter Term	324,228	,000	0	.
Alter	325,135	,907	1	,341
Einkommen	324,827	,599	1	,439
Geschlecht	324,255	,027	1	,868
Online-Einkäufe	324,733	,506	1	,477
Internetrecherchen	326,602	2,374	1	,123

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-37 Parameterschätzer sequentielle LR Schritt 1

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Auswahl Website 1 vs. Website 2	Konstante	-,463	,652	,504	1	,478	,630		
	Alter	,024	,026	,024	1	,346	1,024	,974	1,077
	Geschlecht	-,045	,273	,027	1	,868	,956	,560	1,631
	Einkommen	-,142	,184	,596	1	,440	,867	,605	1,245
	Online-Einkäufe	,007	,010	,503	1	,478	1,007	,987	1,028
	Internetrecherchen	-,009	,007	1,855	1	,173	,991	,978	1,004

Anmerkung: Regressionskoeffizient (B), Standardfehler des Regressionskoeffizienten (SE), Wert der Wald-Statistik (Wald), Anzahl der Freiheitsgrade für die Wald-Statistik (df), Signifikanz der Regressionskoeffizienten (Sign.), Exponentialfunktionswert des Regressionskoeffizienten (Exp(B)).

Schritt 2a der LR: Hinzufügen der ufosV2ra-Scores (Usability reflektiv) für beide Websites

Tab. E-38: Statistiken zur globalen Modellanpassung

1. Likelihood-Quotiententest:				
Modell	-2 Log-Likelihood	χ^2 -Wert	df	Sign.
Nur konstanter Term	331,359			
Endgültig	120,324	211,035	7	,000
2. χ^2 - bzw. Devianz-Statistik:				
	χ^2 -Wert	df	Sign.	
Pearson	444,605	232	,000	
Abweichung	120,324	232	1,000	
3. Pseudo-R ² -Statistiken:				
Cox und Snell	Nagelkerke			
0,585	0,781			

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R²).

Tab. E-39: Klassifikationsmatrix sequentielle LR Schritt 2a

Bebachtet	Vorhergesagt		
	Auswahl Website 1	Auswahl Website 2	% richtig
Auswahl Website 1	113	16	87,6
Auswahl Website 2	9	102	91,9
% insgesamt			89,6

Tab. E-40: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2a

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	χ^2 -Wert	<i>Df</i>	<i>Sign.</i>
Konstanter Term	120,324	,000	0	.
Alter	124,799	4,474	1	,034
Geschlecht	121,589	1,265	1	,261
Einkommen	122,949	2,624	1	,105
Online-Einkäufe	120,833	,509	1	,476
Internetrecherchen	125,615	5,291	1	,021
Usability ufosV2ra Website 1	177,889	57,564	1	,000
Usability ufosV2ra Website 2	267,844	147,520	1	,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-41: Parameterschätzer sequentielle LR Schritt 2a

<i>Block</i>	<i>Regressor</i>	<i>B</i>	<i>SE</i>	<i>Wald</i>	<i>df</i>	<i>Sign.</i>	<i>Exp(B)</i>	<i>95 % Konfidenzintervall Exp(B)</i>	
								<i>Untergrenze</i>	<i>Obergrenze</i>
<i>Auswahl Website 1 vs. Website 2</i>	Konst. Term	-4,549	1,780	6,530	1	,011	,011		
	Alter	,126	,058	4,681	1	,031	1,134	1,012	1,270
	Geschlecht	,550	,493	1,243	1	,265	1,733	,659	4,560
	Einkommen	-,542	,333	2,643	1	,104	,582	,303	1,118
	Online-Einkäufe	,012	,017	,496	1	,481	1,012	,979	1,045
	Internetrecherchen	-,017	,010	2,937	1	,087	,983	,964	1,002
	ufosV2ra Website 1	-1,298	,228	32,294	1	,000	,273	,174	,427
	ufosV2ra Website 2	1,724	,240	51,501	1	,000	5,605	3,500	8,974

Anmerkung: Regressionskoeffizient (B), Standardfehler des Regressionskoeffizienten (SE), Wert der Wald-Statistik (Wald), Anzahl der Freiheitsgrade für die Wald-Statistik (df), Signifikanz der Regressionskoeffizienten (Sign.), Exponentialfunktionswert des Regressionskoeffizienten (Exp(B)).

Schritt 2b der LR: Hinzufügen der ufosV2fa-Scores (Usability formativ) für beide Websites

Tab. E-42: Statistiken zur globalen Modellanpassung sequentielle LR Schritt 2b

1. Likelihood-Quotiententest:				
<i>Modell</i>	<i>-2 Log-Likelihood</i>	χ^2 -Wert	<i>df</i>	<i>Sign.</i>
<i>Nur konstanter Term</i>	331,359			
<i>Endgültig</i>	92,175	239,185	7	,000
2. χ^2 - bzw. Devianz-Statistik:				
	χ^2 -Wert	<i>df</i>	<i>Sign.</i>	
<i>Pearson</i>	226,086	232	,597	
<i>Abweichung</i>	92,175	232	1,000	
3. Pseudo-R ² -Statistiken:				
<i>Cox und Snell</i>	<i>Nagelkerke</i>			
,631	,843			

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R²).

Tab. E-43: Klassifikationsmatrix sequentielle LR Schritt 2b

<i>Bebachtet</i>	<i>Vorhergesagt</i>		
	<i>Auswahl Website 1</i>	<i>Auswahl Website 2</i>	<i>% richtig</i>
<i>Auswahl Website 1</i>	121	8	93,8
<i>Auswahl Website 2</i>	7	104	93,7
<i>% insgesamt</i>			93,8

Tab. E-44: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2b

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	χ^2 -Wert	<i>df</i>	<i>Sign.</i>
Konstanter Term	92,175	,000	0	.
Alter	101,431	9,256	1	,002
Geschlecht	92,279	,104	1	,747
Einkommen	94,290	2,115	1	,146
Online-Einkäufe	93,726	1,551	1	,213
Internetrecherchen	98,400	6,225	1	,013
Usability ufosV2fa Website 1	172,052	79,877	1	,000
Usability ufosV2fa Website 2	252,395	160,220	1	,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-45: Parameterschätzer sequentielle LR Schritt 2b

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Unterer Wert	Oberer Wert
Auswahl Website 1 vs. Website 2	Konstanter Term	-4,042	1,457	7,692	1	,006	,018		
	Alter	,190	,059	10,536	1	,001	1,209	1,078	1,356
	Geschlecht	,178	,550	,104	1	,747	1,194	,406	3,511
	Einkommen	-,573	,392	2,135	1	,144	,564	,261	1,216
	Online-Einkäufe	,027	,022	1,438	1	,231	1,027	,983	1,073
	Internet-recherchen	-,021	,012	2,965	1	,085	,980	,957	1,003
	ufosV2fa Website 1	-3,034	,519	34,233	1	,000	,048	,017	,133
	ufosV2fa Website 2	3,778	,597	40,067	1	,000	43,709	13,570	140,78

Anmerkung: Regressionskoeffizient (B), Standardfehler des Regressionskoeffizienten (SE), Wert der Wald-Statistik (Wald), Anzahl der Freiheitsgrade für die Wald-Statistik (df), Signifikanz der Regressionskoeffizienten (Sign.), Exponentialfunktionswert des Regressionskoeffizienten (Exp(B)).

Schritt 2c der LR: Hinzufügen der Variable Kaufintention für beide Websites

Tab. E-46: Statistiken zur globalen Modellanpassung sequentielle LR Schritt 2c

1. Likelihood-Quotiententest:				
Modell	-2 Log-Likelihood	χ^2 -Wert	df	Sign.
Nur konstanter Term	331,359			
Endgültig	99,296	232,063	7	,000
2. χ^2 - bzw. Devianz-Statistik:				
	χ^2 -Wert	df	Sign.	
Pearson	340,042	231	,000	
Abweichung	99,296	231	1,000	
3. Pseudo-R ² -Statistiken:				
Cox und Snell	Nagelkerke			
,620	,828			

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R²).

Tab. E-47: Klassifikationsmatrix sequentielle LR Schritt 2c

Bebachtet	Vorhergesagt		
	Auswahl Website 1	Auswahl Website 2	% richtig
Auswahl Website 1	120	9	93,0
Auswahl Website 2	8	103	92,8
% insgesamt			92,9

Tab. E-48: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren sequentielle LR Schritt 2c

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	χ^2 -Wert	<i>df</i>	<i>Sign.</i>
Konstanter Term	99,296	,000	0	.
Alter	100,110	,814	1	,367
Geschlecht	100,577	1,281	1	,258
Einkommen	100,121	,825	1	,364
Online-Einkäufe	102,151	2,855	1	,091
Internetrecherchen	109,519	10,222	1	,001
Nutzungsintention Website 1	172,793	73,497	1	,000
Nutzungsintention Website 2	282,039	182,743	1	,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-49: Parameterschätzer sequentielle LR Schritt 2c

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Auswahl Website 1 vs. Website 2	Konstanter Term	-2,215	1,610	1,894	1	,169	,109		
	Alter	,054	,058	,869	1	,351	1,056	,942	1,183
	Geschlecht	,652	,581	1,259	1	,262	1,919	,614	5,993
	Einkommen	-,322	,353	,831	1	,362	,725	,363	1,448
	Online-Einkäufe	,028	,017	2,698	1	,100	1,028	,995	1,062
	Internet-recherchen	-,038	,022	2,993	1	,084	,963	,922	1,005
	Nutzungsint. Website 1	-1,806	,340	28,269	1	,000	,164	,084	,320
	Nutzungsint. Website 2	2,175	,335	42,043	1	,000	8,799	4,560	16,979

Anmerkung: Regressionskoeffizient (B), Standardfehler des Regressionskoeffizienten (SE), Wert der Wald-Statistik (Wald), Anzahl der Freiheitsgrade für die Wald-Statistik (df), Signifikanz der Regressionskoeffizienten (Sign.), Exponentialfunktionswert des Regressionskoeffizienten (Exp(B)).

Tab. E-50: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei reduziertem Datensatz: Ladungen und Gewichte

<i>Latente Variable</i>	<i>Item</i>	<i>ufosV2ra-Modell (reflektive Usability)</i>			<i>ufosV2fa-Modell (formative Usability)</i>		
		<i>Gewicht bzw. Ladung</i>	<i>SE</i>	<i>t-Wert</i>	<i>Gewicht bzw. Ladung</i>	<i>SE</i>	<i>t-Wert</i>
Usability - ufosV2ra (reflektiv)	ur1a	0,889	0,023	39,372			
	ur2a	0,844	0,040	20,926			
	ur3a	0,925	0,015	61,668			
	ur4a	0,871	0,032	27,521			
	ur5a	0,924	0,018	50,192			
	ur6a	0,937	0,012	76,206			
	ur8a	0,911	0,015	59,976			
Usability - ufosV2fa (formativ)	uf in1				0,5982	0,073	8,210
	uf in2				0,6863	0,066	10,410
	uf in3				0,2685	0,070	3,831
	uf in4				0,2354	0,073	3,237
	uf in5				-0,0479	0,064	0,753
Vertrauen (reflektiv)	v1	0,941	0,016	57,326	0,934	0,018	51,810
	v2	0,935	0,017	55,959	0,941	0,017	54,837
	v4	0,925	0,017	53,062	0,926	0,018	52,869
Nutzervergnügen (reflektiv)	vg1	0,873	0,038	23,259	0,946	0,011	89,261
	vg2	0,948	0,011	85,323	0,941	0,013	72,761
	vg3	0,943	0,013	72,147	0,876	0,031	28,067
Involvement (reflektiv)	i1	0,955	0,095	10,054	0,955	0,170	5,624
	i2	0,947	0,086	11,019	0,947	0,171	5,526
	i3	0,966	0,084	11,509	0,966	0,163	5,916
	i4	0,981	0,109	8,984	0,981	0,163	6,004
	i5	0,974	0,109	8,923	0,974	0,159	6,119
Nutzungsintention (reflektiv)	ni1	0,942	0,020	47,174	0,942	0,019	49,753
	ni2	0,950	0,016	61,321	0,930	0,022	41,408
	ni3	0,930	0,020	46,528	0,950	0,014	68,800
Nutzerzufriedenheit (reflektiv)	z1	0,973			0,971	0,009	108,906
	z2	0,972			0,974	0,006	152,817
Wahrgen. Preis (Single-Item)	Subp	1			1		
Bekanntheitsgrad (Single-Item)	b1	1			1		
Kontrollvar.	co1	1			1		
Kontrollvar.	co2	1			1		

Tab. E-50: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei reduziertem Datensatz: Ladungen und Gewichte
(Fortsetzung)

Moderatoreffekte							
Involv. x Nutzerzufr.	z1*i1	0,946	0,136	6,976	0,897	0,171	5,241
	z1*i2	0,953	0,146	6,527	0,910	0,171	5,338
	z1*i3	0,964	0,145	6,648	0,933	0,162	5,763
	z1*i4	0,968	0,160	6,042	0,926	0,173	5,353
	z1*i5	0,970	0,158	6,130	0,934	0,173	5,409
	z2*i1	0,897	0,149	6,023	0,946	0,201	4,709
	z2*i2	0,910	0,152	5,987	0,953	0,198	4,813
	z2*i3	0,933	0,154	6,073	0,964	0,190	5,078
	z2*i4	0,926	0,166	5,569	0,968	0,201	4,813
	z2*i5	0,934	0,164	5,692	0,970	0,199	4,884
Involv. x ufosV2ra	ur1a*i1	0,917	0,023	40,677			
	ur1a*i2	0,919	0,020	45,670			
	ur1a*i3	0,923	0,018	50,691			
	ur1a*i4	0,933	0,016	58,006			
	ur1a*i5	0,931	0,017	54,633			
	ur2a*i1	0,916	0,024	38,841			
	ur2a*i2	0,899	0,026	34,483			
	ur2a*i3	0,911	0,022	41,985			
	ur2a*i4	0,929	0,016	59,153			
	ur2a*i5	0,912	0,019	48,748			
	ur3a*i1	0,932	0,018	50,985			
	ur3a*i2	0,931	0,016	57,836			
	ur3a*i3	0,940	0,015	64,849			
	ur3a*i4	0,934	0,015	61,552			
	ur3a*i5	0,928	0,015	62,124			
	ur4a*i1	0,900	0,031	29,559			
	ur4a*i2	0,904	0,030	30,087			
	ur4a*i3	0,912	0,028	33,184			
	ur4a*i4	0,920	0,022	42,436			
	ur4a*i5	0,913	0,022	42,233			
	ur5a*i1	0,929	0,023	41,244			
	ur5a*i2	0,920	0,021	43,179			
	ur5a*i3	0,934	0,018	52,083			
	ur5a*i4	0,937	0,015	61,293			
	ur5a*i5	0,934	0,015	63,528			
	ur6a*i1	0,940	0,017	55,572			
	ur6a*i2	0,942	0,013	72,572			
	ur6a*i3	0,953	0,011	88,420			
	ur6a*i4	0,949	0,011	86,107			
	ur6a*i5	0,948	0,011	83,999			
	ur8a*i1	0,915	0,020	45,743			
	ur8a*i2	0,927	0,016	59,177			
	ur8a*i3	0,933	0,015	61,653			
	ur8a*i4	0,931	0,015	63,477			
	ur8a*i5	0,922	0,016	56,342			

Tab. E-50: Kennwerte des PLS-Messmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei reduziertem Datensatz: Ladungen und Gewichte
(Fortsetzung)

Involv. x ufosV2fa	uf in1*i1		0,125	0,019	6,544
	uf in1*i2		0,128	0,019	6,787
	uf in1*i3		0,126	0,019	6,514
	uf in1*i4		0,127	0,020	6,505
	uf in1*i5		0,131	0,020	6,599
	uf in2*i1		0,12	0,020	6,288
	uf in2*i2		0,137	0,022	6,392
	uf in2*i3		0,135	0,022	6,319
	uf in2*i4		0,137	0,022	6,230
	uf in2*i5		0,138	0,022	6,209
	uf in3*i1		0,027	0,027	1,037
	uf in3*i2		0,030	0,026	1,170
	uf in3*i3		0,030	0,026	1,163
	uf in3*i4		0,028	0,027	1,073
	uf in3*i5		0,034	0,027	1,280
	uf in4*i1		0,059	0,022	2,681
	uf in4*i2		0,06	0,022	2,777
	uf in4*i3		0,063	0,022	2,925
	uf in4*i4		0,063	0,022	2,975
	uf in4*i5		0,065	0,022	3,019
	uf in5*i1		-0,023	0,025	0,918
	uf in5*i2		-0,027	0,026	1,086
	uf in5*i3		-0,025	0,026	0,994
	uf in5*i4		-0,022	0,027	0,854
	uf in5*i5		-0,024	0,027	0,926

Tab. E-51: Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability-Skala ufosV2ra mit Quadratwurzel des AVE-Maßes (eingeschränkter Datensatz)

	<i>Wahrg. Preis</i>	<i>Vertrauen</i>	<i>Usability * Invol.</i>	<i>Usability reflektiv</i>	<i>Nutzungsintention</i>	<i>Nutzerzufriedenh.</i>	<i>Nutzer-vernügen</i>	<i>Involvement</i>	<i>Bekanntheit</i>
<i>Bekanntheit</i>	0,057	0,009	0,098	-0,021	0,106	0,045	0,081	0,178	
<i>Involvement</i>	-0,054	0,128	0,759	-0,025	0,145	0,033	0,110	0,965	
<i>Nutzer-vernügen</i>	0,161	0,394	0,426	0,584	0,524	0,591	0,922		
<i>Nutzerzufriedenheit</i>	0,010	0,592	0,509	0,836	0,819	0,972			
<i>Nutzungsintention</i>	0,066	0,582	0,513	0,676	0,941				
<i>Usability reflektiv</i>	0,045	0,522	0,560	0,901					
<i>Usability refl. * Involvement</i>	-0,038	0,416	0,926						
<i>Vertrauen</i>	0,001	0,934							
<i>Wahrg. Preis</i>									

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-52: Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability-Skala ufos V2fa mit Quadratwurzel des AVE-Maßes (eingeschränkter Datensatz)

	Vertrauen	Nutzer- vergnügen	Usability * Invol.	Usability formativ	Wahrgen. Preis	Nutzungs- intention	Nutzerzu- friedenheit	Involvement	Bekannt- heit
<i>Bekanntheit</i>	0,009	0,081	-0,023	0,018	0,057	0,106	0,045	0,178	
<i>Involvement</i>	0,128	0,110	0,002	0,029	-0,054	0,145	0,033	0,965	
<i>Nutzerzu- friedenheit</i>	0,592	0,591	0,743	0,834	0,010	0,820	0,972		
<i>Nutzungs- intention</i>	0,582	0,523	0,668	0,705	0,066	0,941			
<i>Wahrgen. Preis</i>	0,001	0,161	0,013	0,028					
<i>Usability formativ</i>	0,614	0,642	0,885						
<i>Usab. form * Involvement</i>	0,521	0,561	0,448						
<i>Nutzerver- gnügen</i>	0,394	0,922							
<i>Vertrauen</i>	0,934								

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-53: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra bei eingeschränktem Datensatz

	<i>Bekanntheit</i>	<i>Involvement</i>	<i>Nutzungs- intention</i>	<i>Usability reflektiv</i>	<i>Vertrauen</i>	<i>Nutzerver- gnügen</i>
<i>b1</i>	1,000	0,178	0,081	0,045	0,106	-0,021
<i>inv1</i>	0,136	0,955	0,117	0,004	0,087	-0,055
<i>inv2</i>	0,169	0,947	0,072	-0,062	0,058	-0,116
<i>inv3</i>	0,166	0,966	0,098	-0,009	0,111	-0,074
<i>inv4</i>	0,194	0,981	0,107	0,076	0,185	0,014
<i>inv5</i>	0,175	0,974	0,118	0,057	0,169	0,015
<i>ni1</i>	0,098	0,121	0,508	0,789	0,942	0,641
<i>ni2</i>	0,092	0,131	0,507	0,772	0,950	0,631
<i>ni3</i>	0,108	0,158	0,463	0,751	0,930	0,635
<i>ur1a</i>	0,018	0,037	0,515	0,766	0,655	0,889
<i>ur2a</i>	-0,102	-0,101	0,417	0,639	0,484	0,844
<i>ur3a</i>	-0,058	-0,060	0,536	0,741	0,599	0,925
<i>ur4a</i>	-0,007	-0,017	0,486	0,727	0,577	0,871
<i>ur5a</i>	0,018	0,006	0,543	0,766	0,622	0,924
<i>ur6a</i>	-0,032	0,012	0,604	0,824	0,680	0,937
<i>ur8a</i>	0,011	-0,054	0,557	0,781	0,618	0,911
<i>v1</i>	0,009	0,059	0,341	0,557	0,562	0,475
<i>v2</i>	0,024	0,139	0,377	0,554	0,549	0,506
<i>v4</i>	-0,009	0,159	0,387	0,546	0,518	0,480
<i>vg1</i>	0,050	0,099	0,873	0,440	0,371	0,422
<i>vg2</i>	0,094	0,116	0,948	0,588	0,505	0,579
<i>vg3</i>	0,076	0,089	0,943	0,586	0,549	0,590
<i>subp1</i>	0,057	-0,054	0,161	0,010	0,066	0,045
<i>z1</i>	0,069	0,029	0,605	0,973	0,811	0,802
<i>z2</i>	0,017	0,035	0,545	0,972	0,782	0,823

Anmerkung: Fett gedruckt dargestellt sind die Ladungen der Indikatoren auf ihren jeweiligen latenten Variablen.

Tab. E-53: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der reflektiven Usability-Skala ufosV2ra bei eingeschränktem Datensatz (Fortsetzung)

	<i>Wahrg. Preis</i>	<i>Nutzer-zufriedenheit</i>
<i>b1</i>	0,009	0,057
<i>inv1</i>	0,130	-0,034
<i>inv2</i>	0,048	0,010
<i>inv3</i>	0,115	-0,048
<i>inv4</i>	0,138	-0,051
<i>inv5</i>	0,137	-0,085
<i>ni1</i>	0,545	0,020
<i>ni2</i>	0,564	0,070
<i>ni3</i>	0,532	0,097
<i>ur1a</i>	0,482	0,036
<i>ur2a</i>	0,339	0,058
<i>ur3a</i>	0,453	0,058
<i>ur4a</i>	0,483	0,065
<i>ur5a</i>	0,466	0,039
<i>ur6a</i>	0,527	-0,011
<i>ur8a</i>	0,514	0,052
<i>v1</i>	0,941	-0,004
<i>v2</i>	0,935	0,012
<i>v4</i>	0,925	-0,007
<i>vg1</i>	0,314	0,145
<i>vg2</i>	0,369	0,156
<i>vg3</i>	0,398	0,145
<i>subp1</i>	0,001	1,000
<i>z1</i>	0,592	-0,004
<i>z2</i>	0,559	0,024

Tab. E-54: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2fa bei eingeschränktem Datensatz

	<i>Bekanntheit</i>	<i>Involvement</i>	<i>Nutzer- zufriedenheit</i>	<i>Nutzungs- intention</i>	<i>Wahrg. Preis</i>	<i>Usability formativ</i>
<i>b1</i>	1,000	0,178	0,045	0,106	0,057	0,018
<i>inv1</i>	0,136	0,955	0,004	0,087	-0,034	0,000
<i>inv2</i>	0,169	0,947	-0,062	0,058	0,010	-0,060
<i>inv3</i>	0,166	0,966	-0,009	0,111	-0,048	-0,012
<i>inv4</i>	0,194	0,981	0,076	0,185	-0,051	0,066
<i>inv5</i>	0,175	0,974	0,056	0,169	-0,085	0,058
<i>z1</i>	0,069	0,029	0,974	0,811	-0,004	0,834
<i>z2</i>	0,017	0,035	0,971	0,782	0,024	0,788
<i>ni1</i>	0,098	0,121	0,790	0,942	0,020	0,673
<i>ni2</i>	0,092	0,131	0,772	0,950	0,070	0,670
<i>ni3</i>	0,108	0,158	0,751	0,930	0,097	0,645
<i>subp1</i>	0,057	-0,054	0,010	0,066	1,000	0,028
<i>Uf ind1</i>	-0,067	-0,036	0,555	0,438	-0,052	0,607
<i>Uf ind2</i>	0,062	0,094	0,527	0,509	0,032	0,692
<i>Uf ind3</i>	0,111	-0,157	0,255	0,111	0,117	0,294
<i>Uf ind4</i>	-0,100	0,097	0,292	0,253	0,023	0,339
<i>Uf ind5</i>	-0,198	-0,114	-0,078	-0,088	-0,007	-0,071
<i>vg1</i>	0,050	0,099	0,440	0,371	0,145	0,502
<i>vg2</i>	0,094	0,116	0,589	0,505	0,156	0,625
<i>vg3</i>	0,076	0,089	0,586	0,549	0,145	0,634
<i>v1</i>	0,009	0,059	0,557	0,562	-0,004	0,563
<i>v2</i>	0,024	0,139	0,555	0,549	0,012	0,577
<i>v4</i>	-0,009	0,159	0,547	0,518	-0,007	0,581

Anmerkung: Fett gedruckt dargestellt sind die Ladungen der Indikatoren auf ihren jeweiligen latenten Variablen.

Tab. E-54: Kreuzladungen der Indikatoren im PLS-Modell bei Verwendung der formativen Usability-Skala ufosV2fa bei eingeschränktem Datensatz (Fortsetzung)

	<i>Nutzervergnügen</i>	<i>Vertrauen</i>
<i>b1</i>	0,081	0,008
<i>inv1</i>	0,116	0,130
<i>inv2</i>	0,071	0,048
<i>inv3</i>	0,096	0,114
<i>inv4</i>	0,106	0,137
<i>inv5</i>	0,117	0,137
<i>z1</i>	0,603	0,591
<i>z2</i>	0,543	0,558
<i>ni1</i>	0,506	0,544
<i>ni2</i>	0,506	0,564
<i>ni3</i>	0,461	0,532
<i>subp1</i>	0,161	0,001
<i>Uf ind1</i>	0,397	0,286
<i>Uf ind2</i>	0,477	0,435
<i>Uf ind3</i>	0,166	0,300
<i>Uf ind4</i>	0,140	0,261
<i>Uf ind5</i>	0,014	-0,037
<i>vg1</i>	0,876	0,313
<i>vg2</i>	0,946	0,369
<i>vg3</i>	0,941	0,398
<i>v1</i>	0,340	0,941
<i>v2</i>	0,376	0,933
<i>v4</i>	0,386	0,926

Tab. E-55: Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen: aufgeklärte Varianz (R^2) bei eingeschränktem Datensatz

<i>Endogene Variable</i>	<i>Modell mit ufosV2ra (reflektive Usability) R^2</i>	<i>Modell mit ufosV2fa (formative Usability) R^2</i>
Vertrauen	0,273	0,377
Nutzungsvergnügen	0,341	0,413
Nutzerzufriedenheit	0,744	0,712
Nutzungsintention	0,687	0,696

Anmerkung: Determinationskoeffizient (R^2).

Tab. E-56: Kennwerte der PLS-Messmodelle zur Validierung der ufosV2a-Skalen: AVE-Werte und Composite Reliability bei eingeschränktem Datensatz

<i>Latente Variable</i>	<i>ufosV2ra-Modell (reflektive Usability)</i>		<i>ufosV2fa-Modell (formative Usability)</i>	
	<i>AVE</i>	<i>Composite Reliability</i>	<i>AVE</i>	<i>Composite Reliability</i>
Usability - ufosV2ra	0,811	0,968	-	-
Nutzervergnügen	0,849	0,944	0,850	0,944
Vertrauen	0,872	0,953	0,872	0,953
Involvement	0,931	0,985	0,931	0,985
Nutzungsintention	0,885	0,958	0,885	0,958
Nutzerzufriedenheit	0,946	0,972	0,946	0,972

Tab. E-57: Ergebnisse des PLS-Strukturmodells für das nomologische Modell zur Validierung der ufosV2a-Skalen bei eingeschränktem Datensatz: Pfadkoeffizienten

<i>Zusammenhang</i>	<i>Modell mit ufosV2ra (reflektive Usability)</i>		<i>Modell mit ufosV2fa (formative Usability)</i>	
	β	<i>t-Wert</i>	β	<i>t-Wert</i>
Usability → Nutzerzufriedenheit	0,654	9,754	0,693	8,521
Usability → Nutzungsintention	-0,072	0,563	-0,108	0,830
Usability → Vertrauen	0,522	6,690	0,614	7,803
Usability → Nutzungsvergnügen	0,584	8,345	0,642	12,457
Vertrauen → Nutzerzufriedenheit	0,196	2,815	0,127	1,735
Nutzungsvergnügen → Nutzerzufriedenheit	0,139	2,090	0,100	1,278
Nutzerzufriedenheit → Nutzungsintention	0,823	8,930	0,756	7,229
Bekanntheit → Vertrauen	0,020	0,248	-0,002	0,030
Wahrgen. Preis → Nutzerzufriedenheit	-0,042	0,890	-0,026	0,446
Involvement → Nutzungsintention	0,033	0,243	0,123	1,890
Involvement x Usability → Nutzungsintention. (originale . Werte)	-0,072	0,204	0,303	1,910
Involv. x Nutzerzufried. → Nutzungsintention. (originale . Werte)	0,205	0,607	-0,193	0,753
Involvement x Usability → Nutzungsintention. (stand. Werte)	-0,028	0,261	0,076	0,692
Involv. x Nutzerzufried. → Nutzungsintention. (stand. Werte)	0,090	0,832	0,004	0,048

Anmerkung: Pfadkoeffizient (β). Der t-Wert der t-Statistik in PLS folgt nur annähernd der t-Verteilung.

Tab. E-58: Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum nomologischen Modell bei eingeschränktem Datensatz

<i>Latente exogene Variable</i>	<i>Latente endogene Variable</i>	<i>Modell mit ufosV2ra (reflektive Usability)</i>		<i>Modell mit ufosV2fa (formative Usability)</i>	
		<i>f²</i>	<i>Beurteilung</i>	<i>f²</i>	<i>Beurteilung</i>
Usability	Nutzungsintention	0,003	keiner	0,033	gering
	Nutzerzufriedenheit	0,930	hoch	0,715	hoch
	Nutzungsvergnügen	0,517	hoch	0,704	hoch
	Vertrauen	0,376	hoch	0,605	hoch
Vertrauen	Nutzerzufriedenheit	0,109	gering	0,038	gering
Nutzungsvergnügen	Nutzerzufriedenheit	0,047	gering	0,021	gering
Nutzerzufriedenheit	Nutzungsintention	0,415	hoch	0,438	hoch
Bekanntheit	Vertrauen	0,001	keiner	-	keiner
Wahrgen. Preis	Nutzerzufriedenheit	0,008	keiner	0,003	keiner
Involvement	Nutzungsintention	0,048	gering	0,076	gering
Involvement x Usability	Nutzungsintention	0,006	keiner	0,030	gering

Anmerkung: Effektgröße (f^2). Die Effektgrößen werden getrennt für das ufosV2r-Modell (Verwendung der reflektiven Usability-Skala ufosV2r) sowie für das ufosV2f-Modell (Verwendung der formativen Usability-Skala ufosV2f) angegeben.

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde zur Erlangung eines akademischen Grades vorgelegt.

Hamburg, den 18.02.2009

Gunther Held