

Timo Christophersen

**Usability im Online-Shopping: Entwicklung  
eines Fragebogeninstrumentes (ufosV2) unter  
Berücksichtigung formativer und reflektiver  
Messmodelle**

Mit einem Geleitwort von Prof. Dr. Udo Konradt

# **Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business**

Herausgegeben von

Prof. Dr. Dr. h.c. Sönke Albers,

Prof. Dr. Birgit Friedl,

Prof. Dr. Daniel Klapper,

Prof. Dr. Achim Walter,

Prof. Dr. Joachim Wolf,

Institut für Betriebswirtschaftslehre,

Christian-Albrechts-Universität zu Kiel

Prof. Dr. Udo Konradt,

Institut für Psychologie,

Christian-Albrechts-Universität zu Kiel

In der Schriftenreihe werden Ergebnisse von Forschungsarbeiten veröffentlicht, die sich in herausragender Weise mit Fragen des Management lose gekoppelter Systeme, virtueller Unternehmen und elektronischer Geschäftsprozesse beschäftigen. Die Reihe richtet sich an Leser in Wissenschaft und Praxis, die Anregungen für die eigene Arbeit und Problemlösungen suchen. Sie ist nicht auf Veröffentlichungen aus den Instituten der Herausgeber beschränkt.

# **Usability im Online-Shopping**

Entwicklung eines Fragebogeninstrumentes (ufosV2)  
unter Berücksichtigung formativer und reflektiver Messmodelle

Dissertation  
zur Erlangung des Doktorgrades  
der Philosophischen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Timo Christophersen

Kiel, 2006

Erstgutachter: Prof. Dr. Udo Konradt

Zweitgutachter: Prof. Dr. Thomas Bliesener

Drittgutachter: Prof. Dr. Roman Ferstl

Tag der mündlichen Prüfung: 22. Januar 2007

Durch den Prodekan Prof. Dr. Ludwig Steindorff zum Druck genehmigt am 30. April 2007

## Geleitwort

Internetdienste werden zunehmend zum warenbezogenen Informationsaustausch, also der Anbahnung, Aushandlung und der Abwicklung von Kaufvorgängen verwendet. Ein spezielles Segment im Rahmen von Business-to-Consumer Transaktionsbeziehungen stellen Online-Shops dar, auf deren Webseiten Produkte und Dienstleistungen angeboten werden. Als wichtiges Merkmal für das Einkaufen in Online-Shops hat sich die Gestaltung der Benutzerschnittstelle erwiesen, die in der Softwareergonomie unter dem Aspekt der Usability oder auch Gebrauchstauglichkeit betrachtet wird. Die überwiegende Zahl der empirischen Studien, die sich mit diesem zentralen Konstrukt befasst, weist jedoch gravierende Defizite hinsichtlich der messtheoretischen Modellierung der Usability sowie der Untersuchungsplanung auf. Insbesondere wird oftmals von messtheoretischen Annahmen ausgegangen, die bei näherer Prüfung kritisch erscheinen, da keine Unterscheidung zwischen formativen und reflektiven Messmodellen getroffen wird. Zudem werden in vielen Untersuchungen zwar Bewertungen von Benutzerschnittstellen sowie Kaufintentionen von Seiten der Nutzer abgefragt, eine Erfassung tatsächlichen Kaufverhaltens erfolgt jedoch zumeist nicht.

In der vorliegenden Arbeit wird entsprechend das Ziel verfolgt, erstmals Skalen zur Erfassung der Gebrauchstauglichkeit zur Verfügung stellen, die messtheoretisch in angemessener Weise modelliert wurden, indem zwischen formativen und der reflektiven Messmodellen der Usability differenziert wird. Dazu erfolgt die Entwicklung eines Usability-Fragebogeninstrumentes, das zur Gestaltung und Optimierung von Online-Shops eingesetzt werden kann. Neben erlebensbezogenen Aspekten des Internetkaufs werden in der Untersuchung auch verhaltensbezogene Kriterien der Kaufentscheidung berücksichtigt, um die Validität der neu entwickelten Usability-Skalen bewerten zu können. Dadurch eröffnet sich die Möglichkeit, die Erfolgswirksamkeit der Usability in Bezug auf die Kaufabsicht bzw. das Kaufverhalten zu betrachten und ihren relativen Einfluss im Verhältnis zu weiteren wahrgenommenen Shopmerkmalen im Rahmen eines nomothetischen Netzwerkes zu untersuchen.

Die Arbeit ist sowohl in Bezug auf die Versuchsplanung als auch die Versuchsdurchführung als beispielhaft zu werten, da der Autor diesbezüglich mit großer Sorgfalt und Akribie vorgegangen ist, wie dem Leser detailliert vermittelt wird. Der oft als unüberwindlich eingeschätzte Spagat zwischen der Optimierung der externen und der internen Validität gelingt hier in hervorragender Form. Eine hohe externe Validität der Laboruntersuchung, also eine Generalisierbarkeit der Ergebnisse auf alltägliche Online-Einkaufssituationen, wurde zum einen durch die hohe Anzahl von nahezu 380 Teilnehmern sowie die adäquate Zusammensetzung der Teilnehmerstichprobe gewährleistet. Zum anderen wurde mit 35 Online-Shops eine breite Auswahl unterschiedlicher Benutzerschnittstellen berücksichtigt, so dass eine Übertragbarkeit auf andere Schnittstellen gegeben ist. Die computergestützte Untersuchungsumgebung, die zur Versuchsdurchführung verwendet wurde, erlaubte zudem die zuverlässige Kontrolle und Parallelisierung von Störeinflüssen, wodurch eine hohe interne Validität der Befunde sicher-

gestellt wurde. Die Tatsache, dass ein großer Anteil der Teilnehmer während des durchgeführten Versuchs in einem der besuchten Online-Shops einkaufte, untermauert zusätzlich, dass es gelungen ist, den tatsächlichen Nutzungskontext des Online-Shoppings gut abzubilden.

Auch in methodischer Hinsicht stellt die Arbeit einen Fortschritt dar. Zum einen wird durch die Unterscheidung zwischen formativen und reflektiven Messmodellen ein innovativer messtheoretischer Ansatz in die Usability-Forschung eingeführt, der in diesem Bereich trotz seiner Vorteile bislang keinerlei Berücksichtigung fand. Zum zweiten werden im Rahmen der Datenanalyse spezielle multivariate Verfahren wie z.B. Partial Least Squares Strukturgleichungsmodelle herangezogen, deren Anwendung ein hohes Maß an methodischer Fachkompetenz erfordert, durch die es jedoch gelingt, den komplexen messtheoretischen Anforderungen der Entwicklung einer formativen Skala gerecht zu werden. Der Autor stellt sowohl die Grundlagen als auch die Durchführung der Analysen detailliert dar, so dass der fachkundige Leser eine gute und anwendungsorientierte Einführung in den Umgang mit dem Messansatz und den Auswertungsmethoden erhält.

Zusammengenommen handelt es sich um eine Arbeit, die im Schnittbereich zwischen Ingenieurpsychologie, Marktpsychologie und betriebswirtschaftlicher Marketingforschung liegt. Sie schließt eine Lücke, da sie „belastbares Wissen“ zur Bedeutung der Schnittstellengestaltung von Online-Shops für das Erleben und Verhalten von Konsumenten bereitstellt, welches für diverse Lesergruppen von Interesse sein kann. So besitzt die Arbeit einerseits für Wissenschaftler aus dem Bereich Human-Computer Interaction eine hohe Relevanz, da neue Maßstäbe zur fragebogenbasierten Erfassung der Usability gesetzt werden. Für Forscher anderer fachlicher Richtungen ist das Buch von Interesse, da die Entwicklung einer formativen Skala in beispielhafter Form geschildert wird. Die Darstellungen zu den verschiedenen kundennahen Erfolgsfaktoren bieten zudem einen umfassenden Überblick, welche Variablen es bei der Planung und Umsetzung von E-Commerce Unternehmungen zu adressieren gilt. Nicht zuletzt werden zwei messtheoretisch fundierte Usability-Skalen zur Verfügung gestellt, die im Zuge der kundenorientierten Gestaltung und Optimierung von Online-Shops von hohem praktischen Wert sein können. Ich wünsche dem Buch eine breite und interessierte Leserschaft.

Professor Dr. Udo Konradt

Christian-Albrechts-Universität zu Kiel

April 2007

## **Vorwort**

Die vorliegende Arbeit befasst sich mit der Usability als ein wesentlicher Erfolgsfaktor im Bereich Online-Shopping. Im Fall von Dissertationsprojekten kann sicherlich die Unterstützung durch andere Personen zu einem erfolgreichen Gelingen verhelfen. Ich wurde bei der Fertigstellung der Arbeit durch eine Reihe von Menschen bzw. Institutionen unterstützt, denen ich an dieser Stelle meinen Dank aussprechen möchte.

An erster Stelle danke ich meinem Doktorvater und Betreuer Prof. Dr. Udo Konradt, dessen wertvolle Unterstützung entscheidend zum Gelingen der Arbeit beigetragen hat. Mein Dank gilt auch dem Zweitgutachter dieser Arbeit Prof. Dr. Thomas Bliesener.

Die Arbeit entstand im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) getragenen Graduiertenkollegs „Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business“. Ich möchte dem Sprecher des Graduiertenkollegs Prof. Dr. Dr. h.c. Sönke Albers für seinen Einsatz bei der Leitung des Kollegs sowie für inhaltliche und methodische Impulse danken. Natürlich gilt mein Dank auch der DFG für die finanzielle Unterstützung dieser Promotion sowie meinen Kollegen aus dem Graduiertenkolleg für viele fachliche Anregungen. Im Rahmen eines durch die DFG finanzierten Forschungsaufenthalts an der University of Maryland hatte ich die Gelegenheit zum Austausch mit Prof. Kent L. Norman, dem ich an dieser Stelle ebenfalls meinen Dank aussprechen möchte.

Danken will ich meinen guten Freunden Henrik Drüner, Arne Holst, Jürgen Kröger und Lars Groeger, auf deren moralische und auch praktische Unterstützung ich in der Entstehungsphase dieser Arbeit setzen konnte.

Ein Dank geht auch an die Teilnehmer meiner Untersuchung, ohne deren Bereitschaft zur Teilnahme am durchgeführten Experiment eine Realisierung des Projektes nicht möglich gewesen wäre. Bei der Durchführung der Versuche konnte ich mich auf die tatkräftige Unterstützung der angehenden Psychologen Lea Schurawitzki, Oliver Harm und Diana Hansen verlassen, die großartige Arbeit geleistet haben. Aus dem Kreis der Kollegen danke ich außerdem Panja Andressen, Christian Sörensen, Jörn Bartels, Oliver Rack und Susanne Geister sowie dem ganzen AOM-Lehrstuhl-Team.

Diese Arbeit baut auf den Ergebnissen eines Projekts auf, das in Zusammenarbeit mit Björn Balazs und Prof. Dr. Hartmut Wandke von der HU Berlin durchgeführt wurde. Mein guter Freund Björn lieferte auch in der Entstehungsphase der vorliegenden Arbeit wertvolle Anregungen, wofür ich sehr dankbar bin.

Nicht zuletzt möchte ich meinen Eltern und meiner Familie ganz herzlich dafür danken, dass ich mich immer und in allen Lebenslagen auf sie verlassen konnte und kann.

Timo Christophersen





# Inhaltsverzeichnis

Abbildungsverzeichnis .....	XVII
Tabellenverzeichnis.....	XIX
Abkürzungsverzeichnis .....	XXI
<b>1. Einleitung.....</b>	<b>1</b>
<b>2. Theoretischer und empirischer Hintergrund .....</b>	<b>5</b>
2.1 Einkaufen im Internet.....	5
2.1.1 Electronic Commerce .....	7
2.1.2 Erscheinungsformen des Bereichs Business-to-Consumer .....	8
2.1.3 Online-Shops.....	8
2.1.4 Vor- und Nachteile des Online-Shoppings aus Kundensicht.....	10
2.1.4.1 Vorteile.....	10
2.1.4.2 Nachteile und Risiken .....	12
2.1.5 Erfolgsfaktorenforschung im Online-Shopping.....	13
2.1.5.1 Begriffsklärung.....	13
2.1.5.2 Grundsätzliche Methodik der Erfolgsfaktorenforschung.....	14
2.1.5.3 Kritik an der Erfolgsfaktorenforschung .....	15
2.1.6 Erfolgskriterien von Online-Shops.....	16
2.1.6.1 Kaufentscheidung.....	17
2.1.6.2 Kaufintention.....	18
2.1.7 Zwischenfazit und Überleitung.....	20
2.2 Usability – Begriffsklärung.....	21
2.2.1 Usability in der DIN EN ISO 9241 .....	21
2.2.1.1 Effektivität, Effizienz und Zufriedenheit im Benutzungskontext.....	21
2.2.1.2 Anforderungen an die Dialoggestaltung .....	24
2.2.1.3 Weitere Usability-bezogene Teile der Norm .....	25
2.2.2 Weitere Auffassungen zur Usability .....	26
2.2.2.1 Jakob Nielsens Usability-Begriff .....	26
2.2.2.2 Mentale Modelle .....	27
2.2.2.3 Perceived Ease of Use und Perceived Usefulness.....	29
2.3 Methoden der Usability-Evaluation .....	29
2.3.1 Formative und summative Evaluation von Software .....	29
2.3.2 Deskriptive vs. prädiktive Usability-Evaluationsmethoden.....	29
2.3.2.1 Deskriptive Methoden.....	30
2.3.2.2 Prädiktive Methoden .....	32
2.3.3 Subjektive vs. objektive Usability-Daten.....	33
2.4 Fragebogenbasierte Usability-Evaluation .....	35

2.4.1	Vorliegende Usability-Fragebögen .....	35
2.4.1.1	IsoMetrics .....	36
2.4.1.2	ISONORM 9241/10 .....	37
2.4.1.3	AttrakDiff 2 .....	38
2.4.1.4	Wahrgenommene Orientierung in Online-Shops – WOOS .....	40
2.4.1.5	Marketingorientierte Evaluationskalen für Online-Shops .....	41
2.4.1.6	Usability Fragebogen für Online-Shops – ufosV1 .....	42
2.4.2	Differenzierende Charakteristika von Usability-Fragebögen .....	43
2.4.2.1	Konstruktfokus/Dimensionalität .....	44
2.4.2.2	Bewertungsgegenstand .....	45
2.4.2.3	Granulationsgrad der Items .....	46
2.4.2.4	Itemform .....	46
2.4.2.5	Antwortskala .....	47
2.4.2.6	Itemanzahl .....	49
2.4.3	Gütebeurteilung von Usability-Fragebögen .....	49
2.4.4	Vor- und Nachteile der fragebogenbasierten Erfassung der Usability .....	50
2.4.4.1	Potenzielle Vorteile .....	50
2.4.4.2	Potenzielle Nachteile .....	52
2.5	Usability im Online-Shopping .....	54
2.5.1	Ausprägung der Usability von Online-Shops .....	54
2.5.2	Usability als Erfolgsfaktor von Online-Shops .....	55
2.5.3	Verschiedene Arten der Nutzung eines Online-Shops .....	58
2.6	Konzeptuelles Erfolgsmodell zur Validierung von Usability-Skalen .....	60
2.6.1	Ästhetik .....	61
2.6.2	Vertrauen .....	62
2.6.3	Wahrgenommene Produktanzahl .....	63
2.6.4	Einkaufsvergnügen .....	64
2.6.5	Wahrgenommener Preis .....	65
2.6.6	Service .....	66
2.7	Zusammenhänge innerhalb des konzeptuellen Erfolgsmodells .....	67
2.7.1	Einfluss der Usability auf das Vertrauen .....	67
2.7.2	Einfluss der Usability auf das Einkaufsvergnügen .....	67
2.7.3	Einfluss der Ästhetik auf die Usability .....	68
2.7.4	Einfluss der Ästhetik auf das Einkaufsvergnügen .....	69
2.7.5	Einfluss der Produktanzahl auf das Vertrauen .....	70
2.7.6	Einfluss des Bekanntheitsgrads auf das Vertrauen .....	70
2.7.7	Moderierende Einflüsse .....	71
2.7.7.1	Produktart .....	71
2.7.7.2	Personenbezogene Merkmale als potenzielle Moderatoren .....	72
2.8	Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle .....	74
2.8.1	Charakteristika reflektiver Messmodelle .....	74

---

2.8.2	Charakteristika formativer Messmodelle .....	76
2.8.3	Indexentwicklung zur Operationalisierung formativer Konstrukte .....	77
2.8.3.1	Schritt 1: Definition des Konstrukts und Festlegung des Messmodells .....	78
2.8.3.2	Schritt 2: Bestimmung der Indikatoren .....	80
2.8.3.3	Schritt 3: Behandlung von Multikollinearität .....	81
2.8.3.4	Schritt 4: Schätzung und Gütebeurteilung des Messmodells .....	82
2.8.3.5	Schritt 5: Indexberechnung .....	83
2.8.4	Fehlspezifizierte Messmodelle .....	84
2.8.5	Folgen fehlerhafter Messmodellannahmen .....	84
<b>3.</b>	<b>Forschungsfragen .....</b>	<b>87</b>
3.1	Ableitung der Ziele .....	87
3.1.1	Formative und reflektive Erfassung der Usability aus Nutzersicht .....	87
3.1.2	Fehlspezifikationen bestehender Usability-Fragebögen .....	89
3.2	Ziele der Untersuchung .....	90
3.3	Inhaltliche Hypothesen .....	91
<b>4.</b>	<b>Entwicklung des Itempools der ufosV2-Skalen .....</b>	<b>97</b>
4.1	Definition des Konstrukts und Festlegung des Messmodells .....	97
4.2	Bestimmung der ufosV2-Items .....	97
4.2.1	Vorüberlegungen zur Itemauswahl der Skala ufosV2f .....	98
4.2.2	Eignung und Modifikation der Items des ufosV1 .....	99
4.2.2.1	Analyse der Anteile fehlender Werte .....	99
4.2.2.2	Treffen von Messmodellannahmen .....	100
4.2.2.3	Optimierung der Itemschwierigkeiten .....	100
4.2.2.4	Umformulierungen von Items .....	101
4.2.2.5	Prüfung auf inhaltliche Redundanz unter den Items .....	101
4.2.3	Gewährleistung der Vollständigkeit des Itempools .....	101
4.2.3.1	Sichtung relevanter Literatur .....	102
4.2.3.2	Ableitung neuer Usability-Items .....	102
4.2.4	Berücksichtigung von Störeffekten durch die Itemformulierung .....	103
4.2.5	Antwortformat der Items .....	103
4.2.5.1	Verwendung einer siebenstufigen Likert-Skala .....	104
4.2.5.2	Zusätzliche Antwortoption „keine Angabe“ .....	104
4.2.6	Voruntersuchung zur Verständlichkeit .....	104
4.2.7	Verwendete formative und reflektive Items .....	105
<b>5.</b>	<b>Untersuchungsplanung .....</b>	<b>107</b>
5.1	Unabhängige Variable Online-Shop .....	107
5.1.1	Kriterien zur Auswahl der Produktarten .....	107
5.1.2	Festlegung der Online-Shops .....	109

5.2	Operationalisierung der Variablen auf Prädiktorebene .....	111
5.2.1	Reflektive Erfassung der Ästhetik .....	111
5.2.2	Reflektive Erfassung des Vertrauens .....	111
5.2.3	Reflektive Erfassung der wahrgenommenen Produktanzahl .....	112
5.2.4	Reflektive Erfassung des Einkaufsvergnügens .....	113
5.2.5	Erfassung des wahrgenommenen Preises .....	113
5.2.6	Formative Erfassung des Service .....	113
5.2.7	Erfassung der Bekanntheit der Shops .....	114
5.3	Operationalisierung der abhängigen Variablen .....	114
5.3.1	Reflektive Erfassung der Kaufintention .....	115
5.3.2	Operationalisierung der Kaufentscheidung .....	116
5.4	Erhebung weiterer Variablen .....	117
5.4.1	Demographische Merkmale der Teilnehmer .....	118
5.4.2	Vorerfahrung im Bereich PC und Online-Shopping .....	118
5.4.3	Kontrollvariablen .....	119
5.4.3.1	Produkt-Involvement .....	119
5.4.3.2	Kontrollvariablen zum experimentellen Kontext .....	120
5.5	Versuchsszenario .....	120
5.6	Berücksichtigung möglicher Reihenfolgeeffekte .....	121
5.6.1	Reihenfolge der formativen Items zur Usability .....	122
5.6.2	Reihenfolge der Usabilityabfrage reflektiv vs. formativ .....	122
5.6.3	Reihenfolge der Shopdarbietung .....	123
5.6.4	Reihenfolge der Konstruktabfrage .....	123
5.7	Versuchsplan .....	124
<b>6.</b>	<b>Statistische Auswertung der Daten .....</b>	<b>127</b>
6.1	Analyse und Behandlung fehlender Werte .....	129
6.1.1	Elimination von Fällen mit zu hohem Anteil fehlender Werte .....	129
6.1.2	Analyse des Fehlendmechanismus .....	129
6.1.3	Mittelwertbasierte Ersetzung bei reflektiven Indikatoren .....	130
6.1.4	Multiple Imputation bei formativen Indikatoren .....	131
6.2	Verteilungsprüfung .....	132
6.3	Prüfung auf Vorliegen eines Common Method Bias .....	133
6.4	Vorgehen zur Identifikation problematischer Fälle .....	133
6.4.1	Teilnehmer mit auffällig langer bzw. kurzer Bearbeitungsdauer .....	133
6.4.2	Fallweise Analyse der Varianz im Antwortverhalten .....	134
6.4.3	Konsistenz des Antwortverhaltens .....	135
6.5	Analyse der Güte reflektiver Maße auf Prädiktorebene .....	136
6.5.1	Explorative Faktorenanalyse .....	137
6.5.1.1	Voraussetzungen und Kriterien der Dateneignung .....	137
6.5.1.2	Beurteilung der Faktorenlösung .....	138

---

6.5.2	Güteprüfung nach Kriterien der klassischen Testtheorie .....	139
6.5.3	Konfirmatorische Faktorenanalyse (CFA) zur Prüfung der Skalenstruktur.....	140
6.5.3.1	Voraussetzungen zur Berechnung einer CFA .....	140
6.5.3.2	Gütekriterien der CFA .....	141
6.6	Strukturgleichungsmodelle auf Basis von Partial Least Squares (PLS) .....	143
6.6.1	Voraussetzungen zur Berechnung von PLS-Modellen .....	143
6.6.2	Gütebeurteilung von PLS-Modellen .....	144
6.6.2.1	Beurteilung des Messmodells.....	144
6.6.2.2	Beurteilung des Strukturmodells.....	146
6.7	Multinomiale logistische Regression (MLR).....	147
6.7.1	Voraussetzungen der MLR.....	147
6.7.2	Ausreißeranalyse bei der MLR .....	148
6.7.3	Gütekriterien der MLR.....	148
6.8	Multikollinearitätsprüfung .....	150
6.9	Weitere angewandete Testverfahren .....	151
6.9.1	t-Tests .....	151
6.9.2	Einfaktorielle Varianzanalyse .....	151
6.9.3	Levene-Test.....	151
<b>7.</b>	<b>Durchführung.....</b>	<b>153</b>
7.1	Teilnehmerstichprobe.....	153
7.1.1	Rekrutierung der Teilnehmer .....	153
7.1.1.1	Wege der Rekrutierung .....	153
7.1.1.2	Ablauf der Rekrutierung.....	154
7.1.2	Probleme bei der Teilnehmerakquise.....	154
7.1.3	Beschreibung der Stichprobe .....	155
7.2	Beschreibung der Versuchsumgebung.....	157
7.3	Computergestützte Durchführung des Versuchs.....	158
7.4	Versuchsablauf.....	160
7.5	Probleme bei der Erhebung.....	166
<b>8.</b>	<b>Ergebnisse .....</b>	<b>169</b>
8.1	Vorbereitende Datenanalysen .....	169
8.1.1	Aufbereitung der Daten: Zusammenfassung der Bewertungen .....	169
8.1.2	Analyse und Behandlung fehlender Werte.....	170
8.1.2.1	Elimination von Fällen mit zu hohem Anteil fehlender Werte .....	170
8.1.2.2	Analyse des Fehlendmechanismus .....	171
8.1.2.3	Ersetzung fehlender Werte .....	171
8.1.3	Verteilungen der erhobenen Indikatoren.....	172
8.1.4	Prüfung auf Vorliegen eines Common Method Bias .....	172

8.1.5	Identifikation problematischer Fälle .....	173
8.2	Erfassung der Variablen auf Prädiktorebene .....	174
8.2.1	Explorative Faktorenanalyse zur Prüfung der Skalenstruktur .....	174
8.2.2	Prüfung der Skalengüte auf Einzelskalenebene .....	177
8.2.3	Konfirmatorische Faktorenanalyse zur Prüfung der Skalenstruktur .....	179
8.2.4	Diskriminationsfähigkeit der reflektiven Skalen .....	183
8.2.5	Erfassung der formativen Variable Service .....	185
8.2.6	Erfassung des wahrgenommenen Preises .....	185
8.3	Erfassung der abhängigen Variablen .....	186
8.3.1	Reflektive Erfassung der Kaufintention .....	186
8.3.2	Erfassung der Kaufentscheidung .....	187
8.4	Festlegung der formativen Skala ufosV2f .....	188
8.4.1	Endgültige Bestimmung der ufosV2f-Indikatoren .....	188
8.4.1.1	Elimination von Indikatoren aufgrund zu vieler fehlender Werte .....	188
8.4.1.2	Elimination inhaltlich redundanter Indikatoren .....	189
8.4.2	Behandlung von Multikollinearität zwischen den ufosV2f-Items .....	190
8.4.3	PLS-Zwei-Konstrukt-Modell zur Validierung .....	192
8.5	Untersuchung des konzeptuellen Erfolgsmodells mit Hilfe von PLS .....	195
8.5.1	Beurteilung der Messmodelle für beide PLS-Modelle .....	195
8.5.2	Beurteilung der Strukturmodelle für beide PLS-Modelle .....	197
8.6	MLR zur Vorhersage der Kaufentscheidung .....	201
8.6.1	Prüfung der Voraussetzungen und Ausreißerdiagnostik .....	202
8.6.2	Sequentielle Berechnung der MLR .....	203
8.7	Übersicht der Befunde .....	206
<b>9.</b>	<b>Diskussion .....</b>	<b>209</b>
9.1	Bewertung des methodischen Vorgehens .....	209
9.1.1	Interne Validität des Quasi-Experiments .....	209
9.1.2	Externe Validität des Quasi-Experiments .....	211
9.1.2.1	Realitätsbezug der experimentellen Einkaufssituation .....	212
9.1.2.2	Operationalisierung der Kaufentscheidung .....	213
9.1.2.3	Repräsentativität der Stichprobe .....	214
9.1.2.4	Abhängigkeit der Ergebnisse von der Shopauswahl .....	215
9.1.3	Vorgehen der statistischen Auswertung .....	216
9.2	Diskussion der Befunde .....	218
9.2.1	Befunde zur Güte der ufosV2-Skalen .....	218
9.2.1.1	Interne Konsistenz der reflektiven ufosV2r-Skala .....	218
9.2.1.2	Faktorielle Validität der reflektiven ufosV2r-Skala .....	218
9.2.1.3	Diskriminationsfähigkeit der reflektiven ufosV2r-Skala .....	219
9.2.1.4	Statistischer Zusammenhang der ufosV2-Skalen .....	220
9.2.1.5	Kriteriumsvalidität der ufosV2-Skalen .....	220

---

9.2.1.6	Nomologische Validität der ufosV2-Skalen .....	222
9.2.2	Befunde zu den weiteren Hypothesen im konzeptuellen Erfolgsmodell .....	223
9.3	Zusammenfassende Bewertung der Ergebnisse zu ufosV2.....	225
9.4	Erfolgswirksamkeit der Usability im Online-Shopping.....	227
9.5	Implikationen für Forschung und Praxis .....	228
9.5.1	Verwendung der ufosV2-Skalen .....	228
9.5.2	Formative und reflektive Erfassung der Usability in Fragebögen .....	229
9.5.3	Theoretisches Fundament der Usability .....	230
9.6	Weiterführende Fragestellungen .....	232
	<b>Zusammenfassung der Untersuchung.....</b>	<b>233</b>
	<b>Summary .....</b>	<b>235</b>
	<b>Literaturverzeichnis.....</b>	<b>237</b>
	<b>Anhang A: Ergänzende Darstellungen zum Hintergrund</b>	<b>281</b>
	<b>Anhang B: Ergänzende Darstellungen zur ufosV2-Entwicklung</b>	<b>293</b>
	<b>Anhang C: Ergänzende Darstellungen zur Untersuchungsplanung</b>	<b>303</b>
	<b>Anhang D: Ergänzende Darstellungen zur Durchführung</b>	<b>309</b>
	<b>Anhang E: Ergänzende Ergebnisdarstellungen</b>	<b>337</b>





## Abbildungsverzeichnis

Abb. 2-1:	B2C-E-Commerce Umsatzentwicklung in Deutschland (Hauptverband des Deutschen Einzelhandels, 2005) .....	6
Abb. 2-2:	Segmente des Electronic Commerce (Hermanns & Sauter, 2001, S. 23) .....	7
Abb. 2-3:	Anzahl der aufgetretenen Sicherheitsprobleme im ersten Quartal 2005 (Statistisches Bundesamt, 2006).....	12
Abb. 2-4:	Methoden zur Identifikation von Erfolgsfaktoren (Haenecke, 2002) .....	15
Abb. 2-5:	Konzeptionalisierung der Usability in der DIN EN ISO 9241-11 (1999).....	22
Abb. 2-6:	Software-ergonomische Teile der DIN EN ISO 9241 (Dzida, 1994, S. 381) ....	26
Abb. 2-7:	Systematisierung verschiedener Usability-Evaluationsmethoden (in Anlehnung an Hamborg & Gediga, 2006).....	30
Abb. 2-8:	Beispiel für die Art der Itemabfrage im ISONORM 9241/10: Item zur Subskala Steuerbarkeit (Prümper & Anft, 1993, S. 149).....	37
Abb. 2-9:	Zwei Items aus dem AttrackDiff-Fragebogen (www.attrakdiff.de) .....	39
Abb. 2-10:	Beispielitem aus dem ufosV1-Fragebogen (www.ufos-online.com) .....	43
Abb. 2-11:	Verschiedene Charakteristika von Evaluationsfragebögen.....	44
Abb. 2-12:	Zwei verschiedene Suchstrategien von Kunden im Online-Shopping (Detlor, Sproule & Gupta, 2003, S. 81) .....	59
Abb. 2-13:	Konzeptuelles Erfolgsmodell zur Validierung von Usability-Fragebogenmaßen im Kontext des Online-Shoppings .....	60
Abb. 2-14:	Latentes Konstrukt mit reflektiven Indikatoren .....	75
Abb. 2-15:	Latentes Konstrukt mit formativen Indikatoren .....	77
Abb. 2-16:	Schematische Darstellung des Indexentwicklungsprozesses formativer Maße .	78
Abb. 2-17:	Zwei-Konstrukt-Modell zur Bestimmung der Validität einer formativen LV (Diamantopoulos & Winklhofer, 2001, S. 273) .....	83
Abb. 3-1:	Zwei-Konstrukt-Modell zur Validierung der formativen Skala ufosV2f anhand des Kriteriums der reflektiven Usability-Skala ufosV2r .....	92
Abb. 3-2:	Konzeptuelles Erfolgsmodell zur Validierung des zu entwickelnden ufosV2...	94
Abb. 5-1:	Schematischer Aufbau des Fragebogens zur Bewertung der Online-Shops ....	123
Abb. 5-2:	Zusammensetzung des Teilnehmercodes .....	125

---

Abb. 7-1:	Screenshot einer Seite der Hypertext-Versuchsumgebung .....	158
Abb. 7-2:	Art der Itemabfrage in der verwendeten Hypertext-Versuchsumgebung.....	159
Abb. 7-3:	Schematische Darstellung des Versuchsablaufs.....	161
Abb. 8-1:	CFA-Modell zur Überprüfung der Skalenstruktur .....	180
Abb. 8-2:	Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala ufosV2f anhand der reflektiven Usability-Skala ufosV2r .....	193
Abb. 8-3:	Ergebnisse des PLS-Strukturmodells des konzeptionellen Modells zur Validierung der ufosV2-Skalen.....	198
Abb. 8-4:	Schritte der sequentiellen multinomialen logistischen Regressionsanalyse unter Angabe der jeweils hinzugefügten Prädiktoren .....	202

## Tabellenverzeichnis

Tab. 2-1:	Häufigkeit der im Internet gekauften Waren und Dienstleistungen in Deutschland (AGOF, 2006) .....	9
Tab. 2-2:	Beispielhafte Aufschlüsselung des Nutzungskontextes für eine Arbeitssoftware (Bevan & Macleod, 1994, S. 137) .....	23
Tab. 2-3:	Skalen und Kennwerte des ufosV1 (Balazs, 2002) .....	43
Tab. 2-4:	Jakob Niensens Top Ten der wichtigsten Design-Fehler des Jahres 2005 (Nielsen, 2005).....	55
Tab. 2-5:	Entscheidungshilfen zur Frage, ob ein reflektives oder ein formatives Messmodell erforderlich ist (Christophersen & Grape, 2006, S. 123; Eberl, 2004; Jarvis, MacKenzie & Podsakoff, 2003).....	79
Tab. 3-1:	Inhaltliche Hypothesen des konzeptuellen Erfolgsmodells zur Skalengüte des ufosV2-Fragebogens (H5 bis H9) .....	93
Tab. 3-2:	Inhaltliche Hypothesen des konzeptuellen Erfolgsmodells zu den Zusammenhängen der weiteren Konstrukte (H10 bis H21) .....	95
Tab. 4-1:	Formative Indikatoren zur Erfassung der Usability .....	105
Tab. 4-2:	Reflektive Indikatoren zur Erfassung der Usability .....	106
Tab. 5-1:	Herangezogene Online-Shops geordnet nach Produktart.....	110
Tab. 6-1:	Übersicht zur Lokalisierung korrespondierender Abschnitte zu den verschiedenen Hypothesen .....	128
Tab. 6-2:	Anspruchsniveaus der herangezogenen Gütemaße zur Beurteilung des CFA-Modells (Homburg & Pflesser, 2000, S. 430; Bühner, 2004, S. 205).....	143
Tab. 7-1:	Alter der Teilnehmer im Vergleich zur Population – Angaben in Prozent .....	155
Tab. 7-2:	Geschlecht der Teilnehmer im Vergleich zur Population – Angaben in Prozent .....	156
Tab. 7-3:	Während des Versuchs genutzte Verbindung ins Internet .....	157
Tab. 7-4:	Anteil der erhobenen Teilnehmer pro Versuchsleiter .....	157
Tab. 8-1:	Missing Data im Datensatz - Aufteilung der Teilnehmer .....	170
Tab. 8-2:	Übersicht über alle berechneten explorativen Faktorenanalysen zur Überprüfung der angenommenen Skalenstruktur.....	175
Tab. 8-3:	Mustermatrix der vierten berechneten Hauptachsenanalyse .....	176

Tab. 8-4:	Mustermatrix der fünften berechneten Hauptachsenanalyse.....	177
Tab. 8-5:	Skalen- und Itemkennwerte der reflektiven Skalen auf Prädiktorenebene .....	178
Tab. 8-6:	Standardisierte Faktorladungen und Messfehlervarianzen der CFA zur Analyse der Skalenstruktur.....	181
Tab. 8-7:	Evaluation des Messmodells der CFA .....	182
Tab. 8-8:	Konstruktinterkorrelationen auf Basis der CFA.....	181
Tab. 8-9:	Prüfung der Normalverteilungsannahme für die reflektiven Skalenwerte .....	183
Tab. 8-10:	Ergebnisse der Levene-Tests auf Varianzhomogenität .....	184
Tab. 8-11:	Ergebnisse der mehrdimensionalen einfaktoriellen Varianzanalyse.....	184
Tab. 8-12:	Kennwerte der formativen Indikatoren zur Erfassung des Service .....	185
Tab. 8-13:	Skalen- und Itemkennwerte der reflektiven Skala zur Kaufintention .....	186
Tab. 8-14:	Wahlhäufigkeit der Gutscheine, Kauf in Shop 1 bzw. Shop 2 nach Produktart als absolute Häufigkeit und in Prozent .....	187
Tab. 8-15:	Formative Usability Indikatoren mit mehr als 10% fehlender Werte .....	189
Tab. 8-16:	Indikatorpaarungen mit hoher bivariater Korrelation und inhaltlicher Redundanz.....	190
Tab. 8-17:	Kennwerte des PLS-Messmodells zum Zwei-Konstrukt-Modell .....	192
Tab. 8-18:	Kennwerte des PLS-Messmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: Ladungen und Gewichte .....	196
Tab. 8-19:	Kennwerte des PLS-Messmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: AVE-Werte und Composite Reliability .....	197
Tab. 8-20:	Ergebnisse des PLS-Strukturmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: aufgeklärte Varianz ( $R^2$ ) .....	199
Tab. 8-21:	Ergebnisse des PLS-Strukturmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: Pfadkoeffizienten .....	199
Tab. 8-22:	Übersicht über die mit Hilfe der PLS-Modelle überprüften Hypothesen.....	200
Tab. 8-23:	Kennwerte der sequentiellen MLR – Schritte 1 bis 2c.....	204
Tab. 8-24:	Items der endgültigen reflektiven Skala ufosV2r.....	206
Tab. 8-25:	Items der endgültigen formativen Skala ufosV2f.....	206
Tab. 8-26:	Hypothesen zur Güte der ufosV2-Skalen .....	208
Tab. 8-27:	Weitere Hypothesen zu Zusammenhängen des konzeptuellen Erfolgsmodells.....	208

## Abkürzungsverzeichnis

$\alpha$	Cronbachs Alpha
$\beta$	Pfadkoeffizient
$\delta$	Delta-Wert in der explorativen Faktorenanalyse
$\zeta$	Messfehler auf Ebene der latenten Variable
$\varepsilon$	Messfehler auf Indikatorebene
$\lambda$	Ladung
$\gamma$	Gewicht
$\eta$	latente Variable, Konstrukt
$\chi^2$	Chi-Quadrat-Wert
$\lambda_j$	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
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
PZW	Proportionale Zufallswahrscheinlichkeit in der MLR
QUIS	Questionnaire for User Interface Satisfaction
r	Korrelationskoeffizient
R <sup>2</sup>	Determinationskoeffizient (aufgeklärte Varianz)
rel(x <sub>i</sub> )	Indikatorreliabilität in der konfirmatorischen Faktorenanalyse
r <sub>it</sub>	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
SUS	Software Usability Scale
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
ufosV2f	formative Skala des ufosV2
ufosV2r	reflektive Skala des ufosV2
ur	Präfix eines reflektiven Items der Usability-Skala ufosV2r
URL	Uniform Resource Locator (Internetadresse)
UV	unabhängige Variable
vgl.	vergleiche
Vpn	Versuchspersonen
vs.	versus
Wald	Wert der Wald-Statistik
WAMMI	Website Analysis and Measurement Inventory
WebQual	Barnes & Vidgen, 2002
WOOS	Wahrgenommene Orientierung in Online-Shops
x	reflektiver Indikator
y	formativer Indikator
zRedsid	standardisierte Residuen





# 1. Einleitung

In den Anfangsjahren des Electronic Commerce boten viele Händler ihre Waren in Online-Shops an, die sich durch eine eher schlecht ausgeprägte Usability (auch Gebrauchstauglichkeit) auszeichneten. Noch immer werden viele Online-Shops betrieben, die es ihren Kunden aufgrund einer mangelnden Gestaltung sehr schwer machen, vorhandene Produkte zu finden und einen Einkauf erfolgreich abzuschließen. Negative Folgen für die Wettbewerbsfähigkeit dieser Unternehmungen sind zu erwarten. In diversen Studien, die sich mit der Frage beschäftigen, was den Erfolg von Online-Anbietern ausmacht, wurde die Erfolgswirksamkeit des Faktors Usability empirisch untermauert. Das Konstrukt ist damit als einer von mehreren kundennahen Erfolgsfaktoren im Segment Business-to-Consumer anzusehen (vgl. Petre, Minocha & Roberts, 2006). Es wird sich für den Betreiber eines Online-Shops folglich in der Regel auszahlen, durch geeignete Evaluationsmaßnahmen in die Gewährleistung einer gut ausgeprägten Usability des eigenen Internetauftritts zu investieren, da sich die generierten Umsätze auf diese Weise steigern lassen.

Zur Evaluation der Usability kann auf eine ganze Reihe verschiedener Methoden zurückgegriffen werden. Die vorhandenen Evaluationsansätze unterscheiden sich in vielerlei Hinsicht – etwa dahingehend, inwiefern sich auf Basis ihrer Anwendung konkrete Gestaltungsempfehlungen der bewerteten Benutzerschnittstelle ableiten lassen oder inwieweit Nutzer in das Verfahren eingebunden werden. Eine wesentliche Verfahrensgruppe stellen Usability-Fragebögen dar, die sich dadurch auszeichnen, dass ein Nutzerurteil zur Gebrauchstauglichkeit einer Software quantitativ erfasst wird. In der Geschichte der Forschungsdisziplin Software-Ergonomie kommt der fragebogenbasierten Erfassung der Usability eine große Bedeutung zu. Als Definition der Usability wird vielfach auf die sieben Prinzipien der Dialoggestaltung in Teil 11 der DIN EN ISO 9241 (1999) verwiesen (Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit etc.). Diese Prinzipien gehen u.a. auf eine faktorenanalytische Untersuchung von Usability-Fragebogendaten durch Dzida, Herda und Itzfeldt (1978) zurück. Seit dieser Zeit wurden diverse standardisierte Fragebogeninstrumente zur Usability-Evaluation von Software entwickelt. Einige von ihnen, wie der QUIS (Chin, Diehl & Norman, 1988) oder der IsoMetrics (Gediga, Hamborg & Düntsch, 1999), haben sich in der Praxis bewährt; ihre Validität wurde in verschiedenen empirischen Untersuchungen bestätigt. Als ein spezifisches Verfahren, das die Erfassung der kundenseitigen Bewertung der Usability von Online-Shops ermöglicht, wurde der *Usability Fragebogen für Online-Shops*, kurz *ufos*, entwickelt (Konradt, Wandke, Balazs & Christophersen, 2003).

Die Usability ist aus messtheoretischer Sicht als ein *latentes Konstrukt* anzusehen, d.h. das Konstrukt *an sich* kann nicht direkt beobachtet werden. Direkt beobachtbar (also manifest) sind jedoch verschiedene dem Konstrukt zugeordnete *Indikatoren*. Sie ermöglichen eine indirekte Erfassung des latenten Konstruktes. Im Fall eines Fragebogens stellen die Items die Indikatoren dar. Aus methodischer Sicht kann zwischen zwei verschiedenen Annahmen zur

Richtung der kausalen Beziehung zwischen einer latenten Variable und ihren Indikatoren differenziert werden (Bollen, 1989). Bei einer *reflektiven Messmodellannahme* wird davon ausgegangen, dass die Indikatoren verschiedene *Folgen* der latenten Variable darstellen. Typische Beispiele für reflektive Variablen aus dem Bereich der Psychologie sind Persönlichkeitseigenschaften wie etwa *Extraversion* (vgl. Eysenck & Eysenck, 1987). Da sich Extraversion in vielen Lebenssituationen gleichermaßen äußern sollte, bestehen hohe Korrelationen zwischen den beobachtbaren Indikatoren dieses Konstruktes. Die zweite Messmodellannahme unterstellt eine entgegengesetzte Kausalbeziehung zwischen der latenten Variable und den Indikatoren. In einem solchen *formativen Messmodell* stellen die Indikatoren die *Verursacher* des Konstruktes dar – sie formen es also. Als ein Beispiel kann der *soziale Status* dienen (Hradil, 2005). Er ergibt sich aus verschiedenen Merkmalen einer Person, die nicht hoch miteinander korrelieren müssen, aber können. Die Erfassung eines formativen Konstruktes ist nur dann als valide anzusehen, wenn alle relevanten formativen Indikatoren berücksichtigt werden, da jeder einzelne die Ausprägung der latenten Variable verändern kann.

Untersucht man die Items vorliegender Usability-Fragebögen, so zeigt sich, dass vielfach sowohl reflektive als auch formative Indikatoren enthalten sind. Die Entwicklung der Instrumente erfolgte zumeist anhand von Kriterien der klassischen Testtheorie und auf Basis der Faktorenanalyse. Dabei wurden in der Regel Items zum Zwecke der Skalenoptimierung eliminiert. Diese Herangehensweise erscheint im Fall reflektiver Indikatoren aufgrund erwarteter hoher Korrelationen zulässig; für formative Konstrukte ist sie jedoch als unangemessen anzusehen, da eine unvollständige und damit invalide Abbildung des Konstruktes resultieren kann. Ein derart inadäquates Vorgehen wurde auch bei der Entwicklung des ufos-Fragebogens verfolgt. Ziel der durchgeführten Untersuchung war es, dieses Instrument zum ufosV2 (V2 für *Version 2*) weiterzuentwickeln. Es sollten sowohl eine reflektive (ufosV2r) und als auch eine formative (ufosV2f) Skala zur Erfassung der Usability von Online-Shops aus Kundensicht entwickelt und bzgl. ihrer Güte untersucht werden.

Die Attraktivität dieses Forschungsvorhabens ist vielfältig. Zum einen kann ein vorhandener messtheoretischer Mangel eines konkreten Usability-Fragebogens ausgeräumt werden. Mit dem ufosV2 sollte ein valides Evaluationsinstrument zur Gebrauchstauglichkeit von Online-Shops aus Kundensicht bereitgestellt werden. Auf Basis des verfolgten Validierungsansatzes kann zudem die Relevanz der Usability im Kontext des Online-Shoppings untermauert werden. Es ergeben sich weiterhin Implikationen für die messtheoretische Fundierung anderer bestehender Usability-Fragebögen. Nicht zuletzt liefert die Arbeit einen Beitrag zur Betrachtung des latenten Konstruktes Usability an sich, indem aufgezeigt wird, dass eine faktorenanalytisch getriebene Definition als problembehaftet anzusehen ist und zwischen reflektiven und formativen Usability-Indikatoren differenziert werden sollte.

Im Folgenden wird der Aufbau der vorliegenden Arbeit skizziert. Das zweite Kapitel stellt den theoretischen und empirischen Hintergrund der Untersuchung vor. Der Fokus liegt auf der Usability und ihrer fragebogenbasierten Erfassung. Der Erfolgsfaktor Usability wird in ein konzeptuelles Erfolgsmodell zur Vorhersage der Kaufintention integriert, das der Validierung der ufosV2-Skalen dienen soll. Schließlich wird näher auf die Unterscheidung zwischen formativen und reflektiven Messmodellen eingegangen.

Im dritten Kapitel wird an die Unterscheidung zwischen beiden Messmodellarten angeknüpft. Es wird abgeleitet, worin die Unterschiede zwischen formativen und reflektiven Usability-Items gesehen werden. Die Ziele der Arbeit werden präzisiert. Die wesentlichen inhaltlichen Hypothesen beziehen sich auf die Einhaltung der Gütekriterien Reliabilität und Validität der zu entwickelnden Usability-Skalen ufosV2r und ufosV2f.

Die Weiterentwicklung des ufos-Fragebogens ist Gegenstand des vierten Kapitels. Aufbauend auf der Unterscheidung zwischen formativen und reflektiven Messmodellen werden die Item-pools für die beiden neuen ufosV2-Skalen zusammengestellt.

Kapitel fünf ist der Schilderung der Versuchsplanung gewidmet. Der ufosV2-Fragebogen wird im Rahmen eines Laborexperiments angewandt. Als abhängige Variable dient die Erfolgsgröße Kaufentscheidung, die an die Entlohnung der Probanden gekoppelt ist. Es wird zudem geschildert, wie die Operationalisierung der Erfolgsfaktoren erfolgt, die neben der Usability im konzeptuellen Erfolgsmodell Berücksichtigung finden.

Um die angenommenen Hypothesen zur Güte der entwickelten Skalen empirisch zu überprüfen, kommen verschiedene statistische Auswertungsverfahren zur Anwendung, deren Voraussetzungen und Gütekriterien in Kapitel sechs dargestellt werden. Zentral ist dabei zum einen der Partial Least Squares (PLS) Ansatz zur Schätzung von Strukturgleichungsmodellen sowie zum anderen die multinomiale logistische Regression (MLR). Neben PLS und MLR werden im Rahmen der Auswertung einige weitere Verfahren angewandt, die ebenfalls vorgestellt werden.

Die Durchführung des Experiments wird in Kapitel sieben dokumentiert. Dazu werden Charakteristika der Stichprobe, die Versuchsumgebung sowie der Ablauf des Experiments beschrieben. In Kapitel acht folgt die Vorstellung der Untersuchungsergebnisse. Dabei kommen die im sechsten Kapitel angeführten statistischen Verfahren zur Anwendung, um zu prüfen, ob sich die angenommenen Hypothesen zur Güte der neuen Usability-Skalen bestätigen.

Anhand dieser Ergebnisse erfolgt im neunten Kapitel eine Diskussion darüber, inwiefern die Ziele der Arbeit erreicht wurden. Es werden Vorzüge bzw. Einschränkungen des verfolgten Untersuchungsansatzes sowie Implikationen für die weitere Forschung und die Praxis besprochen. Die gesamte Arbeit wird schließlich in Kapitel zehn zusammengefasst.



## **2. Theoretischer und empirischer Hintergrund**

Das zentrale Anliegen der vorliegenden Arbeit besteht in der Weiterentwicklung eines Fragebogens zur Erfassung der Usability von Online-Shops aus Nutzersicht (ufos). Das breite Forschungsfeld des Online-Shoppings stellt den Untersuchungsrahmen der Arbeit dar, der in Abschnitt 2.1 umrissen wird. In Abschnitt 2.2 wird eine begriffliche Bestimmung des Kernkonstrukts Usability unter Bezugnahme auf die DIN EN ISO 9241 vorgenommen. Zur Evaluation der Usability von Softwareanwendungen steht eine breite Palette verschiedener Methoden zur Verfügung. In Abschnitt 2.3 erfolgt eine Vorstellung des Spektrums der unterschiedlichen Evaluationsansätze. Als spezifische Methode wird in Abschnitt 2.4 ein besonderes Augenmerk auf die fragebogenbasierte Evaluation der Usability gerichtet. Abschnitt 2.5 beleuchtet die Relevanz des Konstrukts Usability im Kontext des Online-Shoppings. In der vorliegenden Arbeit wird ein Ansatz der Skalvalidierung verfolgt, bei dem das zu erfassende Konstrukt in einen modellhaften Zusammenhang zu weiteren Variablen gestellt wird. Dazu wird in Abschnitt 2.6 ein konzeptuelles Erfolgsmodell vorgestellt, in dem die Usability zu den weiteren Erfolgsfaktoren sowie den Erfolgsgrößen von Online-Shops in Beziehung gesetzt wird. Die in dem konzeptuellen Modell berücksichtigten Zusammenhänge zwischen den Erfolgsfaktoren werden in Abschnitt 2.7 empirisch untermauert. Abschnitt 2.8 stellt vor, dass es bei der Erfassung latenter Konstrukte wie der Usability zwischen zwei Formen von Messmodellen zu differenzieren gilt. Wird diese Unterscheidung nicht berücksichtigt, kann dies negative Auswirkungen auf die Messgüte einer Skala haben.

### **2.1 Einkaufen im Internet**

Die Nutzung des Internets stellt für einen Großteil der deutschen Bevölkerung inzwischen eine Selbstverständlichkeit dar. Einer umfangreichen Studie der Arbeitsgemeinschaft Online-Forschung e.V. (AGOF, 2006) zufolge zählen im ersten Quartal des Jahres 2006 mit 55,3% mehr als die Hälfte der Deutschen über 14 Jahren zur Gruppe der Online-Nutzer. Mittlerweile sind die jungen Altersgruppen nahezu vollständig online, aber auch für Ältere wird die Nutzung immer selbstverständlicher.<sup>1</sup> Die Hauptgründe dafür, ins Internet zu gehen, liegen zumeist in den Bereichen Kommunikation und Information: 86,2% nennen die E-Mail-Nutzung als primären Grund, online zu sein. Weitere gewichtige Gründe sind Online-Recherche (85%), Nachrichten (59,7%), Electronic Commerce (58%) sowie Online-Banking (52,4%). Im ersten Quartal 2006 hatten insgesamt 72,2% der deutschen Internet-Nutzer, also 25,99 Mio Menschen, in den vergangenen zwölf Monaten online eingekauft. Für das Jahr 2006 erwartet der Hauptverband des Deutschen Einzelhandels (2005) einen Gesamtumsatz von 16,3 Milliarden Euro, was gegenüber dem Vorjahr einem Zuwachs von ca. 13%

---

<sup>1</sup> Weitere Charakteristika von Internetnutzen (z.B. Geschlechter- und Altersverteilung) können den Tabellen D-2, D-4, D-6, D-8 und D-10 im Anhang D entnommen werden.

entspricht. In Abbildung 2-1 ist das Wachstum für die Jahre 1999 bis 2006 dargestellt. Wenn auch nicht jeder im Internet einkauft, so nutzt doch nahezu jeder Onliner (94,1%, also 33,86 Mio Menschen) das Internet zum Einholen von Informationen zu Produkten (AGOF, 2006). Besonders beliebt ist die Suche nach Infos zu Urlaubs- und Last-Minute-Reisen. Aber auch über Flug- und Bahntickets, Büchern, Hotels, Autos, Eintrittskarten und Musik-CDs informieren sich Nutzer sehr häufig im Internet.

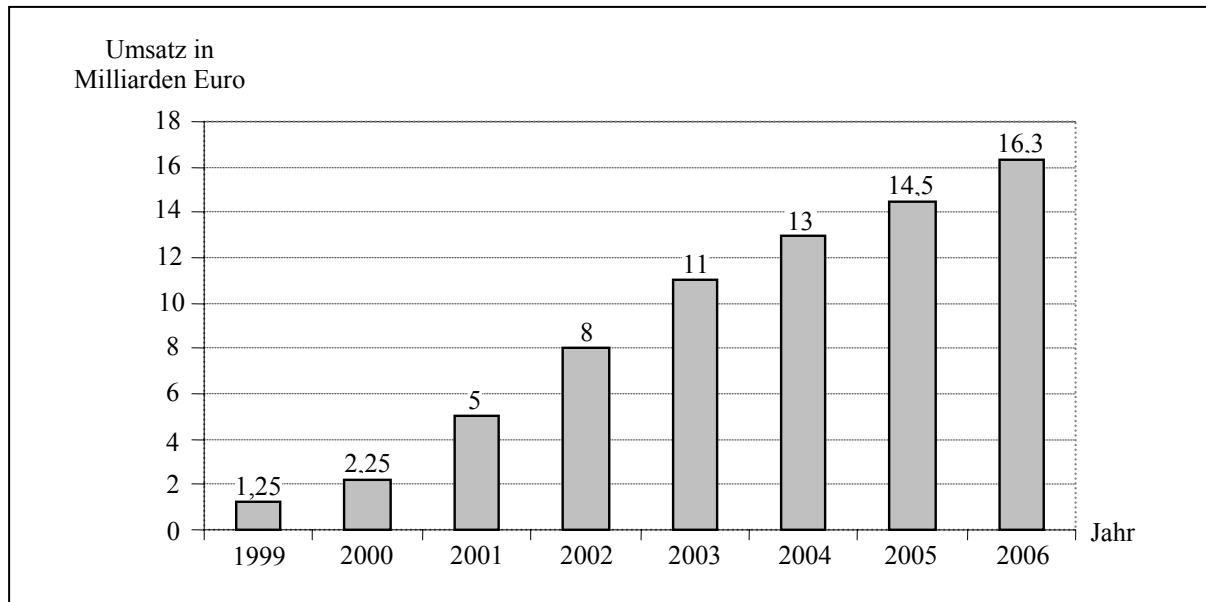


Abb. 2-1: B2C-E-Commerce Umsatzentwicklung in Deutschland (Hauptverband des Deutschen Einzelhandels, 2005)

Offensichtlich erweist sich das Einkaufen im Internet für einen Großteil der Onliner als attraktiv. In den folgenden Abschnitten werden einige generelle Aspekte zum Online-Handel dargestellt. Zunächst wird der Begriff *Electronic Commerce* beleuchtet (Abschnitt 2.1.1). Anschließend folgt in Abschnitt 2.1.2 eine Vorstellung des Unterbereichs *Business-to-Consumer* hinsichtlich seiner verschiedenen Erscheinungsformen. Abschnitt 2.1.3 ist zentralen Charakteristika von *Online-Shops* gewidmet. Schließlich werden in Abschnitt 2.1.4 verschiedene potenzielle *Vor- und Nachteile* bzw. Risiken genannt, die aus Kundensicht mit dem Einkaufen im Internet verbunden sein können.

### 2.1.1 Electronic Commerce

Electronic Commerce umschreibt breit gefasst jede Form der wirtschaftlichen Aktivität, die über elektronische Verbindungswege erfolgt (Wigand, 1996). Clement, Peters und Preiß (1999) fassen unter Electronic Commerce die digitale Anbahnung, Aushandlung und/oder Abwicklung von Transaktionen zwischen Wirtschaftssubjekten. Es kann eine Systematisierung in verschiedene Bereiche vorgenommen werden, bei der danach unterschieden wird, welche Parteien miteinander Handel treiben (vgl. Hermanns & Sauter, 2001, S. 21). In Abbildung 2-2 werden die verschiedenen Handelsformen mit Beispielen aufgeführt.

		<i>Nachfrager der Leistung</i>		
		<i>Consumer</i>	<i>Business</i>	<i>Administration</i>
<i>Anbieter der Leistung</i>	<i>Consumer</i>	Consumer-to-Consumer z.B. Internet-Kleinanzeigenmarkt, Online-Auktionshäuser (Ebay)	Consumer-to-Business z.B. Jobbörsen mit Stellengesuchen	Consumer-to-Administration z.B. Steuerabwicklung von Privatpersonen (Einkommenssteuer etc.)
	<i>Business</i>	Business-to-Consumer z.B. Angebote in Internet-Shopping-Malls	Business-to-Business z.B. Geschäfte zwischen Zulieferern und Unternehmen	Business-to-Administration z.B. Steuerabwicklung von Unternehmen (Umsatzsteuer, Körperschaftssteuer etc.)
	<i>Administration</i>	Administration-to-Consumer z.B. Abwicklung von Unterstützungsleistungen (Sozialhilfe, Arbeitslosenhilfe etc.)	Administration-to-Business z.B. Beschaffungsmaßnahmen öffentlicher Institutionen im Internet	Administration-to-Administration z.B. Transaktionen zwischen öffentlichen Institutionen im In- und Ausland

Abb. 2-2: Segmente des Electronic Commerce (Hermanns & Sauter, 2001, S. 23)

Eine der jeweils zwei beteiligten Parteien stellt im Electronic-Commerce den *Anbieter* von Waren oder Leistungen dar, die jeweils andere den *Nachfrager*. Sowohl der Anbieter als auch der Nachfrager können in einen der drei folgenden Bereiche eingeordnet werden: (1) *Business* (Unternehmen), (2) *Consumer* (Konsument als Privatperson bzw. Verbraucher) oder (3) *Administration* (öffentliche Einrichtungen bzw. Institutionen). Durch eine Paarbildung ergeben sich neun verschiedene Konstellationen bzw. Segmente des Electronic Commerce. Die vorliegende Untersuchung konzentriert sich auf den Bereich *Business-to-Consumer*, bei dem Unternehmen (als Anbieter) mit privaten Konsumenten (als Nachfrager) Handel treiben.

### 2.1.2 Erscheinungsformen des Bereichs Business-to-Consumer

Es finden sich verschiedene Erscheinungsformen des Segmentes *Business-to-Consumer*, die sich hinsichtlich der Art der Transaktion, der angebotenen Waren oder des Geschäftsmodells unterscheiden. Als Formen des Business-to-Consumer sind neben anderen vor allem *Infomediaries*, *Marktplätze*, *Auktionshäuser* und *Online-Shops* zu nennen. Teilweise erweist sich die Zuordnung eines Anbieters zu einer konkreten Form als problematisch.

*Infomediaries*: Der Begriff *Infomediary* setzt sich zusammen aus den beiden Worten *Information* und *Intermediary* (Zwischenhändler) und bezeichnet einen Internetanbieter, der größere Mengen aufbereiteter Informationen für Internetnutzer zum Abruf zur Verfügung stellt (vgl. Hagel III & Rayport, 1997). Zumeist handelt es sich um wertende Informationen zu spezifischen Produkten oder zu den Anbietern. Durch die Weitergabe dieser Infos fungiert ein Infomediary als außenstehender, neutraler Mittler zwischen Kunden und Anbietern.

*Marktplätze*: Online-Marktplätze stellen mehreren Anbietern ein Dach zur Verfügung, unter dem diese ihre eigenen Waren oder Leistungen vorstellen und verkaufen können (vgl. Timmers, 1998). Ein Marktplatz ist somit ebenfalls als ein Vermittler zwischen Nachfrager und Anbieter anzusehen. Ein Beispiel ist die Site [www.atrada.de](http://www.atrada.de) einer T-Online Tochterfirma (Atrada Trading Network AG, 2006); auf diesem Marktplatz finden sich sehr verschiedene Verkäufer, die z.B. Autozubehör, Schmuck oder Möbel anbieten.

*Auktionshäuser*: Ein Online-Auktionshaus wie Ebay kann einerseits dem Bereich Consumer-to-Consumer zugeordnet werden, da auf diesem Wege Konsumenten als Anbieter und Nachfrager via Internet miteinander Handel treiben (vgl. Ariely & Simonson, 2003). Auch nutzen Unternehmen untereinander die Möglichkeiten der Online-Auktion, um miteinander zu handeln (Business-to-Business; vgl. Anandalingam, Day & Raghavan, 2005). Weiterhin werden derartige Sites von vielen professionellen Händlern genutzt, um Waren zu Festpreisen oder zur Versteigerung anzubieten, so dass auch Online-Auktionshäuser dem Bereich Business-to-Consumer zugeschrieben werden können (vgl. Bapna, Goes & Gupta, 2003).

### 2.1.3 Online-Shops

Die im Rahmen der vorliegenden Arbeit relevante Form des Segments Business-to-Consumer stellt der „klassische“ Online-Shop dar (auch bezeichnet als E-Store, Online Store, Internet Retail Store etc.; vgl. Helander & Khalid, 2000). Dieser zeichnet sich im Wesentlichen dadurch aus, dass ein Unternehmen auf einer Site Waren zu Festpreisen anbietet. Aus softwareergonomischer Sicht handelt es sich bei einem Online-Shop um eine graphische Benutzerschnittstelle in Form einer Internetsite, auf der ein Anbieter Waren präsentiert und es seinen Kunden ermöglicht, sich über die Produkte zu informieren und sie zu bestellen. In der DIN EN ISO 9241-110 (2006) wird eine Benutzerschnittstelle definiert als: “Alle Bestandteile eines interaktiven Systems (Software oder Hardware), die Informationen und Steuerelemente



zur Verfügung stellen, die für den Benutzer<sup>2</sup> notwendig sind, um eine bestimmte Arbeitsaufgabe mit dem interaktiven System zu erledigen“.

Wie auch im traditionellen Handel lassen sich Online-Shops verschiedenen Branchen zuordnen. Einige Shops konzentrieren sich lediglich auf eine Art von Produkt (z.B. Blumen bei [www.fleurop.de](http://www.fleurop.de)), andere bieten Waren und Dienstleistungen aus einem breiten Spektrum an (z.B. Textilien, Haushaltsgeräte, Unterhaltungselektronik etc. bei [www.karstadt.de](http://www.karstadt.de)). Tabelle 2-1 stellt dar, welche Waren und Dienstleistungen im ersten Quartal des Jahres 2006 am häufigsten in Deutschland online gekauft wurden. An erster Stelle stehen seit einigen Jahren Bücher (AGOF, 2006). Weiterhin sind Veranstaltungstickets, Reisetickets (Flug und Bahn), Hotelbuchungen, Musik-CDs, Mode und Schuhe, Computer-Hardware und -Zubehör, Urlaubsreisen sowie Filme sehr begehrt. Daneben werden auch diverse andere Arten von Produkten im Internet erworben. Es kann unterschieden werden zwischen Online-Händlern, die Waren ausschließlich im Internet anbieten, und solchen, die auch Geschäfte im traditionellen Handel betreiben (vgl. Van Vliet & Pota, 2000). Weiterhin finden sich Katalog-Händler, die ihr Geschäft auf das Internet ausgeweitet haben, sowie ehemalige Vertriebs-händler.

Tab. 2-1: Häufigkeit der im Internet gekauften Waren und Dienstleistungen in Deutschland (AGOF, 2006)

Ware/Dienstleistung	%
Bücher	33,8
Eintrittskarten für Kino, Theater, Klassische Konzerte, Popkonzerte	29,2
Flug- und Bahntickets	25,2
Hotels für Urlaubs- oder Geschäftsreisen	23,8
Musik-CDs	21,8
Mode oder Schuhe	20,9
Computer-Hardware und -Zubehör	20,8
Urlaubsreisen und Last-Minute-Reisen	19,9
Filme auf DVDs, Videos	17,2
Computer-Software ohne Games	15,6
Telekommunikationsprodukte, wie z.B. Handys oder schnurlose Telefone	14,6
Computer- und Videogames	13,6
Unterhaltungselektronik, wie z.B. HiFi-, TV- oder DVD-Geräte, Satellitenreceiver	13,2
Gebührenpflichtige Musik oder Filme aus dem Internet als Download	11,6
Sportartikel, Sportgeräte	10,8
Elektrische Haushaltsgeräte	10,2
Möbel, Wohnungseinrichtung	8,4
Schmuck und Uhren	7,7
Heimwerkerbedarf, Heimwerkergeräte	7,7
Gesundheitsprodukte, Medikamente	7,2
Autos	6,6
Parfums, Düfte für Damen oder Herren	6,4

**Anmerkung:** Lesebeispiel „33,8% der deutschen Internetnutzer haben in den letzten 12 Monaten Bücher online gekauft“; Erhebungszeitraum der Internet Facts 2006-I-Studie: 01.01. bis 31.03.2006.

<sup>2</sup> Im Bereich Software-Ergonomie wird der Einfachheit halber zumeist von *dem* Nutzer gesprochen, wobei natürlich sowohl weibliche als auch männliche Personen gemeint sein können. Der Grund für die Eingrenzung von Aussagen auf einen *konkreten* Nutzer liegt in der Kontextabhängigkeit der Usability (siehe Abschnitt 2.2.1.1).

Softwaretechnisch betrachtet, handelt es sich bei Online-Shops oftmals um sehr komplexe Sites (vgl. Lee, Lee, Kim & Stout, 2004). Wie auch im traditionellen Handel kann die Anzahl der angebotenen Produkte stark variieren (vgl. Brynjolfsson, Smith & Hu, 2003). Zudem können die Preise zwischen den Shops aufgrund unterschiedlicher Preisstrategien der Anbieter verschieden ausfallen (Chun & Kim, 2005; Sotgiu & Ancarani, 2005). Wie den diversen Online-Sites zum Vergleich von Anbietern (z.B. [www.geizkragen.de](http://www.geizkragen.de)) zu entnehmen ist, bestehen weiterhin Unterschiede bzgl. der Versandgebühren, Zahlungsmöglichkeiten, Garantiegewährleistungen, Rückgaberegelungen und weiteren spezifischen Serviceleistungen. Angesichts der vielfältigen Möglichkeiten der internetbasierten Darstellung von Informationen erstaunt es wenig, dass Online-Shops sehr unterschiedlich gestaltet sind (vgl. Helander & Khalid, 2000). Einige Sites bieten umfangreiche multimediale Informationen (z.B. Video und Audio-Streams oder 3-D Ansichten von Produkten, Downloads von Demo-Versionen etc.), wohingegen andere Shops kaum mehr sind als eine HTML-basierte Produktauflistung (vgl. Karat, Karat, Vergo, Pinhanez, Riecken & Cofino, 2002).

Trotz der vielfältigen Gestaltungsmöglichkeiten haben sich inzwischen Standards des gestalterischen Aufbaus von Online-Shops entwickelt, die die Site-Struktur, die verwendeten Symbole und Benennungen sowie die Lokalisierung von Seitenelementen betreffen (vgl. Adkisson, 2002). Weitgehend durchgesetzt hat sich die Verwendung der Warenkorb- oder Einkaufswagen-Metapher, die eine Anlehnung an traditionelle Formen des Handels darstellt (vgl. Fletcher & Flechter, 2003; siehe auch Abschnitt 2.2.2.2). Auf Kundenseite bestehen entsprechend allgemeine Erwartungen bzgl. der Site-Gestaltung, die sich auf das Nutzungsverhalten beim Erstbesuch eines Shops auswirken (vgl. Bellman & Rossiter, 2004). Beispielsweise gehen Nutzer davon aus, sowohl den Warenkorb als auch Hilfe-Links oben rechts auf den Seiten eines Shops zu finden (Bernard, 2001, 2002; Bernard & Sheshadri, 2004). Ein weiterer Grund für eine ähnliche Gestaltung von Shops ist darin zu sehen, dass vielfach Standard-Plattformen wie von Strato ([www.strato.de](http://www.strato.de)) verwendet werden.

#### **2.1.4 Vor- und Nachteile des Online-Shoppings aus Kundensicht**

Im Verhältnis zu anderen Transaktionsformen zeichnet sich das Einkaufen im Internet im Segment Business-to-Consumer aus Sicht der Kunden durch eine Reihe von Vor- und Nachteilen aus (Alba et al., 1997). Diese werden im Folgenden erläutert.

##### **2.1.4.1 Vorteile**

Bei der (Weiter-)Entwicklung eines Online-Shops sollte der jeweilige Anbieter sich stets vor Augen führen, worin aus Sicht seiner Kunden die Vorteile des Einkaufens im Internet liegen. Denn werden diese nicht angemessen adressiert, ist zu fürchten, dass ein Nutzer eine andere Möglichkeit des Einkaufs wahrnehmen wird. Als Vorteile aus Sicht von Kunden sind die Folgenden zu nennen:

*Bequemlichkeit:* Viele Kunden sehen das hohe Maß an Bequemlichkeit als einen besonderen Vorteil des Online-Shoppings an (vgl. Bellman, Lohse & Johnson, 1999; Jarvenpaa & Todd, 1997; Kau, Tang & Ghose, 2003; Keeney, 1999; Maguire & Pearce, 2001; Wolfinbarger & Gilly, 2001). Der Einkauf im Internet erfordert wenig Zeit und kann schnell vom eigenen Rechner aus getätigt werden (Keeney, 1999). Die Waren werden direkt nach Hause geliefert. Es besteht also keine Notwendigkeit, einen Shop persönlich aufzusuchen, was vielfach als angenehm erlebt wird (Jarvenpaa & Todd, 1997). Zudem schätzen Kunden die Möglichkeit, zeitunabhängig einzukaufen, da keine Ladenschlusszeiten bestehen (Wolfinbarger & Gilly, 2001). Die Bequemlichkeit des Online-Einkaufs erweist sich daher vor allem für Personen als vorteilhaft, die unter Zeitmangel leiden oder in ihrer Mobilität eingeschränkt sind.

*Informationsdichte:* Zumeist stehen umfangreiche Informationen zu den Waren zur Verfügung, was aus Sicht vieler Kunden als vorteilhaft erlebt wird (Wolfinbarger & Gilly, 2001). Zwar muss dies nicht auf Ebene des einzelnen Anbieters der Fall sein, da viele Shops die umfangreichen Möglichkeiten der multimedialen Produktdarstellung wenig nutzen (Fiore, Kim & Lee, 2005). Jedoch eröffnet sich das gesamte Internet zur Suche nach Produktinformationen (vgl. Chang, 2003). Durch Peer-to-Peer Infos können Kunden sowohl zu Anbietern als auch zu Produkten die Meinungen anderer Kunden und Benutzer einholen.

*Produktauswahl:* Die Produktauswahl im Internet ist sehr groß, und Online-Kunden wissen dies zu schätzen (Wolfinbarger & Gilly, 2001). Oft sind Produkte verfügbar, die im traditionellen Handel am Wohnort nicht zu finden sind, zumal online sogar internationale Anbieter besucht werden können. Die Anzahl der in Online-Shops angebotenen Produkte ist oftmals sehr viel größer als in vergleichbaren Ladengeschäften (Brynjolfsson et al., 2003).

*Suchkosten, Wechselkosten und Anbietervergleich:* Die Suchkosten (*Search Costs*) von Kunden sind im Internet gering (vgl. Bakos, 1997; Schmitz & Latzer, 2002). Es ist verhältnismäßig leicht, zu einem anderen Anbieter zu wechseln, falls ein Shop nicht den eigenen Bedürfnissen genügt. Aufgrund der Vielzahl verschiedener Anbieter kann ein Vergleich verschiedener Shops schnell gezogen werden, etwa mit Hilfe von Preisvergleichsdiensten (sogenannten Shopbots), um den billigsten Händler zu identifizieren (vgl. Iyer & Pazgal, 2003). Aufgrund dessen werden dem Einkauf im Internet oftmals Preisvorteile gegenüber dem traditionellen Handel zugeschrieben (vgl. Forsythe, Liu, Shannon & Gardner, 2006).

*Anonymität:* Einige Online-Kunden empfinden es als angenehm und vorteilhaft, dass beim Einkauf im Internet Anonymität herrscht (Wolfinbarger & Gilly, 2001). Hintergrund ist dabei, dass Verkäufer im traditionellen Handel oftmals als unfreundlich und wenig hilfbereit erlebt werden und zudem sozialer Druck zum Kaufen ausgeübt werden kann.

Zusammenfassend besteht aus Kundensicht ein finanzieller (z.B. Preisvergleich), informationeller (z.B. Produktinfos), quantitativer (Produktauswahl), logistischer (z.B. Ortsunabhängigkeit), zeitlicher (z.B. Suchkosten) und sozialer (Anonymität) Mehrwert des Online-Shoppings (vgl. Bucher, Büffel & Wollscheid, 2002, S. 136).

### 2.1.4.2 Nachteile und Risiken

Allgemein wird das Online-Shopping als risikoreicher als das Einkaufen im traditionellen Handel empfunden (Lee & Tan, 2003). Dies ist auf ein deutlich höheres Maß an Unsicherheit für die Kunden zurückzuführen, das zum Teil darin begründet ist, dass in einer computermedierten Beziehung diverse persönliche Hinweisreize wie Sprache, Gestik und Mimik nicht oder allenfalls eingeschränkt zur Beurteilung der Situation bzw. zur Einschätzung des Gegenübers herangezogen werden können (Bryant & Colledge, 2002; Winand & Pohl, 2000). Als Risiken bzw. Nachteile des Einkaufens im Internet sind zu nennen:

*System-inhärente Risiken:* Ein Großteil der deutschen Internetnutzer sah sich im ersten Quartal des Jahres 2005 mit Sicherheitsproblemen beim Surfen im WWW konfrontiert (Statistisches Bundesamt - Pressestelle, 2006; siehe Abbildung 2-3). Zwar können internetinhärente Risiken sowohl von Seiten der Kunden als auch des Anbieters reduziert werden, z.B. durch verschlüsselte Datenübertragung, Firewalls sowie Authentifizierungsverfahren (vgl. Furnell & Karweni, 1999). Die generelle Bereitschaft der Kunden, das Internet zum Einkaufen zu nutzen, wird jedoch nicht durch die objektive sondern durch die subjektiv empfundene Sicherheit des Internets bestimmt.

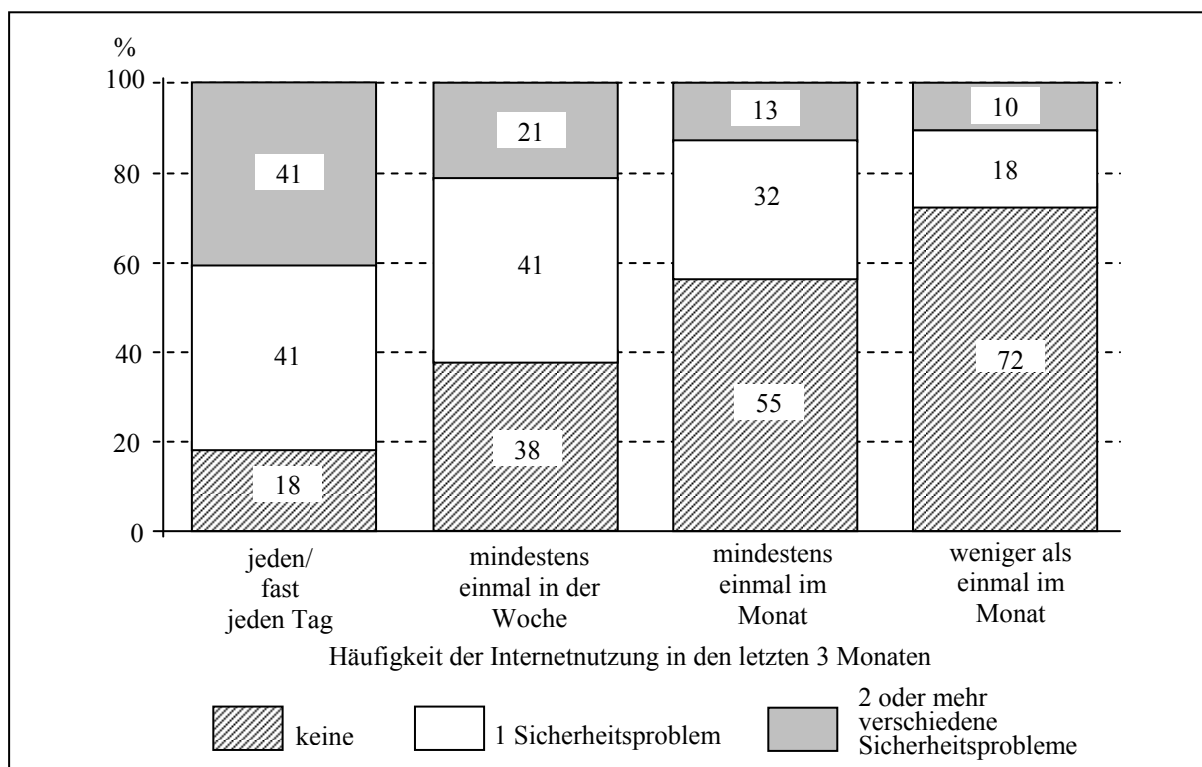


Abb. 2-3: Anzahl der aufgetretenen Sicherheitsprobleme im ersten Quartal 2005 (Statistisches Bundesamt, 2006)

*Transaktionsspezifische Risiken:* Die verschiedenen Arten transaktionsspezifischer Risiken beruhen auf der asymmetrischen Informationsverteilung seitens der Transaktionsakteure (Weiber & Adler, 1995). *Kompetenzrisiken* beziehen sich auf die Wahrnehmung der Kunden, inwiefern ein Anbieter überhaupt in der Lage ist, die Transaktion angemessen abzuwickeln

und sie somit zufrieden zu stellen (vgl. Jarvenpaa & Tiller, 2001). Weiterhin besteht aus Kundensicht eine Gefahr des *Anbieter-Opportunismus*, der dann vorliegt, wenn ein Verkäufer intentional betrügt. Dies ist z.B. der Fall, wenn unsachgemäß abgerechnet wird.

*Unbequemes Einkaufen*: Die Bequemlichkeit des Einkaufs kann aufgrund einer mangelnden Usability eingeschränkt sein, was vielfach ein Kaufhemmnis darstellt (siehe Abschnitt 2.5.2; vgl. Tognazzini, 2004).

*Keine Produkte zum Anfassen*: Viele Kunden sehen es als ein Hemmnis an, die Produkte nicht in die Hand nehmen zu können (Bucher et al., 2002, S. 136), denn dies birgt das Risiko, dass das Produkt eventuell nicht den eigenen Vorstellungen entspricht.

*Bedenken zur Privatsphäre und Sicherheit*: Aus Sicht der Kunden erweisen sich vor allem die Wahrung der eigenen Privatsphäre sowie die Gewährleistung einer sicheren Transaktion als wichtig. Entsprechend kommt den Aspekten *Privacy* und *Security* in der Literatur ein vermehrtes Interesse zu, wobei einige Autoren nicht zwischen beiden Konstrukten differenzieren (vgl. Brown & Muchira, 2004; Diller, Lin & Tashjian, 2003; Dinev & Hart, 2005; Liu, Marchewka, Lu & Yu, 2005; Resnick & Montania, 2003).

So wie es bei der Entwicklung einer kundenfreundlichen Benutzerschnittstelle Anspruch sein muss, die Vorteile des Online-Shoppings besonders zu berücksichtigen, gilt es ebenso, den genannten potenziellen Risiken entgegen zu treten. Wesentlich ist dabei die Vermittlung von Vertrauen (siehe Abschnitt 2.6.2).

### **2.1.5 Erfolgsfaktorenforschung im Online-Shopping**

Ziel der vorliegenden Untersuchung ist die Entwicklung von Skalen zur Usability im Kontext des Online-Shoppings. Die Usability wird als ein Erfolgsfaktor betrachtet, der zum Zwecke der Skalvalidierung in Verbindung zu Erfolgskriterien gesetzt wird. Entsprechend ist die Arbeit dem Bereich der internetbezogenen *Erfolgsfaktorenforschung* zuzuordnen. Im Folgenden wird skizziert, was gemeinhin unter diesem Begriff verstanden wird (Abschnitt 2.1.5.1), welche allgemeinen methodischen Ansätze verfolgt werden (Abschnitt 2.1.5.2) und in welcher Hinsicht die Erfolgsfaktorenforschung kritisiert wird (Abschnitt 2.1.5.3).

#### **2.1.5.1 Begriffsklärung**

Im Rahmen wirtschaftlicher Forschung wird vielfach nach gültigen Einflussfaktoren, so genannten *Schlüsselfaktoren* oder *Erfolgsfaktoren* gesucht, die unternehmerischen Erfolg verursachen (Steinle, Kirschbaum & Kirschbaum, 1996). Anliegen der *Erfolgsfaktorenforschung* ist es, Erkenntnisse zu liefern, die sich in der Praxis von Managementseite aufgreifen lassen, um unternehmerische Ziele zu erreichen. Dabei sind Studien, die die Erfolgswirksamkeit einzelner Aspekte untersuchen, nicht der Erfolgsfaktorenforschung zuzuordnen (Fritz, 1990). Empirische Untersuchungen beziehen sich entweder auf einzelne Branchen oder sind branchenübergreifend orientiert.

Die Erfolgsfaktorenforschung entstammt ursprünglich der Beratungspraxis auf Einzelfall-ebene und zielte ausschließlich darauf ab, unternehmensspezifische Faktoren zu identifizieren. Die damit verbundene ideographische Entscheidungsorientierung trat im Laufe der Zeit gegenüber einer stärkeren Forschungsorientierung in den Hintergrund, die nomothetisch orientiert ist (Grünig, Heckner & Zeus, 1996), indem sich der Anspruch an die Reichweite der abzuleitenden Empfehlungen auf ganze Branchen ausweitete (Nicolai & Kieser, 2002). Im Rahmen der Forschung zum unternehmerischen Erfolg steht gemeinhin ein hypothesengeleiteter Erkenntnisgewinn im Vordergrund, der die kausalen Zusammenhänge zwischen Stellgrößen und dem Erfolg betrifft. Dabei besteht der Anspruch einer Verallgemeinerung der erlangten Ergebnisse, was eine Übertragung auf die Einzelbetriebsebene oft erschwert. Der entscheidungsorientierte Ansatz ist dagegen eher explorativer Natur und sucht nach Implikationen auf Ebene der einzelnen Unternehmung. Beide Orientierungen sind schwer miteinander vereinbar, da einerseits möglichst generalisierbare Aussagen abgeleitet werden sollen und andererseits die Erfolgsmaximierung des spezifischen Einzelfalls im Vordergrund steht.

#### 2.1.5.2 Grundsätzliche Methodik der Erfolgsfaktorenforschung

Es lassen sich grob fünf verschiedene methodische Zugänge der Erfolgsfaktorenforschung unterscheiden (Haenecke, 2002; in Anlehnung an Grünig et al., 1996; siehe Abbildung 2-4). Diese werden zunächst dahingehend abgegrenzt, inwiefern die Erfolgsfaktoren *direkt* oder *indirekt* ermittelt werden. Eine direkte Ermittlung erfolgt durch eine Befragung von Experten; hierbei wird wiederum zwischen einer *methodisch* (z.B. Kreativitätstechniken, heuristische Verfahren) und einer *methodisch materiell* (z.B. Nutzung von Checklisten) gestützten Ermittlung differenziert. Ein indirekter Zugang beruht dagegen auf der Anwendung statistischer Verfahren oder gedanklicher Analysen zur Untersuchung des Erfolgseinflusses von Faktoren. Beim *qualitativen* indirekten Ansatz werden typischerweise mehrere erfolgreiche Unternehmen ausgewählt und auf gemeinsame Eigenschaften hin untersucht, die als Erfolgsfaktoren angesehen werden (z.B. Peters & Waterman, 2000; Albers, Panten & Schäfers, 2002). Dagegen erlaubt eine *quantitative* Herangehensweise die Anwendung statistischer Auswertungsverfahren, die entweder *explorativ* (ein entdeckendes Vorgehen ohne strenge Vorannahmen) oder *konfirmatorisch* (im Vorfeld angenommene Zusammenhänge prüfend) orientiert sein kann.

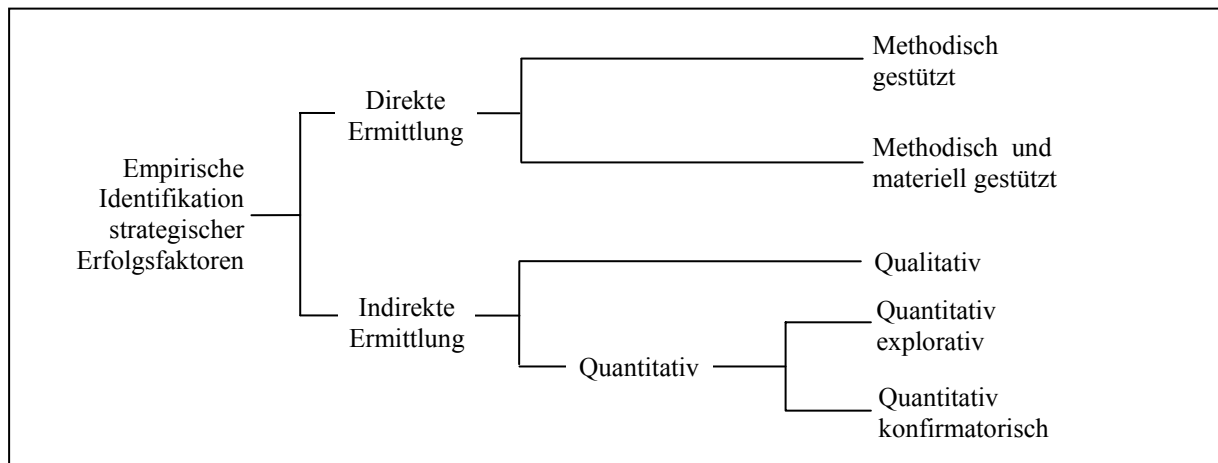


Abb. 2-4: Methoden zur Identifikation von Erfolgsfaktoren (Haenecke, 2002)

Gängige Praxis der quantitativen Erfolgsfaktorenforschung ist es, ein Ergebnismaß zu wählen, das als Indikator für den Erfolg einer Unternehmung fungiert (vgl. Rohn, 2006); ein typisches Beispiel für eine solche Erfolgsgröße ist der ROI (Return on Investment; vgl. Newman, 1966). Weiterhin werden Variablen auf Prädiktorebene identifiziert und operationalisiert, von denen erwartet wird, dass sie in einem Zusammenhang zur Erfolgsgröße stehen. Beispiele für typische Faktoren sind Produktivität, Marktposition, Innovation, vertikale Integration oder Kostendruck. Mit Hilfe korrelativer, regressionsanalytischer oder pfadanalytischer statistischer Verfahren wird schließlich empirisch geprüft, inwiefern die erwarteten Zusammenhänge Bestätigung finden (Haenecke, 2002). Vorausgesetzt wird somit einerseits, dass die Erfolgsgröße einen validen Indikator für den unternehmerischen Erfolg darstellt, sowie andererseits, dass die prädiktierenden Größen einer validen Messung zugänglich sind.

### 2.1.5.3 Kritik an der Erfolgsfaktorenforschung

An den gebräuchlichen Formen der Erfolgsforschung wird seit einiger Zeit in verschiedener Hinsicht Kritik geübt. Die kritischen Argumente werden von March und Sutton (1997) sowie Nicolai und Kieser (2002) zusammengefasst. Sie führen an, dass auf wissenschaftlicher Seite scheinbar kein tatsächliches Interesse an der Praxisrelevanz der vermeintlich gewonnenen Erkenntnisse besteht. Zudem ist eine starke Methodenabhängigkeit zu beobachten, wonach die Art der statistischen Methoden bestimme, was für Zusammenhänge „entdeckt“ würden (Vanderwerf & Mahon, 1997). Bzgl. des gängigen methodischen Vorgehens werden weitere Vorbehalte geäußert: (1) zugrunde liegende Stichproben seien oft unrepräsentativ, (2) die oft nur auf Basis von Querschnittsdaten betrachteten Beziehungen zwischen Variablen würden vielfach kausal interpretiert werden, obwohl dies in der Regel nicht zulässig ist, und (3) Daten unterlägen häufig einem Key Informant Bias - um nur einige Punkte zu nennen. Eine zentrale Kritik richtet sich gegen den Anspruch, auf Grundlage gewonnener Erkenntnisse die Performance einzelner Unternehmen optimieren zu können. Dabei wird argumentiert, die Verbreitung des Wissens zu Erfolgsfaktoren mache diese gleichermaßen unwirksam (March & Sutton, 1997). Angesichts schneller Marktentwicklungen sei zudem oft unklar, ob die in

der Vergangenheit identifizierten Erfolgsfaktoren auch in der Zukunft eine Relevanz besitzen würden. Auch kann der Erfolgsfaktorenforschung insbesondere aus Sicht des gewissermaßen „theorie-verwöhnten“ Psychologen eine Art Theorieentleertheit zugesprochen werden: kaum einer Studie liegt eine *Theorie* im eigentlichen Sinne zugrunde, die den Ansprüchen genügt, um von einer solchen sprechen zu können.

Innerhalb der Betriebswirtschaftslehre ist auf Grundlage der genannten Kritikpunkte eine rege Diskussion entbrannt (vgl. Bauer & Sauer, 2004; Fritz, 2004; Homburg & Krohmer, 2004; Nicolai & Kieser, 2004; siehe auch [www.dialog-erfolgsfaktorenforschung.de](http://www.dialog-erfolgsfaktorenforschung.de)). Verfechter des Forschungsstranges argumentieren, es sei unangemessen, der Erfolgsfaktorenforschung Erfolglosigkeit zu unterstellen, ohne zuvor klar zu definieren, worin sich ihr potenzieller Erfolg widerspiegeln sollte (Fritz, 2004). Zudem bestünde bzgl. der Erfolgswirksamkeit einzelner Konstrukte wie der *Marktorientierung* durchaus breiter Konsens, was die Kritik entkräfte, der Forschungsansatz habe bislang keine gesicherten Erkenntnisse erbracht. Weiterhin würden die Schlussfolgerungen von March und Sutton (1997) sowie Nicolai und Kieser (2004) auf einem Bild der Erfolgsfaktorenforschung beruhen, das nicht der Realität entspreche sondern stark pauschalisiert sei (Homburg & Krohmer, 2004).

Als Fazit für die vorliegende Arbeit ist festzuhalten, dass eine Studie im Bereich der Erfolgsfaktorenforschung diversen methodischen Anforderungen genügen muss (Repräsentativität der Stichprobe etc.), eine Praxisrelevanz der Untersuchung nachzuweisen ist und eine angemessene Definition von Erfolg zugrunde zu legen ist. Im folgenden Abschnitt wird an die letztgenannte Forderung angeknüpft.

### **2.1.6 Erfolgskriterien von Online-Shops**

Im Rahmen einer Arbeit, die dem Bereich der Erfolgsfaktorenforschung zugeordnet werden kann, erweist sich eine Beschäftigung mit möglichen Erfolgsgrößen als unerlässlich. Zwar kommt dem Erfolg von Unternehmen in der betriebswirtschaftlichen Forschung eine zentrale Bedeutung zu, allerdings bestehen keine allgemeinen Konventionen bzgl. der Frage, welche Erfolgsgrößen die wesentlichen darstellen bzw. wie diese zu messen sind (Mellewigt & Matiaske, 2000; Rohn, 2006). Die Erfassung des Erfolgs kommerzieller Websites ist allgemein als recht schwierig anzusehen (vgl. Auger, 2005). Bezogen auf Informationstechnologien wird oftmals die tatsächliche Nutzung als angemessenes Kriterium angesehen (vgl. DeLone & McLean, 1992, 2003; Devaraj, Fan & Kohli, 2003). Auch im Kontext des Electronic Commerce finden sich Studien, denen die Nutzung als Erfolgsgröße zugrunde liegt (Lohse & Spiller, 1998; Kim & Prabhakar, 2004). Die vielfach als Erfolgsgröße verwendete Conversion-Rate ergibt sich aus dem Verhältnis der generierten Käufe zur Anzahl der Besucher einer Site (Moe & Fader, 2004; Montgomery, Li, Srinivasan & Liechty, 2002). Dieses Maß trägt dem Umstand Rechnung, dass eine Site, die häufig genutzt wird (etwa für die Suche nach Produktinformationen), nicht unbedingt erfolgreich sein muss – nämlich dann, wenn die Nutzer nicht zu Käufern werden. Neben Nutzungsmaßen finden sich in Studien auch



diverse andere Erfolgsgrößen wie die subjektive Performance-Einschätzung von Seiten firmeninterner Manager (Wade & Saggi, 2005) oder die Loyalität von Online-Kunden (Anderson & Srinivasan, 2003; Bhattacharjee, 2001; Danaher, Wilson & Davis, 2003; Kim, Lee & Moon, 2001; Shankar, Smith & Rangaswamy, 2003). Weiterhin wird als abhängige Variable die Wahl zwischen mehreren Shops (Lowengart & Tractinsky, 2001) oder die Weiterempfehlungsintention abgefragt (Liu et al., 2005). Im Rahmen der vorliegenden Untersuchung wird auf die Erfolgsgrößen Kaufentscheidung und Kaufintention fokussiert, die in den folgenden beiden Kapiteln näher betrachtet werden.

### 2.1.6.1 Kaufentscheidung

Eine Typologie aus dem Bereich der Markt- und Konsumpsychologie unterscheidet drei verschiedene Kaufhandlungen auf Grundlage von Unterschieden bzgl. des Ausmaßes an Planung, des Grades an Impulsivität und der kognitiven Beteiligung seitens des Käufers (Wiswede, 1995). Ein *Spontankauf* wird aufgrund interner und externer Stimuli getätigt. Im Fall eines *überlegten Kaufs* liegt der Kaufentscheidung eine gewisse aktive kognitive Beteiligung zugrunde. Dagegen stellt ein *Gewohnheitskauf* zumeist das Ergebnis vorausgehender überlegter Käufe dar. Diese drei Arten von Kaufhandlungen sind gewissenmaßen als Idealtypen anzusehen, wobei in der Realität häufig Mischformen auftreten. Von einer Kaufentscheidung kann nach Wiswede (1995) nur dann die Rede sein, wenn dem Käufer mindestens drei verschiedene Alternativen zur Verfügung stehen und ein Mindestmaß an kognitiver Beteiligung vorliegt. Folglich kann nur ein überlegter Kauf mit einer Kaufentscheidung assoziiert sein.

Aus Sicht von Helander und Khalid (2000) ergibt sich ein Online-Kauf aus einer Reihe verschiedener Entscheidungen seitens der Kunden. Eine erfolgreiche Bestellung erfordert nach ihrer Auffassung insgesamt fünf wesentliche Entscheidungen: (1) *den Shop zu besuchen*, (2) *durch den Shop zu navigieren*, (3) *den Bestellvorgang zu beginnen*, (4) *zu kaufen*<sup>3</sup> und nach erfolgter Lieferung die (5) *Entscheidung das Produkt zu behalten*. Als besonders interessant erweisen sich im Kontext der geplanten Untersuchung die Schritte (3) und (4). Die *Entscheidung den Bestellvorgang zu beginnen* wird zu dem Zeitpunkt getroffen, wenn ein Kunde sich entschließt, den Inhalt des Warenkorb zu erwerben, indem der Bestellvorgang durch entsprechenden Buttonklick initiiert wird. Dies entspräche im traditionellen Handel dem Entschluss, sich an der Kasse anzustellen. Die *Entscheidung zu kaufen* bildet den Abschluss des Bestellvorganges, indem ein Kunde die Bestellung unter Angabe des Bezahlungsweges endgültig abschickt. Übertragen auf den Einkauf im traditionellen Ladengeschäft käme dies dem Aushändigen des Geldes, der Eingabe einer EC-Kartenummer etc. an der Kasse gleich.

---

<sup>3</sup> Helander und Khalid (2000) verwenden für die *Entscheidung den Bestellvorgang zu beginnen* und die *Entscheidung zu kaufen* andere Bezeichnungen: Entscheidung zu kaufen bzw. Entscheidung zu bezahlen. Dies könnte aus Sicht des Autors der vorliegenden Arbeit jedoch für Verwirrung sorgen.

Für den Erfolg eines Shops dürfte die *Entscheidung zu kaufen* wohl die wichtigste der fünf genannten Entscheidungen darstellen, da hierdurch bestimmt wird, wie viele Käufe insgesamt getätigt werden. Im Folgenden wird sie entsprechend als die *Kaufentscheidung* bezeichnet. Zwar ließe sich streng genommen argumentieren, letztlich determiniere die *Entscheidung das Produkt zu behalten* die endgültige Anzahl der Kaufabschlüsse. Diese wird jedoch nicht primär von den verschiedenen Qualitätsdimensionen des Webauftritts eines Online-Shops beeinflusst, die im Vordergrund der Untersuchung stehen, sondern eher von den Produkteigenschaften und Kontextfaktoren. Die Zahl der Studien, in denen die Kaufentscheidung als abhängige Variable dient, ist im Vergleich zur Anzahl an Untersuchungen, in denen Intentionsmaße als Erfolgsgrößen von Online-Anbietern herangezogen werden, eher gering (u.a. Konradt et al., 2003; Liang & Lai, 2002).

#### 2.1.6.2 Kaufintention

In sehr vielen marktpsychologischen Untersuchungen zum Online-Shopping wird als Erfolgsgröße die Kaufintention erhoben (z.B. Bhattacharjee, 2002; Gefen & Straub, 2000; Jarvenpaa et al., 2000; Khalifa & Limayem, 2003; Konradt et al., 2003; Koufaris & Hampton-Sosa, 2004). Nach Ajzen (1991) fassen Intentionen allgemein die motivationalen Faktoren zusammen, die ein Verhalten beeinflussen. Sie stellen Anzeichen dafür dar, wie viel Aufwand ein Individuum betreiben wird, um ein Verhalten auszuführen. Generell gilt, dass die Wahrscheinlichkeit einer Verhaltensausführung umso höher ist, je stärker die entsprechende Intention ausfällt. Die Kaufintention beschreibt somit die Motivation eines Kunden, einen Einkauf zu tätigen.

Die Güte eines Intentionsmaßes wird maßgeblich durch die Stärke des Zusammenhangs zwischen der Intention und dem tatsächlichen Verhalten bestimmt. Einflussreiche psychologische Theorien zur Vorhersage konkreten Verhaltens sind die Theorie des überlegten Handelns (Fishbein & Ajzen, 1975) bzw. die Theorie des geplanten Verhaltens (Ajzen, 1985, 1991). In beiden Theorien wird die Verhaltensintention eines Individuums als direkter Prädiktor für ihr tatsächliches Verhalten angesehen. Ein starker Zusammenhang zwischen Intention und Verhalten besteht gemäß den Theorien, wenn die folgenden drei Voraussetzungen erfüllt sind:

- Die Intensionsmessung sollte dem Ausmaß an Generalität des Verhaltens entsprechen. Gilt es also, ein spezifisches Verhalten vorherzusagen, so ist die Intention in gleichem Maße spezifisch zu erheben („Korrespondenzannahme“ nach Ajzen & Fishbein, 1977).
- Ein mögliches Problem besteht darin, dass sich die Intention zwischen den Zeitpunkten der Erfassung und der Verhaltensbeobachtung ändert. Negative mit dem Verhalten assoziierte Aspekte werden beim Individuum umso salienter, je näher das Verhalten zeitlich rückt. Eine solche Umgewichtung positiver und negativer Erwägungen hat unter Umständen zur Folge, dass die Intention im Vorfeld der Handlungsausführung revidiert

wird (Ajzen, 1985). Damit steigt die Wahrscheinlichkeit einer Intensionsänderung, je länger das Zeitintervall zwischen Intentionserhebung und Verhaltensbeobachtung wird.

- Das fragliche Verhalten muss unter der willentlichen Kontrolle des Handlungsausführenden stehen (Ajzen, 1985).

Die drei genannten Voraussetzungen sind auf die Kaufintention zu übertragen. Die erste impliziert, dass ein möglichst spezifisches Verhalten angesprochen werden muss, um eine hohe Kongruenz zwischen Intention und dem prädiktierten Verhalten zu erreichen. Dies dürfte dann gegeben sein, wenn dem Käufer zum Zeitpunkt der Intensionsmessung bereits klar ist, welches Produkt er erwerben möchte. Hat ein Kunde diesbezüglich keine konkreten Vorstellungen, besteht die Gefahr, dass die Intensionsabfrage einen sehr hypothetischen Charakter erhält.

In Hinblick auf die zweite Voraussetzung führen Grabner-Kräuter und Kaluscha (2003) Zweifel bezüglich der prädiktiven Validität von Kaufintensionsmaßen an. Sie weisen darauf hin, dass Intentionen keineswegs automatisch Verhalten implizieren müssen. Es sei unsicher, inwiefern Kunden, die eine hohe Intention zum Online-Einkauf äußern, auch tatsächlich die damit verbundenen Risiken<sup>4</sup> eingehen. Tatsächlich ist im Sinne der zweiten Voraussetzung zum Zusammenhang zwischen Intention und Verhalten anzunehmen, dass potenzielle Risiken des Einkaufs kurz vor dem endgültigen Treffen einer Kaufentscheidung besonders salient sind. Eine hohe Salienz möglicher Risiken erhöht die Wahrscheinlichkeit einer Revidierung der Verhaltensplanung.

Bzgl. der dritten Voraussetzung ist als Einschränkung zu berücksichtigen, dass so wie im traditionellen Handel auch im Internet vielfach Spontankäufe getätigt werden (vgl. Wiswede, 1995). Für diese Art von Kaufverhalten erweist sich eine Prädiktion durch ein Intensionsmaß sicherlich als problematisch. Eine weitere Einschränkung bzgl. der willentlichen Kontrolle ist darin zu sehen, dass Internet-Kunden häufig zunächst eine Entscheidung zu einem Kauf treffen und den Bestellvorgang des Online-Shops daraufhin beginnen, ihn jedoch wieder abbrechen, bevor sie die Bestellung endgültig abgeschlossen haben (vgl. Helander & Khalid, 2000). Die Gründe für diese Abbrüche liegen zumeist in mangelhafter Usability der Shops, denn aus Sicht der Nutzer gestaltet sich der Bestellvorgang oft als besonders problematisch bzw. unangenehm (siehe auch Abschnitt 2.5.1). Beispielsweise wird in manchen Shops erst zu diesem Zeitpunkt deutlich, welche zusätzlichen Kosten mit einer Bestellung verbunden sind (vgl. Schindler, Morrin & Bechwati, 2005). Weiterhin sind Kunden oft nicht bereit, dem Anbieter spezifische Daten wie ihre Telefonnummer mitzuteilen, deren Angabe jedoch von Seiten des Anbieters in Pflichtfeldern der Bestellformulare gefordert wird. Helander und Khalid (2000) unterscheiden angesichts der hohen Anzahl an Abbrechern zwischen der *Entscheidung den Bestellvorgang zu beginnen* und der *Entscheidung zu kaufen* (siehe Vorab-

---

<sup>4</sup> Verschiedene Risiken des Online-Shoppings aus Kundensicht werden in Abschnitt 2.1.4.2 dargestellt.

schnitt). Im Fall von Intentionsmaßen, die *vor* Beginn des Bestellvorgangs erhoben werden, kann somit aufgrund von Usability-Problemen eine verringerte Prädiktionsgüte für die endgültige Kaufentscheidung resultieren.

In diversen Untersuchungen im Bereich „traditionellen“ Handels erwies sich die Kaufintention bei vorhandener willentlicher Kontrolle als signifikanter Prädiktor für tatsächliches Kaufverhalten (Ajzen & Driver, 1992; Clawson, 1971; East, 1993; Ferber & Piskie, 1965; Notani, 1997; Pickering & Isherwood, 1974; Tauber, 1975; Warshaw, 1980). Pavlou (2003) verwendete in einer Online-Befragungsstudie die *Kaufintention* als abhängige Variable. Zusätzlich zog der Autor ein Item heran, bei dem die Befragten angaben, wie häufig sie in der Vergangenheit in dem Shop gekauft hätten. In einem PLS-Strukturgleichungsmodell ergab sich ein signifikanter Pfadkoeffizient von  $\beta=0,46$  zwischen der Kaufintention und dem genannten Item. Wenngleich es sich hierbei nicht um eine *Prädiktion zukünftigen Verhaltens* handelt, so kann dieser Befund doch als ein Hinweis für die Güte der Kaufintention als Erfolgsgröße gewertet werden.

Vielfach setzt sich eine Skala zur Erfassung der Kaufintention im Zusammenhang des Online-Shoppings aus mehreren Items zusammen, die strenggenommen unterschiedliche Intentionen betreffen. Bei Hampton-Sosa und Koufaris (2005) wird Verhaltensintention als Kombination aus der Intention zu kaufen und der Intention zum Shop zurückzukehren gebildet.

### **2.1.7 Zwischenfazit und Überleitung**

Online-Shopping ist für viele Internet-Nutzer mittlerweile eine Selbstverständlichkeit geworden. Der im Internet erzielte Gesamtumsatz in Deutschland ist hoch. Erworben werden sehr unterschiedliche Waren, wobei sich beispielsweise Bücher als besonders attraktiv erweisen. Begünstigt wird die positive Entwicklung im Online-Handel durch die zahlreichen Vorteile aus Kundensicht gegenüber traditionellen Handelsformen, wie etwa ein hohes Maß an Bequemlichkeit, eine große Produktauswahl und geringe Suchkosten. Ein recht großer Unterschied besteht zwischen dem Anteil an Internet-Nutzern, die online kaufen (72,2%), und denen, die sich online über Produkte informieren (94,1%). Die aus Kundensicht mit dem Internet-Einkauf verbundenen Nachteile bzw. Risiken stellen potenzielle Gründe für einen Nicht-Kauf dar. Die Erfolgsfaktorenforschung im Online-Shopping befasst sich u.a. mit der Frage, welche Erfolgsfaktoren bestimmen, ob Kunden zu einer positiven Kaufentscheidung kommen bzw. eine hohe Kaufintention ausbilden. Es drängt sich der Schluss auf, dass diejenigen Anbieter erfolgreich sein dürften, die in der Lage sind, die Vorteile und Nachteile des Online-Shoppings adäquat zu adressieren. Ein wesentlicher Faktor stellt dabei die Usability dar, die im nächsten Abschnitt zunächst auf allgemeiner Ebene betrachtet wird.

## 2.2 Usability – Begriffsklärung

Im folgenden Abschnitt soll das Konstrukt Usability einer näheren Betrachtung unterzogen werden. Zunächst stellt Abschnitt 2.2.1 dar, wie der Begriff Usability bzw. Gebrauchstauglichkeit im Rahmen der DIN EN ISO 9241 gefasst wird. In Abschnitt 2.2.2 werden drei Ansätze zur Betrachtung der Usability vorgestellt, die keinen direkten Bezug zur besagten DIN Norm herstellen, jedoch in der Literatur vielfach zur theoretischen Fundierung des Konzeptes angeführt werden.

### 2.2.1 Usability in der DIN EN ISO 9241

Zur Beschreibung des Konstrukts *Usability* kann die internationale Normreihe DIN EN ISO 9241 (1992-2006) herangezogen werden, in der die „ergonomischen Anforderungen für die Büroarbeit mit Bildschirmgeräten“ international festgelegt sind. In der deutschen Version der Norm findet sich für den Begriff *Usability* die Übersetzung *Gebrauchstauglichkeit*. Dzida (1994, S. 387), der wesentliche Impulse zur wissenschaftlichen Betrachtung des Konstrukts gesetzt hat, spricht sich für eine synonyme Verwendung der Begriffe *Usability*, *Gebrauchstauglichkeit* und *Benutzbarkeit* aus.<sup>5</sup> Ursprünglich entstammt das Konstrukt der anwendungsorientierten Wissenschaftsdisziplin *Software-Ergonomie*, die einen Teilbereich der Forschung zur *Mensch-Computer-Interaktion* (Human-Computer Interaction, HCI) darstellt (Dzida & Wandke, 2006). Software-ergonomische Fragestellungen betreffen das Zusammenwirken zwischen Mensch und einer Software-Benutzerschnittstelle in der Arbeitswelt. Das Ziel der Disziplin liegt in der Anpassung der Eigenschaften von technischen Artefakten (Software) an die Eigenschaften und Bedürfnisse von Benutzern.

#### 2.2.1.1 Effektivität, Effizienz und Zufriedenheit im Benutzungskontext

Teil 11 der DIN EN ISO 9241 (1999) stellt die übergeordnete Ebene der softwareergonomischen Normung dar (Dzida, 1994) und wird entsprechend häufig als Definition der Usability herangezogen:

*Die Gebrauchstauglichkeit ist das Ausmaß, in dem ein Produkt von einem bestimmten Benutzer verwendet werden kann, um bestimmte Ziele in einem bestimmten Nutzungskontext effektiv, effizient und zufriedenstellend zu erreichen.*

Die Gebrauchstauglichkeit ergibt sich gemäß dieser Anschauung aus der Effektivität, Effizienz und Zufriedenheit einer Zielerreichung in einem konkreten Nutzungskontext (siehe auch Abbildung 2-5). Dabei betrifft die *Effektivität* das Maß an Genauigkeit und Vollständigkeit, mit der ein Benutzer ein spezifisches Ziel mit Hilfe des Softwareproduktes erreichen kann. Die *Effizienz* ist als das Verhältnis zwischen der Effektivität (also der

---

<sup>5</sup> Eine inhaltliche Abgrenzung der Begriffe Usability, Gebrauchstauglichkeit und Benutzbarkeit erscheint auch im Zusammenhang der vorliegenden Untersuchung nicht sinnvoll. Zugunsten des Leseflusses wird auf eine Festlegung auf lediglich einen Begriff verzichtet.

Genauigkeit und Vollständigkeit) und dem mit der Zielerreichung verbundenen Aufwand aufzufassen. Entsprechend stellt die Effektivität eine Voraussetzung für die Effizienz dar: Ein Produkt kann nur dann als effizient bewertet werden, wenn die gewünschten Arbeitsergebnisse vollständig und genau erreicht werden. Mit der *Zufriedenheit* wird im Rahmen des Konzeptes Gebrauchstauglichkeit zudem die nutzerseitige Produktbewertung und -akzeptanz berücksichtigt. Diese hängt von der subjektiven Wahrnehmung eines Nutzers bzgl. der Effizienz und Effektivität ab.

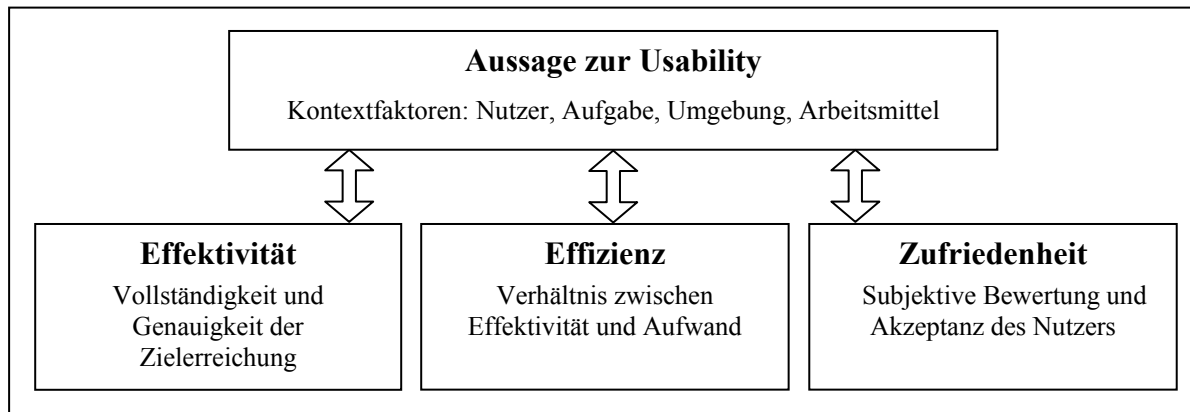


Abb. 2-5: Konzeptionalisierung der Usability in der DIN EN ISO 9241-11 (1999)

Wie oben beschrieben, besagt die DIN EN ISO 9241-11 (1999, S. 4), eine Aussage zur Usability könne nur für einen konkreten Kontext der Nutzung getroffen werden (vgl. Whiteside, Bennet & Holtzblatt, 1988, S. 796). Die Usability kann somit nicht als eine feststehende Produkteigenschaft angesehen werden, sondern ist als ein *Attribut einer Interaktion eines Nutzers im konkreten Nutzungskontext* aufzufassen. Bevan und Macleod (1994) stellen beispielhaft vor, welche Einzelaspekte bei der Nutzung einer im Arbeitskontext genutzten Software zu berücksichtigen sind (siehe Tabelle 2-2).

Die aus dem Nutzungskontext erwachsenden Anforderungen lassen sich mit Hilfe der Anforderungsanalyse (*Requirements Analysis*) im Sinne des Usability Engineering identifizieren (vgl. Hamborg & Gediga, 2006; Mayhew, 2003). Diese gliedert sich in die *Nutzeranalyse*, die *Arbeits- und Aufgabenanalyse* sowie die *Bedarfsanalyse* (vgl. Hix & Hartson, 1993). Der Gesamtkontext der Benutzung wird durch den *Nutzer*, die *Arbeitsaufgaben*, die *Arbeitsmittel* sowie die *physische und soziale Umgebung* bestimmt.

Tab. 2-2: Beispielhafte Aufschlüsselung des Nutzungskontextes für eine Arbeitssoftware (Bevan &amp; Macleod, 1994, S. 137)

<i>Nutzer</i>	<i>Arbeitsaufgabe</i>	<i>Physische &amp; soziale Umgebung</i>	<i>Arbeitsmittel</i>
<i>Fähigkeiten &amp; Wissen</i> Produkterfahrung Systembezog. Wissen Aufgabenbezog. Wissen Organisationale Erfahrung Training Tipp- & Eingabeleistungen Qualifikationen Sprachl. Fähigkeiten Allgemeinbildung  <i>Individuelle Merkmale</i> Alter Geschlecht Physische Leistung Physische Einschränkungen & Behinderung Intellektuelle Fähigk. Einstellung Motivation	Aufgabenaufschlüsselung Aufgabenname Aufgabenziel Aufgabenhäufigkeit Aufgabendauer Häufigkeit von Ereignissen Aufgabenflexibilität Physische & mentale Anforderungen Aufgabenabhängigkeiten Aufgabenergebnisse Risiko im Fehlerfall	<i>Organisationale Umgebung &amp; Struktur</i> Arbeitsstunden Stellung im Unternehmen Assistenz Unterbrechungen Managementstruktur Kommunikationsstruktur Unternehmenskultur Organisationale Ziele Werte & Normen Arbeitsflexibilität Performanz Monitoring Performanz Feedback Pacing Autonomie Diskretion  <i>Arbeitsplatzdesign</i> Freiraum & Möbel Körperliche Haltung  <i>Arbeitsplatzsicherheit/ Gesundheitsrisiken</i>	<i>Basis Beschreibung</i> Produktidentifikation Produktbeschreibung Wesentliche Bereiche der Applikation Hauptfunktionen  <i>Spezifikation</i> Hardware Software Materialien Andere Items

- *Nutzer*: Ein Interface, das auf eine bestimmte Nutzergruppe perfekt zugeschnitten ist, kann für eine andere Gruppe unbrauchbar sein. Nutzer sind daher sowohl bzgl. ihrer soziodemographischen Merkmale (Alter, Geschlecht etc.), ihrer Fähigkeiten (intellektuell, soziale und körperliche Skills etc.) als auch ihres Wissens (aufgaben-, domänen-, systembezogenes Wissen, Allgemeinbildung etc.) zu beschreiben.
- *Aufgabe*: Es gilt die Aufgabe näher zu charakterisieren. Dazu gehört unter anderem, die einzelnen Handlungen aufzuschlüsseln, die Ziele zu beschreiben und Zielkonflikte zu benennen. Kognitive und psychische Beanspruchungen sowie potenzielle Arbeitsrisiken sind zu ermitteln (vgl. Hacker & Richter, 1980). Die Norm DIN EN ISO 9241-2 (1992) nennt aus ergonomischer Sicht wünschenswerte Aufgabencharakteristika.
- *Arbeitsmittel*: Einen weiteren Teil des Nutzungskontextes stellt die technische Ausstattung dar, die bei der Aufgabenerledigung zum Einsatz kommt. Es sind die Charakteristika der Hardware zu beschreiben. Neben der zu evaluierenden Software gilt es auch die weiteren Softwareprodukte zu berücksichtigen, die im Zuge der Aufgabe genutzt werden (z.B. bei einer Website-Evaluation: Browser, Popup-Blocker, Antiviren-Software, Firewalls etc.).
- *Physische & soziale Umgebung*: Die Umgebung muss sowohl hinsichtlich ihrer physischen als auch ihrer sozialen Eigenschaften beschrieben werden. Im Arbeitskontext gilt dabei, unternehmensbezogene Charakteristika zu analysieren. Von Relevanz können beispielsweise die Temperatur, die Luftfeuchtigkeit sowie die Raumgestaltung sein.

### 2.2.1.2 Anforderungen an die Dialoggestaltung

Neben Teil 11 der DIN EN ISO 9241 (1999) wird am häufigsten der Teil 10 (1996) derselben Normreihe herangezogen, um das Konstrukt Usability näher zu charakterisieren bzw. aufzuschlüsseln. Teil 10 wurde jüngst durch eine aktualisierte Fassung, den Teil 110 ersetzt (DIN EN ISO 9241-110, 2006). Der Geltungsbereich wurde von der Büroarbeit mit Bildschirmgeräten auf alle Formen der Mensch-Maschine-Interaktion ausgeweitet. In diesem Teil der Norm werden sieben Grundsätze der Dialoggestaltung genannt, die allgemein auf jede Form der Dialogtechnik (unabhängig davon, ob menübasiert, kommandobasiert etc.) anzuwenden sind:

- *Aufgabenangemessenheit*: Ein Dialog ist aufgabenangemessen, wenn er einen Benutzer unterstützt, die Arbeitsaufgabe effektiv und effizient zu erledigen.
- *Selbstbeschreibungsfähigkeit*: Ein Dialog ist selbstbeschreibungsfähig, wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Benutzer auf Anfrage erklärt wird.
- *Steuerbarkeit*: Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist.
- *Erwartungskonformität*: Ein Dialog ist erwartungskonform, wenn er konsistent ist und den Merkmalen des Benutzers entspricht, z.B. seinen Kenntnissen aus dem Arbeitsgebiet, seiner Ausbildung und seiner Erfahrung sowie den allgemein anerkannten Konventionen.
- *Fehlertoleranz*: Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens eines Benutzers erreicht werden kann.
- *Individualisierbarkeit*: Ein Dialog ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe sowie an die individuellen Fähigkeiten und Vorlieben eines Benutzers zulässt.
- *Lernförderlichkeit*: Ein Dialog ist lernförderlich, wenn er einen Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet.

Die genannten Gestaltungsgrundsätze machen den „Feel“ einer Benutzerschnittstelle aus (Oppermann & Reiterer, 1997). Sie gehen auf empirische Studien von Dzida et al. (1978) zurück (vgl. Beimel, Schindler & Wandke, 1994). Ziel der Autoren war es herauszufinden, was Nutzer in der Interaktion mit Dialogsystemen möchten und erwarten bzw. die nutzerseitig wahrgenommene Qualität durch distinkte Qualitätsaspekte zu beschreiben. Zu diesem Zweck stellten sie einen Katalog mit insgesamt 100 Qualitätsanforderungen zusammen. Diesen Anforderungen kann aufgrund ihrer Formulierungen ein gewisser Guideline-Charakter zugeschrieben werden (z.B. „be tolerant towards erroneous user input“). Die Itemsammlung



wurde von Nutzern (Mitglieder des *German Chapter of the ACM* bzw. der *Gesellschaft für Informatik*) auf einer sieben-stufigen Skala bewertet. Die Befragung ergab eine hohe Relevanzeinschätzung bei 57 der berücksichtigten Aspekte. Diese wurden faktorenanalytisch (Hauptachsenanalyse mit Varimax-Rotation) untersucht. Es ergab sich eine Struktur mit sieben Faktoren, die insgesamt 44% der Gesamtvarianz aufklärten und wie folgt interpretiert wurden: (1) Eignung für das Problem, (2) Selbstbeschreibung, (3) Erlernbarkeit, (4) Reliabilität, (5) Kontrollierbarkeit, (6) Fehlertoleranz und (7) Flexibilität. Die Autoren gaben an, die Faktorenlösung erweise sich über verschiedene Nutzergruppen nicht als gänzlich valide. Die inhaltliche Interpretation dieser Faktoren, die Dzida et al. (1978) vornahm, bildete zunächst die Grundlage für den ehemaligen deutschen Standard der ergonomischen Dialoggestaltung (DIN 66234-8, 1988) und später für die sieben Gestaltungsgrundsätze der DIN EN ISO 9241-11 (1999) bzw. DIN EN ISO 9241-110 (2006).

### 2.2.1.3 Weitere Usability-bezogene Teile der Norm

Aus software-ergonomischer Sicht sind insgesamt acht Teile der DIN EN ISO 9241 von Relevanz (vgl. Dzida, 1995). Neben den genannten Qualitätsfaktoren der Gebrauchstauglichkeit aus den Teilen 110 und 111, handelt sich dabei weiterhin um die Teile 12 bis 17. Dzida (1994) ordnet die Normbestandteile hierarchisch an (siehe Abbildung 2-6). Auf oberster Ebene wird in Teil 111 die Konzeptionalisierung der Usability in Effektivität, Effizienz und Zufriedenheit vorgenommen (DIN EN ISO 9142-11, 1999). Dem direkt unterzuordnen sind die abstrakt formulierten Teile 110 und 112. In Teil 112 werden *Gestaltungsgrundsätze der Informationsdarstellung* angeführt (DIN EN ISO 9142-12, 2000): Klarheit, Unterscheidbarkeit, Prägnanz, Konsistenz, Erkennbarkeit, Lesbarkeit und Verstehbarkeit. Diese Eigenschaften präsentierter Informationen stellen statische Aspekte einer Schnittstelle dar und bedingen damit ihren „Look“ (vgl. Oppermann & Reiterer, 1997). *Anforderungen an die Benutzerführung* werden in Teil 113 beschrieben (Eingabeaufforderungen, Feedback, Statusanzeigen, Fehlersupport und Hilfsfunktionen; DIN EN ISO 9142-13, 2000). Die Teile 114 bis 117 befassen sich mit spezifischeren Anforderungen bei verschiedenen Formen der Dialoggestaltung (DIN EN ISO 9241-14, 2000; DIN EN ISO 9241-15, 1999; DIN EN ISO 9241-16, 2000; DIN EN ISO 9241-17, 2000).

Nach Dzida (1994) kann ein Software-Produkt im Fall einer Normenkonformität (also bei Erfüllung der ISO-Normen) nicht grundsätzlich als gebrauchstauglich angenommen werden. Zudem wird kritisiert, die Formulierung der Anforderungen falle in Folge des breiten Normgeltungsbereichs derart allgemein aus, dass eine Konformitätsprüfung oftmals kaum möglich wäre (Bevan & Macleod, 1994).

Gebrauchstauglichkeit (Teil 11)				
Informationsdarstellung (Teil 12)				
Grundsätze der Dialoggestaltung (Teil 110)				
Benutzerführung (Teil 13)	Menü-Dialogführung (Teil 14)	Kommando-Dialogführung (Teil 15)	Direkte Manipulation (Teil 16)	Formular-Dialogführung (Teil 17)

Abb. 2-6: Software-ergonomische Teile der DIN EN ISO 9241 (Dzida, 1994, S. 381)

Anmerkung: prinzipielle Anforderungen in Zeilen, spezielle Anforderungen in Spalten.

## 2.2.2 Weitere Auffassungen zur Usability

Neben der DIN EN ISO 9241 wird zur Begriffsklärung vielfach auf die Ausführungen zur Usability von Jakob Nielsen (1993, S. 26) zurückgegriffen, die Gegenstand von Abschnitt 2.2.2.1 ist. Vereinzelt wird in der vorliegenden Arbeit auf den Ansatz der Mentalen Modelle verwiesen, der als eine Art theoretisches Fundament der Usability angesehen werden kann. Um dem Leser eine inhaltliche Einordnung zu ermöglichen, erfolgt in Abschnitt 2.2.2.2 eine knappe Darstellung dieses Ansatzes. Da in der für diese Arbeit relevanten Literatur anstelle der Usability oft auf die zwei verwandten Konstrukte Perceived Ease of Use und Perceived Usefulness Bezug genommen wird, befasst sich Abschnitt 2.2.2.3 mit den inhaltlichen Beziehungen dieser Variablen.

### 2.2.2.1 Jakob Niensens Usability-Begriff

Nielsen (1993, S. 26) nennt als wesentliche Merkmale der Usability die fünf Attribute (1) Erlernbarkeit, (2) Effizienz, (3) Einprägsamkeit, (4) Fehler und (5) Zufriedenheit. Im Gegensatz zur DIN-Norm nimmt Nielsen (1993) Stellung zum Begriff *Benutzerfreundlichkeit*. Er äußert, Nutzer würden keine Maschinen benötigen, die freundlich zu ihnen seien, sondern solche, die ihnen bei der Arbeit nicht im Wege stünden (Nielsen, 1993, S. 23). Zudem sei es erstrebenswerter, sich auf messbare Aspekte der Usability zu beziehen als auf ein warmes, verschwommenes Gefühl der Benutzerfreundlichkeit (Nielsen, 1993, S. 27). Diese Position steht in einem Widerspruch zum inzwischen starken Interesse an emotionalen Aspekten in der Literatur zur Mensch-Maschine-Interaktion (vgl. Norman, 2004b).

### 2.2.2.2 Mentale Modelle

Gediga und Hamborg (2002) beklagen, dass eine „Theorie der Gebrauchstauglichkeit“ noch immer ausstünde. Angesichts dieser Theorielosigkeit wird als kognitionspsychologische Grundlage für die Interaktion von Nutzern mit einer Benutzerschnittstelle oftmals die Theorie der mentalen Modelle herangezogen (z.B. Satzinger & Olfman, 1998; Stibel, 2005). Mit Hilfe dieses Ansatzes lässt sich modellhaft veranschaulichen und begründen, inwiefern konkrete gestalterische Lösungen einer Software gebrauchstauglich ausfallen. Ein mentales Modell stellt eine gedankliche Repräsentation eines genutzten Systems dar, die ein Nutzer aufbaut und während des Umgangs mit dem System heranzieht (Van der Veer & Del Carmen Puerta Melguizo, 2003). Die Bildung mentaler Modelle erfolgt im Regelfall durch Analogieschluss auf Grundlage bereits vorhandenen Wissens, das in einem Bezug zum benutzten System steht (vgl. Dutke, 1994). Zentral ist dabei die Verwendung von Metaphern (vgl. Benyon & Imaz, 1999). So werden etwa Nutzer, die zum ersten Mal einen Online-Shop besuchen, mit hoher Wahrscheinlichkeit auf ihr Wissen über das Einkaufen im physikalischen Geschäft sowie zur Benutzung von Computern und Internetseiten zurückgreifen. Folglich gilt es, sich an übliche Formen der Gestaltung von Software bzw. von Internetseiten zu orientieren und den Online-Einkaufsvorgang an den im traditionellen Geschäft anzulehnen. Eine solche Anlehnung stellt die Metapher des Warenkorbes dar. Jeder weiß, wie man mit einem Warenkorb im physikalischen Geschäft umgeht. Die Funktionsweise eines Online-Warenkorbes ist daran angelehnt.

### 2.2.2.3 Perceived Ease of Use und Perceived Usefulness

Außerhalb der Disziplin Mensch-Maschine-Interaktion wird bei der Untersuchung von Software-Benutzungsproblemen vielfach das von Davis (1989, 1993) postulierte Technology Acceptance Model (TAM) herangezogen, dessen Ziel in der Vorhersage von Akzeptanz zur Nutzung neuer Technologien liegt. Zwei Kernvariablen des Modells sind die Konstrukte Perceived Ease of Use (PEOU) und Perceived Usefulness (PU). Davis definiert die PEOU als den Grad, mit dem eine Person glaubt, die Benutzung eines bestimmten Systems erfordere keinen Aufwand (Davis, 1989, S. 320). Daneben wird die PU als der Grad ausgefasst, mit dem eine Person glaubt, dass die Benutzung eines bestimmten Systems seine oder ihre Arbeitsperformanz erhöht (Davis, 1989, S. 320). Im Rahmen der vorliegenden Arbeit ist weniger das gesamte TAM von Interesse. Es stellt sich jedoch die Frage, wie die Konstrukte PEOU und PU gegenüber dem Konstrukt Usability abzugrenzen sind, um eine inhaltliche Einordnung TAM-bezogener Publikationen vornehmen zu können.

Hinsichtlich der inhaltlichen Abgrenzung der drei Konstrukte herrscht in der Literatur Uneinigkeit. Calisir und Calisir (2004) sind offenbar der Auffassung, die drei Konstrukte PEOU, PU und Usability würden sich als inhaltlich diskriminant erweisen und betrachten ihren Einfluss auf weitere Variablen getrennt voneinander. Ein Grund für die empirische Diskriminanz der Variablen in ihrer Studie ist nach Auffassung des Verfassers der

vorliegenden Arbeit die mangelnde Operationalisierung der Skalen. Venkatesh (2000; Venkatesh & Davis, 2000) nennt die objektive Usability (siehe dazu Abschnitt 2.3.3) als eine von mehreren Determinanten der PEOU. Dagegen sieht Nielsen (1993) die Usability als einen unter mehreren Faktoren an, die die Akzeptanz eines Produktes beeinflussen, wobei die Usability und die Utility gemeinsam die Usefulness formen. Die *Utility* betreffe dabei die Frage, inwiefern ein System prinzipiell in der Lage sei, das zu tun, was es soll; die Usability beziehe sich hingegen auf die Frage, wie gut Nutzer ein System nutzen können. Eine deutliche inhaltliche Parallele besteht zwischen dem Usability-Kriterium Aufgabenangemessenheit (DIN ISO EN 9241-10, 1996) und der oben genannten Definition der PU von Davis (1989). Gemäß der Norm ist ein Dialog aufgabenangemessen, wenn er einen Benutzer unterstützt, die Arbeitsaufgabe effektiv und effizient zu erledigen.

Die Sichtung der genannten Quellen bringt folglich keine eindeutige Klärung, wie die Konstrukte inhaltlich abzugrenzen sind. Offensichtlich bestehen inhaltliche Parallelen bzw. Zusammenhänge, die sich ad hoc nicht genauer spezifizieren lassen. Im Folgenden werden entsprechend auch Studien zur PEOU und PU herangezogen, wenn es gilt, die Zusammenhänge zwischen der Usability und anderen Variablen zu beleuchten.

## 2.3 Methoden zur Evaluation der Usability

Die Evaluation der Usability kann mit Hilfe sehr unterschiedlicher Methoden erfolgen (vgl. Hamborg & Gediga, 2006; Nielsen & Mack, 1994). Diese lassen sich hinsichtlich diverser Kriterien wie der Breite der Evaluation, des Datenniveaus, der benötigten Expertise der Anwender, der Kosten und des Ortes der Durchführung unterscheiden (Oppermann & Reiterer, 1997). In Abschnitt 2.3.1 werden zunächst zwei unterschiedliche Zielsetzungen der Usability-Evaluation gegeneinander abgegrenzt. Um dem Leser eine Einordnung der fragenbasierten Evaluation zu erleichtern, folgt in Abschnitt 2.3.2 eine kurze Einführung zur Bandbreite der verschiedenen Evaluationsmethoden. In Abschnitt 2.3.3 werden die Unterschiede zwischen subjektiven und objektiven Usabilitydaten erläutert.

### 2.3.1 Formative und summative Evaluation von Software

Aus software-ergonomischer Sicht kann die Evaluation von Software zwei unterschiedliche Ziele verfolgen (vgl. Hamborg & Gediga, 2006):

- Die *summative Evaluation* stellt eine abschließende Form der Bewertung dar. Sie liefert Antworten auf die Fragen, welche Software besser ist als andere (z.B. im Rahmen des Vergleichs konkurrierender Prototypen) und wie gut eine Software ist (z.B. Abgleich mit zuvor gesteckten, quantifizierbaren Zielen; Bewertung eines Softwareendproduktes; Prüfung auf Konformität mit Normen).
- Die *formative Evaluation* hingegen wird mit der Zielsetzung durchgeführt, Probleme in der Nutzung zu identifizieren, zu beschreiben und Verbesserungsvorschläge abzuleiten. Es handelt sich somit um eine (weiter-)entwicklungsorientierte Form der Bewertung von (unfertiger) Software im Rahmen iterativer Produktverbesserungen oder korrektiver Arbeitsgestaltung. Ihr Nutzen kann damit gegenüber dem summativen Ansatz höher eingeschätzt werden (Carroll, 1997). Die verschiedenen Usability-Methoden eignen sich in unterschiedlichem Maße zur formativen oder summativen Evaluation. Entscheidend für die Eignung sind die Art der zu gewinnenden Ergebnisse sowie der zur Durchführung des Verfahrens nötige Aufwand. Im weiteren Verlauf der Arbeit wird häufig von „formativen“ Messmodellen die Rede sein. Es sei darauf hingewiesen, dass keine direkte inhaltliche Beziehung zur „formativen“ Form der Usability-Evaluation besteht.

### 2.3.2 Deskriptive vs. prädiktive Usability-Evaluationsmethoden

Einer Systematisierung von Gediga und Hamborg (2002) folgend lassen sich Usability-Evaluationsmethoden in zwei Hauptkategorien einteilen: in *deskriptive* und *prädiktive Methoden* (vgl. auch Hamborg & Gediga, 2006; siehe Abbildung 2-7). Angesichts diverser ausführlicher schriftlicher Darstellungen der einzelnen methodischen Ansätze soll im Folgenden lediglich ein Überblick gegeben werden. Die jeweils angegebene Literatur bietet die Möglichkeit einer vertieften Beschäftigung mit dem genauen methodischen Vorgehen

sowie den spezifischen Vor- und Nachteilen. Vergleiche zwischen mehreren Verfahren bzgl. verschiedener Kriterien ziehen beispielsweise Jeffries, Miller, Wharton und Uyeda (1991), John und Marks (1997), Lavery, Cockton und Atkinson (1997), Molich, Ede, Kassgaards und Karyukin (2004) oder Virzi (1997).

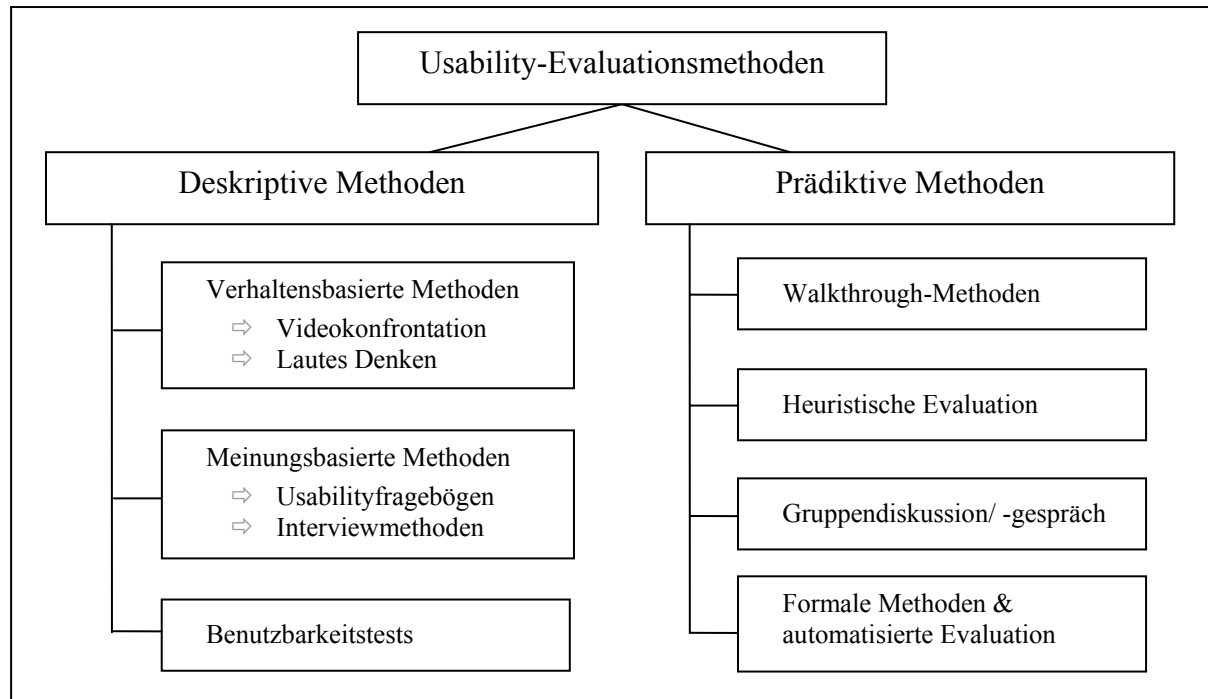


Abb. 2-7: Systematisierung verschiedener Usability-Evaluationsmethoden (in Anlehnung an Hamborg & Gediga, 2006)

### 2.3.2.1 Deskriptive Methoden

Mit Hilfe *deskriptiver Usability-Evaluationsmethoden* wird das Ziel verfolgt, die ergonomische Qualität einer Software objektiv, reliabel und valide aus Benutzersicht *beschreibend* zu erfassen, also den aktuellen Zustand und die akuten Problembereiche der Software darzulegen (vgl. Hamborg & Gediga, 2006). Es lassen sich drei unterschiedliche Ansätze deskriptiver Evaluationsmethoden unterscheiden: *verhaltensbasierte Methoden*, *meinungsbasierte Methoden* sowie *Benutzbarkeitstests*.

*Verhaltensbasierte Methoden*: Die verhaltensbasierten Methoden beziehen sich auf die Erfassung tatsächlichen Verhaltens von Nutzern in der Interaktion mit der Software durch Beobachtung. Als konkrete Methoden werden die *Videokonfrontation* sowie die Technik der *Lauten Denkens* im Folgenden knapp vorgestellt.

- *Videokonfrontation*: Bei der Methode der Videokonfrontation wird Benutzern eine Videoaufzeichnung seiner Interaktion mit der Software vorgestellt (Döhl, 2001; Hamborg & Greif, 1999; Moll, 1987). Dabei besteht die Möglichkeit, auf spezifische Sequenzen zu fokussieren, um diese mit Hilfe eines Interviews näher zu analysieren. Durch eine (Halb-) Standardisierung der Fragen wird eine kriterienorientierte Evaluation der Anforderungskonformität ermöglicht.

- *Lautes Denken*: Die Methode des lauten Denkens dient der Erfassung von Emotionen und Kognitionen, die Benutzern in der Interaktion mit der Software bewusst werden (vgl. Dumas & Redish, 1993; Lewis, 1982; Nielsen, 1993). Dazu werden Nutzer angewiesen, ihre Gedanken entweder direkt vor, während oder nach der Handlungsausführung (also prä-, peri- oder postaktional) zu verbalisieren. Die Äußerungen werden erfasst und später durch den Evaluator bzgl. verwertbarer Informationen analysiert. Hertzum und Jacobsen (2003) weisen auf einen Evaluatoreffekt hin: Mehrere Evaluatoren kommen aufgrund der Methode oft zu sehr unterschiedlichen Ergebnissen. Nachteilig kann weiterhin sein, dass die Aufgabe, laut zu denken, von Nutzern mitunter als ungewohnt empfunden wird und Auswirkungen auf ihr Verhalten angesichts der Bewusstmachung des Nutzungsvorgangs nicht auszuschließen sind.

*Meinungsbasierte Methoden*: Im Gegensatz zu den verhaltensbasierten Methoden steht bei der meinungsbasierten Evaluation nicht die Beobachtung konkreten Verhaltens sondern vielmehr die Meinung bzw. Zufriedenheit der Nutzer im Vordergrund (vgl. Gediga & Hamborg, 2002). Der Vorteil dieses Ansatzes besteht darin, dass sich Daten wie z.B. Befindlichkeiten erfassen lassen, die mit Hilfe der Verhaltensbeobachtung nicht oder nur eingeschränkt zugänglich sind.

- *Usability-Fragebögen*: Die fragebogenbasierte Evaluation ist Gegenstand von Abschnitt 2.4. Da das Ziel der vorliegenden Arbeit in der (Weiter-)Entwicklung von Usability-Skalen liegt, wird auf diesen methodischen Ansatz besonders fokussiert.
- *Interviewmethoden*: Die interviewbasierte Evaluation von Nutzermeinungen bietet gegenüber der Fragebogenmethode den deutlichen Vorteil eines höheren Grades an Flexibilität, insofern ein halbstrukturiertes Vorgehen gewählt wird (vgl. Kirakowski & Corbett, 1990; Nielsen, 1993).

*Benutzbarkeitstests*: Die wohl geläufigste Form der formativen Usabilityevaluation stellt der Benutzbarkeitstest (Usability Test) dar (vgl. Caulton, 2001; Dumas, 2003, S. 1096 ff.; Fu, Salvendy & Turley, 2002; Nielsen, 1993; Rubin, 1994). Dieser basiert auf der Aufzeichnung bzw. Beobachtung der Interaktion zwischen Benutzer und Software. Allgemeines Ziel ist in der Regel die Aufdeckung konkreter Schwachstellen (vgl. Hassenzahl & Seewald, 2004) und darauf aufbauend die Ableitung von Verbesserungsvorschlägen. Der mit der Durchführung von Usability Tests verbundene Aufwand ist vergleichsweise groß; entsprechend hoch sind die Kosten. Uneinigkeit besteht in der Literatur bzgl. der Frage, wie viele Nutzer herangezogen werden sollten, um einen Großteil von Gestaltungsproblemen zu identifizieren. Vielfach wird auf die Ergebnisse einer Studie von Nielsen und Landauer (1993; siehe auch Nielsen, 2006) Bezug genommen, wonach es in der Regel ausreicht, fünf Teilnehmer zu untersuchen. Andere Autoren kommen zu dem Schluss, es wäre eine größere Zahl an Nutzern nötig (vgl. Faulkner, 2003).

### 2.3.2.2 Prädiktive Methoden

Die Zielsetzung der prädiktiven Usability-Evaluationsmethoden liegt nicht darin, den Zustand einer Software beschreibend darzustellen, sondern vielmehr Schwachstellen vorherzusagen bzw. Gestaltungsempfehlungen für die Weiterentwicklung zu treffen (vgl. Hamborg & Gediga, 2006). Der Usability-Experte spielt entsprechend eine zentrale Rolle bei der Anwendung dieser Verfahren, allerdings werden auch häufig Benutzer in die Durchführung miteinbezogen. Als wesentliche Methoden innerhalb dieser Kategorie werden im Folgenden die Walkthrough-Methoden, die Heuristische Evaluation sowie das Gruppengespräch skizziert. Die von Hamborg und Gediga (2006) ebenfalls in der Gruppe der prädiktiven Ansätze angeführten formalen und automatisierten Evaluationsmethoden finden keine Berücksichtigung in der Darstellung.

*Walkthrough-Methoden:* Eine Untergruppe der prädiktiven Evaluationsansätze stellen die Walkthrough-Methoden dar. Dabei werden Handlungssequenzen gemäß konkreter Vorgaben schrittweise durchlaufen; währenddessen auftretende Benutzungsprobleme werden dokumentiert, um darauf aufbauend Vorschläge zur Verbesserung zu erarbeiten (vgl. Cockton, Lavery & Woolrych, 2003). Hertzum und Jacobsen (2003) bescheinigen auch dieser Evaluationsmethode einen Evaluatoreffekt, dass nämlich mehrere Experten auf Basis derselben Methode zu sehr unterschiedlichen Ergebnissen kommen können. Zwei unterschiedliche Ansätze der Methode liegen im Cognitive Walkthrough und im Pluralistic Walkthrough. Im Falle des *Cognitive Walkthrough* wird von Seiten des Evaluators das Benutzungsverhalten von Nutzern simuliert, welche im Vorfeld genau charakterisiert werden (Lewis & Wharton, 1997; Wharton, Rieman, Lewis & Polson, 1994). Als effektiver gegenüber dieser individuellen Methodenform erweist sich der *Pluralistic Walkthrough* (Preece, 1994, S. 706), bei dem eine Walkthrough-Evaluation in Gruppen durchgeführt wird (Bias, 1994, S. 63). Die Gruppen setzen sich aus unterschiedlichen Parteien zusammen; neben Usability-Experten können Benutzer, Programmierer, Entwickler oder Designer beteiligt sein. Walkthrough Methoden können zur Identifikation von Nutzungsproblemen als geeignet angesehen werden, erweisen sich jedoch als recht zeitaufwändig und entsprechend kostenintensiv.

*Gruppendiskussion/ -gespräch:* Bei dieser Methode findet eine Evaluation einer Software im gemeinsamen Gespräch mehrerer Personen statt (Virzi, 1997). Mindestens ein Teilnehmer ist Usability Experte, bei den übrigen kann es sich z.B. um Benutzer, Designer oder Programmierer handeln. Ziel ist wiederum die Identifikation konkreter Benutzungsprobleme und die Ableitung von Verbesserungen in der Softwaregestaltung. Der Pluralistic Walkthrough (s.o.) stellt eine spezielle Form dieses allgemeinen Evaluationsansatzes dar.



*Heuristische Evaluation:* Durch die Methode der Heuristischen Evaluation erfolgt eine strukturierte Bewertung einer Software durch Usability-Experten (vgl. Cockton, Lavery & Woolrych, 2003; Gediga & Hamborg, 1997; Muller, Matheson, Page & Gallup, 1998; Nielsen, 1994; Nielsen & Molich, 1990). Als Beurteilungsrichtlinien fungieren sogenannte Heuristiken, die abstrakte Regeln zur Gestaltung gebrauchstauglicher Software darstellen. Wiederum können die Ergebnisse dieses Untersuchungsansatzes einem Evaluatoreffekt unterliegen (Hertzum & Jacobsen, 2003). Eine kleine Anzahl von Experten zieht diese zur einzeln durchgeführten Begutachtung der Software heran. Vielfach zugrunde gelegt werden die folgenden Heuristiken von Nielsen (1994), die auf den von Nielsen und Molich (1990) veröffentlichten aufbauen: (1) visibility of system status, (2) match between system and the real world, (3) user control and freedom, (4) consistency and standards, (5) error prevention, (6) recognition rather than recall, (7) flexibility and efficiency of use, (8) aesthetic and minimalist design, (8) help users recognise, (9) diagnose and recover from errors, (10) help and documentation.

### 2.3.3 Subjektive vs. objektive Usability-Daten

Usability-Evaluationsmethoden lassen sich weiterhin bzgl. der Qualität der erhobenen Daten klassifizieren. Eine wesentliche Unterscheidung wird durch die Trennung von subjektiven und objektiven Daten vorgenommen (vgl. Hix & Hartson, 1993, S. 286; Hornbæk, 2006; Oppermann & Reiterer, 1994, S. 342 ff.). Unterschiede zwischen objektiven und subjektiven Maßen bestehen auch bzgl. anderer Konstrukte; beispielsweise kann der wahrgenommene Preis gegenüber dem tatsächlichen Preis abgegrenzt werden (siehe Abschnitt 2.6.5).

*Objektive Daten:* Objektive Usability-Daten sind direkt beobachtbar. Üblicherweise handelt es sich um Resultate des Benutzerverhaltens, welches während der Nutzung des zu evaluierenden Softwareproduktes erfasst wird. Ein subjektiver Güteindruck soll nicht mit einfließen. Entsprechend werden Leistungs-, Aufwands- oder Benutzungsparameter erhoben, z.B. die benötigte Zeit zur Erledigung vorgegebener Aufgaben sowie Fehler- und Klickraten. Die gängigste Methode ist die Nutzerbeobachtung. Objektive Daten bieten den Vorteil, dass sie Aspekte der Gebrauchstauglichkeit beleuchten, die sich mit Hilfe subjektiver Methoden oft nicht aufdecken lassen, z.B. aufgrund von Problemen der Verbalisierbarkeit seitens des Nutzes. Als nachteilig erweist sich, dass die Erhebung objektiver Daten oft mit einem hohen Aufwand verbunden ist, da sie in der Regel im Usability-Labor stattfindet (Oppermann & Reiterer, 1994, S. 344).

*Subjektive Daten:* Im Gegensatz zu den objektiven Daten basieren subjektive Daten auf Meinungen bzgl. der Gebrauchstauglichkeit eines Produktes. Gemeinhin stellen Nutzer die Quelle für diese Art von Daten dar, die z.B. mittels eines Usability-Fragebogens oder eines Think-Aloud-Protokolls erhoben werden. Subjektive nutzerbasierte Daten bieten den Vorteil, dass Rückschlüsse zur Akzeptanz gezogen werden können (Oppermann & Reiterer, 1994, S. 343). Nach Meinung einiger Autoren beziehen sich subjektive Daten stets auf die Nutzer-

wahrnehmung oder -einstellung zur Interaktion, zur Schnittstelle oder zum Ergebnis (Hornbæk, 2006). Jedoch kann auch dem Expertenurteil, das z.B. auf einer heuristischen Evaluation beruht, ein gewisses Maß an Subjektivität nicht abgesprochen werden. So zeigte etwa eine Studie von Hertzum und Jacobsen (2003), dass unterschiedliche Experten oftmals zu deutlich verschiedenen Bewertungen von ein und derselben Software kommen. Tractinsky (2004) stellt fest, dass in der Literatur weitgehend Einigkeit bzgl. der Frage besteht, was die relevanten Aspekte der objektiven Usability sind, allerdings herrscht eine gewisse Unklarheit, was unter der subjektiven Usability zu verstehen ist.

Es stellt sich die Frage nach dem Zusammenhang zwischen der objektiven und der subjektiven Usability. In einer Untersuchung von Fu und Salvendy (2002) wurde systematisch die objektive Usability einer Website zweigestuft variiert (schlecht vs. gut ausgeprägt). Als subjektive Maße dienten standardisierte Usability-Fragebögen QUIS (Chin, Diehl & Norman, 1988; Shneiderman, 1987) und PSSUQ (Lewis, 2002). Insgesamt weisen die Ergebnisse der Studie darauf hin, dass ein Zusammenhang zwischen der objektiven Usability und subjektiven Fragebogendaten besteht. Andere Autoren stellen fest, dass objektive und subjektive Usability-Maße zu unterschiedlichen Schlussfolgerungen zur Gebrauchstauglichkeit einer Schnittstelle führen können (vgl. Hornbæk, 2006). So ergaben sich in einer Studie von Tractinsky und Meyer (2001) Unterschiede in der subjektiv wahrgenommenen Dauer bei konstanter (objektiver) Dauer. Everard und Galletta (2003) kommen auf Basis der Ergebnisse einer experimentellen Untersuchung zu dem Schluss, dass eher die Wahrnehmung von Designfehlern als deren tatsächliches Vorhandensein die nutzerseitige Qualitätsbewertung eines Online-Shops bestimme.

Neben der Unterscheidung zwischen subjektiven und objektiven Daten kann eine Differenzierung zwischen quantitativen und qualitativen Daten vorgenommen werden (vgl. Hix & Hartson, 1993, S. 286).

*Quantitative Daten:* Bei *quantitativen* (numerischen) Usability-Daten kann es sich einerseits um objektive Daten handeln wie z.B. Klickraten oder die Dauer einer Aufgabenerledigung. Andererseits können quantitative Daten auch subjektiver Natur sein, z.B. im Falle eines Usability-Fragebogens, bei dem eine Ratingskala zum Einsatz kommt.

*Qualitative Daten:* Im Gegensatz dazu sind qualitative Daten nicht-numerisch. Zumeist handelt es sich um Beschreibungen konkreter Nutzungsprobleme (vgl. Hamborg & Gediga, 2006).

Mit Hilfe von fragebogenbasierten Usability-Verfahren werden subjektive Daten erhoben, die in der Regel quantitativer Natur sind. Diese Verfahrensgruppe ist Gegenstand des folgenden Kapitelabschnitts.

## 2.4 Fragebogenbasierte Usability-Evaluation

Da das Ziel der vorliegenden Arbeit in der Weiterentwicklung eines vorhandenen Usability-Fragebogens liegt, wird im Folgenden eine genauere Betrachtung dieser Evaluationsmethode vorgenommen. Wie bereits in Abschnitt 2.3.3 dargestellt, ist zwischen einer objektiven und einer subjektiven Usability zu unterscheiden. Mit Hilfe von Fragebögen wird die nutzerseitige Wahrnehmung der Usability erfasst. Folglich liefern Usability-Fragebögen subjektive Daten. Die Fragebogenmethode stellt eine weit verbreitete Art der Erfassung der Usability dar (Chignell, 1990). Einige Autoren sehen darin eine Abfrage der Zufriedenheitsdimension der DIN EN ISO 9241-11 (1999). Im Rahmen des Software-Entwicklungsprozesses werden Fragebögen vor allem in der Spezifikations-, Design- und Redesignphase verwendet (vgl. Gediga & Hamborg, 2002). Die fragebogenbasierte Erfassung der Usability geht zurück auf die bereits erwähnte Studie von Dzida et al. (1978; siehe Abschnitt 2.2.1.2), die einen Katalog mit insgesamt 100 Qualitätsanforderungen zusammenstellten und diese von Nutzern bewerten ließen.

In Abschnitt 2.4.1 werden zunächst einige der Software-Ergonomie entstammende Usability-Fragebögen vorgestellt. Aufbauend auf Sichtung der verschiedenen Instrumente werden in Abschnitt 2.4.2 einige differenzierende Merkmale von Usability-Fragebögen angeführt. Abschnitt 2.4.3 stellt die Vor- und Nachteile dieser Verfahrensgruppe vor. Gütekriterien und Ansätze zur Validierung von Usability-Fragebögen sind Gegenstand von Abschnitt 2.4.4.

### 2.4.1 Vorliegende Usability-Fragebögen

Obwohl eine Reihe standardisierter Usability-Fragebögen entwickelt wurde, kommen in empirischen Untersuchungen vielfach Fragebogenmaße zur Anwendung, die keiner ausführlichen Güteprüfung unterzogen wurden (vgl. Hornbæk, 2006). In Folge dessen kann die Validität der Messung bzw. der Untersuchungsergebnisse oftmals nicht beurteilt werden (Kirakowski, 1997). Zudem erweist es sich als problematisch, Vergleiche zwischen mehreren Studien zu ziehen, da nur selten einheitliche Maße verwendet werden. Standardisierte Fragebögen bieten demgegenüber den Vorteil, dass Erkenntnisse zur Messgüte vorliegen. Im Folgenden werden verschiedene standardisierte Usability-Fragebögen vorgestellt. Ziel der Darstellung ist es, einen Überblick zur gängigen Praxis der fragebogenbasierten Usability-Erfassung zu geben. Auf diverse Verfahren wird nicht im Detail eingegangen, z.B. der *After Scenario Questionnaire* (ASQ; Lewis, 1991), der *Benutzerfragebogen ErgoNorm* (Dzida et al., 2000), das *Computer Satisfaction Tool* (Bailey & Pearson, 1983), der *Computer System Usability Questionnaire* (CSUQ; Lewis, 1995), das *Computer User Satisfaction Inventory* (CUSI; Kirakowski, 1987), das *End-User Computing Satisfaction Instrument* (EUCS; Doll & Torkezadeh, 1988) oder die *Software Usability Scale* (SUS; Brooke, 1996).

#### 2.4.1.1 IsoMetrics

Der *IsoMetrics*-Fragebogen dient der nutzerbasierten Erfassung der Usability von Software-Systemen (Gediga & Hamborg, 1997, 1999; Gediga et al., 1999; Gediga, Hamborg & Willumeit, 1999; Hamborg, 2002; Willumeit, Gediga & Hamborg, 1995; siehe auch [www.isometrics.uni-osnabrueck.de/index.htm](http://www.isometrics.uni-osnabrueck.de/index.htm)). Das Verfahren bezieht sich auf die sieben Prinzipien der Dialoggestaltung, die in der DIN EN ISO 9241-11 gefordert werden (1999, vgl. Abschnitt 2.2.1.2).

*Aufbau:* Insgesamt 90 Items dienen der Erfassung der sieben Gestaltungsprinzipien, wobei die Items in entsprechende Subskalen eingeteilt wurden. Die Beantwortung der Items erfolgt mit Hilfe einer fünf-stufigen Rating-Skala von „stimme gar nicht zu“ bis „stimme voll zu“ mit einer zusätzlichen Antwortoption „keine Meinung“. Es existieren zwei Versionen des Fragebogens, wobei die eine der formativen und die andere der summativen Evaluation dient. Die formative Version des Fragebogens *IsoMetrics*<sup>L</sup> unterscheidet sich von der summativen Version *IsoMetrics*<sup>S</sup> darin, dass a) für jedes Item zusätzlich eine Wichtigkeitseinschätzung (ebenfalls fünfstufig, von „unwichtig“ bis „wichtig“ mit „keine Meinung“-Option) abgefragt wird und b) ein offenes Antwortfeld zur Verfügung steht, in dem der Befragte angeben soll, auf welchen Interaktionserfahrungen seine Bewertung des Items beruht. Einerseits finden sich in den Subskalen vereinzelt sehr globale Items (z.B. „Die Software ist auf die Erfordernisse meiner Arbeit gut zugeschnitten“; Skala Aufgabenangemessenheit, Item A10). Die meisten Indikatoren sind jedoch recht spezifischer Natur (z.B. „Die von der Software verwendeten Begriffe sind für mich eindeutig und unmittelbar verständlich“; Skala Selbstbeschreibungsfähigkeit, Item S12).

*Entwicklung:* Das grobe Vorgehen zur Entwicklung des Fragebogens war wie folgt (Gediga et al., 1999): Zunächst erstellten die Autoren auf Basis vorhandener Literatur bzw. bereits existierender Usability-Verfahren einen großen Itempool, den sie anhand mehrerer Kriterien reduzierten. Es wurden Items eliminiert, die a) sich als nicht verständlich erwiesen, b) zu spezifisch auf eine bestimmte Applikation zugeschnitten waren sowie c) keiner der sieben Gestaltungsprinzipien zuzuordnen waren. Anspruch war es, eher konkrete als abstrakte Items zu verwenden, sich einer einfachen Semantik und Syntax zu bedienen und keine negativen Formulierungen vorzunehmen. Die Zuordnung der Items zu den sieben Dialoggestaltungsprinzipien erfolgte auf Grundlage des Urteils von sechs deutschen Usability-Experten. Insgesamt wurde der Fragebogen im Sinne der klassischen Testtheorie auf Basis von Cronbachs Alpha konstruiert.

*Gütenachweis:* Gediga et al. (1999) geben im Falle einer Evaluation einer konkreten Software Skalenreliabilitäten zwischen  $\alpha=0,71$  und  $\alpha=0,86$  an. Eine Validierung erfolgte anhand eines Samples mit  $n=229$  Mitarbeitern der Uni Osnabrück, die gebeten wurden, ein Software System zu bewerten, das sie häufig nutzen. Zum Zwecke der Optimierung der internen Konsistenz eliminierten die Autoren einige wenige Items aus einzelnen Subskalen. Danach

ergaben sich Reliabilitäten zwischen  $\alpha=0,47$  und  $\alpha=0,86$ . Die Validität der Skalen sehen die Autoren einerseits durch signifikant unterschiedliche Skalen-Scores für verschiedene Software-Systeme bzw. zwei konkurrierende Prototypen untermauert. Weiterhin analysieren sie das Antwortprofil der Subskala *Aufgabenangemessenheit* auf Einzelitemebene für die formative Version des Fragebogens (IsoMetrics<sup>L</sup>). Dabei bestätigt sich die folgende Hypothese: Bei Items, die einerseits sehr schlecht bewertet, andererseits jedoch als sehr wichtig eingeschätzt werden, finden sich die meisten Usability-Problemschilderungen von Nutzerseite. Zudem stehen die angeführten Nutzungsprobleme im Einklang mit der Bewertung der verschiedenen Skalen.

*Kritische Beurteilung:* Es lässt sich kritisieren, dass einige der Items eher Fakten abfragen, deren Beantwortung auch von Seiten eines Evaluators vorgenommen werden können. So erscheint es beispielsweise zweifelhaft, ob es Sinn macht, mehrere Nutzer zu fragen, ob die Software eine Wiederhol-Funktion für sich wiederholende Arbeitsschritte bietet (Item A14). Des Weiteren setzen sich einige wenige Items jeweils aus mehreren Bewertungsaspekten zusammen (z.B. „Die Software bietet mir Möglichkeiten, Kommandonamen, Objekte und Aktionen individuell zu benennen“).

#### 2.4.1.2 ISONORM 9241/10

Ein weiterer Fragebogen, der sich streng an den Kriterien der Dialoggestaltung nach DIN EN ISO 9241-10 (1996) orientiert, ist der *ISONORM 9241/10* (Prümper & Anft, 1993; Prümper, 1997; Online: [www.bao.de/seue/evaluation/index.html](http://www.bao.de/seue/evaluation/index.html)). Der Fragebogen stellt nach Angaben von Prümper (1997, S. 254) ein nützliches formatives Evaluationsverfahren dar, wenn er im Rahmen eines partizipativen Designkonzeptes genutzt wird, um in Kombination mit einer visualisierten Gruppendiskussion erste Schwachstellen einer Software zu identifizieren.

Können Sie die Art und Weise, wie Sie mit der Software arbeiten, beeinflussen?								
Die Software ...	---	--	-	-/+	+	++	+++	Die Software ...
bietet keine Möglichkeit, die Arbeit an jedem Punkt zu unterbrechen und dort später ohne Verluste wieder weiterzumachen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	bietet die Möglichkeit, die Arbeit an jedem Punkt zu unterbrechen und dort später ohne Verluste wieder weiterzumachen.

Abb. 2-8: Beispiel für die Art der Itemabfrage im ISONORM 9241/10: Item zur Subskala Steuerbarkeit (Prümper & Anft, 1993, S. 149)

*Aufbau:* Jeder der sieben Gestaltungsgrundsätze wird durch fünf Items abgedeckt, die mit Hilfe einer siebenstufigen Antwortskala („---“ bis „+++“) bewertet werden. Das Ausfüllen des Fragebogens erfordert ca. 10 Minuten. Jeder der sieben Fragenblöcke beginnt mit den Worten „Die Software...“. In Abbildung 2-8 wird die Art der Abfrage visualisiert. Die einzelnen Items decken recht spezifische Aspekte der Usability ab.

*Entwicklung:* Bei der Itemgenerierung war es das Ziel von Prümper und Anft (1993), Items mit mittlerem „Granulationsgrad“ zu entwerfen (vgl. Rauterberg, 1992), so dass einerseits zwar spezifische Aussagen getroffen werden, andererseits eine Anwendung des Fragebogens über möglichst viele Oberflächentypen hinweg möglich ist.

*Gütenachweis:* In einer summativen Evaluationsstudie ergaben sich signifikante Skalenunterschiede für Software-Anwendungen mit und ohne GUI<sup>6</sup> (Prümper, 1993). Die Retest-Reliabilität des Verfahrens auf Basis einer zweimaligen Bewertung derselben Software durch dieselben 49 Benutzer liegt nach Angaben von Prümper (1997) zwischen  $r_{t1/t2}=0,59$  und  $r_{t1/t2}=0,68$ , wobei das Zeitintervall zwischen beiden Bewertungen im Mittel bei 6,7 Monaten lag. Zur Bestimmung der internen Konsistenz zog der Entwickler 1265 ISONORM-Beantwortungen zu 178 Softwareprogrammen heran. Das Cronbachs Alpha der verschiedenen Subskalen liegt zwischen  $\alpha=0,81$  und  $\alpha=0,89$ . Als Nachweis der inneren kriterienbezogenen Validität werden hohe Korrelationen zwischen den Ergebnissen des ISONORM mit anderen Verfahren angegeben (QUIS von Chin et al., 1988; BBD-Verfahren von Spinax, 1987). Die äußere kriterienbezogene Validität sieht Prümper (1997) in einer hohen Korrelation zwischen ISONORM-Ergebnissen und denen des leitfadenorientierten Evaluationsverfahrens EVADIS II (Oppermann, Murchner, Reiterer & Koch, 1992) bestätigt. In einer weiteren Validierungsuntersuchung fallen die Subskalenwerte für ältere Versionen der Software MS-Word und MS-Excel großteils niedriger aus als für neuere (4.0 vs. 5.0). Zwecks einer Normierung berechnet Prümper (1997) über eine Reihe verschiedener Softwareprodukte hinweg Mittelwert, Standardabweichung, Minimum und Maximum für die verschiedenen Subskalen.

#### 2.4.1.3 AttrakDiff 2

Der Fragebogen AttrakDiff 2 stellt kein klassisches Verfahren zur Erfassung der Usability dar, sondern beansprucht, Qualitätsaspekte zu messen, die über die reine Gebrauchstauglichkeit hinausgehen (Hassenzahl, Burmester & Koller, 2003; Online: <http://www.attrakdiff.de>). Die Autoren weisen darauf hin, dass es neben der Gebrauchstauglichkeit weitere Qualitätsanforderungen zu berücksichtigen gelte. Es stelle ein Bedürfnis von Nutzern dar, dass ein Produkt stimulierend und identitätsstiftend sei (Hassenzahl, 2004b). Die *wahrgenommene pragmatische Qualität* bezeichnet, inwieweit ein interaktives Produkt geeignet ist, um die Umwelt zu manipulieren. Eine *hedonische Qualität* besitzt ein interaktives Produkt, wenn es Möglichkeiten der Nutzer erweitert, sie herausfordert, stimuliert und eine gewünschte Identität (etwa Coolness, Moderne, Professionalität etc.) kommuniziert.

---

<sup>6</sup> Die Abkürzung GUI steht für „Graphical User Interface“, also eine graphische Benutzerschnittstelle.

*Aufbau:* Der AttrackDiff 2 umfasst 21 Items, die mit Hilfe eines siebenstufigen semantischen Differentials beantwortet werden. Als Pole werden gegensätzliche Adjektivpaare verwendet (z.B. „verwirrend“ vs. „übersichtlich“). Der Fragebogen dient der Erhebung der von Seiten seiner Entwickler eingeführten Konstrukte *wahrgenommene pragmatische Qualität* (PQ), *hedonische Qualität – Stimulation* (HQ-S) und *hedonische Qualität – Identität* (HQ-I), die die Subskalen des Instrumentes darstellen.

### 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. 2-9: Zwei Items aus dem AttrackDiff-Fragebogen ([www.attrakdiff.de](http://www.attrakdiff.de))

*Entwicklung:* Den Ausgangspunkt zur Entwicklung des Instrumentes stellte der Fragebogen AttrackDiff 1, der die drei Skalen *pragmatische Qualität* (PQ), *hedonische Qualität* (HQ) und *Attraktivität* messen soll (vgl. Hassenzahl, Platz, Burmester & Lehner, 2000). Er setzt sich aus 23 Items zusammen. Der AttrackDiff 1 vermochte nach Angaben von Hassenzahl et al. (2003) nicht zwischen den Aspekten *Identität* (HQ-I) und *Stimulation* (HQ-S) zu differenzieren, so dass eine neue Version des Instruments entwickelt wurde. Der Itempool des neuen Fragebogens setzte sich aus den sieben Items der AttrackDiff 1 Skala *Attraktivität* sowie 50 weiteren Items zur Erfassung von PQ, HQ-S und HQ-I, die im Rahmen eines Expertenworkshops ermittelt wurden, zusammen. In einer ersten Anwendung bewerteten 23 Probanden drei verschiedene Websites. Mit Hilfe einer HKA (Varimax-Rotation) wurden 21 der ursprünglich 50 Items identifiziert, die die Subskalen PQ, HQ-S und HQ-I erfassen sollen.

*Gütenachweis:* Skalenunterschiede für drei bewertete Websites interpretieren die Entwickler als erste Hinweise auf eine gegebene Konstruktvalidität. In einer zweiten Untersuchung bewerteten Studenten verschiedene Versionen derselben Software mit Hilfe des AttrackDiff 2. Die internen Konsistenzen der Skalen wurden getrennt für jedes einzelne Interface berechnet und liegen zwischen  $\alpha=0,73$  und  $\alpha=0,90$ . Eine mittlere Korrelation von  $r=0,55$  ergibt sich zwischen den beiden Skalen HQ-S und HQ-I zur hedonischen Qualität. Unterschiede in den Skalenwerten für die vier verschiedenen Softwareversionen werden von Seiten der Autoren unter Einschränkung als Hinweise auf die Konstruktvalidität des Messinstrumentes gewertet.

#### 2.4.1.4 Wahrgenommene Orientierung in Online-Shops – WOOS

Nachdem Skalen zur Desorientierung bei der Nutzung von Websites bereits von Ahuja und Webster (2001) bzw. Otter und Johnson (2000) entwickelt wurden, stellen Yom und Wilhelm (2004) den Fragebogen zur *wahrgenommenen Orientierung in Online-Shops* vor (WOOS, siehe auch Yom, 2003). Die Autoren definieren das zu erfassende Konstrukt als „das subjektiv-globale Empfinden, eine Lagebestimmung durchführen sowie zur Verfügung stehende Navigationsoptionen im Shop erkennen und zielführend nutzen zu können“ (Yom & Wilhelm, 2004, S. 44) und betonen die Nähe dieses Konzeptes zur Usability, wobei sie sich auf die Usability-Definition der DIN EN ISO 9241-11 beziehen (1999).

*Aufbau:* Der Fragebogen umfasst sieben positiv formulierte Items zur „Orientierungsfreundlichkeit“ im Statement-Format ab. Einerseits sind Items enthalten, die eine globale Aussage zum abgefragten Konstrukt erfassen (z.B. „Man fühlt sich im Angebot orientiert“); andere Items sind spezifischer gehalten (z.B. „Auf den Seiten werden zu viele Verweise (Links) angeboten“).

*Entwicklung:* Auf Basis vorhandener Skalen zur Erfassung der Usability leiten die Entwickler der Skala zehn Items ab, die sie an den Kontext des Online-Shoppings anpassten. Eine erste Datenerhebung wurde mit 40 Teilnehmern durchgeführt. Auf Grundlage der Faktorenstruktur einer Hauptkomponentenanalyse bzw. im Zuge einer Alpha-Maximierung eliminieren die Autoren drei Items.

*Gütenachweis:* In einer experimentellen Untersuchung mit 40 Teilnehmern wandten die Autoren die verkürzte WOOS-Skala zur Bewertung eines Bücher-Online-Shops an. Als abhängige Variablen wurden erhoben: a) eine Skala von Donovan und Rossiter (1982) zur affektiven Bewertung des Shopping-Erlebens, b) die Klickanzahl und Bearbeitungsdauer in einer Suchaufgabe sowie c) ein Usability-Globalbewertungsmaß, dessen Operationalisierung jedoch unklar bleibt. In einer HKA ergibt sich ein einziger Faktor. Die Reliabilität der Skala liegt bei  $\alpha=0,90$ . Hohe Korrelationen des WOOS-Maßes mit den erhobenen AVs untermauern nach Angaben von Yom und Wilhelm (2004) die Konstrukt- bzw. Kriteriumsvalidität der Skala. In zusätzlichen Online-Erhebungen zur Bewertung mehrerer Reise-Online-Shops bestätigte sich die Eindimensionalität der Skala mittels Faktorenanalysen (HKA, die Stichprobengröße variierte zwischen  $n=32$  und  $n=178$ ), für Cronbachs Alpha ergaben sich Werte zwischen  $\alpha=0,91$  und  $\alpha=0,97$ .

*Kritische Beurteilung:* Kritisch anzumerken ist, dass Yom und Wilhelm (2004) zum Zwecke der Skalenentwicklung die Faktorenanalyse anwenden, obwohl die zugrunde liegenden Stichprobengrößen sehr klein ausfallen (vgl. Bühner, 2004, S. 157).



#### 2.4.1.5 Marketingorientierte Evaluationsskalen

Im Bereich der Marketing-Literatur wurden diverse Skalen zur Evaluation von E-Commerce-Auftritten entwickelt. Diese beziehen neben einigen wenigen, oft recht spezifischen Aspekten der wahrgenommenen Usability diverse andere Bewertungskriterien mit ein, die aus Sicht der jeweiligen Autoren relevant für die Zufriedenheit von Kunden beim Einkaufen in einem Online-Shop sind. Da das Vorgehen bei der Entwicklung dieser Skalen sich zumeist sehr ähnelt, werden im Folgenden exemplarisch zwei Instrumente genauer vorgestellt. Neben diesen wurden von Charitou, Georgiou und Soteriou (2003), McKinney, Yoon und Zahedi (2002), Parasuraman, Zeithaml und Malhotra (2005), Schubert (2002) Srivihok (1999), Wang, Tang und Tang (2001) sowie Yang, Cai, Zhou und Zhou (2005) ähnliche Skalen entwickelt.

*WebQual 4.0* ist ein 23 Items umfassender Fragebogen zur Erfassung der Nutzerwahrnehmungen zur Qualität von E-Commerce Sites, der beansprucht die Kategorien *Usability*, *Informationsqualität*, *Serviceinteraktion* und *Gesamtbeurteilung* mittels einer siebenstufigen Ratingskala zu messen (Barnes & Vidgen, 2001, 2002, 2003). Es finden sich sowohl allgemeine („I find the Site easy to use“) als auch eher spezielle („I find the site easy to navigate“) Items zur Usability. Es wurde eine Faktorenanalyse zur Untersuchung der Dimensionalität gerechnet; Cronbachs Alpha diente als Kriterium der internen Konsistenz (Werte zwischen  $\alpha=0,76$  und  $\alpha=0,93$ ). Positiv ist zu bewerten, dass das Instrument inhaltlich verschiedene Aspekte der Qualitätswahrnehmung aufgreift (Ästhetik, Reputation, Usability, Sicherheit etc.). Allerdings drängt sich der Verdacht auf, dass sich die Faktorenstruktur vor allem aufgrund der oft sehr ähnlichen Itemformulierungen ergeben (sechs Items beginnen mit „Provides...“ und jeweils drei mit „The site...“ bzw. „I find...“).

Der von Wolfinbarger und Gilly (2003) entwickelte Fragebogen *eTailQ* wurde ebenfalls zur Messung der nutzerseitig wahrgenommenen Qualität von E-Commerce entwickelt. Wie auch im Falle des *WebQual 4.0* finden sich in diesem Instrument neben einigen Usability-Aspekten auch Bewertungskriterien, die anderen Konstrukten zuzuordnen sind. Die Entwicklung des Instrumentes erfolgte ebenfalls faktorenanalytisch. Auf Basis der Faktorladungen sowie der Trennschärfe wurden diverse Items eliminiert (z.B. „The search function gives me useful results“), so dass nur 14 von ursprünglich 40 Indikatoren übrigbleiben. Die Autoren extrahieren letztlich vier Skalen, die sie wie folgt interpretieren: a) Website Design (z.B. „The site doesn't waste my time“), b) Verlässlichkeit (z.B. „You get what you ordered from this site“), c) Kundenservice („Inquiries are answered promptly“) sowie d) Security/Privacy („I feel safe in my transactions with this website“). Die internen Konsistenzen der Skalen liegen zwischen  $\alpha=0,79$  und  $\alpha=0,88$ .

#### 2.4.1.6 Usability Fragebogen für Online-Shops – ufosV1

Da im Rahmen der vorliegenden Arbeit eine Weiterentwicklung dieses Instruments vorgenommen wird, fällt die Vorstellung des Usability Fragebogen für Online-Shops (ufos) zur Erfassung der Usability von Online-Shops aus Kundensicht entsprechend ausführlich aus (Konradt et al., 2003; Online: <http://www.ufos-online.com/>). Diese alte Version des Fragebogens wird im Folgenden mit ufosV1 bezeichnet; die neue, zu entwickelnde Fassung wird mit ufosV2 abgekürzt.

*Entwicklung:* Zunächst wurden Verfahren zur Erfassung der Usability identifiziert, auf Grundlage derer sich Items ableiten ließen. Gemäß eines allgemein üblichen Vorgehens zur Erstellung des Itempools von Usability-Fragebögen (vgl. Dumas, 2003, S. 1094) wurden relevante Items spezifiziert und in ihrer Formulierung an den Kontext angepasst. Als einheitliches Itemformat wurde die Aussageform gewählt. Der initiale Itempool von 400 Items wurde in einem nächsten Schritt von Seiten der Entwickler auf inhaltliche Redundanz und Verständlichkeit der Formulierungen geprüft. Redundante Items wurden eliminiert und mangelhafte Formulierungen verbessert. Auf diese Weise konnte der Pool auf 235 Items reduziert werden. Diese wurden im Rahmen einer online durchgeführten Expertenbefragung, an der sich insgesamt elf Experten im Bereich Usability beteiligten, bzgl. ihrer Verständlichkeit und Relevanz untersucht. Auf Basis dieser Vorbefragung ließ sich der Itempool auf 103 Items reduzieren. In einem letzten Schritt zur Festlegung des Itempools wurde eine weitere Voruntersuchung in Form einer Befragung einer studentischen Stichprobe zur Verständlichkeit der Items durchgeführt.

*Datenerhebung:* In die Hauptuntersuchung gingen insgesamt 80 Items ein; sie können dem Anhang A entnommen werden. Abbildung 2-10 enthält ein Beispielitem. An der Datenerhebung nahmen  $N=60$  Probanden teil, von denen jeder im Rahmen eines computergestützten Experimentes zwei von insgesamt sechs Online-Shops zum Verkauf von Musik-CDs kennen lernte und bewertete. Gegen Ende des Versuchs wurde eine Kaufentscheidung operationalisiert: die Teilnehmer hatten die Wahl, eine im Vorfeld des Untersuchungstermins selbst gewählte CD in einem der beiden Shops zu bestellen oder alternativ einen Gutschein über dieselbe CD zu erhalten. Im Fall einer Bestellung wurden ihnen als Belohnung zur Teilnahme der sich ergebende Rechnungsbetrag in bar ausgezahlt.

*Skalenfestlegung:* Die Festlegung des ufosV1-Fragebogens erfolgte faktorenanalytisch. Es wurde eine Hauptkomponentenanalyse (HKA) mit schiefwinkliger Rotation berechnet. Einige Indikatoren gingen aufgrund zu hoher Anteile fehlender Werte nicht in diese Analyse ein. Die weitere Itemselektion erfolgte nach dem Kriterium, eine möglichst zufrieden stellende Faktorenlösung zu erreichen. Letztlich gingen 47 Items in die HKA und damit in den Fragebogen ein. Die sieben extrahierten Komponenten ließen sich inhaltlich recht plausibel interpretieren. In Tabelle 2-3 werden die Skalen benannt und die Skalenkennwerte aufgeführt.

1. Dieser Shop macht das Einkaufen im Netz leicht.		
Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

Abb. 2-10: Beispielitem aus dem ufosV1-Fragebogen (www.ufos-online.com)

*Validierung:* Zum Zwecke wurde eine hierarchische multiple Regressionsanalyse mit dem Kriterium Kaufintention berechnet (Christophersen, 2002). Ein Großteil der Varianz in der AV (42,2%) wurde im ersten Schritt der Analyse durch die Usability aufgeklärt. Weiterhin wurde eine Diskriminanzanalyse berechnet, bei der vorhergesagt werden sollte, in welchen von beiden Shops ein Teilnehmer kauft, *wenn* er kauft. Die erste Subskala des Fragebogens allgemeiner Usabilityeindruck sowie die Variable wahrgenommene Größe erwiesen sich als signifikante Prädiktoren und führten zu einer korrekten Klassifikation von 93,8% der Fälle.

Tab. 2-3: Skalen und Kennwerte des ufosV1 (Balazs, 2002)

Subskala	Item- anzahl	$\lambda_j$	Anteil aufgekl. Varianz	Homo- genität	Korrelation mit Gesamtscore	$\alpha$
Allgemeiner Usability-Eindruck	14	18,27	33,88%	0,563	0,92	0,948
Informationen zu den Rahmenbedingungen	7	3,14	6,67%	0,426	0,62	0,832
Produktsuche	7	2,74	5,84%	0,614	0,84	0,918
Warenkorbhandling / Bestellung	7	1,83	3,89%	0,451	0,76	0,851
Produktüberblick	4	1,66	3,53%	0,511	0,71	0,812
Selbstbeschreibungsfähigkeit	4	1,48	3,14%	0,536	0,79	0,821
Produktbeschreibungen	4	1,27	2,70%	0,466	0,72	0,773

Anmerkung: Cronbachs Alpha ( $\alpha$ ), Eigenwert ( $\lambda_j$ ).

Der endgültige ufosV1-Fragebogen findet sich in Anhang A. Er steht auf der Internetsite [www.ufos-online.com](http://www.ufos-online.com) zur Verfügung. Auf dieser Site besteht für Anbieter von Online-Shops die Möglichkeit, mit ufosV1 eine Online-Kundenbefragung zur *Usability* des eigenen Shops durchzuführen. Büttner, Schulz und Silberer (in Druck) wendeten die ufosV1-Skala *Allgemeiner Usabilityeindruck* in einer Untersuchung zur Nutzung von Internet-Apotheken an. Die Autoren berichten eine interne Konsistenz der Skala von  $\alpha=0,94$  bzw.  $\alpha=0,94$  für zwei verschiedene Online-Anbieter. Schulz, Mau und Silberer (2006) zogen den ufosV1-Fragebogen heran, um ihn in abgewandelter Form zur Evaluation der Benutzerfreundlichkeit von Reise-Printkatalogen zu nutzen. Der ufosV1-Fragebogen fand seit seiner Entwicklung zudem Anwendung im Rahmen diverser nicht-forschungsorientierter Usability-Projekte (z.B. Apliki, 2006).

## 2.4.2 Differenzierende Charakteristika von Usability-Fragebögen

Nachdem diverse konkrete Usability-Evaluationsskalen vorgestellt wurden, soll im Folgenden genauer analysiert werden, hinsichtlich welcher Charakteristika Unterschiede zwischen den verschiedenen Verfahren bestehen. Auf Grundlage dieser Analyse kann zum einen eine Einordnung des Fragebogeninstrumentes erfolgen, das im Rahmen der vorliegenden Arbeit (weiter-)entwickelt wird. Zum anderen lassen sich später Implikationen für den Aufbau des Verfahrens ziehen.

Die angeführten Merkmale lassen sich grob in zwei verschiedene Bereiche einteilen (siehe Abbildung 2-11): (1) Auf einer inhaltlichen Ebene bestehen zwischen Usability-Fragebögen Unterschiede bzgl. des Konstrukt-Fokus bzw. der Dimensionalität (Abschnitt 2.4.2.1), des Bewertungsgegenstandes (2.4.2.2) sowie des Granulationsgrades der Items (2.4.2.3). (2) Die formale Ebene des Fragebogendesigns betrifft Eigenschaften bzgl. der Itemform (2.4.2.4), der Antwortskala (2.4.2.5) sowie die Itemanzahl (2.4.2.6). Bei der Neuentwicklung eines Usability-Fragebogens gilt es, Entscheidungen bzgl. der gewünschten Eigenschaften auf beiden Ebenen zu treffen.

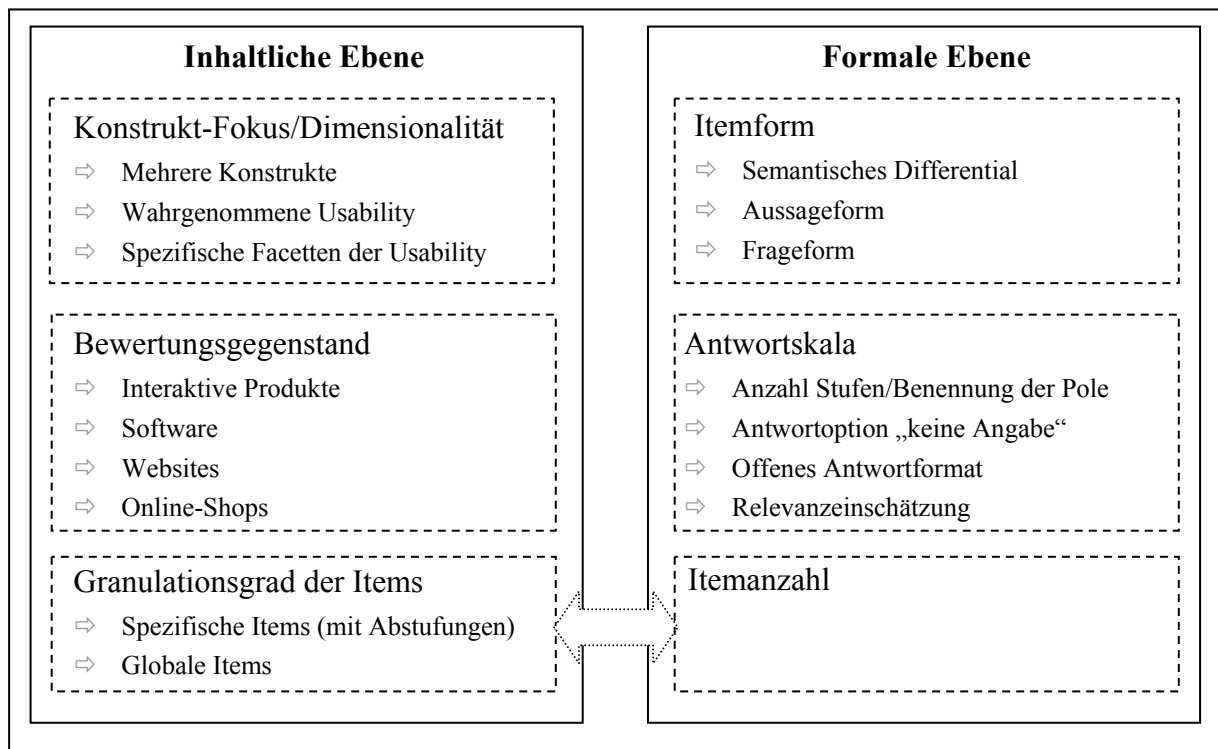


Abb. 2-11: Verschiedene Charakteristika von Evaluationsfragebögen

### 2.4.2.1 Konstruktfokus/Dimensionalität

Die inhaltliche Spannweite von Fragebögen im Bereich der Mensch-Maschine-Interaktion ist groß (vgl. Dumas, 2003, S. 1094). Es können grob drei Gruppen unterschieden werden:

*Mehrere verschiedene Konstrukte:* Diverse Instrumente fokussieren nicht auf ein einzelnes Konstrukt, sondern beziehen neben der wahrgenommenen Usability andere Bewertungs-

dimensionen bzw. -aspekte mit ein. Beispielsweise dient der AttrakDiff-Fragebogen der Erfassung der hedonischen Qualität eines interaktiven Produktes (Hassenzahl et al., 2003; siehe Abschnitt 2.4.1.3). Einige Fragebögen beanspruchen zwar die wahrgenommene Usability zu erfassen, enthalten jedoch auch Items zu anderen Konstrukten – ohne ausdrücklich darauf hinzuweisen. Beispielsweise findet sich im SUMI ein Item, das die Intention zur Weiterempfehlung der Software abfragt.

*Wahrgenommene Usability:* Einige Verfahren sind ausschließlich auf die Erfassung der nutzerseitig wahrgenommenen Gebrauchstauglichkeit in einem klassisch software-ergonomischen Sinne ausgerichtet (z.B. IsoMetrics von Gediga et al., 1999). Den verschiedenen Methoden zur Evaluation der Usability liegen oftmals unterschiedliche Konzeptionalisierungen der Usability zugrunde. Einige (insbesondere aus Deutschland stammende) Verfahren stellen einen direkten Bezug zur DIN EN ISO 9241 her (z.B. ErgoNorm von Dzida et al., 2000; IsoMetrics von Gediga et al., 1999). Bei einer Fülle von Verfahren werden jedoch auch Ansätze zugrunde gelegt, die diese Normvorgaben nicht direkt oder gar nicht mit einbeziehen.

*Spezifische Facetten der wahrgenommenen Usability:* Andere Fragebögen beanspruchen, eine spezifische Facette der subjektiven Usability zu erfassen. Ein Beispiel stellt die WOOS-Skala zur Erfassung der *wahrgenommenen Orientierung* bei der Nutzung von Online-Shops von Yom und Wilhelm (2004) dar (siehe Abschnitt 2.4.1.4).

Unterschiede zwischen den Instrumenten bestehen zudem dahingehend, dass einige ein-dimensional angelegt sind und somit nur eine Skala umfassen (z.B. WOOS). Andere Verfahren erfassen die Gebrauchstauglichkeit auf mehreren Dimensionen, indem sie sich aus mehreren Subskalen zusammensetzen (z.B. IsoMetrics).

#### 2.4.2.2 Bewertungsgegenstand

Ein weiterer Unterschied zwischen standardisierten Usability-Fragebögen liegt im Bewertungsgegenstand (Kirakowski, 1997). Während einige Verfahren sehr allgemein orientiert sind (Bezug auf interaktive Produkte), richten sich andere auf konkrete Gruppen von Benutzerschnittstellen (Bezug auf Online-Shops).

*Interaktive Produkte:* Einen sehr breiten Anwendungsfokus besitzt der Attrak-Diff, der auf die Erfassung der hedonischen und pragmatischen Qualität bei jeglicher Art von interaktiven Produkten ausgerichtet ist (Hassenzahl et al., 2003).

*Software:* Viele Fragebögen beziehen sich auf Software im Allgemeinen (z.B. PUTQ von Lin et al., 1997).

*Websites:* Weiterhin liegt eine Reihe von Instrumenten zur Erfassung der Zufriedenheit bei der Nutzung von Websites vor (z.B. WAMMI von Kirakowski & Cierlik, 1998).

*Online-Shops*: Andere Verfahren fokussieren auf Online-Shops als spezielle Untergruppe von Websites (z.B. WOOS von Yom & Wilhelm, 2004; ufosV1 von Konradt et al., 2003).

#### 2.4.2.3 Granulationsgrad der Items

Usability-Fragebögen unterscheiden sich auf inhaltlicher Ebene dahingehend, wie spezifisch die verwendeten Items formuliert wurden, also in Bezug auf den „Granulationsgrad“ der Items (vgl. Rauterberg, 1992). Einerseits können Items sehr global sein, andererseits aber auch (in unterschiedlichem Maße) spezifisch.

*Spezifische Items*: Andererseits finden sich in Usability-Fragebögen auch Items, die sehr spezifische Bewertungsaspekte berücksichtigen, wie z.B. „Die Software bietet gute Möglichkeiten, sich häufig wiederholende Bearbeitungsvorgänge zu automatisieren“ (ISONORM 9241/10 von Prümper, 1997). Der Grad an Spezifität kann sehr unterschiedlich ausfallen, auch innerhalb desselben Fragebogens. Ein Beispiel ist die Subskala „Terminology and System Information“ des QUIS-Fragebogens (Chin et al., 1988). Sie enthält zum einen ein Item zur *angemessenen Verwendung von Begrifflichkeiten*. Zum anderen wurde auch ein Item zur *angemessenen Verwendung arbeitsbezogener Begrifflichkeiten* in die Skala aufgenommen. Das zweitgenannte Item stellt einen Unteraspekt des ersten dar.

*Globale Items*: Beispiele für Usability-Items, die eine Art Globalurteil seitens der Nutzer abfragen, sind die folgenden: „Die Software ist unkompliziert zu bedienen“ (ISONORM 9241/10 von Prümper, 1997) oder „I thought the system was easy to use“ (aus dem SUS von Brooke, 1996).

Nahezu immer werden innerhalb derselben Skalen sowohl spezifische als auch globale Items abgefragt. Ausnahmen stellen der PUTQ (Lin et al., 1997) sowie die Subskala „Overall reactions to the system“ des QUIS dar. Im PUTQ finden sich ausschließlich spezifische Items. In der genannten Subskala des QUIS werden ausschließlich Items zur Gesamtbewertung einer Software verwendet. Wird ein neuer Usability-Fragebogen entwickelt, so hat eine Entscheidung zum Granulationsgrad der Items Konsequenzen für die Itemanzahl, die ein Merkmal auf der formalen Ebene des Instrumentes darstellt. Je spezifischer die Items ausfallen, umso mehr Items werden benötigt, um die Usability einer Benutzerschnittstelle insgesamt abzudecken.

#### 2.4.2.4 Itemform

Es finden sich in vorliegenden Usability-Fragebögen sehr unterschiedliche Itemformate.

*Semantisches Differential*: Eine Reihe von Instrumenten nutzt das semantische Differential (z.B. AttrakDiff2 von Hassenzahl et al., 2003; QUIS von Chin et al., 1988). Dabei werden den Nutzern gegensätzliche Adjektivpaare vorgelegt.

*Aussageform*: Ein weiteres geläufiges Itemformat ist die Aussage wie z.B. „I found the various functions in this system were well integrated“ aus dem SUS-Fragebogen von Brooke (1996).

*Frageform:* Andere Fragebögen werden dem Namen ihrer Verfahrensklasse gerecht, indem sie sich aus Frageitems zusammensetzen, wie z.B. „Do you get the information you need in time?“ im EUCS von Doll und Torkzadeh (1988).

#### 2.4.2.5 Antwortskala

*Anzahl der Stufen:* In der Regel kommen bei Rating-Skalen in den empirischen Wirtschafts- und Sozialwissenschaften entweder fünf oder sieben Abstufungen zur Anwendung (Greving, 2006, S. 82). Vielfach werden in Studien keine Begründungen für die verwendete Anzahl der Stufen angeführt – so auch bei Skalen zur Evaluation von E-Commerce Auftritten (z.B. Doll, Deng, Raghunathan, Torkzadeh & Xia, 2004; Sing, 2004; Srivihok, 1999; Wang et al., 2001). Unter den vorhandenen Usability-Fragebögen finden sich sowohl Skalen mit fünf Stufen (ufosV1 von Konradt et al., 2003; SUS von Brooke, 1996; IsoMetrics von Gediga et al., 1999; Sing, 2004; Wang et al., 2001) als auch mit sieben Stufen (ISONORM 9241/10 von Prümper, 1997; PSSUQ von Lewis, 2002). Weiterhin liegen Skalen vor, die drei (SUMI von Porteous, Kirakowski & Corbett, 1993; Kirakowski, 1997) oder auch neun (QUIS von Chin et al., 1988) Abstufungen vorgeben. Die Entscheidung, wie viele Stufen zur Verfügung gestellt werden, sollte sich stets danach richten, in welchem Maße die Respondenten in der Lage sind, eine differenzierte Bewertung bzgl. der abgefragten Aspekte abzugeben (vgl. Bühner, 2004, S. 51). Werden zu wenig Stufen genutzt, so besteht im Sinne der klassischen Testtheorie die Gefahr, dass der wahre Wert nicht auf einer der Stufen, sondern eher dazwischen liegt und dadurch ein Messfehler produziert wird. Eine zu hohe Anzahl an Abstufungen kann andererseits ebenfalls zu einer Erhöhung des Messfehlers führen, da ein zu hohes Maß an Genauigkeit vorge täuscht wird, obwohl sich auf Seiten des Antwortenden keine spezifische Meinung ausgebildet hat. Einer Untersuchung von Lewis (1993) zufolge führt eine siebenstufige Antwortskala bei der Erfassung des Konstrukts Usability zu besseren Ergebnissen als eine fünfstufige.

*Benennung der Pole bzw. Stufen:* Neben der Anzahl der Stufen unterscheiden sich Usability-Fragebögen auf Ebene der Antwortskala hinsichtlich der Benennung der Pole. Bei Aussage-Items wird das Maß an Zustimmung im Englischen zumeist durch die Pole „Strongly Agree“ und „Strongly Disagree“ abgedeckt (z.B. SUS von Brooke, 1996; WAMMI; Kirakowski & Cierlik, 1998); dies kann als eine Art Standard angesehen werden. Im ufosV1 werden die deutschen Übersetzungen „stimme voll zu“ bzw. „stimme gar nicht zu“ verwendet (Konradt et al., 2003). Im Fall eines semantischen Differentials erscheint es trivial, dass sich die Bezeichnungen der Pole von Item zu Item ändern (z.B. „isolierend“ vs. „verbindend“ im AttrakDiff2 von Hassenzahl et al., 2003; „terrible“ vs. „wonderful“ im QUIS von Chin et al., 1988). Gelegentlich werden alle einzelnen Stufen einer Usability-Skala bezeichnet. Im QUIS werden die Stufen des semantischen Differentials von 1 bis 9 durchnummeriert. Die sieben Stufen des ISONORM 9241/10 (Prümper & Anft, 1993) werden jeweils durch die Verwendung von Minus- und Plus-Zeichen gekennzeichnet („---“ bis „+++“; siehe Abbildung 2-8). Für die drei Stufen der SUMI- Antwortskala verwendet Kirakowski (1997) die drei Bezeich-

nungen „agree“, „undecided“ und „disagree“. Die Wahl der Benennung von Polen hat selbstverständlich Auswirkungen auf die Beantwortung der Items (Bühner, 2004, S. 51 ff.).

*Antwortoption „keine Angabe“:* Vielfach wird den Respondenten in Usability-Fragebögen die Option eingeräumt, ein Item als nicht beantwortbar zu kennzeichnen (z.B. im ufosV1 von Konradt et al., 2003; im QUIS von Chin et al., 1988). Dadurch soll gewährleistet werden, dass nur solche Aspekte bewertet werden, die einer Bewertung auch tatsächlich zugänglich sind (vgl. DeRouvray & Couper, 2002). Hawkins und Coney (1981) zeigten in einer Untersuchung, dass Respondenten eines Fragebogens auch dann Bewertungen zu Aspekten abgeben, wenn sie über keinerlei relevantes Hintergrundwissen verfügen; die Autoren bezeichnen dies als einen *Uninformed Response Error* (auch *Tendenz im Zweifel zu Raten*; Nunnally & Bernstein, 1994). Ein solcher Störeffekt tritt dann auf, wenn ein Respondent eine Bewertung bzgl. eines Aspektes abgibt, zu dem er eigentlich gar keine Meinung besitzt. Dies kann bei der Bewertung von Usability-Aspekten dann der Fall sein, wenn Nutzer keine entsprechende Interaktion mit dem Interface eingegangen sind, die durch ein Item abgefragt wird. So haben Nutzer sicher keine Meinung zu dem Item „Der Warenkorb ist übersichtlich“, wenn sie den Warenkorb gar nicht gesehen haben. Graef (2003) zeigte in einer empirischen Untersuchung, dass der *Uninformed Response Error* u.a. davon abhängt, wie vertraut der abgefragte Aspekt dem Antwortenden erscheint. Klingt ein Item, zu dem kein Wissen vorliegt, ähnlich wie andere Items, zu denen der Rater eine Meinung besitzt, so wird er seine Antwort entsprechend anpassen.

*Offene Antwortmöglichkeit:* In einigen Usability-Fragebögen wird zusätzlich zur Antwortskala ein offenes Antwortformat verwendet, das den Nutzern ermöglicht, eine deskriptive Schilderung ihrer Interaktionserfahrungen vorzunehmen bzw. darzustellen, woran sie ihre quantitative Bewertung festmachen. Im IsoMetrics<sup>L</sup> (Gediga et al., 1999) findet sich ein offenes Antwortfeld für jedes Item; im QUIS (Chin et al., 1988) wird jeweils ein offenes Feld für jede der zehn Subskalen zur Verfügung gestellt. Durch diese qualitative Abfrage soll es möglich werden, konkrete Gestaltungsmängel zu identifizieren und auf Grundlage dessen Gestaltungsempfehlungen abzuleiten. Entsprechend kann ein solcher Usability-Fragebogen auch im Rahmen formativer Evaluation von Nutzen sein. Als großer Nachteil der Verwendung offener Antwortfelder ist anzumerken, dass die Beantwortung ein deutliches Mehr an Zeitaufwand erfordert. Zudem gestaltet sich die Auswertung sehr aufwendig und lässt einen hohen Interpretationsspielraum. Eine Anwendung dieser Art von Usability-Abfrage im Rahmen einer Online-Erhebung erscheint folglich fraglich. Es stellt sich angesichts des hohen Aufwandes die Frage, ob es nicht sinnvoller ist, Nutzer im Rahmen eines Usability-Tests direkt zu befragen.

*Skala zur Relevanzeinschätzung:* Zusätzlich zur regulären Bewertung eines Items wird in einigen Fragebögen eine Wichtigkeitseinschätzung seitens der Nutzer abgefragt. Auf Grundlage dieser Beurteilung soll es möglich sein, die einzelnen inhaltlichen Aspekte unter-



schiedlich zu priorisieren (z.B. IsoMetrics<sup>L</sup> von Gediga et al., 1999; WebQual von Barnes & Vidgen, 2002). Die Einbeziehung einer solchen Relevanzskala birgt wiederum den Nachteil, dass die Beantwortung des Fragebogens mehr Zeit erfordert – allerdings nicht in dem Maße wie bei der Verwendung offener Antwortfelder.

#### 2.4.2.6 Itemanzahl

Usability-Fragebögen variieren in hohem Maße in der Anzahl der verwendeten Items. Während beispielsweise der After-Scenario-Questionnaire (ASQ) von Lewis (1991) lediglich drei Items umfasst, werden im PUTQ (Lin et al., 1997) insgesamt 100 Items verwendet. Es gilt bei der Festlegung der Itemanzahl einen Trade-Off zwischen dem Aufwand der Fragebogenerhebung und dem Nutzen des inhaltlichen Informationsgewinns zu ziehen. So ist es umso wahrscheinlicher, eine große Anzahl an Respondenten zu erreichen, je weniger Items gestellt werden. Die Kosten und der Aufwand der Erhebung werden eher gering sein. Andererseits bietet eine hohe Anzahl an Items die Option, eine spezifischere Analyse durchzuführen; Kosten und Aufwand sind jedoch im Allgemeinen höher.

#### 2.4.3 Gütebeurteilung von Usability-Fragebögen

Die Güte psychologischer Testverfahren wird gemeinhin auf Grundlage psychometrischer Qualitätsmerkmale beurteilt. Herangezogen werden dabei die allgemein bekannten Kriterien der klassischen Testtheorie (vgl. Lienert & Raatz, 1998). Dzida et al. (2000, S. 88) weisen darauf hin, dass bestimmte Kriterien der klassischen Testtheorie jedoch zur Gütebestimmung eines Usability-Benutzerfragebogens „eigentlich nicht gar nicht anwendbar ist“, da diese eine zeitliche Stabilität des zu messenden Merkmals voraussetzt. Aufgrund der in Abschnitt 2.2.1.1 dargestellten Kontextabhängigkeit des Konstrukts Usability muss diese Voraussetzung als unerfüllt angesehen werden. Folglich erweist sich das Kriterium der Retest-Reliabilität zur Gütebeurteilung der Messgenauigkeit im Falle eines Usability-Fragebogens als unsinnig – es sei denn, es wird gewährleistet, dass der Kontext einer Benutzung für die zu verschiedenen Zeitpunkten durchgeführten Messungen identisch ist. Eine solche hohe Vergleichbarkeit des Kontextes erscheint jedoch nur schwer umsetzbar. Dzida et al. (2000) führen an, es werde trotz dieses Einwands versucht, psychometrische Gütekriterien anzuwenden, um einen Vergleich verschiedener Usability-Fragebögen herstellen zu können.

Zum Zwecke der Validierung von Usability-Fragebögen stellt Ollermann (2004) eine Prozedur vor, bei der die Ergebnisse im Fragebogen mit der Anzahl an Problemnennungen in einem formativen Evaluationsverfahren abgeglichen werden. Als einen weiteren Ansatz der Bestimmung der Kriteriums-Validität nennt der Autor die Bestimmung des korrelativen Zusammenhangs zwischen den Fragebogenergebnissen und denen einer expertenbasierten Evaluation. Zudem führt er die „Technik der bekannten Gruppen“ an. Diese besteht darin, dass zwei Gruppen von Software, die Unterschiede in der Usability aufweisen, mit Hilfe des Fragebogens bewertet werden. Ist das Instrument in der Lage, zwischen beiden Gruppen zu

differenzieren, so wird dies als Hinweis auf eine Validität gewertet. Ein derartiges Vorgehen verfolgten z.B. Kirakowski, Claridge und Whitehand (1998) zur Validierung des WAMMI, Hassenzahl et al. (2003) im Fall des AttrakDiff2 sowie Lin, Choong und Salvendy (1997) für den PUTQ. In einer Validierungsuntersuchung zum WAMMI-Fragebogen von Kirakowski und Cierlik (1998) führen die Autoren Übereinstimmungen der WAMMI-Ergebnisse mit einem Performanzmaß (relativierte Zeit der Aufgabenerledigung) und einem nutzerseitig eingeschätzten Stressmaß als Beleg für die konvergente Validität des Instruments an. Analog gingen Yom und Wilhelm (2004) vor, um die Validität der WOOS-Skala zu belegen, indem sie Korrelationen zu einer Skala von Donovan und Rossiter (1978) zur affektiven Bewertung des Shopping-Erlebens berechnen. Im Fall des WOOS wurden zudem die Zusammenhänge der Skalenwerte zu objektiven Usability-Daten bestimmt (Klickanzahl und Bearbeitungsdauer in einer Suchaufgabe).

#### **2.4.4 Vor- und Nachteile der fragebogenbasierten Usability-Erfassung**

Wie jede Usability-Evaluationsmethode zeichnen sich Fragebögen durch eine Reihe von Vor- und Nachteilen aus. Anhand der genannten Vor- und Nachteile kann geschlossen werden, in welchem Rahmen die Anwendung dieser Verfahren sinnvoll ist. Ihr Nutzen richtet sich in erster Linie nach der Zielsetzung, die mit der Usability-Evaluation verbunden ist. Entsprechend können sich spezifische Eigenschaften in dem einen Kontext als vorteilhaft und in einem anderen als nachteilig erweisen.

##### **2.4.4.1 Potenzielle Vorteile**

Im Folgenden werden zunächst die möglichen Vorteile von Usability-Fragebögen angeführt.

*Subjektives Nutzerurteil:* Mit einem Fragebogen wird ein Nutzerurteil erhoben. Eine nutzerseitige Bewertung kann ausschlaggebend sein, inwiefern eine Produkteigenschaft als bedeutsame Abweichung von der Norm anzusehen ist (Dzida et al., 2000). Dieser Vorteil gilt selbstverständlich für alle nutzerbasierten Usability-Erhebungsansätze gegenüber einer rein expertenbasierten Evaluation.

*Geringer Aufwand:* Fragebogenerhebungen sind in der Regel schnell und ohne viel Aufwand durchführbar (vgl. Gediga & Hamborg, 2002). Insbesondere durch die Möglichkeit, einen Fragebogen auch online präsentieren zu können, wird der Aufwand einer Erhebung deutlich reduziert (vgl. Richter, 1999; Shneiderman, 1998, S. 133). Generell bergen Online-Umfragen gegenüber anderen Erhebungsmethoden diverse Vorteile (Evans & Mathur, 2005).

*Geringe Kosten:* Die mit der Anwendung verbundenen Kosten sind vergleichsweise gering, da eine Anwesenheit eines Usability-Experten in der Regel nicht notwendig ist.

*Hohe Anzahl an Bewertungen:* Es kann mit Usability-Fragebögen eine große Anzahl an Bewertungen erreicht werden, was in der Außenwirkung eine Art Autorität gegenüber Ergebnissen ausstrahlt, die auf einer kleinen Anzahl an Nutzern beruhen (vgl. Shneiderman,

1998, S. 132). Eine große Anzahl an Respondenten ermöglicht eine Querschnittsbildung über verschiedene Interaktionsformen und Nutzertypen. Andererseits kann auch eine vergleichende Betrachtung verschiedener Nutzertypen oder auch eine Clusterbildung von Nutzern vorgenommen werden.

*Hinweise auf Gestaltungsmängel:* Insofern in einem Fragebogen Items mit eher hohem Granulationsgrad enthalten sind bzw. ein offenes Antwortformat verwendet wird, können auf Grundlage der Ergebnisse auf Itemebene erste Hinweise für konkrete Usability-Probleme gesammelt werden (vgl. Barnes & Vidgen, 2003).

*Anwendung statistischer Methoden:* Die Erhebung quantitativer Fragebogendaten ermöglicht die Anwendung statistischer Auswertungsverfahren. Dies macht diese Verfahrensgruppe vor allem zur Untersuchung von Forschungsfragen attraktiv, beispielsweise wenn es gilt, die Usability zu anderen Größen in Beziehung zu setzen.

*Vergleichende Evaluation bzw. Benchmarking:* Ein Vergleich verschiedener Software-Produkte kann leicht durchgeführt werden (vgl. Gediga & Hamborg, 2002; Bevan & Macleod, 1994). Dies erweist sich zum einen im Rahmen des Rapid-Prototyping bzw. der fortschreitenden Weiterentwicklung von Schnittstellen als vorteilhaft (vgl. Barnes & Vidgen, 2003). Zum anderen bietet es sich an, Fragebögen zum Vergleich verschiedener Produkte im Rahmen eines Benchmarking anzuwenden (vgl. Barnes & Vidgen, 2002; Kim, Shaw & Schneider, 2003), etwa um zu prüfen, inwiefern Nutzer die Gebrauchstauglichkeit von Konkurrenzprodukten höher bewerten. Zum Vergleich ähnlicher Schnittstellen kann eine Referenzdatenbank angelegt werden (vgl. Kirakowski & Cierlik, 1998).

*Sicherung von Mindestanforderungen der ergonomischen Qualität:* Es kann ermittelt werden, inwiefern die Qualität eines Software-Produktes im Vorfeld definierten Mindestanforderungen bzw. Richtwerten genügt. Dazu wird geprüft, inwiefern eine bestimmte Ausprägung des jeweiligen Fragebogenmaßes erreicht wird (vgl. Dzida et al, 2000). Im Sinne einer Tätigkeitsanalyse erfolgt also ein Vergleich zwischen Ist-Zustand und Soll-Zustand (vgl. Ulich, 1981).

*Anwendung im Rahmen der summativen Evaluation:* Die Anwendung bietet sich vor allem im Rahmen einer summativen Evaluation an, da Fragebögen in der Regel eine Aussage zur Gesamtzufriedenheit von Nutzern ermöglichen (vgl. Gediga & Hamborg, 2002).

*Gewährleistung von Anonymität:* Fragebögen können anonym ausgefüllt werden. Dies bietet den Vorteil, dass die Wahrscheinlichkeit einer Bewertung im Sinne der sozialen Erwünschtheit reduziert wird, die bei Anwesenheit eines Experten unter Umständen gegeben wäre (vgl. Dzida et al., 2000).

#### 2.4.4.2 Potenzielle Nachteile

Andererseits zeichnen sich Usability-Fragebögen gegenüber anderen Methoden durch einige Nachteile aus. Aufgrund dieser Aspekte erweist sich eine Anwendung dieser Verfahren in bestimmten Untersuchungs- bzw. Evaluationskontexten als unattraktiv:

*Identifikation spezifischer Gestaltungsmängel:* Fragebogenergebnisse lassen in der Regel nicht darauf schließen, welches exakte Gestaltungsmerkmal von Nutzern als problematisch erlebt wird. Je geringer der Granulationsgrad der Items, umso weniger lassen sich spezifische Mängel identifizieren.

*Anwendung im Rahmen der formativen Evaluation:* Aufgrund der Einschränkungen bzgl. der Identifikation einzelner Gestaltungsmängel ist die Möglichkeit der Ableitung konkreter Verbesserungsmaßnahmen auf Grundlage von Usability-Fragebögen deutlich eingeschränkt (vgl. Yom, 2003, S. 162), so dass eine Anwendung im Rahmen einer formativen Evaluation fraglich erscheint. Um diesem Nachteil zu begegnen, werden in einigen Usability-Fragebögen offene Antwortfelder verwendet (siehe Abschnitt 2.4.2.5).

*Unkontrollierte Erhebung:* Die Usability stellt gemäß der DIN ISO EN 9241-11 (1999) ein in hohem Maße kontextabhängiges Konstrukt dar. Die Interaktion, also z.B. die Art einer erledigten Aufgabe, kann einen starken Einfluss auf die Bewertung seitens der Nutzer nehmen. Dies erweist sich im Fall von Usability-Fragebögen dann als problematisch, wenn eine Anwendung nicht in einem kontrollierten Rahmen wie im Usability-Labor stattfindet, da unklar bleibt, auf was für einer Interaktion die Bewertung beruht.

*Unflexibler Einsatz:* Standardisierte Fragebögen sind derart konzipiert, dass sie allgemein anwendbar sind oder sich auf eine Gruppe von Systemen beziehen. Je nach Kontext einer Anwendung ist es jedoch vielfach wünschenswert, sehr spezifische Aspekte zu berücksichtigen. Diese sozusagen spontan in Items umzusetzen, erweist sich insofern als problematisch, als dass die Messgüte eines Fragebogens vor seiner Anwendung evaluiert werden muss. Andere Evaluationsverfahren sind dieser Einschränkung der Freiheitsgrade nicht unterlegen.

*Spezifische Messfehler:* Ein Problem fragebogenbasierter Usability-Evaluation kann darin liegen, dass Befragte vereinzelt meinen, sie selber und nicht die jeweilige Benutzerschnittstelle seien Gegenstand der Bewertung (Xenos, 1999). Insbesondere Respondenten mit geringer Nutzungsexpertise bzw. Vorerfahrung neigen zudem dazu, eine negative Bewertung einer Schnittstelle mit einer Art Eingeständnis ihrer vermeintlichen Inkompetenz gleichzusetzen. Entsprechend fällt ihre Bewertung nicht in dem Maße negativ aus, wie es angesichts der tatsächlich aufgetretenen Nutzungsprobleme zu erwarten wäre.

Angesichts der genannten Nachteile kann die Usability-Evaluation einer Software nicht ausschließlich auf der Anwendung fragebogenbasierter Verfahren fußen (Dzida, Wiethoff & Arnold, 1993). Schließlich stellt die nutzerseitige Zufriedenheit im Sinne der DIN EN ISO 9241-11 (1999) nur einen von drei Faktoren der Usability dar. Die Entwickler des SUMI-

Fragebogens (Porteous & Kirakowski, 1992) empfehlen daher ein zweistufiges Vorgehen: Zunächst lasse man eine Software mit Hilfe eines Usability-Fragebogens von einer Stichprobe von Nutzern (mindestens  $N=10$ ) beantworten, um ein Antwortprofil zu erstellen. Im zweiten Schritt solle ein Interview mit einigen Nutzern desselben Samples stattfinden, in dem der Interviewer auf die Beantwortungen im Fragebogen Bezug nimmt, um herauszufinden, woran die Nutzer ihre Bewertungen festmachen. In diesem Sinne können Usability-Fragebögen als Ausgangspunkt für halbstrukturierte Interviews mit Nutzern dienen (vgl. Dzida et al., 1993).

## 2.5 Usability im Online-Shopping

Im Folgenden wird zunächst aufgezeigt, dass Bedarf an reliablen und validen Instrumenten zur Usability-Evaluation von Online-Shops besteht, da viele Shops gravierende Gestaltungsmängel aufweisen (Abschnitt 2.5.1). Dass sich eine mangelnde Gebrauchstauglichkeit in erheblichem Maße negativ auf den Erfolg eines Online-Anbieters auswirken kann, wird in Abschnitt 2.5.2 geschildert. Schließlich wird in Abschnitt 2.5.3 eine Unterscheidung zwischen verschiedenen Formen der Online-Shop-Nutzung vorgenommen.

### 2.5.1 Ausprägung der Usability von Online-Shops

In Studien zur Usability des Online-Shoppings wurden eine Reihe gravierender Gestaltungsmängel aufgedeckt (Magenat, 1999; Rehman, 2000; Tilson & Dong, 1998). Zumeist liegen diesen Publikationen Usability-Tests zugrunde, in denen Nutzer konkrete einkaufsbezogene Aufgaben erledigen sollten, etwa ein bestimmtes Produkt im Sortiment eines Shops finden, es in den Warenkorb legen und / oder eine Bestellung aufgeben. Es zeigte sich, dass derartige Aufgaben oft nicht bewältigt werden konnten. Nielsen, Molich, Snyder und Farrell (2001) berichten von einer durchschnittlichen Fehlschlagrate von nahezu 50%. Besonders schwerwiegende Usability-Probleme sehen die Autoren als „Sales Catastrophes“ an, da sie dazu führen, dass die Abwicklung einer Bestellung für Nutzer unmöglich wird (Nielsen et al., 2001, S. 26). Neben besonders folgenschweren Gestaltungsfehlern treten in Nutzertests sehr viel häufiger weniger kritische Probleme auf, die in ihrer störenden Wirkung auf das Einkaufserlebnis der Kunden jedoch nicht unterschätzt werden sollten. Eine Reihe konkreter Beispiele führen Christophersen und Konradt (2004) an.

Es ließe sich annehmen, dass die Gebrauchstauglichkeit vieler Online-Shops inzwischen besser ausgeprägt sein dürfte als früher – zumal in diversen Publikationen konkrete Usability-Aspekte im Electronic Commerce dokumentiert sind (z.B. Nielsen et al., 2001; Kim, Lee, Han & Lee, 2002). Ein Besuch auf Jakob Niensens Website „Alertbox“ (Online: <http://www.useit.com/alertbox/>) lässt jedoch darauf schließen, dass noch immer viele Fehler bei Gestaltung von Online-Shops gemacht werden. Beispielhaft werden in Tabelle 2-4 die von Nielsen aufgestellte Top Ten der Design-Fehler des Jahres 2005 angeführt (vgl. Nielsen, 2005b)<sup>7</sup>. Alle in dieser Liste genannten Gestaltungsfehler wurden bereits in den Vorjahren als Probleme angemahnt. Die Fülle der in der Literatur genannten Usability-Verstöße von Online-Shops legt nahe, die Usability in diesem Bereich als einen Erfolgsfaktor zu betrachten (vgl. Chaparro, 2002; Coyle & Gould, 2002; Öörni, 2003).

---

<sup>7</sup> Zu berücksichtigen ist dabei, dass eine Aussage zur Usability eigentlich nur auf einen Benutzungskontext bezogen werden kann (DIN EN ISO 9241-10, 1999; siehe Abschnitt 2.2.1.1). Es widerspricht insofern genau genommen dem definitorischen Hintergrund, einzelne Probleme losgelöst aufzuführen.

Tab. 2-4: Jakob Niensens Top Ten der gewichtigsten Design-Fehler des Jahres 2005  
(Nielsen, 2005b)

(1)	Mangelnde Lesbarkeit der verwendeten Schrift (zu kleine Font-Größe, schlechte Farbkontraste etc.)
(2)	Keine standardgemäße Verwendung von Links (nicht als Links identifizierbar, keine Kennzeichnung bereits besuchter Seiten etc.)
(3)	Unangemessene Einbindung von Flash-Elementen (für Intros und zur Navigation)
(4)	Verwendung von Inhalten, die nicht aufs Web abgestimmt wurden (zu lang, nicht „scannable“ etc.)
(5)	Schlechte Suchfunktionen
(6)	Keine Gewährleistung von Browser Kompatibilität
(7)	Unangemessene Formulare (unnötige Abfrage von Angaben, Pflichtangaben, keine flexible Eingabe von Daten wie Telefonnummern möglich)
(8)	Keine Angabe von Kontakt- oder anderen Informationen zum Unternehmen
(9)	Starre Layouts mit festgelegter Seitenbreite
(10)	Inadäquate Vergrößerung von Fotos (nur ausschnittsweise, keine Zoom-Funktionen).

### 2.5.2 Usability als Erfolgsfaktor von Online-Shops

Innerhalb der Forschungsdisziplin der Mensch-Maschine-Interaktion besteht seit langem Einigkeit bzgl. des Umstandes, dass der Erfolg einer Software in hohem Maße von ihrer Gebrauchstauglichkeit beeinflusst wird (vgl. Bias & Mayhew, 2005). Beispielsweise werden im Arbeitskontext durch den Einsatz benutzerfreundlicher Software Produktivität und Umsatz gesteigert und Kosten reduziert. Eine schlechte Usability von Arbeitssoftware führt dazu, dass die Nutzung schwer erlernt wird und die Mitarbeitermotivation bzw. -zufriedenheit sinkt (Bevan & Macleod, 1994). „Wettbewerbsvorteile können in einem Markt, in dem Produkte mit vergleichbarer Funktionalität angeboten werden, durch bessere ergonomische Qualität erzielt werden“ (Dzida, 1994, S. 374).

Die Anzahl der Publikationen, die sich zur Erfolgswirksamkeit der Usability im Kontext des Einkaufens im Internet heranziehen lässt, ist groß (vgl. Constantinides, 2004). Im Folgenden erfolgt eine Einteilung der Befunde in grobe Aussagenbereiche. Dabei wird an dieser Stelle nur auf den Zusammenhang zwischen der Usability und dem Erfolg eingegangen. Medierte Einflüsse werden in den Abschnitten 2.7.1 und 2.7.2 dargestellt. Im Folgenden wird zunächst der Einfluss der objektiven Usability auf den Erfolg von Online-Anbietern betrachtet (siehe Abschnitt 2.3.3 zur Unterscheidung zwischen der subjektiven und objektiven Usability).

*Schwerwiegende Usability-Mängel führen zu Kaufabbrüchen:* Wie bereits in Abschnitt 2.5.1 erwähnt, erweisen sich einige Gestaltungsmängel von Online-Shops als derart gravierend, dass Kunden gar nicht in der Lage sind, ihren Einkauf abzuschließen. Sie sehen sich gezwungen, ihren Kaufvorgang aufgrund dieser „Sales Catastrophes“ abubrechen (Nielsen et al., 2001, S. 5 und S. 26; Rehman, 2000). Es ist trivial: Ein Shop, in dem Kunden aufgrund besonders gravierender Gestaltungsfehler ihren Einkauf abbrechen, *kann* nicht erfolgreich sein, da keine Bestellungen im Shop getätigt werden.

*Gebrauchstaugliche Elemente eines Shops erweisen sich als erfolgswirksam:* Einige Autoren stellen Zusammenhänge zwischen dem Vorhandensein konkreter Designmerkmale und Erfolgsgrößen her. Auch auf Grundlage dieser Untersuchungen kann geschlossen werden: Shops, die sich durch besonders gebrauchstaugliche Elemente auszeichnen, erweisen sich gegenüber anderen Shops als erfolgreicher. Eine Reihe empirischer Studien deutet darauf hin, dass die Einbindung interaktiver Elemente eine erfolgssteigernde Wirkung besitzt (Auger, 2005; Bhatt, 2004; Fiore, Kim & Lee, 2005; Gammack & Hodgkinson, 2003). Zudem wirkt sich die Art und Qualität der Informationsdarstellung auf die Kaufentscheidung sowie die Kundenzufriedenheit aus (Lee & Lee, 2004; Palmer, 2002; Shankar et al., 2003); entscheidend ist dabei, dass kein „Information Overload“ entsteht (Lee & Lee, 2004). In einer Untersuchung von Weathers und Makienko (2006) nimmt eine hohe Anzahl verschiedener Bestellwege eines Shops („Ease of Ordering“) signifikanten Einfluss auf den Erfolg des Anbieters. In einer frühen Studie von Lohse und Spiller (1999) erweisen sich Online-Shops umso erfolgreicher, je mehr benutzerfreundliche Seitenbestandteile vorhanden waren. Kohavi und Round (2004) berichten, dass konkrete Änderungen der Usability einzelner Seitenbestandteile beim Online-Anbieter Amazon.com signifikante Anstiege der Verkäufe zur Folge hatten.

Neben diesen Befunden, die für den Einfluss der objektiven Usability auf den Erfolg von Online-Anbietern sprechen, lassen sich einige weitere zur Erfolgswirksamkeit der subjektiven Usability heranziehen:

*Usability-Mängel führen zu einem unbequemen und zeitintensiven Online-Einkauf:* Neben den Sales Catastrophes können sich auch weniger gravierende Gestaltungsmängel negativ auf die Kaufbereitschaft auswirken. Sie führen dazu, dass der Online-Einkauf unbequem wird und mehr Zeit erfordert (vgl. Ceaparu, Lazar, Bessiere, Robinson & Shneiderman, 2004; Öörni 2003). Viele Internet-Kunden schätzen gerade die Bequemlichkeit und Zeitersparnis des Online-Shoppings (siehe Abschnitt 2.1.4.1). Online-Shops, die diese Vorteile nicht bieten, werden entsprechend seltener in Anspruch genommen. Begünstigt wird dies durch die niedrigen Such- bzw. Wechselkosten im Internet (vgl. Yang & Peterson, 2004): die Konkurrenz ist stets „only one click away“ (Helander & Khalid, 2000; Smith, Bailey & Brynjolfsson, 1999).

*Gebrauchstaugliche Shoppelemente bzw. -eigenschaften sind Online-Kunden besonders wichtig:* In verschiedenen Untersuchungen wurden Internetnutzer dahingehend befragt, was ihnen im Bereich Online-Shopping besonders wichtig ist. Nach Zhang, von Dran, Blake und Pipithsuksunt (2001) sind die wichtigsten Features aus Sicht der Nutzer: sichere Datenübertragung, einfache Navigation, angemessene Erklärungen, gute Suchfunktionen, die Produktdarstellung sowie Service-Erwägungen. Auch in einer frühen Untersuchung von Tilson und Dong (1998) erweisen sich diverse Aspekte der Usability aus Sicht der Kunden als besonders wichtig (Feedback nach erfolgter Bestellung; leicht zu findende und verständliche



Bestellbuttons etc.). Die Notwendigkeit, umfangreiche Produktinformationen zur Verfügung zu stellen, wird durch Ergebnisse von Schlosser (2003) unterstrichen. Die Autorin zeigte experimentell, dass Personen am Computer mehr Produktinformationen verlangen als bei papierbasierter Darstellung und diese Informationen kognitiv elaborierter verarbeiten.

*Die Usability nimmt Einfluss auf die Einstellung zum Einkaufen im Internet:* Da im TAM angenommen wird, die Wirkung der PEOU und PU auf die Nutzungsintention werde durch die Einstellung zur Nutzung mediiert, untersuchen einige Autoren diesen indirekten Einfluss auf die Kaufintention (Davis, 1989, 1993; siehe Abschnitt 2.2.2.3). In empirischen Untersuchungen von Pavlou und Fygenson (2006) sowie Van der Heijden, Verhagen und Creemers (2003) ergeben sich signifikante Zusammenhänge zwischen der PEOU und der Einstellung zum Online-Kauf. In Studien von Hassanein und Head (2005) sowie Pavlou und Fygenson (2006) erweist sich auch der Zusammenhang zwischen der PU und der Einstellung zum Online-Kauf als signifikant.

*Die Usability nimmt Einfluss auf die Kaufintention:* Aufgrund einer Reihe von Publikationen kann geschlossen werden, dass die Usability Einfluss auf die Kaufintention von Online-Kunden nimmt. In der bereits angeführten Untersuchung von Konradt et al. (2003) erweist sich die Usability als signifikanter Prädiktor für diese Erfolgsgröße. Vielfach beziehen sich Autoren nicht auf das Konstrukt Usability sondern auf die verwandten Konstrukte Perceived Ease of Use (PEOU) bzw. Perceived Usefulness (PU) im Sinne des TAM (Davis 1989, 1993; siehe Abschnitt 2.2.2.3). In einer empirischen Studie von Muthitacharoen, Palvia, Brooks, Krishnan, Otondo und Retzlaff-Robert (2006) erwies sich der direkte Einfluss der PEOU sowie der PU auf die Kaufintention als signifikant. Auch bei Gefen und Straub (2000) ergab sich ein signifikanter Einfluss der PU auf die Kaufintention.

*Die Usability nimmt Einfluss auf die Intention, zum Shop zurückzukehren:* In weiteren Untersuchungen zeigte sich, dass die Usability nicht nur Einfluss auf die Einstellung bzw. die Intention zum Internetkauf nimmt, sondern auch auf die Bereitschaft einen Online-Shop erneut aufzusuchen bzw. dort wiederholt einzukaufen (Koufaris, Kambil & LaBarbera, 2002; Liang & Lai, 2002; Magnenat, 1999). Je anstrengender Kunden die Benutzung empfinden, umso weniger sind sie geneigt, den jeweiligen Shop wieder zu besuchen. Eine langfristige Kundenbindung wird somit durch eine mangelnde Usability deutlich erschwert.

*Die Usability nimmt Einfluss auf das Kaufverhalten:* Nur wenige Studien liegen vor, in denen die Usability in einen Zusammenhang zum tatsächlichen Kaufverhalten gesetzt wurde. In der Studie von Konradt et al. (2003) erwies sich die Usability als ein signifikanter Prädiktor für die Kaufentscheidung. In einer experimentellen Untersuchung von Liang und Lai (2002) kauften studentische Teilnehmer eher in Shops mit einer „guten Design-Qualität“, wobei die Autoren der Studie schließen, die Design-Qualität erweise sich als genauso wichtig wie der Verkaufspreis und einflussreicher als die Reputation des Shops. Signifikante Zusammenhänge

zwischen PEOU bzw. PU mit Parametern des Nutzungsverhaltens im Kontext des Einkaufens im Internet ergeben sich bei Henderson und Divett (2003).

In den Abschnitten 2.7.1 und 2.7.2 wird dargestellt, dass die Usability darüber hinaus Einfluss auf die zwei Variablen Einkaufsvergnügen und Vertrauen nimmt, die ihrerseits ebenfalls als Erfolgsfaktoren angesehen werden können. Die dargestellten Zusammenhänge und Befunde untermauern die Erfolgswirksamkeit der Usability im Kontext des Online-Shoppings. Es kann zusammenfassend geschlossen werden:

*Je gebrauchstauglicher ein Shop gestaltet ist, umso höher wird die Intention von Online-Kunden ausfallen, im Shop zu kaufen bzw. umso höher ist die Wahrscheinlichkeit, dass positive Kaufentscheidungen getroffen werden.*

### 2.5.3 Verschiedene Arten der Nutzung eines Online-Shops

In der DIN EN ISO 9241 wird auf die starke Kontextabhängigkeit des Konstrukts Usability hingewiesen (siehe Abschnitt 2.2.1.1). Eine Aussage zum Konstrukt sollte stets nur vor dem Hintergrund des Nutzers, der Arbeitsaufgabe, der physischen und sozialen Umgebung sowie den genutzten Arbeitsmitteln getroffen werden. Auf verschiedene Charakteristika der Nutzer wird in Abschnitt 2.7.7.2 näher eingegangen. Im Folgenden wird dem Aspekt der Arbeitsaufgabe besondere Aufmerksamkeit geschenkt, indem betrachtet wird, welche unterschiedlichen Formen der Nutzung von Online-Shops grob auseinander gehalten werden können.

Sowohl im Bereich des Online-Shoppings als auch im traditionellen Handel wird oftmals zwischen zwei Formen der Produktsuche unterschieden, die das Verhalten bei der Suche nach Informationen bestimmen: *Stöbern (Browsen)* und *gezieltes Suchen* (Rowley, 2000).

- *Stöbern* stellt ein ungerichtetes Interaktionsverhalten von Kunden dar, bei dem sie Informationen zu Produkteigenschaften wahrnehmen, ohne ein explizites Kaufziel vor Augen zu haben (Toms, 2000). Nutzer haben dabei eine wenig konkrete Vorstellung bzgl. der Information, die sie benötigen bzw. die verfügbar sein könnte, so dass diese Suchform einen starken explorativen Charakter besitzt (Detlor, Sproule & Gupta, 2003; vgl. auch Wolfenbarger & Gilly, 2001).
- Im Gegensatz dazu stellt eine *gezielte Suche* eine Verhaltenskategorie von Nutzern dar, bei der aktiv nach Informationen zu einem Produkt oder dem Produkt selber gesucht wird. Diese Kunden wissen also, was sie finden möchten und verfügen für gewöhnlich bereits über eine gewisse Information zum Produkt – sei es die Marke oder der Titel (Rowley, 2000).

Kunden zeigen während eines Shopbesuchs nicht immer nur eine der beiden genannten Verhaltensweisen. Suchstrategien können sich zum Zeitpunkt der Interaktion ändern, indem die Nutzer die gesammelten Informationen verarbeiten und als neuen Ausgangspunkt wählen

(Hodkinson, Kiel & McColl-Kennedy, 2000). Oftmals starten Kunden mit einer explorativen Suche, bei der sie durch den Shop stöbern, und richten ihr Suchverhalten im Laufe der Interaktion immer gezielter aus (Shim, Eastlick, Lotz & Washington, 2001). Detlor et al. (2003) untersuchten, welche Unterschiede sich für die Wichtigkeit verschiedener Informationsaspekte eines Shops beim Stöbern im Vergleich zur gezielten Suche ergeben. Die Autoren kamen zu dem Ergebnis, dass in beiden Fällen Informationen zu den Preisen, den Produkteigenschaften, zur Produktauswahl sowie Entscheidungshilfen (z.B. FAQs, Peer-to-Peer-Infos, Empfehlungen, Geschenkideen) von hoher Relevanz sind. Ein benutzerfreundliches Interfacedesign erscheint somit bei beiden Suchstrategien gleichermaßen wichtig zu sein, um Nutzer zu unterstützen. In Abbildung 2-12 werden die Gemeinsamkeiten und Unterschiede zwischen beiden Suchstrategien visualisiert. Moe (2003) erweitert die Typologie an Suchstrategien beim Online-Shop-Besuch, indem er zusätzlich zur Dichotomie gezielte Suche vs. Stöbern unterscheidet, inwiefern ein Einkauf unmittelbar oder in der Zukunft stattfindet. Entsprechend ergeben sich vier Formen von Kaufstrategien, die er als *Gezieltes Kaufen*, *Suche/Überlegen*, *Hedonisches Browsen* und *Wissensbildung* bezeichnet.

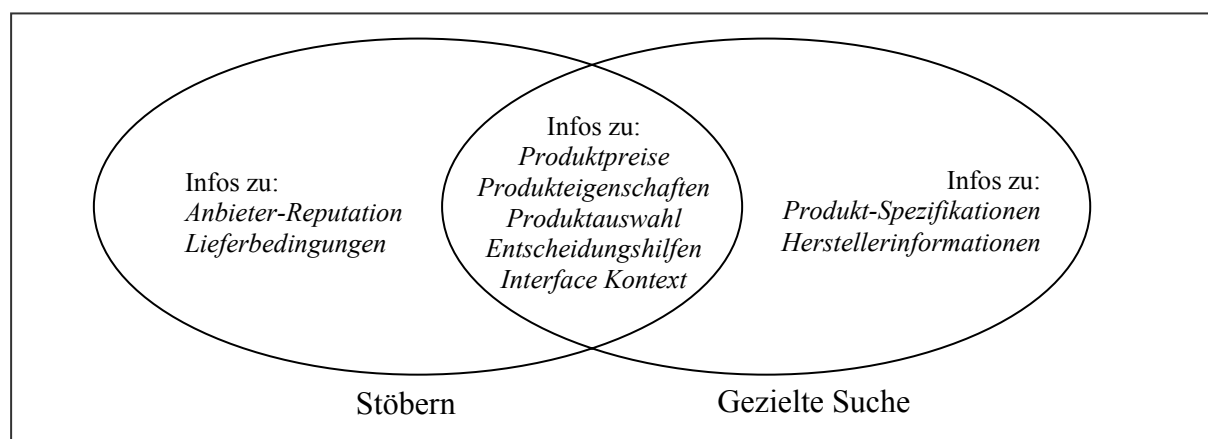


Abb. 2-12: Zwei verschiedene Suchstrategien von Kunden im Online-Shopping (Detlor, Sproule & Gupta, 2003, S. 81)

Die angeführten Formen der „Arbeitsaufgabe“ bei der Benutzung eines Online-Shops stellen lediglich grobe Kategorien dar, die im Rahmen der Usabilityuntersuchung zur Klassifikation von Benutzergruppen und der Auswahl von Kontexten verwendet werden können. Das individuelle Nutzungsverhalten setzt sich hingegen aus sehr vielen verschiedenen Schritten und Entscheidungen seitens der Kunden zusammen, die es im Einzelfall zu berücksichtigen gilt, deren Betrachtung im Kontext der vorliegenden Arbeit jedoch nicht zielführend wäre (vgl. Travis, 2003).

## 2.6 Konzeptuelles Erfolgsmodell zur Validierung von Usability-Skalen

Wie in Abschnitt 2.4.4 genannt wurde, bietet sich ein Ansatz zur Validierung von Usability-Fragebögen darin, die Skalenmaße im Rahmen eines komplexen Modells in einen nomologischen<sup>8</sup> Zusammenhang zu weiteren Variablen zu stellen. Es erscheint somit sinnvoll, die neuen Fragebogenmaße des ufosV2 in einem konzeptuellen Erfolgsmodell ins Verhältnis zu anderen kundennahen Erfolgsfaktoren von Online-Shops zu setzen. In dem Modell werden Zusammenhänge zwischen der Usability und weiteren Variablen berücksichtigt, die auf Grundlage zurückliegender Befunde als gesichert anzusehen sind. Finden diese Zusammenhänge empirische Bestätigung, so spricht dies für die Validität des neuen Fragebogens. Zudem wird eine Abschätzung dahingehend möglich, welche Relevanz den Maßen des ufosV2-Fragebogen in Hinblick auf den Erfolg eines Online-Anbieters zukommt.

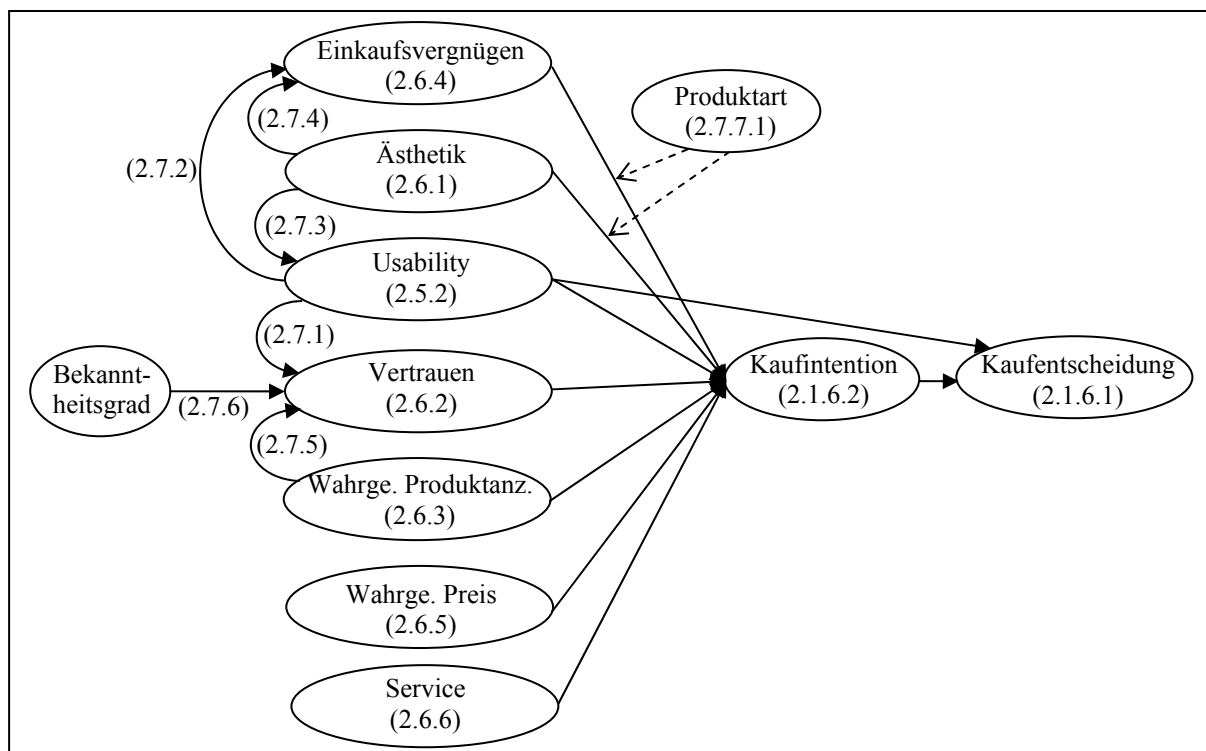


Abb. 2-13: Konzeptuelles Erfolgsmodell zur Validierung von Usability-Fragebogenmaßen im Kontext des Online-Shoppings

Anmerkung: In Klammern wird jeweils angegeben, in welchen Abschnitten dieser Arbeit die Konstrukte sowie die entsprechenden Zusammenhangsbefunde vorgestellt werden.

In Abbildung 2-13 ist das konzeptuelle Modell zur Validierung von Fragebogenmaßen im Kontext des Online-Shoppings dargestellt. Als Erfolgsgrößen (also Kriterien bzw. abhängige Variablen) dienen die *Kaufentscheidung* bzw. die *Kaufintention*, die in den Abschnitten 2.1.6.1 und 2.1.6.2 dargestellt und diskutiert wurden. Gemäß der Annahme der Theorie des geplanten Verhaltens dient die Erfolgsgröße Kaufintention als Prädiktor für die Kaufent-

<sup>8</sup> Die nomologische Validität betrifft also die Frage, inwiefern sich theoretisch angenommene Zusammenhänge zu weiteren Konstrukten in einem Modell bestätigen (Fritz, 1995, S. 138 f.).

scheidung (Ajzen, 1991). Neben der Usability werden die kundennahen Erfolgsfaktoren Ästhetik, Vertrauen, wahrgenommene Produktanzahl, Einkaufsvergnügen, wahrgenommener Preis sowie Service-Leistungen als Prädiktoren herangezogen. In den Abschnitten 2.6.1 bis 2.6.6 werden diese Konstrukte vorgestellt. Dabei wird durch vorliegende Befunde untermauert, dass es sich bei ihnen um Erfolgsfaktoren handelt. Zwischen einigen der Erfolgsfaktoren werden innerhalb des konzeptuellen Modells gerichtete Beziehungen angenommen. Diese Zusammenhänge werden in den Abschnitten 2.7.1 bis 2.7.6 näher beschrieben und empirisch begründet. Neben den genannten Erfolgsgrößen finden im Modell moderierende Einflüsse der Produktart Berücksichtigung, die in den Abschnitt 2.7.7.1 erläutert werden. Abschnitt 2.7.7.2 ist personenbezogenen Merkmalen als potenziellen Moderatoren gewidmet.

### 2.6.1 Ästhetik

Der ästhetische Eindruck eines Produktes, im Folgenden auch kurz *Ästhetik*, stellt im Sinne der Konsumentenforschung eine Reaktion auf die Symmetrie, Proportionen und Einheitlichkeit eines Objektes, eines Gedichtes oder einer Performance dar (Olsen, 1981; Verzyer, 1993). Bei Webseiten wird dieser Eindruck aus Sicht von Mathwick, Malhotra und Rigdon (2001) durch die Kombination der verwendeten Farben, des graphischen Layouts und der bildlichen Darstellung gebildet.

Tractinsky und Rao (2001) nennen die Ästhetik als eine der sozialen Dimensionen von Online-Shops. Kim, Williams und Lee (2003) sind ähnlicher Auffassung. Ihnen zufolge stellt die Ästhetik (hier: *Visual Appeal*) eines Online-Shops eine der wesentlichen Dimensionen der *Website Qualität* dar (vgl. auch Mathwick et al., 2001). Schenkman und Jönsson (2000) zeigten in einer experimentellen Studie, dass der Gesamteindruck bzgl. einer Website auf Seiten der Nutzer vor allem dadurch bestimmt wird, als wie *schön* sie die Site wahrnehmen. Park, Choi und Kim (2004) stellen fest, der ästhetische Eindruck derselben Webpage könne zwischen den Nutzern stark variieren. Auch sei es möglich, dass ein Nutzer zu verschiedenen Zeitpunkten zu unterschiedlichen Bewertungen der Ästhetik komme. Kwon, Kim und Lee (2002) betrachteten den Einfluss konkreter Designfaktoren auf die nutzerseitige Gesamtbewertung von Online-Auktionshäusern. Die Bewertung des Designs übte gegenüber den Konstrukten *Usability* und *Informationsgehalt* den stärksten Einfluss auf den Gesamteindruck aus. Hall und Hanna (2004) führten ein Experiment durch, bei dem sie die farbliche Gestaltung von Internetseiten systematisch variierten. Ihren Ergebnissen zufolge wirkt sich sowohl die Verwendung präferierter Farben als auch die von Nutzerseite eingeschätzte *Ästhetik* einer Site positiv auf die *Kaufintention* aus. Die dargestellten Befunde legen folgenden Schluss nahe:

*Je ästhetischer Kunden die Gestaltung eines Online-Shops empfinden, umso höher ist ihre Intention, einen Einkauf bei dem Anbieter zu tätigen.*

### 2.6.2 Vertrauen

Vertrauen kann als ein Mechanismus eines Individuums angesehen werden, in einer mit Unsicherheit verbundenen Situation Komplexität zu reduzieren (Luhmann, 1989). Vertrauen besteht folglich nur in Situationen, die mit Unsicherheiten bzw. Risiken behaftet sind (vgl. Moorman, Deshpandé & Zaltman, 1993). Bei vollkommener Sicherheit ist es dagegen nicht nötig, Vertrauen aufzubringen. Grabner-Kräuter und Kaluscha (2003) beklagen, in der Literatur werde vielfach undifferenziert mit dem Konstrukt umgegangen. Die im Rahmen der vorliegenden Arbeit relevante Form des Konstrukts ist das *Vertrauen in den Anbieter*: Jarvenpaa und Tractinsky (1999) definieren es als die Bereitschaft von Kunden, dem Anbieter zu vertrauen und in Situationen zu handeln, in denen sie verletzlich gegenüber dem Anbieter sind. Während sich das Vertrauen im sozialen Kontext gemeinhin auf andere Personen bezieht, stellt im Bereich Online-Shopping der Shop bzw. der Anbieter das Objekt des Vertrauens dar. Gegenüber dem Vertrauen in den Anbieter können einige weitere Vertrauensformen abgegrenzt werden, die im Zusammenhang dieser Untersuchung jedoch nicht von besonderem Interesse sind. So bezeichnet etwa das *Initiale Vertrauen* das Vertrauen in einen Shop, der einem Kunden zuvor unbekannt war (Hampton-Sosa & Koufaris, 2005). Das *System-Vertrauen* ist als das Vertrauen in den Vertriebsweg Internet anzusehen (Grabner-Kräuter & Kaluscha, 2003). Weitere Formen stellen das *Dispositionale Vertrauen* (McKnight, Cummings & Chervany, 1998), *Trusting Beliefs* (Schlosser, White & Lloyd, 2006) sowie *Trusting Intentions* (Trifts & Häubl, 2003) dar.

In der Forschung zu traditionellen Handelsbeziehungen wird dem Konstrukt Vertrauen seit langem eine bedeutende Rolle zugeschrieben (z.B. Doney & Cannon, 1997; Hawes, Mast & Swan, 1989; Schurr & Ozanne, 1985). Auch in der Literatur zu den Erfolgsfaktoren im Online-Shopping besteht weitgehende Einigkeit bzgl. des starken Einflusses dieser Variable. Mangelndes Vertrauen auf Seiten von Kunden führt oftmals dazu, dass Personen keinen Online-Einkauf tätigen (z.B. George; 2004; Hoffman, Novak & Peralta, 1998; Lee & Turban, 2001). Vertraut ein Kunde dem Anbieter, so führt dies zu einer Minderung der mit einer Online-Transaktion verbundenen Unsicherheiten und Risiken (siehe Abschnitt 2.1.4.2). Entsprechend steigt die Wahrscheinlichkeit, dass eine positive Kaufentscheidung getroffen wird. Der positive Einfluss des Vertrauens in den Anbieter auf die Kaufintention findet in diversen empirischen Untersuchungen Bestätigung: Einwiller (2003), Gefen und Straub (2003), Hampton-Sosa und Koufaris (2005), Hassanein und Head (2005), Jarvenpaa et al. (2000), Pavlou (2003), Schlosser et al. (2006) sowie Van der Heijden und Verhagen (2004). In Untersuchungen von Harris und Goode (2004) sowie Gefen (2002) bestätigte sich zudem ein Einfluss des Vertrauens auf die Loyalität von Online-Kunden. Auf Grundlage dieser Befunde kann die folgende Aussage getroffen werden:

*Je stärker Online-Kunden einem Anbieter vertrauen, umso höher ist die Intention der Kunden, einen Einkauf im jeweiligen Shop zu tätigen.*

### 2.6.3 Wahrgenommene Produktanzahl

Marketingorientierte Studien zum traditionellen Handel deuten darauf hin, dass ein Unterschied zwischen der tatsächlichen Auswahl an Produkten und der Wahrnehmung seitens von Kunden bestehen kann (Hoch, Bradlow & Wansink, 1999; Van Herpen & Pieters, 2002). Die *wahrgenommene Produktanzahl* stellt den subjektiven Eindruck von der Anzahl der Produkte seitens der Kunden dar. Eine Divergenz zwischen tatsächlicher und wahrgenommener Anzahl kann auf unterschiedliche Faktoren wie die räumliche Anordnung, die Aufteilung sowie auf Eigenschaften der Produkte zurückzuführen sein. Auch im Kontext des Einkaufens im Internet kann die Art der Gestaltung von Online-Shops zur Folge haben, dass ein Unterschied zwischen der tatsächlichen und der wahrgenommenen Produktanzahl besteht (vgl. Brynjolfsson, Smith & Hu, 2003). Eine gute Gebrauchstauglichkeit führt dazu, dass den Online-Kunden schnell klar wird, welche Produkte verfügbar sind. Dagegen erhöht eine schlechte Usability die Wahrscheinlichkeit, dass viele vorhandene Produkte gar nicht erst gefunden und damit wahrgenommen werden (Nielsen et al., 2001, S. 116).

Es ist anzunehmen, dass Kunden u.a. deswegen bevorzugt große Shops auswählen, da die Wahrscheinlichkeit eines Erfolgs bei der Suche nach Produkten höher ist als in kleinen Shops. Im traditionellen Handel besteht die Erkenntnis einer Erfolgswirksamkeit einer hohen Produktanzahl seit langem (Baumol & Ide, 1956; Kahn & Lehmann, 1991). Ergebnisse verschiedener empirischer Untersuchungen lassen darauf schließen, dass der tatsächlichen Produktanzahl auch im Bereich des Online-Shoppings eine hohe Erfolgswirksamkeit zuzumessen ist (Lohse & Spiller, 1999; Menon & Kahn, 2002; Weathers & Makienko, 2006). Der Branchenprimus Amazon.com nennt eine große Anzahl angebotener Produkte als eine seiner wesentlichen Business-Strategien (Kohavi & Round, 2004).

Die Wichtigkeit der wahrgenommenen Produktanzahl wird dadurch unterstrichen, dass viele Kunden eine große Produktvielfalt als einen wesentlichen Vorteil des Online-Shoppings ansehen (vgl. Wolfinger & Gilly, 2001). Befragungsstudien weisen darauf hin, dass Online-Kunden besonderen Wert auf eine große Produktvielfalt legen (Donthu & Garcia, 1999; Jarvenpaa & Todd, 1997; Levin, Levin, & Weller, 2005). Ergebnisse einer Conjoint Analyse von Schaupp und Bélanger (2005) sprechen für einen Einfluss der Produktvielfalt auf die *Kundenzufriedenheit*. Bei Konradt et al. (2003) erwies sich die wahrgenommene Produktanzahl als ein signifikanter Prädiktor für die Kaufintention sowie für die Kaufentscheidung. Angesichts der in diesem Abschnitt genannten Befunde kann geschlossen werden:

*Je höher aus Sicht von Kunden die Anzahl der in einem Online-Shop angebotenen Produkte ausfällt, umso höher ist ihre Intention, einen Einkauf bei dem Anbieter zu tätigen.*

### 2.6.4 Einkaufsvergnügen

Das Einkaufsvergnügen kann als eine unspezifische Form des Affekts auf Seiten der Kunden angesehen werden, den sie während des Einkaufs in einem Shop erleben. *Emotionen*, *Stimmungen*<sup>9</sup> und *Sentiments*<sup>10</sup> stellen verschiedene Formen von Affekt dar (Frijada, 1994). Eine *Emotion* ist zielgerichtet auf ein bestimmtes Objekt und intentional. Sie stellt eine Reaktion eines Individuums auf ein relevantes Geschehen dar und enthält affektive, verhaltensbezogene sowie kognitive Komponenten (Kleinginna & Kleinginna, 1981). Im Gegensatz dazu sind *Stimmungen* nicht gerichtet und unintentional. Stimmungen werden vor allem durch Emotionen ausgelöst oder übernommen, aber auch durch weitere Einflüsse wie z.B. Musik, Wetter oder Temperatur (vgl. Levy, 1984; Thayer, 1989). Bei einem *Sentiment* handelt es sich nicht um einen Gefühlszustand, sondern vielmehr um eine affektive Bewertung eines Objektes. So können Nutzer einer Software die Bedienung im Sinne eines Sentiments als frustrierend bezeichnen (vgl. Clore, 1994, S. 108). Sentiments können im Gegensatz zu Emotionen und Stimmungen zeitlich bestehen. Das im Marketing oft verwendete Konstrukt Einkaufsvergnügen ist aus psychologischer Sicht insofern undifferenziert, als dass keine Aussage zur konkreten Form von Affekt getroffen wird.

Diverse Studien im Bereich Mensch-Maschine-Interaktion weisen darauf hin, dass sich Affekt auf Seiten von Nutzern in hohem Maße auf ihr Nutzungsverhalten und ihre Entscheidungsprozesse auswirken kann (Hudlicka, 2003; Karat et al., 2002; Krone, Hamborg & Gediga, 2002; Partala & Surakka, 2004). In der Erfolgsfaktorenforschung zum traditionellen Handel wird affektiven Zuständen auf Seiten von Kunden ein wesentlicher Einfluss auf die Wahrscheinlichkeit zu kaufen beigemessen (vgl. Babin, Darden & Griffin, 1994; Belk, 1975; Dawson, Bloch & Ridgway, 1990; Holbrook & Hirschman, 1982). Das mit einem Besuch im traditionellen Laden verbundene Einkaufsvergnügen führt zu einer größeren Anzahl gekaufter Produkte, einer höheren ausgegebenen Geldsumme, einem längeren Shopaufenthalt sowie einer höheren Bereitschaft zurück zu kehren (Donovan & Rossiter, 1982; Sherman & Smith, 1986). Diverse empirische Studien untermauern den Einfluss des Einkaufsvergnügens für den Bereich des Online-Shoppings (Eroglu, Machleit & Davis, 2003; Hassanein & Head, 2005; Liu & Arnett, 2000; Lynch, Kent & Srinivasan, 2001). Eine hohe Ausprägung der Variable wirkt sich insofern positiv auf das Kaufverhalten aus, als dass die Bereitschaft zur weiteren Erkundung des Shops und die Anzahl ungeplanter Käufe steigen (Menon & Kahn, 2002). Zusammenfassend wird auf Grundlage der angeführten Befunde die Annahme getroffen:

*Je höher das Einkaufsvergnügen auf Seiten von Kunden eines Online-Shops ausfällt, umso höher ist ihre Intention, bei dem Anbieter zu kaufen.*

---

<sup>9</sup> Engl. *Moods*.

<sup>10</sup> Englischer Ausdruck; in der deutschen Sprache findet sich keine adäquate Entsprechung.



### 2.6.5 Wahrgenommener Preis

Sobald Kunden ihre Zufriedenheit mit einem Produkt oder einem Service einschätzen, beziehen sie für gewöhnlich den *Preis* in ihre Bewertung mit ein (z.B. Anderson & Sullivan, 1993; Cronin, Brady & Hult, 2000). Den Verkaufspreis eines Produktes aufbringen zu müssen, wird von Kunden vielfach als eine Art Verlust angesehen (Anderson, Fornell & Lehmann, 1994; Athanassopoulos, 2000). Dabei kann die Höhe dieses Verlusts individuell unterschiedlich wahrgenommen werden, was eine Unterscheidung zwischen dem tatsächlichen und dem wahrgenommenen Preis nahe legt. Beide Formen des Preises können entscheidend für die Qualitätsbewertung sein (Zeithaml, 1988). Der wahrgenommene Preis ist aus Sicht des Konsumenten jedoch wichtiger (Bei & Chiao, 2001). Kunden können den genauen Preis durchaus bewusst wahrnehmen, viele werden ihn jedoch eher als mehr oder minder „billig“ bzw. „teuer“ verschlüsseln und erinnern. Wieder andere nehmen gar keine Notiz vom Preis (Zeithaml, 1988). Dabei können Konsumenten kognitiv auf eigene Standards für angemessene Beträge zurückgreifen, mit denen sie Produktpreise vergleichen (Monroe, 1973).

Bereits in einer Untersuchung von Jarvenpaa und Todd (1997) aus der frühen Phase des E-Commerce zeigte sich eine hohe Relevanz des Verkaufspreises aus Sicht der Kunden. Zhang, et al. (2001) stellten mittels einer Befragung Ranglisten der wichtigsten Design Features für verschiedene Arten von Websites auf. Angemessene und transparente Informationen zu den Preisen der Produkte sind für Nutzer von E-Commerce-Seiten von besonders hoher Relevanz. Im Jahr 2001 gaben Internetnutzer an, ihnen sei der Preis der online angebotenen Produkte sehr wichtig, wobei die Relevanz dieses Kriteriums jedoch mit steigendem Einkommen sank (Lightner, 2003). Die Ergebnisse einer Untersuchung von Reibstein (2002) deuten daraufhin, dass der *Produktpreis* der angebotenen Produkte im Verhältnis zu anderen Variablen eine große Rolle bei der Wahl des Online-Shops spielt. Kotha, Rajgopal und Venkatachalam (2004) zeigen, dass sich eine Preisführerschaftsstrategie bei Internetunternehmen in den Jahren 1999 bis 2000 als erfolgreich erwies; es machte sich also für Online-Anbieter bezahlt, die Verkaufspreise günstiger anzusetzen, als die Konkurrenz es tat, um eine höhere Anzahl an Produkten zu verkaufen (vgl. Porter, 1980). Für Amazon.com stellt es eine drei wesentlichen Business-Strategien dar, die Preise ihrer Produkte niedrig zu halten (Kohavi & Round, 2004). In einer Untersuchung von Cho, Im, Fjermestad und Hiltz (2003) bestätigte sich die Annahme, dass die Unzufriedenheit von Online-Kunden umso größer ist, je höher der wahrgenommene Preis ausfällt. Auf Grundlage der dargestellten Befunde kann zusammenfassend die Annahme getroffen werden:

*Ein subjektiv als niedrig wahrgenommener Preis wirkt sich positiv auf den Erfolg eines Online-Anbieters aus.*

### 2.6.6 Service

Der Begriff *Service* findet in der Literatur sehr unterschiedliche Verwendung. In der betriebswirtschaftlich orientierten Literatur zu den Erfolgsfaktoren im Internethandel werden dem Service oftmals Gestaltungsmerkmale der Sites zugeordnet (Crisp, Jarvenpaa & Todd, 1997; Harris & Goode, 2004). Gelegentlich werden auch die Verfügbarkeit und Güte von Informationen zum Produkt, Unternehmen etc. als Aspekte des Service aufgefasst (Lohse & Spiller, 1999). In einer Skala zur Service-Qualität von Harris und Goode (2004) finden sich selbst Items zur Abfrage des ästhetischen Eindrucks einer Site. Eine derartige konzeptuelle Zuordnung dieser Bewertungsdimensionen zum Konstrukt Service erscheint im Lichte des zugrundegelegten Usabilitybegriffs wenig sinnvoll. Vielmehr gilt es, Service-Aspekte zu identifizieren, die in der vorliegenden Arbeit nicht bereits durch andere Konstrukte abgedeckt werden und eine hohe Relevanz für die Kaufentscheidung eines Online-Shoppers besitzen.

Den Ergebnissen verschiedener Studien zufolge messen Kunden dem *Vorhandensein alternativer Möglichkeiten zur Abwicklung der Bezahlung* (Optionen zur Zahlung per Kreditkarte, Scheck, Debit-Card sowie in Bar) eine hohe Relevanz zu (Burke, 2002; Liang & Lai, 2002; Weathers & Makienko, 2006). Weiterhin erwiesen sich bei Burke (2002) die *Möglichkeiten zur persönlichen Kontaktaufnahme mit dem Verkaufspersonal* aus Kundensicht als besonders wichtig (kompetente Ansprechpartner, persönlicher Kontakt per Telefon, kostenfreier Telefonkontakt). Die Ergebnisse einer Conjoint Analyse von Schaupp und Bélanger (2005) lassen auf eine hohe Relevanz kurzer *Lieferfristen* von Online-Shops für die Zufriedenheit von Kunden schließen. Lwin und Williams (2006) zeigten, dass die von Seiten des Anbieters gegebenen *Garantien* zu einer Reduktion der Risikowahrnehmung der Kunden führen können und sich insofern als erfolgswirksam erweisen. Als ein weiterer Serviceaspekt wird der „Follow-Up“ oder „Post-Sale“-Service angesehen, der die Leistungen des Verkäufers nach erfolgter Bestellung umfasst (vgl. Liang & Lai, 2002; Liu & Arnett, 2000). Darunter fällt auch die *Regelung zur Rückgabe von Produkten* für den Fall, dass ein Produkt nicht den Erwartungen von Kunden genügt. Wie Lewis, Singh und Fay (2006) sowie Schindler et al. (2005) zeigten, können außerdem die von Seiten des Shops erhobenen *Kosten für Porto und Verpackung* das Online-Kaufverhalten stark beeinflussen. Neben den genannten Service-Aspekten finden sich diverse weitere Leistungen auf Anbieterseite (z.B. Geschenkservices, Trackingmöglichkeiten, Store Locators, mehrsprachige Versionen etc.; vgl. Nielsen et al., 2001). Bzgl. der Relevanz dieser Facetten liegen jedoch kaum verwertbare Befunde vor. Zusammenfassend kann auf Grundlage der Darstellung die Annahme getroffen werden:

*Das Bereitstellen verschiedener Service-Leistungen seitens eines Online-Anbieters (Möglichkeiten zur Abwicklung der Bezahlung, Möglichkeiten der persönlichen Kontaktaufnahme, Lieferfristen der Produkte, Garantien, Rückgaberegelungen sowie die Kosten für Porto und Verpackung) wirkt sich positiv auf den Erfolg des Shops aus.*

## **2.7 Zusammenhänge innerhalb des konzeptuellen Modells**

In dem in Abschnitt 2.6 vorgestellten konzeptuellen Modell werden zwischen einigen der einbezogenen Konstrukte Zusammenhänge angenommen. In den folgenden Kapiteln werden diese Beziehungen zwischen den Erfolgsfaktoren im Online-Shopping konzeptionell und empirisch begründet.

### **2.7.1 Einfluss der Usability auf das Vertrauen**

Vielfach wird in der Literatur die These vertreten und empirisch untersucht, die Usability beeinflusse das Vertrauen in einen Anbieter. Das Vertrauen ist neben der Usability als ein wesentlicher Erfolgsfaktor im Online-Shopping anzusehen (siehe Abschnitt 2.5.2); ein positiver Einfluss des Vertrauens auf die Kaufintention findet in diversen Studien empirische Unterstützung. Der Einfluss der Usability auf das Vertrauen in den Anbieter wird gemeinhin wie folgt begründet: Der Eindruck der Gebrauchstauglichkeit diene Nutzern als Beurteilungsgrundlage, inwiefern einem Anbieter vertraut werden könne (vgl. Egger, 2001). Einerseits bewirkt eine schlechte Gebrauchstauglichkeit eines Shops einen Vertrauensverlust. Andererseits vermittelt eine gute Usability den Eindruck, der Anbieter sei in der Lage, eine Bestellung im Interesse der Kunden abzuwickeln; durch eine hohe Gebrauchstauglichkeit werden somit die sogenannten *Ability Beliefs* gestärkt. Bucher et al. (2002, S. 134) sowie Büffel (2002, S. 163 ff.) schließen auf Basis umfangreicher qualitativer Nutzertests zum Verhalten von Online-Kunden, dass Elemente und Eigenschaften der Website von Nutzern zur Vertrauensprüfung herangezogen werden (siehe auch Nielsen et al., 2001, S. 77 ff.). In verschiedenen empirischen Studien erweist sich der Zusammenhang zwischen beiden Konstrukten als signifikant (Büttner et al., 2006; Egger, 2003; Everard & Galletta, 2003; Konradt et al., 2003; Roy, Dewit & Aubert, 2001). Der Zusammenhang zwischen der PEOU bzw. der PU und dem Vertrauen findet empirische Bestätigung bei Pavlou (2003), Featherman und Pavlou (2003) sowie Pavlou und Fygenson (2006). Auf Grundlage der angeführten Untersuchungen kann die folgende Annahme getroffen werden:

*Je gebrauchstauglicher ein Online-Shop aus Sicht der Kunden gestaltet ist, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.*

### **2.7.2 Einfluss der Usability auf das Einkaufsvergnügen**

Allgemein erleben Personen Frustration, wenn sie davon abgehalten werden, erwartete zufriedenstellende Ergebnisse zu erreichen (Berkowitz, 1978). Mängel der Gebrauchstauglichkeit können dazu führen, dass Nutzer ihre Ziele nicht in erwarteter Weise erreichen. Dies kann negative affektive Erlebniszustände und Reaktionen in Form von Frustration oder Ärger auf Seiten der Nutzer zur Folge haben (vgl. Ceaparu et al., 2004; Kim & Moon, 1998; Lazar, Jones, Hackley & Shneiderman, 2006). Beim Browsen im Internet erweist sich insbesondere ein Gefühl der Desorientierung („Lostness“) als ein frustrierendes Erlebnis

(Otter & Johnson, 2000). Andererseits macht eine hoch ausgeprägte Usability den Aufbau eines Flowerlebnisses (Pilke, 2004) in der Benutzung wahrscheinlicher (Chan & Ahern, 1999).

Kommentare von Nutzern in Usability-Tests lassen darauf schließen, dass auch beim Einkaufen im Internet Frustration ausgelöst wird, wenn Erwartungen nicht erfüllt werden (Coyle & Gould, 2002). Andererseits kann eine gut ausgeprägte Usability einem positiven Einkaufserlebnis förderlich sein. In einer Fragebogenuntersuchung von Van der Heijden und Verhagen (2004) bestätigt sich eine Hypothese, wonach die Variablen PEUO und PU einen signifikanten Einfluss auf das Einkaufsvergnügen nehmen. Insgesamt weisen die angeführten Befunde auf einen sehr starken Zusammenhang zwischen der wahrgenommenen Usability eines Online-Shops und dem bei der Nutzung erlebten Vergnügen hin, wobei anzunehmen ist, dass die wahrgenommene Usability Einfluss auf das Einkaufsvergnügen nimmt:

*Je gebrauchstauglicher ein Online-Shop von Seiten der Kunden wahrgenommen wird, umso mehr Vergnügen bereitet ihnen der Einkauf in dem Shop.*

### **2.7.3 Einfluss der Ästhetik auf die Usability**

Diversen Untersuchungen zufolge besteht ein enger Zusammenhang zwischen dem ästhetischen Eindruck einer Software auf Seiten der Nutzer und der Usability des Interfaces (Kurosu & Kashimura, 1995; Tractinsky, 1997; Tractinsky, Katz & Ikar, 2000; van der Heijden, 2003). Gewisse Uneinigkeit besteht in der Literatur jedoch bzgl. der Richtung des kausalen Zusammenhangs zwischen den Konstrukten (vgl. Frohlich, 2004; Hassenzahl, 2004a, 2004b; Norman, 2004a; Overbeeke & Wensveen, 2004; Tractinsky, 2004).

Zumeist wird die Annahme vertreten, der ästhetische Eindruck bestimme die nutzerseitig wahrgenommene Usability. Tractinsky und Rao (2001) erklären eine Einflussnahme der Ästhetik auf andere Variablen wie das Einkaufsvergnügen und die Usability (das sie als das "Schön ist gut"-Phänomen bezeichnen), indem sie drei mögliche kognitive Prozesse auf Seiten von Nutzern anführen. Einerseits könnte eine Art Stereotypenbildung stattfinden, bei der *Schönheit* mit anderen positiven Attributen einer Person oder eines Objektes assoziiert wird; die Autoren beziehen sich bei diesem Erklärungsansatz auf Studien aus dem Bereich der Sozialpsychologie (Dion, Berscheid & Walster, 1972; Eagly, Ashmore, Makhijani & Longo, 1991). Andererseits ließe sich auch der *Halo Effekt* als Erklärung heranziehen (auch *Hof-Effekt*; Thorndike, 1920; Nisbett & Wilson, 1977). Dieser besagt, dass besonders hervortretende Eigenschaften eines Objektes andere Attribute bei ihrer Bewertung überstrahlen können (vgl. Murphy, Anhalt & Jako, 1993; Wirtz & Bateson, 1995). Da sich ein ästhetischer Eindruck zumeist schnell bilde, wirke er auf den Gesamteindruck bzw. auf die Bewertung anderer Aspekte desselben Objektes. Eine weitere Erklärung für den Zusammenhang zwischen Ästhetik und Usability sieht Karvonen (2000) in der Einfachheit (engl. *Simplicity*) bei der Gestaltung von Schnittstellen. Einerseits wird gefordert, ein System solle möglichst

simpel gestaltet sein, damit ein hohes Maß an Usability erreicht werden könne (z.B. Nielsen, 1993). Andererseits wird Einfachheit vom Betrachter vielfach auch als ästhetischer erlebt.

Wie in einigen Studien gezeigt wurde, muss nicht in jedem Fall ein Zusammenhang zwischen den beiden Variablen Ästhetik und Usability von Software-Anwendungen bestehen. In einem Experiment von Lindgaard und Dudek (2003) wurde eine Website als sowohl benutzerfreundlich als auch optisch ansprechend angesehen; in einem weiteren Experiment wurde eine Site jedoch als *schön* erlebt, obwohl die Bewertung der Usability von Nutzerseite mangelhaft ausfiel. In einem von Hassenzahl (2004b) durchgeführten Experiment ergab sich bei der Bewertung diverser Interfaces keine hohe bzw. teilweise sogar eine negative Korrelation zwischen der *Schönheit* und der *pragmatischen Qualität*, welche einem nutzerseitigen Eindruck der Usability entspricht. Trotz dieser Befunde scheint zumeist der folgende Zusammenhang plausibel:

*Je höher Nutzer die Ästhetik einer Software-Anwendung bewerten, als umso höher nehmen sie die Ausprägung der Usability der Benutzerschnittstellen wahr.*

#### **2.7.4 Einfluss der Ästhetik auf das Einkaufsvergnügen**

Einige Untersuchungen sprechen dafür, dass es sich positiv auf das Vergnügen einer Nutzung von Software auswirkt, wenn ein Benutzer die Schnittstelle als ästhetisch gestaltet empfindet. In Abschnitt 2.7.3 wurde bereits eine Argumentation von (Tractinsky & Rao, 2001) angeführt, die als Erklärung für den Einfluss der Ästhetik auf die Variablen Usability und Einkaufsvergnügen dienen kann. Tractinsky et al. (2000) kommen auf Grundlage eines Experimentes zu dem Schluss, die Ästhetik beeinflusse durch eine affektive Reaktion die Stimmung der Nutzer, welche wiederum auf die Gesamtbeurteilung wirke. Kim, Lee und Choi (2003) führten eine umfangreiche experimentelle Untersuchung zum Zusammenhang zwischen Designfaktoren einer Homepage und den emotionalen Reaktionen auf Seiten der Nutzer durch. Zunächst identifizierten die Autoren emotionale Dimensionen, die Nutzer bei der Betrachtung von Internetseiten erleben. Anschließend konnten den Dimensionen konkrete Designmerkmale zugeordnet werden. Beispielweise erleben Nutzer ein hohes Maß an Spannung bei einer Webseite, wenn ihr Hintergrund keine nicht scharfen Abgrenzungen und keine gelbe Färbung aufweist. Diese Ergebnisse legen es nahe, dass ein enger Zusammenhang zwischen der ästhetischen Gestaltung einer Site und dem affektiven Erleben auf Seiten von Nutzern besteht. Den Ergebnissen einer Studie von Jordan (1997) zufolge beeinflusst der ästhetische Eindruck eines Produktes in hohem Maße das Vergnügen, das mit der Nutzung verbunden ist. In der bereits angeführten Studie von van der Heijden (2003) nimmt die visuelle Attraktivität einer Website signifikanten Einfluss auf die Variablen PU, PEOU sowie auf das Einkaufsvergnügen. Dabei sprechen die Ergebnisse dafür, dass sich der Einfluss auf das Einkaufsvergnügen als stärker erweist als im Falle der PU bzw. der PEOU. Angesichts dieser Befunde kann die folgende Annahme getroffen werden:

*Je ästhetischer ein Online-Shop aus Sicht der Kunden gestaltet ist, umso mehr Vergnügen empfinden sie bei der Nutzung des Shops.*

### **2.7.5 Einfluss der Produktanzahl auf das Vertrauen**

Jarvenpaa, Tractinsky und Vitale (2000) postulierten, die wahrgenommene Größe eines Shops (operationalisiert durch die wahrgenommene Anzahl der Produkte) nehme Einfluss auf das Vertrauen in den Shop, welches sich wiederum als einflussreich auf die Kaufintention von Kunden auswirke. Je größer ein Shop sei, umso eher würden Internetnutzer dies als einen Hinweis dafür werten, dass andere dem Unternehmen trauen und erfolgreich Käufe abwickeln. Diese vermeintlich positiven Erfahrungen anderer Kunden begründen Vertrauen. Zudem vermittelt eine gewisse Größe, der Anbieter verfüge über die Expertise und die Ressourcen, eine Bestellung angemessen abzuwickeln. Die Ergebnisse der Studie von Jarvenpaa et al. (2000) lassen jedoch darauf schließen, dass der durch das Vertrauen medierte Zusammenhang zwischen der wahrgenommenen Produktanzahl und dem Erfolgsmaß abhängig ist von der Produktart: Im Fall von Buchshops erweist sich der entsprechende Pfad als nicht signifikant, wohingegen ein signifikanter Zusammenhang bei Reiseshops vorliegt. Da die Autoren keine schlüssige Erklärung für diesen Befund liefern, erscheint eine generelle Annahme eines moderierenden Einflusses der Produktart auf den Zusammenhang zwischen der wahrgenommenen Produktanzahl und Vertrauen nicht gerechtfertigt. Angesichts der vagen Befundlage kann unter Vorbehalt die folgende Schlussfolgerung getroffen werden:

*Je mehr Produkte aus Sicht von Online-Kunden in einem Shop erworben werden können, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.*

### **2.7.6 Einfluss des Bekanntheitsgrads auf das Vertrauen**

In der Literatur besteht weitgehende Einigkeit dahingehend, dass das Ausmaß des Vertrauens seitens der Kunden vom Bekanntheitsgrad bzw. der Reputation des Anbieters beeinflusst wird (Jarvenpaa et al., 2000). Dabei wird argumentiert, eine hohe Reputation stelle für Kunden eine Vertrauensbekundung Dritter dar, die ihnen signalisiere, auch sie könnten dem Shop sein Vertrauen schenken (Einwiller, 2003). In diversen empirischen Studien bestätigt sich der Einfluss der Reputation auf das Vertrauen in den Anbieter (De Ruyter, Wetzels & Kleijnen, 2001; Einwiller, 2003; Jarvenpaa et al., 2000; Van der Heijden & Verhagen, 2004). Entsprechend wird der folgende Zusammenhang im konzeptionellen Modell berücksichtigt:

*Je bekannter ein Shop aus Sicht von Online-Kunden im Allgemeinen ist, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.*

### 2.7.7 Moderierende Einflüsse

In der Literatur zum Online-Shopping werden gelegentlich moderierende Einflüsse zwischen Erfolgsfaktoren und dem Erfolg diskutiert. Es stellt sich die Frage, inwiefern einzelne Moderatoren im Rahmen des konzeptuellen Modells Berücksichtigung finden sollten. In Abschnitt 2.7.7.1 wird die Produktart als Moderator vorgestellt. Um einen Eindruck von der Vielfältigkeit der Befunde zu vermitteln, wird in Abschnitt 2.7.7.2 ein Überblick gegeben, welche personenbezogenen Merkmale verschiedentlich als Moderatoren genannt werden.

#### 2.7.7.1 Produktart

Im Zusammenhang des vorgestellten konzeptuellen Modells (Abbildung 2-13) kann ein moderierender Einfluss durch die Art des Produktes angenommen werden. Zur Einteilung von Produkten, die online angeboten werden, werden verschiedene Taxonomien vorgeschlagen. Häufig wird zwischen *Search* und *Experience* Produkten unterschieden. Demzufolge unterscheiden sich Produkte dahingehend, inwiefern sie über Such- oder Erfahrungseigenschaften (Search und Experience Attributes) verfügen (Nelson, 1970; siehe auch Klein, 1998). Burke (2002) nimmt eine alternative Einteilung vor, und schlägt drei Klassen von im Internet verkauften Produkten vor. Eine Gruppe wird durch *selten gekaufte, langlebige Gebrauchsgüter gebildet* (z.B. Möbel, Autos). In die zweite Gruppe der *Verbrauchsgüter* fallen Produkte wie Nahrungsmittel, Büromaterialien und Arzneimittel. Weiterhin grenzt er als dritte Kategorie *Entertainmentprodukte* wie Bücher, Spielzeuge und Computerspiele bzw. Kleidungsstücke ab.

Hassanein und Head (2005) beklagen, vielen Shops mangle es an einer emotionalen und sozialen Note. Sie stellen in ihrer Untersuchung fest, dass ein positiver Einfluss auf die kundenseitige Kaufbereitschaft erreicht werden kann, wenn die soziale Präsenz eines Shops durch die Verwendung einer sozial ansprechenden Informations- und Produktdarstellung geschaffen wird. Dabei besteht jedoch eine Abhängigkeit von der Produktart. Im Fall von Produkten wie Kleidung, bei denen Kunden sich einen vergnüglichen und unterhaltsamen Einkauf wünschen, wirkt sich eine Betonung des Einkaufsvergnügens positiv auf die Kaufbereitschaft aus. Auch Burke (2002) stellt auf Basis einer umfangreichen Diskriminanzanalyse fest, der Online-Kauf von Entertainmentprodukten und Kleidungsstücken solle aus Kundensicht unterhaltsam sein und Spaß machen. Bei Verbrauchsgütern kann die Gestaltung hingegen durchaus nüchtern ausfallen. Levin et al. (2005) kommen zu einem ähnlichen Ergebnis, indem sie empirisch feststellen, ein hohes Einkaufsvergnügen sei besonders wichtig im Falle von Experience-Produkten wie Kleidung. Zeichnen sich Produkte hingegen vor allem durch Such-Eigenschaften aus (z.B. Flugtickets), so erweise sich der Einfluss der Variable als weniger stark. Zusammenfassend können zwei Annahmen bzgl. des moderierenden Einflusses der Produktart im Fall der Variablen Ästhetik und Einkaufsvergnügen getroffen werden:

*Die Wirkung des Einkaufsvergnügens auf die Online-Kaufintention wird durch die Art des Produktes moderiert. Im Fall von Entertainmentprodukten bzw. Produkten, die primär Experience-Eigenschaften besitzen, erweist sich die Variable als einflussreich. Bei Verbrauchsgütern bzw. Produkten, die primär Sucheigenschaften aufweisen, besitzt das Einkaufsvergnügen keine besonders hohe Relevanz.*

*Die Wirkung der Ästhetik auf die Online-Kaufintention wird durch die Art des Produktes moderiert. Im Fall von Entertainmentprodukten bzw. Produkten, die primär Experience-Eigenschaften besitzen, erweist sich die Variable als einflussreich. Bei Verbrauchsgütern bzw. Produkten, die primär Sucheigenschaften aufweisen, ist es weniger relevant, ob der Online-Shop als ästhetisch wahrgenommen wird oder nicht.*

#### 2.7.7.2 Personenbezogene Merkmale als potenzielle Moderatoren

Forschungsergebnisse aus dem Bereich der Software-Ergonomie zeigen, dass sich die Bedürfnisse und Präferenzen von weiblichen und männlichen Nutzern unterscheiden können (vgl. Cassel, 2003). Zwar liegt eine ganze Reihe verschiedener Publikationen vor, die sich mit den Unterschieden zwischen Männern und Frauen in der Nutzung und Präferenz im Bereich Online-Shopping beschäftigen (Garbarino & Strahilevitz, 2004; Kumar, Smith & Bannerjee, 2004; Kwon et al., 2002; Levin et al., 2005; Rodgers & Harris, 2003; Schimmel & Morris, 2003; Teo & Lim, 2000; Venkatesh & Morris, 2000). Eine Sichtung dieser Quellen erbringt jedoch kein einheitliches Bild bzgl. des Einflusses des Geschlechts, da sich spezifische Unterschiede nicht über mehrere Studien hinweg als signifikant erweisen.

Die Ergebnisse einiger Studien deuten hingegen darauf hin, dass Persönlichkeitseigenschaften von Kunden im Rahmen des Online-Shoppings eine Rolle spielen können (Jahng, Jain & Ramamurthy, 2002; Ranaweera, McDougall & Bansal, 2005; Schiffman, Sherman & Long, 2003). Weiterhin sind auch Einflüsse des sozialen Umfeldes auf das Kaufverhalten von Kunden in Betracht zu ziehen (vgl. Cole, O'Keefe & Siala, 2000; Foucault & Scheufele, 2002; Garbarino & Strahilevitz, 2004). Zudem bestehen Unterschiede zwischen Internetnutzern offenbar hinsichtlich ihrer kognitiven Verarbeitungsstrategien (Graff, 2005). Schließlich befassen sich diverse Autoren mit kulturellen Einflüssen im Kontext des Online-Shoppings (Cole, O'Keefe & Siala, 2000; Nielsen et al., 2001, S. 289 ff.; Stafford, Turan & Raisinghani, 2004). Der kulturelle Hintergrund von Nutzern nimmt Einfluss auf ihr Ästhetikempfinden (Becker, 2002), ihre Neigung anderen zu vertrauen (Mahmood, Bagchi & Ford, 2004) sowie auf die wahrgenommene Usability (Lightner, Yenisey, Ozok & Salvendy, 2002; Noiwan & Norcio, 2006). Im Bereich des traditionellen Handels wird vielfach versucht, dem Zusammenwirken mehrerer personenbezogener Merkmale dadurch Rechnung zu tragen, dass Typologien von Käufern aufgestellt werden (Westbrook & Black, 1985). Diverse Autoren widmen sich dieser Aufgabe für den Bereich des Online-Shoppings (Arnold & Reynolds, 2003; Assael, 2005; Jayawardhena, Wright & Masterson, 2003; Warden, Wu & Tsai, 2006).



In verschiedenen Studien wurde der Einfluss der Vorerfahrung von Nutzern bzw. Online-Kunden auf das Nutzungsverhalten bzw. die Präferenzen untersucht. Allgemein ist der Vorerfahrung von Nutzern bei der Betrachtung der Usability ein hohes Maß an Aufmerksamkeit zu schenken; es können unterschiedliche Dimensionen der Vorerfahrung wie a) das Wissen über Computer im Allgemeinen, b) die Expertise in der Benutzung der spezifischen Software sowie c) der Kenntnisstand über die jeweilige Domäne der Software betrachtet werden (Nielsen, 1993). Unerfahrene Nutzer werden bei einem hohen Komplexitätsgrad einer Software bzw. eines Online-Shops mehr Probleme in der Interaktion haben. Entsprechend läge die Vermutung nahe, dass Unerfahrene eine komplexe Site schlechter bewerten, als erfahrene Nutzer dies tun würden. Zwar bestätigt sich diese Annahme beispielsweise in Untersuchungen von Bruner und Kumar (2000) sowie Hampton-Sosa und Koufaris (2005). Andere Studien stützen jedoch eine entgegen gesetzte These: Nutzer, die viel Zeit im Internet verbringen, bewerten diverse Aspekte einer Internet Site offenbar kritischer als unerfahrene Nutzer. Sie erweisen sich somit als sensitiver bzgl. der Usability (Bucher & Jäckel, 2002, S. 42; Jacko, Sears & Borella, 2000). Als Erklärung ließe sich anführen, dass sich die Bewertungsmaßstäbe von Nutzern verändern, je größer ihre Erfahrung im Umgang mit dem Internet wird. Als ähnlich inkonsistent erweist sich die Befundlage zum Einfluss der Vorerfahrung von Kunden auf die Vertrauenswirkung von Online-Shops (Dinev & Hart, 2005; Gefen, Karahanna & Straub, 2003; Jarvenpaa & Tractinsky, 1999).

Meuter, Ostrom, Bitner und Roundtree (2003) schlussfolgern, dass demographische Merkmale zwar generell einen Einfluss auf die Nutzung von Self-Service-Technologien nehmen würden, die Art der Einflussnahme sich jedoch nicht spezifizieren lasse, da die Befundlage nicht konsistent sei (vgl. auch Ranaweera et al., 2005). Zusammenfassend erscheint es schwierig, klare Schlüsse aus den vorhandenen Befunden zum moderierenden Einfluss personenbezogener Merkmale zu ziehen, die im Rahmen der vorliegenden Arbeit von Relevanz wären.

## 2.8 Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle

Eine latente Variable ist ein hypothetisches, nicht *direkt* beobachtbares bzw. messbares Konstrukt (vgl. Homburg & Dobratz, 1998).<sup>11</sup> Um eine indirekte Erfassung einer latenten Variable vornehmen zu können, gilt es, ein Modell zur Messung zugrunde zu legen. Die Operationalisierung latenter Variablen kann mit Hilfe formativer oder reflektiver Messmodelle erfolgen (vgl. Bollen, 1989). Diese zunächst vor allem im Bereich der betriebswirtschaftlich orientierten Marketingforschung aufgegriffene und mittlerweile auch in psychologischen Arbeiten verwendete Dichotomie betrifft die Richtung des Kausalzusammenhangs zwischen der latenten Variable und den dazugehörigen manifesten Variablen (Indikatoren; in Fragebögen Items), die das Konstrukt unter Zugrundelegung eines Messmodells erfassbar machen. Im Falle eines reflektiven Messmodells wird von der Prämisse ausgegangen, die latente Variable verursache die Indikatoren (vgl. Nunnally & Bernstein, 1994). Im Gegensatz dazu gehen formative Messmodelle von der Voraussetzung aus, dass die Indikatoren Verursacher der latenten Variable sind (vgl. Diamantopoulos & Winklhofer, 2001). Der Unterschied zwischen beiden Modellansätzen liegt somit in einer gegensätzlichen Kausalitätsannahme zwischen latenter Variable und Indikatoren.

### 2.8.1 Charakteristika reflektiver Messmodelle

Bei einem reflektiven<sup>12</sup> Messmodell stellen die Indikatoren die Folge der latenten Variable dar (vgl. Fornell & Bookstein 1982; Edwards & Bagozzi, 2000; siehe Abbildung 14). Eine Veränderung der Konstruktausprägung geht daher mit einer entsprechenden Veränderung aller Indikatoreausprägungen einher. Gemäß der klassischen Testtheorie lassen sich theoretisch unendlich viele Indikatoren für ein reflektives Konstrukt heranziehen, die als Folgen der latenten Variablen anzusehen sind. Zur Erfassung des Konstrukts wird im Sinne des domain samplings eine Stichprobe an Indikatoren aus dem theoretischen Indikatoren-Universum gezogen (Cronbach, 1971). Es handelt sich bei reflektiven Indikatoren also um grundsätzlich austauschbare Variablen. Da jeder einzelne Indikator das gesamte Konstrukt *reflektiert*, sind hohe Korrelationen zwischen den Indikatoren zu erwarten. Bei einer perfekten Messung ohne jeglichen Messfehler  $\varepsilon$  auf Seiten der Indikatoren ergäbe sich sogar eine perfekte Korrelation.

---

<sup>11</sup> Die Darstellungen in diesem Abschnitt orientieren sich in großen Teilen an denen von Christophersen und Grape (2006).

<sup>12</sup> Für die englischsprachige Bezeichnung „reflective“ der Originalliteratur finden sich in deutschsprachigen Publikationen verschiedene Übersetzungen. Am häufigsten wird die Bezeichnung „reflektiv“ verwendet (z.B. Albers & Hildebrandt, 2006; Krafft, Götz & Liehr-Gobbers, 2005); gleichbedeutend verwendet werden die Begriffe „reflexiv“ (Scholderer & Balderjahn, 2006) oder „reflektierend“ (Temme, 2006).

Die Spezifikation eines reflektiven Messmodells erfolgt gemäß der Formel:

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

wobei:  $x_i$  = der i-te reflektive Indikator,

$\eta$  = die latente Variable,

$\lambda$  = die Faktorladung,

$\varepsilon$  = der Messfehler auf Indikatorebene.

Als Kriterium der Güte des Messmodells kann Cronbachs Alpha als Maß für die interne Konsistenz herangezogen werden (Finn & Kayande, 1997). Es lassen sich bei reflektiven Maßen traditionelle Methoden des Skalenbereinigungsprozesses angemessen anwenden (vgl. Churchill, 1979), d.h. zur Optimierung der Reliabilität können einzelne, besonders messfehlerbehaftete Indikatoren aus der Skala eliminiert werden, ohne dass dies Änderungen im Bedeutungsgehalt des Konstrukts oder eine Einschränkung der Validität zur Folge hat.

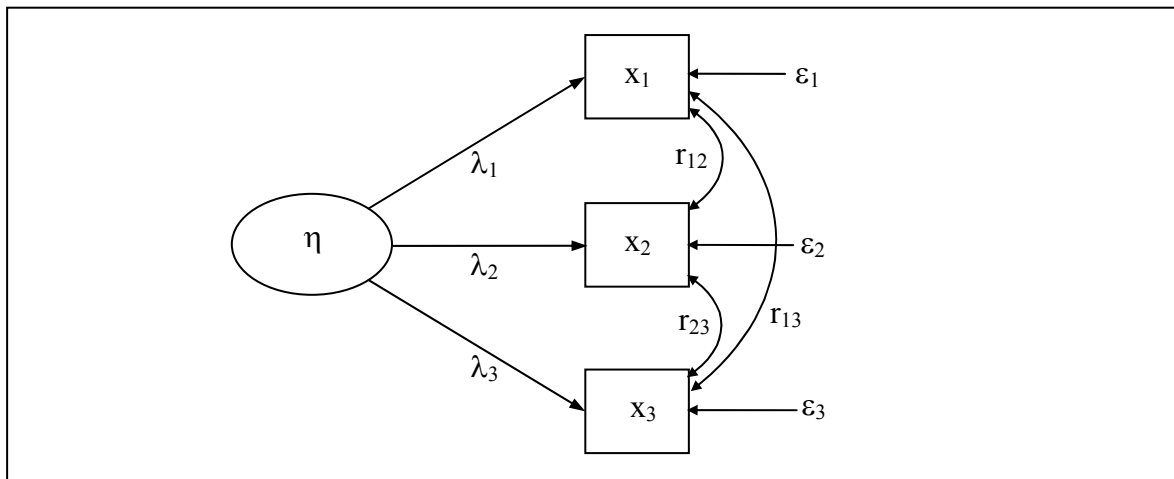


Abb. 2-14: Latentes Konstrukt mit reflektiven Indikatoren

Anmerkung: Latente Variable ( $\eta$ ), Faktorladung ( $\lambda$ ), reflektiver Indikator ( $x$ ), Messfehler auf Indikatorebene ( $\varepsilon$ ), Korrelation zwischen den Indikatoren ( $r$ ).

Wie in der Einleitung bereits angeführt, stellen Persönlichkeitseigenschaften wie etwa *Extraversion* typische Beispiele für reflektive Variablen aus dem Bereich der Psychologie dar (vgl. Eysenck & Eysenck, 1987). Die Variable Extraversion ist latent – man kann also nur aufgrund von Verhaltensbeobachtungen, der Verbalisierung von Kognitionen oder den Ergebnissen in einem Test darauf schließen, wie extravertiert ein Individuum ist. Da sich die Persönlichkeitseigenschaft in vielen Lebenssituationen gleichermaßen äußert, bestehen hohe Korrelationen zwischen ihren Indikatoren. So zeigen sich stark extravertierte Personen in der Regel in sehr selbstsicher, aktiv, gesprächig, energisch, heiter und optimistisch.

### 2.8.2 Charakteristika formativer Messmodelle

Formative Messmodelle zeichnen sich dadurch aus, dass die Indikatoren die Verursacher der latenten Variable sind (vgl. Bollen, 1989; siehe Abbildung 2-15). Formative Indikatoren nehmen nicht zwangsläufig in gleich starkem Maße Einfluss auf die latente Variable; die unterschiedliche Einflusstärke wird dadurch berücksichtigt, dass jede manifeste Variable ein individuelles Gewicht  $\gamma$  erhält (vgl. Williams, Edwards & Vandenberg, 2003). Dies eröffnet die Möglichkeit, Einflusstärken der einzelnen Facetten zu identifizieren. Entsprechend können die besonders gewichtigen Facetten in der Praxis adressiert werden, um eine Veränderung der Konstruktausprägung in eine gewünschte Richtung zu erreichen. Ein formatives Messmodell wird gemäß der folgenden Formel spezifiziert:

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

wobei:  $\eta$  = die latente Variable,

$\gamma_n$  = das n-te Gewicht,

$y_n$  = der formative n-te Indikator

$\zeta$  = der Messfehler auf Ebene der latenten Variable.

Hohe Interkorrelationen zwischen den manifesten Variablen sind im Falle formativer Konstrukte zwar möglich, jedoch nicht zwangsläufig zu erwarten (Diamantopoulos & Winklhofer, 2001). Je nach Konstrukt kann die Bewegung eines Indikators durch die Gegenbewegung eines anderen Indikators substituiert werden, so dass die Konstruktausprägung unverändert bleibt. Entsprechend verbietet sich eine Eliminierung eines einzelnen Indikators auf Grundlage eines Kriteriums wie der internen Konsistenz sowie die Anwendung der Faktorenanalyse. Ein formatives Messmodell erlaubt somit keine Beurteilung der Modellgüte, wie es bei reflektiven Messmodellen der Fall ist (Rossiter, 2002). Vielmehr gilt es, zur Messgüteeinschätzung eine Einschätzung zur Genauigkeit der Skalenentwicklung zu treffen (Rossiter, 2005). Die Entwicklung eines formativen Maßes kann nicht auf Basis eines Skalenbereinigungsprozesses im Sinne der klassischen Testtheorie erfolgen und erfordert im Verhältnis zu reflektiven Maßen in der Regel einen höheren Aufwand (Christophersen & Grape, 2006).

Ausgehend von einer präzisen Konstruktdefinition, die Aussagen zu den verschiedenen Facetten und deren Beziehung untereinander macht, muss dazu zunächst die Menge der formativen Indikatoren, die einen Einfluss auf die Ausprägung des Konstrukts ausüben, erschöpfend identifiziert werden. Eine Nichtberücksichtigung einzelner Indikatoren führt zu einer Änderung des konzeptionellen Inhalts der formativen Variable und somit zu einer nicht intendierten Verringerung der Messgüte (Eggert & Fassott, 2003). Da sich eine nachträgliche Elimination einzelner Indikatoren verbietet, muss entsprechend ein hohes Maß an Genauigkeit bei ihrer Operationalisierung angesetzt werden. Die Zuordnung des Messfehlers erfolgt im formativen Messmodell auf Konstruktebene und nicht wie bei reflektiven

Konstrukten auf Ebene der Einzelindikatoren. Eine ausführliche Darstellung zum Vorgehen der Entwicklung formativer Maße findet sich in Abschnitt 2.8.3.

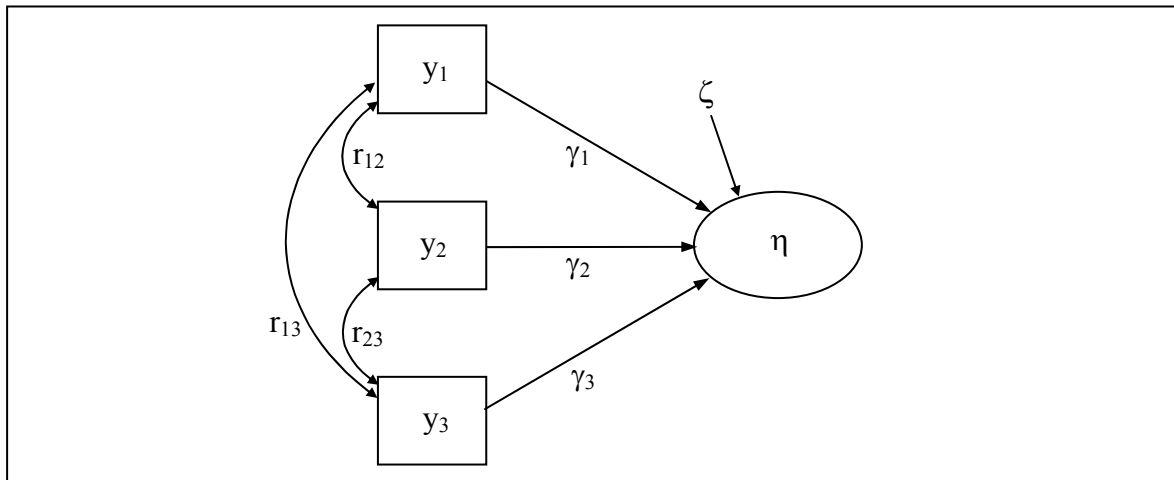


Abb. 2-15: Latentes Konstrukt mit formativen Indikatoren

Anmerkung: Latente Variable ( $\eta$ ), Gewicht ( $\gamma$ ), formativer Indikator ( $y$ ), Messfehler auf Ebene der latenten Variable ( $\zeta$ ), Korrelation zwischen den Indikatoren ( $r$ ).

Ein typisches Beispiel für ein formatives Konstrukt stellt die *Service-Qualität* dar (Rossiter, 2002). Sie setzt sich aus mehreren konkreten Leistungen seitens eines Anbieters zusammen, die summiert eine Service-Performance ergeben und je nach Kontext unterschiedlich ausfallen können. Im Bereich Online-Shopping müssen beispielsweise Facetten wie die Gewährleistung einer sicheren Transaktion, die Dauer der Lieferung, die Rückgaberegulungen sowie die Möglichkeiten zur persönlichen Kontaktaufnahme als formative Aspekte der *Service-Qualität* Berücksichtigung finden (Parasuraman et al., 2005; dort wird das Konstrukt allerdings als reflektiv fehlspezifiziert). Die manifesten Indikatoren zur Erfassung dieser Service-Facetten können, aber müssen nicht miteinander korrelieren. Zudem wäre es denkbar, dass Online-Kunden lange Lieferzeiten eines Shops tolerieren, wenn dieser besonders gute Garantiegewährleistungen bietet. Es ist also möglich, dass sich die Wirkungen mehrerer formativer Aspekte gegenseitig nivellieren.

### 2.8.3 Indexentwicklung zur Operationalisierung formativer Konstrukte

Der folgende Abschnitt widmet sich der Darstellung einer Prozedur zur Entwicklung formativer Maße. Im Wesentlichen orientiert sich das vorgestellte Vorgehen an Vorgaben von Diamantopoulos und Winklhofer (2001), weicht in einigen Punkten jedoch von dieser Vorlage ab. Vorschläge anderer Autoren werden integriert (Rossiter, 2002; Albers & Hildebrandt, 2006). Die Abbildung 2-16 stellt den Ablauf der Prozedur schematisch dar. Das Vorgehen zur Entwicklung *reflektiver* Skalen soll nicht dezidiert dargestellt werden, da es sich aus den bisherigen und folgenden Darstellungen indirekt erschließt und den allgemein verbreiteten Ansatz der Skalenentwicklung darstellt (z.B. Churchill, 1979; Homburg & Giering, 1996).

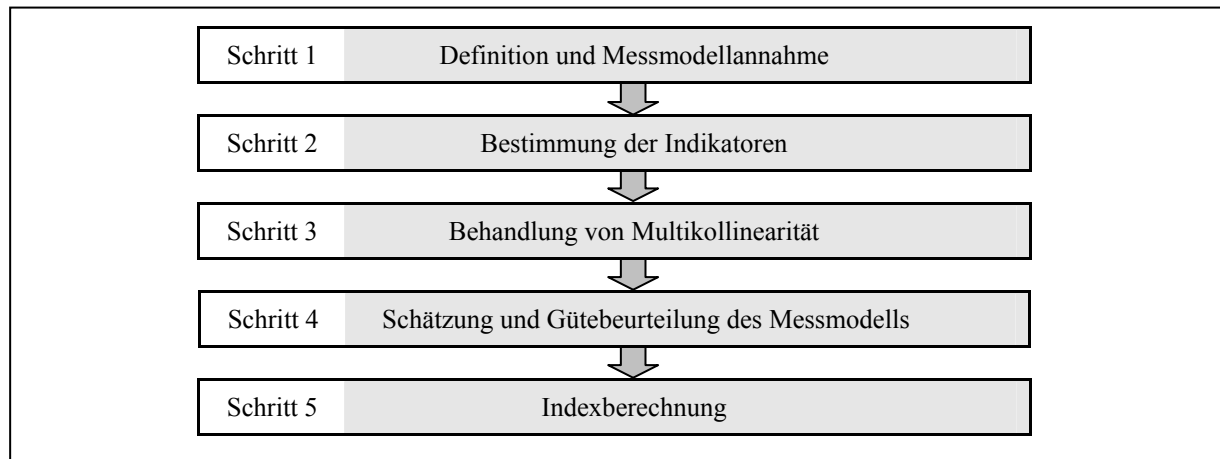


Abb. 2-16: Schematische Darstellung des Indexentwicklungsprozesses formativer Maße

Eine Einbindung formativer Indikatoren bzw. Konstrukte ist im Fall kovarianzbasierter Verfahren wie AMOS und LISREL (vgl. Arbuckle & Wothke, 1999 bzw. Jöreskog & Sörbom 1996) prinzipiell möglich und wird in bestimmten Zusammenhängen sogar empfohlen (vgl. Temme, 2006). Allerdings erweist es sich zumeist als eher aufwendig und in vielen Fällen als nicht möglich (vgl. Albers & Hildebrandt, 2006; Scholderer & Balderjahn, 2006). Aufgrund der besonderen Eignung des Strukturgleichungsverfahrens Partial Least Squares (PLS), formative latente Variablen in Strukturmodelle einzubinden, wird im Folgenden auf dieses Verfahren konkret Bezug genommen (vgl. Panten & Thies, 2006; Wold, 1966).

### 2.8.3.1 Schritt 1: Definition des Konstrukts und Festlegung des Messmodells

Sowohl im Falle formativer als auch reflektiver Maße gilt es in einem ersten Schritt einer Skalenentwicklung, eine Definition des zu erfassenden Konstrukts zugrunde zu legen. Dabei kann es hilfreich sein, zunächst eine Bestimmung von *Objekt*, *Attribut* und *Rater* vorzunehmen (vgl. Rossiter, 2002). Diese drei Aspekte betreffen die Fragen, (a) welcher Bewertungsgegenstand im Mittelpunkt des Interesses steht (Objekt), (b) welche Qualität bzw. Eigenschaft dieses Gegenstands bewertet werden soll (Attribut) sowie (c) welche Person bzw. welche Personengruppe die Beantwortung der Skala vornehmen wird (Rater). Ein Beispiel für eine Konstruktclassifikation (nicht *Konstruktdefinition*) nach diesem Schema wäre die Zufriedenheit aller T-Mobile-Kunden bzgl. der T-Mobile-Serviceleistungen. Eine Sichtung der relevanten Literatur zum jeweiligen Forschungsgegenstand sollte in jedem Fall erfolgen. Durch eine schriftliche Fixierung der Definition in ihrer ganzen Komplexität lassen sich unklare Aspekte leichter identifizieren; darüber hinaus wird eine wichtige Grundlage für Diskussionen bzw. zur Erstellung der Items geschaffen.

Im Fall formativer Maße besteht eine besondere Notwendigkeit, die Definition des Konstrukts präzise und in dem Sinne differenziert vorzunehmen, dass alle wesentlichen Facetten Berücksichtigung finden. Zur definatorischen Bestimmung können Voruntersuchungen sinnvoll sein (Fallstudien, Fokus-Gruppen, Interviews bzw. schriftliche Befragungen). Einzubinden sind Experten der Domäne oder repräsentative Rater. Insofern es möglich ist,

sämtliche Facetten des Konstrukts bereits in der Definition konkret zu benennen, sollten Aussagen zu den Beziehungen zwischen den Facetten getroffen werden.

Tab. 2-5: Entscheidungshilfen zur Frage, ob ein reflektives oder ein formatives Messmodell erforderlich ist (Christophersen & Grape, 2006, S. 123; Eberl, 2004; Jarvis, MacKenzie & Podsakoff, 2003)

<i>Kausalität zwischen latenter Variable (LV) und Indikatoren</i>	
⇒	Stellen die Indikatoren die Ursache der LV dar? → <i>ja: formativ</i>
⇒	Ergeben sich die Indikatoren als Folge (Konsequenz) aus der LV? → <i>ja: reflektiv</i>
⇒	Bestimmen die Indikatoren in ihrer Kombination den Inhalt der LV? → <i>ja: formativ</i>
⇒	Sind die Indikatoren als Manifestationen der LV anzusehen? → <i>ja: reflektiv</i>
⇒	Stehen die Indikatoren zeitlich gesehen vor der LV? → <i>ja: formativ</i>
⇒	Steht die LV zeitlich gesehen vor den Indikatoren? → <i>ja: reflektiv</i>
<i>Elimination und Austauschbarkeit der Indikatoren</i>	
⇒	Verändert sich bei Elimination eines Indikators die inhaltliche Aussage der, die hinter der LV steht? → <i>ja: formativ</i>
⇒	Lassen sich einzelne Indikatoren durch andere Indikatoren ersetzen? Sind die Indikatoren also austauschbar? → <i>ja: reflektiv</i>
⇒	Besitzen die Indikatoren inhaltlich denselben Kern, messen also gewissermaßen dasselbe? → <i>ja: reflektiv</i>
<i>Kovariation von Indikatoren und LV</i>	
⇒	Sind Konstellationen denkbar, bei denen die Indikatoren untereinander nicht hoch korrelieren? → <i>ja: formativ</i>
⇒	Sind in allen möglichen Untersuchungszusammenhängen hohe Korrelationen der Indikatoren untereinander zu erwarten? → <i>Keine Aussage möglich!</i>
⇒	Angenommen einer von mehreren (gleich kodierten) Indikatoren ändert plötzlich seine Ausprägung in eine bestimmte Richtung. Verändern sich alle übrigen Indikatoren notwendigerweise in gleicher Weise? → <i>ja: reflektiv</i>
⇒	Geht die Veränderung der LV notwendigerweise mit einer Veränderung aller Indikatoren einher? → <i>ja: reflektiv</i>
⇒	Geht die Veränderung eines Indikators notwendigerweise mit einer Veränderung der LV einher? → <i>ja: formativ</i>

Neben der Zugrundelegung einer Definition gilt es eine Aussage bzgl. der Frage zu treffen, ob die Erfassung des Konstrukts mit Hilfe eines formativen oder reflektiven Messmodells erfolgen muss. Oft stellt es sich als schwierig dar, diese Entscheidung spontan auf Grundlage der Theorie zu treffen. Liegen bereits potenzielle Indikatoren zur Erfassung des Konstrukts vor, so erleichtern möglicherweise die in Tabelle 2-5 zusammengefassten Fragen zu verschiedenen Charakteristika des Konstrukts und der Indikatoren diese Entscheidung (Christophersen & Grape, 2006; Eberl, 2004; Jarvis, MacKenzie & Podsakoff, 2003). Diese Entscheidungshilfen leiten sich aus den Eigenschaften beider Messmodellarten ab. Eine Bearbeitung der Fragen kann vom Forscher selbst oder von externen Experten der jeweiligen Domäne erfolgen.

Als Alternative zur theoriebasierten bzw. heuristischen Bestimmung der Kausalrichtung zwischen Indikatoren und dem latenten Konstrukt können in bestimmten Fällen Experimente durchgeführt werden (Edwards & Bagozzi, 2000, S. 159). Da dies jedoch gemeinhin mit einem sehr hohen Aufwand verbunden ist, erweist sich diese Herangehensweise als eher unattraktiv. Zudem bieten nur wenige Forschungsgegenstände überhaupt die Möglichkeit, auf experimentellem Wege Aussagen zur Messmodellannahme abzuleiten. Gelegentlich liegen bereits empirische Daten aus Voruntersuchungen vor, auf Grundlage derer Korrelationen zwischen den Indikatoren des Konstrukts bestimmt werden können. Zeigen sich niedrige Korrelationskoeffizienten, so kann dies einen Hinweis darauf darstellen, dass eine formative Beziehung zwischen der latenten Variable und den Indikatoren besteht. Weisen die Daten aus Voruntersuchungen hingegen hohe Korrelationen zwischen den reflektiven Indikatoren auf, so lässt sich keine Aussage zur Art des Messmodells treffen, da sowohl formative als reflektive Indikatoren hoch miteinander korrelieren können. Im Übrigen ließe sich ein solcher Ansatz der Skalenentwicklung dahingehend kritisieren, dass er empiriegetrieben sei und einer theoretischen Fundierung entbehrt.

#### 2.8.3.2 Schritt 2: Bestimmung der Indikatoren

Auf Basis der im ersten Schritt der Entwicklungsprozedur erarbeiteten Definition gilt es im Fall eines formativen Konstrukts, anschließend die Indikatoren zu bestimmen. Dabei ist es von Vorteil, wenn die Definition bereits konkrete Facetten benennt. Vielfach ist dies jedoch nicht der Fall, so dass eine theoriebasierte Ableitung der Indikatoren erforderlich wird. Wieder kann es sinnvoll sein, Experten zu befragen und Literatur in Bezug auf bereits vorhandene und bewährte Items zu sichten. Empirische Ergebnisse können dazu dienen, einzelne Facetten zu identifizieren, die dem Konstrukt zuzuordnen sind. Es erscheint problematisch eine Standardprozedur zur Ableitung der Indikatoren aufzustellen, da sich Konstrukte in ihrer Natur zu stark unterscheiden, als dass eine Verallgemeinerung möglich wäre. Rossiter (2002) schlägt vor, den Bewertungsgegenstand (Objekt) in einzelne *Bewertungsbereiche* einzuteilen und diese (quasi multiplikativ) mit den verschiedenen Attributen, die dem Konstrukt zugehörig sind, zu verknüpfen. Auch wenn dieser Ansatz eine gewisse Augenscheinvalidität besitzt, so zeigt sich jedoch in der praktischen Umsetzung, dass dieses Vorgehen im Fall vieler Konstrukte nicht möglich ist. Wichtig ist in jedem Fall, dass für alle relevanten Facetten des Konstrukts Indikatoren entwickelt werden, so dass eine valide Abbildung des Bedeutungsinhalts erreicht werden kann.

Vielfach erweisen sich Konstrukte als derart komplex und schwer zu greifen, dass sich der Anspruch, alle Facetten durch entsprechende Indikatoren abzubilden, als nicht umsetzbar darstellt. Folglich muss es Ziel sein, ein Set an formativen Indikatoren zusammenzustellen, das das Konstrukt in einem angemessenen Maße abdeckt. So kann es beispielsweise sinnvoll sein, sich auf eine begrenzte Anzahl wesentlicher Indikatoren zu beschränken, wenn durch viele weitere formative Indikatoren nur ein geringer Beitrag zur Erklärung des Konstrukts



erbracht wird. Hier gilt es folglich, den Aufwand, der mit der Erhebung vieler einzelner, marginal einflussreicher Indikatoren verbunden ist, gegenüber der durch ihre Aufnahme erreichten Steigerung der Messgüte und ihren zusätzlichen Informationsgehalt abzuwägen.

Bei der Formulierung der Items gilt es, die üblichen Maßnahmen zur Verringerung eines Method Bias anzuwenden. Diese betreffen einerseits die Verständlichkeit aus Sicht des Antwortenden. Unbekannte oder unklare Begriffe sollten vermieden werden oder – falls ihre Einbindung unvermeidbar erscheint – im Vorfeld für den Respondenten nachvollziehbar definiert werden. Mitunter kann die Einbindung kurzer Beispiele zur Erklärung vager Konzepte dienen. Allgemein sollten Items möglichst kurz und spezifisch ausfallen. Dazu gehört, dass jedes Einzelitem stets nur einen Beurteilungsaspekt adressiert und die Syntax nicht zu komplex ausfällt. Ein Item sollte sich stets klar auf den jeweiligen Bewertungsgegenstand bzw. einen konkreten Teilaspekt dieses Gegenstandes beziehen. Zudem ist darauf zu achten, dass die Formulierungen der Items untereinander keine starken Ähnlichkeiten aufweisen. Vielfach werden Items verwendet, die zwar sehr unterschiedliche formative Bewertungsfacetten ansprechen, aufgrund einer sehr ähnlichen Formulierung von Seiten der Respondenten jedoch als inhaltlich zusammengehörig wahrgenommen und entsprechend gleich beantwortet werden. Folglich ergeben sich hohe Korrelationen für Bewertungsaspekte, die eigentlich nicht mit einander einhergehen (z.B. Ahuja & Webster, 2001).

Zur Gewährleistung einer hohen Qualität der erstellten Items sollten Voruntersuchungen durchgeführt werden. Repräsentative Rater können im direkten Gespräch befragt werden, was sie unter den einzelnen Items verstehen. Im Idealfall sollte eine empirische Vorerhebung stattfinden. Liegen Daten vor, so wird zum einen die *Itemschwierigkeit* analysiert. Zeigt sich für ein Item eine zu hohe bzw. zu niedrige Schwierigkeit, kann die Formulierung entsprechend „ent- bzw. verschärft“ werden. Zum anderen betrachte man für jedes Item den *Anteil fehlender Werte*. Unter Umständen kann zur Sicherstellung der inhaltlichen Validität eine Befragung von Experten und/oder repräsentativen Ratern sinnvoll sein, die eine Zuordnung von Indikatoren zu Konstrukten vornehmen sollen (vgl. Anderson & Gerbing, 2001; Eggert & Fassott, 2003).

### 2.8.3.3 Schritt 3: Behandlung von Multikollinearität

Wie der Formel (2) zur Berechnung eines formativen Messmodells in Abschnitt 2.8.2 zu entnehmen ist, basiert es auf einer multiplen Regression. Entsprechend ist zu berücksichtigen, dass Multikollinearität zwischen den formativen Indikatoren ein potenzielles Problem bei der Schätzung der Parameter im Strukturgleichungsmodell darstellt.<sup>13</sup> Multikollinearität bezeichnet den Grad linearer Anhängigkeit zwischen den Indikatoren (Schneider, 2006). Je höher die Multikollinearität zwischen formativen Indikatoren ausfällt, umso ungenauer

---

<sup>13</sup> Da die Messung reflektiver Konstrukte stets auf einer einfachen Regression beruht, stellt sich die Frage nach einer Multikollinearität bei reflektiven Messmodellen nicht.

werden ihre Gewichte geschätzt. In Abschnitt 6.8 wird ausgeführt, anhand welcher Kennwerte eingeschätzt werden kann, inwiefern die Multikollinearität ein kritisches Maß erreicht.

Zur Bewältigung eines Multikollinearitätsproblems empfehlen Albers und Hildebrandt (2006) eine Zusammenfassung der untereinander multikollinearen Indikatoren zu einem Index. Dieser Index wird anschließend im formativen Messmodell als ein eigenständiger formativer Indikator eingebunden. Die Gewichtung der Indikatoren eines solchen Multikollinearitätsindizes richtet sich nach der Art der Beziehung der Indikatoren untereinander. Stehen sie in einem kompensatorischen Verhältnis, sollte der Mittelwert der normalisierten Indikatoren gebildet werden. Möglich ist auch, die Korrelationen zwischen den einzelnen Indikatoren und einer abhängigen Variable zur Gewichtung heranzuziehen, wenn sich die Güte des Messmodells dadurch verbessert. Liegt keine kompensatorische Beziehung zwischen den Indikatoren vor, so berechne man das geometrische Mittel der zuvor normalisierten Indikatoren, um den Index zu erstellen (siehe z.B. Vorgehen bei Homburg, Hoyer & Fassnacht, 2002, S. 94).

#### 2.8.3.4 Schritt 4: Schätzung und Gütebeurteilung des Messmodells

Im Fall von reflektiven Variablen ist es möglich, ein isoliertes Messmodell zu schätzen, bei dem lediglich die Effektindikatoren in Verbindung zur latenten Variable gesetzt werden (siehe Abbildung 2-14). Aufgrund einer statistischen Unteridentifizierung ist dies im Falle formativer Messmodelle nicht möglich (Bollen & Lenox, 1991). Stattdessen muss die formative Variable in ein größeres Strukturmodell eingebunden werden.

Dabei kann es sich um ein Zwei-Konstrukt-Modell handeln, bei dem die formative latente Variable als Prädiktor (exogene Ebene) zu einem Kriterium (endogene Ebene) in Beziehung gesetzt wird. Ein Beispiel für ein Zwei-Konstrukt-Modell wird in Abbildung 2-17 dargestellt. Weiterhin ist eine Einbindung der fraglichen formativen Variable in ein komplexes Modell möglich, also in eines mit mehr als nur zwei latenten Variablen. Dieses Vorgehen entspricht dem in Abschnitt 2.4.4 skizzierten Ansatz zur Bestimmung der nomologischen Validität.

Im Folgenden wird grob dargestellt, welche Kriterien im Falle einer Validierung mit Hilfe von PLS zu beachten sind. Auf Ebene des Strukturmodells sind zwei Parameter zu betrachten: Zum einen gilt es, die Signifikanz und Höhe des Pfadkoeffizienten  $\beta$  zu prüfen, der die Stärke des Zusammenhangs zwischen zwei Konstrukten beziffert. Die Signifikanz dieses Parameters ist anhand eines t-Wertes abzulesen, der sich mit Hilfe einer Resampling-Prozedur wie Bootstrapping bestimmen lässt (Hermann, Huber & Kressmann, 2006). Wie hoch der

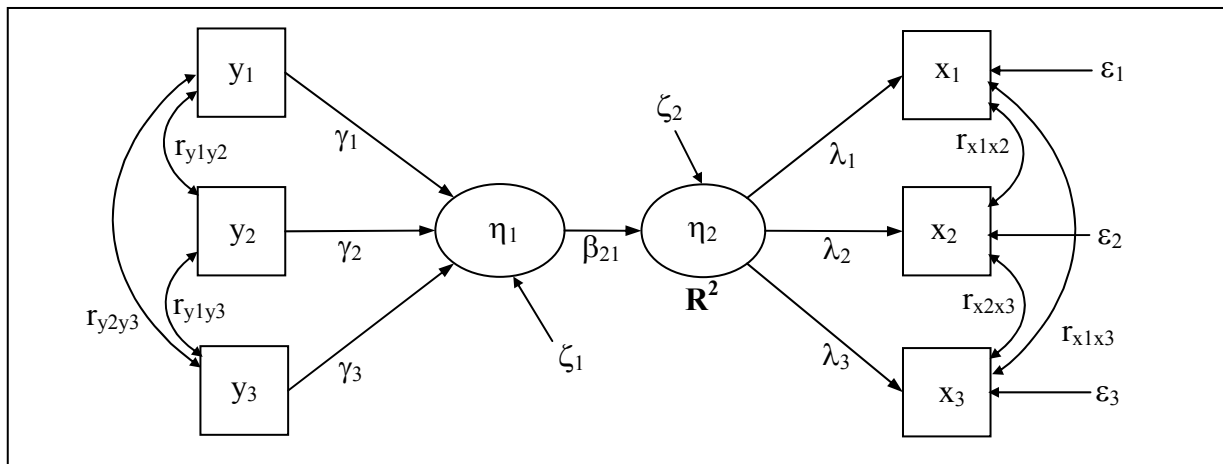


Abb. 2-17: Zwei-Konstrukt-Modell zur Bestimmung der Validität einer formativen LV (Diamantopoulos & Winklhofer, 2001, S. 273)

Anmerkung: Latente Variable ( $\eta$ ), Gewicht ( $\gamma$ ), Faktorladung ( $\lambda$ ), reflektiver Indikator ( $x$ ), formativer Indikator ( $y$ ), Messfehler auf Indikatorebene ( $\varepsilon$ ), Messfehler auf Ebene der latenten Variable ( $\zeta$ ), Korrelation zwischen den Indikatoren ( $r$ ), Pfadkoeffizient ( $\beta$ ), Determinationskoeffizient ( $R^2$ ).

Pfadkoeffizient ausfallen sollte, hängt von der Art des postulierten Zusammenhangs ab. Dabei sind vorhandene empirische Ergebnisse zur selben Konstruktbeziehung als Vergleichsdaten heranzuziehen. Zum anderen ist der Determinationskoeffizient  $R^2$  heranzuziehen, der sich für die nachgeordnete latente Kriteriumsvariable ergibt und den Anteil aufgeklärter Varianz dieser Variable benennt. Der Wert für diesen Parameter sollte bei mindestens  $R^2 \geq 0,3$  liegen (Hermann et al., 2006). Eine ausführliche Darstellung der Kriterien zur Gütebeurteilung von PLS-Modellen findet sich in Abschnitt 6.6.

Für den Fall, dass sich ein und dieselbe Variable sowohl formativ als auch reflektiv messen lässt, kann die reflektive Skala in einem Zwei-Konstrukt-Modell als abhängige Variable für das zu validierende formative Maß dienen. Gleiches gilt dann, wenn neben der formativen Erfassung einer Variable auch eine direkte Beobachtung desselben Konstrukts möglich ist. Eine ausreichende Inhaltsvalidität ist gegeben, wenn sich für den Pfadkoeffizienten ein Wert von  $\beta > 0,80$  ergibt (Chin, 1998, S. 324). Dieser Ansatz der Validierung besitzt den Nachteil, dass er sich als recht aufwändig erweist, da stets zwei unterschiedliche Operationalisierungen desselben Konstrukts erfasst werden müssen. Dies ist aus forschungsökonomischer Sicht wenig praktikabel.

#### 2.8.3.5 Schritt 5: Indexberechnung

Nach Bestimmung der Gewichte für die einzelnen Indikatoren der formativen latenten Variable lässt sich der Index berechnen, der die Latenten Variablen Scores angibt. Die Berechnung des Index erfolgt mit Hilfe der bereits angeführten Formel (2) zur Spezifikation formativer Messmodelle (Abschnitt 2.8.2). Danach werden alle Indikatorvariablen mit ihren jeweiligen Gewichten multipliziert und anschließend aufaddiert. Da in der Regel keine Werte für den Messfehler  $\zeta$  vorliegen, wird er nicht in die Berechnung einbezogen. Die Latenten Variablen Scores werden von der PLS-Software ausgegeben. Für den Fall, dass ein formativer

Index nicht mit Hilfe eines Strukturgleichungsmodells wie PLS erstellt werden soll, müssen die Gewichte der Einzelindikatoren auf Grundlage der Theorie bzw. bereits vorliegender empirischer Daten bestimmt werden.

#### **2.8.4 Fehlspezifizierte Messmodelle**

Wie bereits erörtert, fällt die Entscheidung, ob ein Konstrukt ein reflektives oder formatives Messmodell erfordert, vielfach recht schwer. Betrachtet man ausschließlich eindimensionale Messmodelle, so können zwei verschiedene Arten von Fehlspezifikationen vorliegen (vgl. Eberl, 2004). Zum einen lässt sich fälschlicherweise ein formatives Messmodell annehmen, wenn die latente Variable die Ursache der Indikatoren darstellt und somit eigentlich ein reflektive Messmodell angemessen wäre. Zum anderen kann ein in Wahrheit formativer Zusammenhang unberechtigtweise als reflektiv spezifiziert werden. Die Erfahrung zeigt, dass es insbesondere im Nachhinein oftmals schwer fällt, die Korrektheit einer Messmodellannahme zu beurteilen, zumal in vielen Beiträgen weder Konstruktdefinitionen noch Indikatoren genannt werden (vgl. Jarvis et al., 2003, S. 206). Um eine nachträgliche Prüfung vorzunehmen, lassen sich wiederum die in Tabelle 2-5 dargestellten Entscheidungskriterien heranziehen.

Einige Metastudien widmen sich der Frage nach der Häufigkeit des Auftretens fehlspezifizierter Messmodelle in wissenschaftlichen Publikationen (vgl. Eberl, 2004; Eggert & Fassott, 2003; Jarvis et al., 2003). Wird nur die Häufigkeit beider Messmodellarten betrachtet – unabhängig von der Frage nach Korrektheit einer Modellspezifikation –, so zeigt sich eine deutliche Dominanz des reflektiven Ansatzes. Der Anteil an Untersuchungen mit formativen Messmodellannahmen liegt bei weniger als 5%. Entsprechend werden kaum Beiträge identifiziert, in denen irrtümlicherweise formative Messmodelle zugrunde gelegt werden. Der Anteil an Studien, in denen formative Konstrukte fehlerhaft als reflektiv spezifiziert werden, fällt dagegen recht groß aus.

Ein Grund für das häufige Auftreten von Fehlspezifikationen kann in einer Unkenntnis bzgl. des Unterschiedes zwischen beiden Messmodellarten auf Seiten vieler Forscher gesehen werden. Auch kann eine mangelnde Akzeptanz auf Autorensseite eine Rolle spielen. Zudem sind die weit verbreiteten kovarianzbasierten Verfahren LISREL und AMOS eher ungeeignet, um formative Konstrukte zu berücksichtigen. Das diesbezüglich besser geeignete Verfahren PLS wird hingegen deutlich seltener angewandt.

#### **2.8.5 Folgen fehlerhafter Messmodellannahmen**

Angesichts der im Folgenden genannten Befunde muss es als eine wesentliche Gefahr angesehen werden, dass Fehlspezifikationen von Messmodellen zu falschen Interpretationen und Schlussfolgerungen bzgl. der postulierten Beziehungen zwischen betrachteten Konstrukten führen.

Es zeigt sich einerseits, dass die Ergebnisse von Strukturgleichungsmodellen stark verzerrt ausfallen können, wenn für eine in Wahrheit formative latente Variable fälschlicherweise ein reflektives Messmodell angenommen wird (Law & Wong, 1999; Jarvis et al., 2003). Law und Wong (1999) stellten zwei komplexe kovarianzbasierte Strukturgleichungsmodelle auf, die sich lediglich darin unterschieden, dass für eine exogene Variable in dem einen Fall ein reflektives Messmodell und in dem anderen Fall ein formatives Messmodell verwendet wurde. Die Pfadkoeffizienten zwischen den Kernkonstrukten in den beiden berechneten Modellen wichen signifikant voneinander ab. In einer ähnlichen Untersuchung von Jarvis et al. (2003) ergaben sich ebenfalls substantielle Verzerrungen der Strukturparameter aufgrund von Fehlspezifikationen. Die Ergebnisse der Autoren weisen auf eine erhebliche Unterschätzung von Koeffizienten hin, deren Pfade von der als reflektiv fehlspezifizierten latenten Variable wegführen. Dagegen fallen die Koeffizientenschätzungen von Pfaden, die auf die latente Variable wirken, also zu ihr hinführen, zu niedrig aus. Je niedriger die Korrelationen zwischen den Indikatoren der fehlspezifizierten latenten Variable sind, umso mehr werden die Parameter über- bzw. unterschätzt (MacKenzie, Podsakoff & Jarvis, 2005). Zudem zeigen die gängigen Fit-Indizes kovarianzbasierter Verfahren trotz Fehlspezifikation oft einen adäquaten Fit zwischen Modell und den zugrundeliegenden Daten an.

Aufgrund der in Abschnitt 2.8.4 berichteten großen Häufigkeit an fälschlicherweise reflektiv angenommenen Konstrukten, sieht man sich in der Forschungspraxis mit einem Problem konfrontiert: Im Fall formativer latenter Variablen kann nur selten auf vorhandene Maße aus früheren Studien zurückgegriffen werden, die dem formativen Charakter der Konstrukte tatsächlich gerecht werden. In dieser Situation erweist es sich jedoch nicht als sinnvoll oder gar wiedergutmachend, eine Skala im Strukturgleichungsmodell als formativ zu behandeln, obwohl diese fälschlicherweise auf Basis der traditionellen reflektiven Entwicklungsprozedur entstanden ist. Schließlich kann es sein, dass wesentliche formative Facetten des Konstrukts aufgrund niedriger Interkorrelationen der entsprechenden Indikatoren eliminiert wurden, um eine möglichst hohe interne Konsistenz der Skala zu erreichen (Rossiter, 2002). Eine solche Nichtberücksichtigung wichtiger inhaltlicher Aspekte eines Konstrukts wirkt sich wie bereits dargestellt negativ auf die inhaltliche Validität eines Maßes aus. Schließlich kann es sowohl in kovarianzbasierten (LISREL-) als auch in varianzbasierten (PLS-)Modellen zu starken Abweichungen der Parameterschätzungen auf Strukturmodellebene kommen, wenn aufgrund einer Fehlspezifikation ein reflektiver Skalenbereinigungsprozess stattfindet, obwohl eigentlich ein formatives Konstrukt vorliegt (Albers & Hildebrandt, 2006). Dieses unangemessene Vorgehen kann zu einem deutlich anderen Bedeutungsinhalt des Konstrukts sowie einer eingeschränkten Kriteriumsvalidität der Messung führen (Diamantopoulos & Siguaw, 2002).



### **3. Forschungsfragen**

Aufbauend auf den Inhalten des vorangehenden Kapitels werden im Folgenden zunächst die Ziele der Untersuchung abgeleitet, indem aufgezeigt wird, worin der konkrete Forschungsbedarf besteht (siehe Abschnitt 3.1). Anschließend werden die Ziele der Untersuchung präziser formuliert (Abschnitt 3.2), auf Basis derer eine Festlegung der inhaltlichen Hypothesen erfolgt (Abschnitt 3.3).

#### **3.1 Ableitung der Ziele**

Wie in Abschnitt 2.8 dargestellt wurde, können Messmodelle zur Erfassung latenter Variablen entweder formativer oder reflektiver Natur sein. Im Fall einer Fehlspezifikation des Messmodells sind negative Auswirkungen auf die Güte der Erfassung zu erwarten (siehe Abschnitt 2.10.6). Primäres Interesse der vorliegenden Arbeit war die (Weiter-) Entwicklung eines Fragebogens zur Erfassung des Konstruktes Usability von Online-Shops aus Kundensicht. Im Folgenden wird die für die Arbeit zentrale Annahme getroffen, dass es möglich ist, dieses Konstrukt sowohl unter Zugrundelegung eines reflektiven als auch eines formativen Messmodells zu erfassen (Abschnitt 3.1.1). Ausgehend von dieser Annahme wird dargelegt, dass die Messmodelle des ufosV1-Fragebogens und vieler weiterer Usability-Fragebögen als fehlspezifiziert anzusehen sind (Abschnitt 3.1.2).

##### **3.1.1 Formative und reflektive Erfassung der Usability aus Nutzersicht**

Wie bereits angedeutet wurde, ist es möglich, im Fall der fragebogenbasierten Erfassung der Usability aus Nutzersicht entweder ein formatives oder ein reflektives Messmodell zugrunde zu legen. Im Folgenden wird erläutert, worin die Unterschiede beider Messmodellannahmen im Fall der Usability zu sehen sind.

In der Literatur bestehen vereinzelt Zweifel, dass es Nutzern und auch Experten möglich sei, ein Globalurteil bzgl. der Usability abzugeben, da ein derart holistisches Urteil eine Zusammenfassung und Gewichtung der Wahrnehmungen erfordern würde (Hassenzahl, 2004a, S. 333). Allerdings finden sich in vorhandenen Usability-Fragebögen diverse Items, die den Gesamteindruck zur Usability auf Seiten des Nutzers abfragen (z.B. „I thought the system was easy to use“ aus dem SUS von Brooke, 1996). Berichte darüber, dass Nutzer bei Globalitems vermehrt die in vielen Usability-Fragebögen vorhandene Antwortoption „keine Angabe“ ankreuzen, sind nicht bekannt. Auch die Erfahrungen mit dem ufosV1-Fragebogen sprechen gegen die vage Annahme, Nutzer wären nicht in der Lage, Globalurteile zur Usability abzugeben. Neben Items zum Gesamteindruck finden sich weitere Items, die die Usability aus Nutzersicht reflektieren und sich auf Folgen einer Usabilityausprägung wie die Erlernbarkeit beziehen (z.B. „It was easy to learn to use this system“ aus dem PSSUQ von Lewis, 2002). Diese Formen der Usability-Abfrage erfordern ein reflektives Messmodell, da die Indikatoren eine Folge der latenten Variable Usability darstellen bzw. die Variable in aller

Gänze reflektieren. Die reflektive Usability aus Nutzersicht lässt sich somit als eine Bewertung des Nutzers zur Gebrauchstauglichkeit einer Benutzerschnittstelle auffassen, die ein Globalurteil bzw. eine Folge der Usability betrifft. Da auf Grundlage einer reflektiven Usability-Skala keine Aussagen zur Gebrauchstauglichkeit spezifischer Gestaltungsbereiche gezogen werden können, handelt es sich um ein summatives Evaluationsinstrument (siehe Abschnitt 2.3.1).

Neben Items zum Gesamteindruck bzw. den Folgen der Usability können sich Items eines Usability-Fragebogens auf einzelne Aspekte der Gebrauchstauglichkeit beziehen, z.B. „Die Verweise zu den Inhaltsrubriken (Produktkatalog) sind sinnvoll benannt“ oder „Auf den Seiten werden zu viele Verweise (Links) angeboten“ aus dem WOOS (Yom & Wilhelm, 2004). Hohe Korrelationen zwischen derart spezifischen Indikatoren sind zwar denkbar und werden auch häufig auftreten, sind aus theoretischer Sicht jedoch keinesfalls *zwangsläufig* zu erwarten (vgl. Hornbæk, 2006). Auch bezogen auf die beiden genannten Items aus dem WOOS wären niedrige Korrelationen durchaus denkbar. So können Verweise zu den Inhaltsrubriken einer Website sinnvoll benannt sein, während die Anzahl der Links aus Sicht des Nutzers zu hoch ausfällt. Bestimmte Gestaltungsaspekte einer Software können aus Nutzersicht also gebrauchstauglich gestaltet sein, wohingegen andere Elemente derselben Schnittstelle als schlecht bewertet werden. Es ist anzunehmen, dass sich der Gesamteindruck der Usability auf Seiten des Nutzers durch eine Summe der einzelnen spezifischen Bewertungsaspekte ergibt. Dabei werden die Einflussstärken der verschiedenen Aspekte auf den Gesamteindruck abhängig sein vom Nutzer, der Arbeitsaufgabe und der Nutzungsumwelt. Die Aspekte sind somit bzgl. ihres Einflusses zu gewichten. Diese Form der Usability-Abfrage erfordert ein formatives Messmodell, da die Indikatoren als Ursachen der Usability aus Nutzersicht anzusehen sind. Es gilt, alle wesentlichen Facetten der Gebrauchstauglichkeit zu berücksichtigen, um eine angemessene formative Abbildung des Konstruktes zu erreichen (vgl. Diamantopoulos & Winklhofer, 2001). Dabei ist zu berücksichtigen, dass die Usability ein sehr komplexes Konstrukt darstellt. Es besteht a) eine starke Kontextabhängigkeit (siehe Abschnitt 2.2.1.1), und b) handelt es sich bei vielen Benutzerschnittstellen um sehr komplexe Bewertungsgegenstände. Entsprechend sind häufig sehr viele verschiedene formative Aspekte der Usability denkbar. Diese Aussage wird durch die sehr große Anzahl verschiedener Usability-Guidelines untermauert (Nielsen, 2005a).

Es kann also zwischen zwei gleichermaßen plausiblen Messmodellannahmen unterschieden werden: einem formativen und einem reflektiven Ansatz der fragebogenbasierten Usability-Erfassung. Wie im Folgeabschnitt erläutert wird, ist diesem Umstand in bestehenden Usability-Fragebögen bislang nicht Rechnung getragen worden, was Einschränkungen der Messgüte dieser Instrumente zur Folge haben könnte.



### 3.1.2 Fehlspezifikationen bestehender Usability-Fragebögen

Bei keinem dem Autor bekannten Usability-Fragebogen wird auf die Unterscheidung zwischen formativen und reflektiven Messmodellen bzw. Indikatoren Bezug genommen. Entsprechend ist zu befürchten, dass es sich in vielen Fällen um fehlspezifizierte Skalen handelt. Exemplarisch wird das Vorliegen einer Fehlspezifikation für den ufosV1-Fragebogen von Konradt et al. (2003) erläutert. In diesem Instrument finden sich einerseits Items, die als formative Indikatoren der Usability aufzufassen sind, wie z.B. die beiden Items „Ich kann gut auf die Suchfunktionen des Shops zugreifen“ sowie „Der Warenkorb ist übersichtlich“. Als Begründung für den formativen Charakter dieser Items ist anzuführen, dass sich für einen konkreten Online-Shop nicht zwangsläufig hohe Korrelationen zwischen beiden zeigen müssen, aber können. So ist sicherlich denkbar, dass der Warenkorb eines Shops sehr übersichtlich ist, die Suchfunktionen jedoch aus Sicht des Nutzers kaum aufzufinden sind. In einem solchen Fall werden die beiden Items nur geringfügig miteinander korrelieren. Neben diversen formativen enthält der ufosV1-Fragebogen auch reflektive Items, wie etwa „Der Einkauf kann zügig erledigt werden“. Dieser Indikator kann als eine Folge der Usability-Ausprägung angesehen werden bzw. spiegelt die Gebrauchstauglichkeit insgesamt wider. Es handelt sich somit um ein Globalurteil, das vom Nutzer abgefragt wird.

Obwohl sich sowohl formative und reflektive Indikatoren im ufosV1 finden, kam bei der Entwicklung der Skalen die Faktorenanalyse zum Einsatz – wie auch bei nahezu allen anderen Usability-Fragebögen (siehe Abschnitt 2.4.1). Dabei fand eine Skalenbereinigungsprozedur statt, in deren Rahmen diverse Items eliminiert wurden. Die Anwendung dieses an der klassischen Testtheorie orientierten Skalenentwicklungsansatzes ist im Fall formativer Indikatoren jedoch nicht gerechtfertigt (siehe Abschnitt 2.8.2). Es wurden auch formative Items eliminiert, die für die nutzerseitig wahrgenommene Usability von Online-Shops von hoher Relevanz sind, so etwa das Item „Die Texte sind nur schwer lesbar“. Aus inhaltlicher Sicht ist der Ausschluss dieses Usability-Items nicht zu rechtfertigen; gleiches gilt für einige weitere Items aus dem ursprünglichen ufosV1-Itempool. Folglich dürfte die Inhaltsvalidität des ufosV1-Fragebogens angesichts der Nicht-Berücksichtigung relevanter Facetten eingeschränkt sein.

Es wurde bereits angemerkt, dass nicht nur der ufosV1-Fragebogen als fehlspezifiziert zu kritisieren ist. Nach Auffassung des Autors der vorliegenden Arbeit liegt nahezu allen vorhandenen Usability-Fragebögen ein fehlspezifiziertes Messmodell zugrunde, wobei zwei unterschiedliche Arten der Fehlspezifikation unterschieden werden können. Die eine Art besteht darin, dass sich innerhalb derselben Skala sowohl formative als auch reflektive Indikatoren finden (z.B. WOOS von Yom & Wilhelm, 2004). Die andere Art der Fehlspezifikation ist darin zu sehen, dass ausschließlich formative Usability-Indikatoren verwendet werden, als Gütekriterium jedoch das Maß Cronbachs Alpha herangezogen wird (PUTQ von Lin et al., 1997).

Die einzigen dem Autor bekannten Skalen, deren Messmodelle nicht als fehlspezifiziert gelten müssen, sind die erste Subskala des QUIS (Chin et al., 1988)<sup>14</sup>, der SUS (Brooke, 1996) sowie der AttrackDiff2 (Hassenzahl et al., 2003) – wobei bzgl. der Items des AttrackDiff2 eine gewisse Unklarheit besteht, inwiefern es sich um formative oder reflektive Indikatoren handelt. Um die angeführte Kritik an bekannten Usability-Fragebögen zu untermauern, müsste streng genommen eine ausführliche Analyse durch mehrere Rater sowie eine entsprechende inhaltliche Darstellung vorgenommen werden. Beides erfolgt an dieser Stelle nicht, da es nicht primäres Ziel der vorliegenden Arbeit war, die Fehlspezifikationen anderer Fragebögen zu belegen, sondern vielmehr den ufosV1 weiterzuentwickeln.

### **3.2 Ziele der Untersuchung**

Das primäre Vorhaben der vorliegenden Arbeit ergibt sich angesichts der Fehlspezifikation des Messmodells, das dem ufosV1-Fragebogen zugrunde liegt. Ziel war es, das Instrument dahingehend weiterzuentwickeln, dass eine klare Unterscheidung zwischen einem formativen und einem reflektiven Messmodell der fragebogenbasierten Usability aus Nutzersicht vorgenommen wird. Einerseits sollte eine reflektive Skala erarbeitet werden (ufos Version 2 reflektiv, im Folgenden kurz: ufosV2r), die das Konstrukt in Form eines Globalurteils des Nutzers erfassbar macht. Andererseits war es Ziel, eine formative Skala zu entwickeln, die alle wesentlichen formativen Facetten berücksichtigt, um eine valide Abdeckung der latenten Variable zu erreichen (ufos Version 2 formativ: ufosV2f). Es wurde also angestrebt, die bestehende Fehlspezifikation des ufosV1-Fragebogens auszuräumen.

Im Fall reflektiver Maße besteht weitgehende Einigkeit, wie und auf Grundlage welcher Kriterien eine Skalenentwicklung erfolgen sollte. Beim Vorhaben, eine formative Skala zur Erfassung der fragebogenbasierten Usability von Online-Shops aus Nutzersicht ufosV2f zu entwickeln, kann die im Theorieteil vorgestellte Prozedur zur Neuentwicklung formativer Skalen herangezogen werden (siehe Abschnitt 2.8.3). Im vorliegenden Fall handelt es sich jedoch nicht um eine Neu-, sondern vielmehr um eine Weiterentwicklung eines Fragebogens; entsprechend war es sinnvoll, im Detail von der Prozedur abzuweichen. Nachdem die formative und die reflektive Skala des ufosV2 zusammengestellt wurden, sollte eine Datenerhebung erfolgen, auf Grundlage derer die Güte der Skalen beurteilt werden konnte. Die Güteanforderungen lassen sich in Form inhaltlicher Hypothesen formulieren, die dem folgenden Abschnitt zu entnehmen sind.

---

<sup>14</sup>Alle weiteren Subskalen des QUIS beinhalten nach Ansicht des Autors sowohl formative als auch reflektive Items und sind folglich ebenfalls als fehlspezifiziert anzusehen.

### 3.3 Inhaltliche Hypothesen

Die Entwicklung reflektiver Maße stellt aus psychometrischer Sicht kein Novum dar. Die Überprüfung wesentlicher Gütekriterien einer reflektiven Skala kann gemäß der klassischen Testtheorie erfolgen. Zu fordern ist somit einerseits, dass das Maß eine reliable Erfassung der latenten Variable ermöglicht. Wie in Abschnitt 2.4.3 erläutert, erscheint es angesichts der starken Kontextabhängigkeit des Konstruktes Usability nicht angemessen, sich auf das Kriterium der Retest-Reliabilität zu beziehen. Als Hinweis auf die Reliabilität lässt sich jedoch im Fall einer reflektiven Usability-Skala die interne Konsistenz in Form des Cronbachs Alpha-Koeffizienten heranziehen. Entsprechend ergibt sich die erste inhaltliche Hypothese der Untersuchung:

H1: Die reflektive Skala zur fragebogenbasierten Erfassung der Usability von Online-Shops aus Nutzersicht ufosV2r erweist sich als reliabel im Sinne einer hohen internen Konsistenz.

Sicherlich ist auch im Fall der formativen Usability-Skala ufosV2f zu fordern, dass sie eine reliable Messung ermöglicht. Wie jedoch in Abschnitt 2.8.2 dargestellt, kann eine Erfassung dieses Gütekriteriums bei einem formativen Konstrukt nicht auf Basis eines internen Konsistenz-Maßes erfolgen. Aufgrund der starken Kontextabhängigkeit der Usability erscheint eine Bestimmung der Retest-Reliabilität ebenfalls unangemessen (siehe Abschnitt 2.4.3). Entsprechend kann eine Überprüfung der Reliabilität für die formativen Skala ufosV2f nicht vorgenommen werden.

Da sich im Falle einer reflektiven Skala die klassische Testtheorie anwenden lässt, eröffnet sich die Möglichkeit, die faktorielle Validität der Skala ufosV2r zu untersuchen (vgl. Bühner, 2004, S. 32). Es kann folglich die Hypothese aufgestellt werden:

H2: Die reflektive Usability-Skala erweist sich gegenüber weiteren reflektiven Skalen, die andere kundennahe Erfolgsfaktoren erfassen, als faktoriell valide.

Wie in Abschnitt 3.1.1 angeführt, ist eine reflektive Usability-Skala als ein summatives Evaluationsinstrument aufzufassen, da eine Gesamtbeurteilung zur Gebrauchstauglichkeit einer Benutzerschnittstelle erhoben wird. Eine Bewertung einzelner Usability-Facetten kann auf Basis einer solchen Skala nicht erfolgen. Ein typisches Anwendungsziel eines summativen Evaluationsverfahrens liegt im Vergleich mehrerer Benutzerschnittstellen. Gebrauchstaugliche Systeme sollten im Verfahren besser abschneiden als gebrauchsuntaugliche. Es ist also zu fordern, dass die reflektive ufosV2r-Skala in der Lage ist, zwischen mehreren Online-Shops zu differenzieren (vgl. Abschnitt 2.4.3). Es ergibt sich somit die folgende inhaltliche Hypothese bzgl. der Konstruktvalidität der reflektiven Skala:

H3: Mit Hilfe der reflektiven Skala zur fragebogenbasierten Erfassung der Usability von Online-Shops aus Nutzersicht ufosV2r kann zwischen mehreren Online-Shops differenziert werden.

Da sowohl eine reflektive als auch eine formative Skala zur Erfassung der Usability von Online-Shops entwickelt werden soll, bietet es sich an, den in der Prozedur angeführten Validierungsansatz des Zwei-Konstrukt-Modells zu verfolgen, bei dem die formative latente Variable als Prädiktor für die reflektive dient (siehe Abschnitt 2.8.3.4). Danach besteht eine hohe Inhaltsvalidität des formativen Maßes, wenn sich ein starker Zusammenhang zwischen beiden latenten Variablen zeigt. In Abbildung 3-1 wird diese Annahme visualisiert. Die entsprechende inhaltliche Hypothese lautet:

H4: Zwischen der formativen ufosV2f und der reflektiven Skala ufosV2r zur fragebogenbasierten Erfassung der Usability von Online-Shops aus Nutzersicht zeigt sich in einem Zwei-Konstrukt-Modell ein signifikanter positiver Zusammenhang.

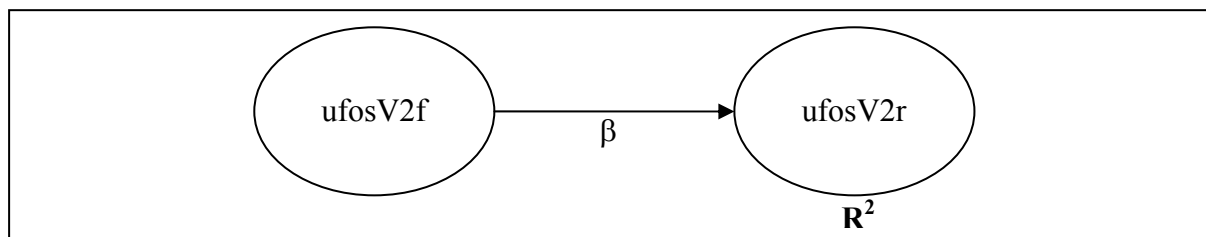


Abb. 3-1: Zwei-Konstrukt-Modell zur Validierung der formativen Skala ufosV2f anhand des Kriteriums der reflektiven Usability-Skala ufosV2r

Anmerkung: Angesichts der Tatsache, dass die Indikatoren der Skalen zum Zeitpunkt der Hypothesenformulierung noch nicht vorliegen, wird auf eine Darstellung des Messmodells verzichtet (vgl. Abbildung 2-17). Pfadkoeffizient ( $\beta$ ), Determinationskoeffizient ( $R^2$ ).

Gemäß der dargestellten Prozedur zur Entwicklung formativer Maße kann neben dem Zwei-Konstrukt-Modell auch ein anderer Zugang zur Prüfung der Validität des formativen Maßes verfolgt werden. Dieser besteht darin, das formative Maß in einem „komplexen Modell“ in einen Zusammenhang zu anderen latenten Variablen zu stellen. Die verschiedenen im zweiten Kapitel dargestellten Zusammenhänge zwischen der Usability und weiteren Konstrukten bilden ein konzeptuelles Modell (siehe Abschnitt 2.6 und 2.7), das als ein komplexes Validierungsmodell herangezogen werden kann. Erweisen sich die in dem Modell angenommenen Zusammenhänge zwischen der Usability und den weiteren kundennahen Erfolgsfaktoren als statistisch signifikant, so spricht dies für die nomologische Validität der Usability-Erfassung (vgl. Krafft, Götz & Liehr-Gobbers, 2005, S. 82). Dient die Usability im Rahmen des Modells als ein signifikanter Prädiktor für die Kaufintention (Hypothese 5), so kann dies darüber hinaus als ein Hinweis auf die konkurrente Kriteriumsvalidität der Messung gewertet werden. Wie in Abschnitt 2.5.2 gezeigt wurde, wirkt sich die Usability auch auf das tatsächliche Kaufverhalten von Kunden und entsprechend auf den Umsatz eines Online-Shops aus. Daher ist zu fordern, dass sich für die ufosV2-Skalen nicht nur ein starker Zusammenhang zu einem Intentionsmaß ergibt, sondern auch eine Vorhersage der kundenseitigen Kaufentscheidung mit Hilfe beider Skalen möglich ist (Hypothese 6). Diese Annahme betrifft also die prädiktive Validität des ufosV2. Es sei an dieser Stelle ausdrücklich darauf hingewiesen, dass dem besagten Modell kein Theoriestatus zugewiesen wird.

Tab. 3-1: Inhaltliche Hypothesen des konzeptuellen Erfolgsmodells zur Skalengüte des ufosV2-Fragebogens (H5 bis H9)

Nr.	Annahme	Kapitel 2
H5	<p>Je gebrauchstauglicher ein Shop gestaltet ist, umso höher wird die Intention von Online-Kunden ausfallen, im Shop zu kaufen.</p> <p>H5a: Zwischen der reflektiven Usability-Skala ufosV2r und der Kaufintention von Kunden besteht ein signifikanter positiver Zusammenhang.</p> <p>H5b: Zwischen der formativen Usability-Skala ufosV2f und der Kaufintention von Kunden besteht ein signifikanter positiver Zusammenhang.</p>	2.5.2
H6	<p>H6a: Die reflektive Skala zur fragebogenbasierten Erfassung der Usability von Online-Shops ufosV2r besitzt in signifikantem Maße eine Vorhersagekraft für die Kaufentscheidung von Kunden.</p> <p>H6b: Die formative Skala zur fragebogenbasierten Erfassung der Usability von Online-Shops ufosV2f besitzt in signifikantem Maße eine Vorhersagekraft für die Kaufentscheidung von Kunden.</p>	2.5.2
H7	<p>Je gebrauchstauglicher ein Online-Shop aus Sicht von Kunden gestaltet ist, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.</p> <p>H7a: Zwischen der reflektiven Usability-Skala ufosV2r und dem Vertrauen in den Anbieter besteht ein signifikanter positiver Zusammenhang.</p> <p>H7b: Zwischen der formativen Usability-Skala ufosV2f und dem Vertrauen in den Anbieter besteht ein signifikanter positiver Zusammenhang.</p>	2.7.1
H8	<p>Je gebrauchstauglicher ein Online-Shop von Seiten der Kunden wahrgenommen wird, umso mehr Vergnügen bereitet ihnen der Einkauf in dem Shop.</p> <p>H8a: Zwischen der reflektiven Usability-Skala ufosV2r und dem Einkaufsvergnügen besteht ein signifikanter positiver Zusammenhang.</p> <p>H8b: Zwischen der formativen Usability-Skala ufosV2f und dem Einkaufsvergnügen besteht ein signifikanter positiver Zusammenhang.</p>	2.7.2
H9	<p>Je höher Nutzer die Ästhetik einer Software-Anwendung bewerten, als umso höher nehmen sie die Ausprägung der Usability der Benutzerschnittstelle wahr.</p> <p>H9a: Zwischen der Ästhetikbewertung von Online-Kunden und der reflektiven Usability-Skala ufosV2r besteht ein signifikanter positiver Zusammenhang.</p> <p>H9b: Zwischen der Ästhetikbewertung von Online-Kunden und der formativen Usability-Skala ufosV2f besteht ein signifikanter positiver Zusammenhang.</p>	2.7.3

Anmerkung: Der Spalte „Kapitel 2“ kann entnommen werden, in welchen Teil des zweiten Kapitels die Befunde dargestellt werden, die die jeweilige Annahme des konzeptuellen Modells untermauern.

In Tabelle 3-1 werden alle Annahmen des konzeptuellen Erfolgsmodells zusammengefasst, die das Konstrukt Usability betreffen. Dabei wird auf den jeweiligen Abschnitt im zweiten Kapitel verwiesen, in dem der theoretische und empirische Hintergrund der einzelnen Hypothesen dargestellt wurde. Die im Modell angenommenen Zusammenhänge zwischen der Usability und den übrigen Konstrukten sollten für beide ufosV2-Skalen erfüllt werden. Entsprechend sind für jede Konstruktbeziehung zwei inhaltliche Hypothesen zu stellen. Die Abbildung 3-2 stellt das konzeptuelle Modell unter Zuordnung der Hypothesen zu konkreten Pfaden dar.

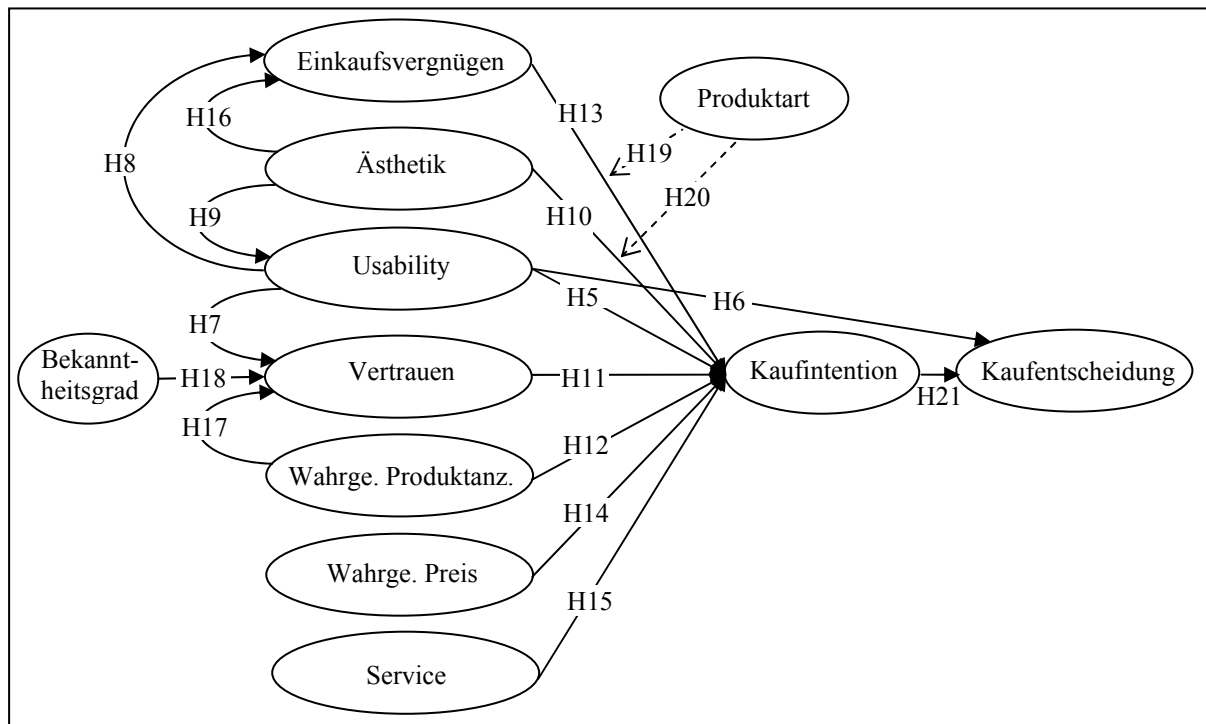


Abb. 3-2: Konzeptuelles Erfolgsmodell zur Validierung des zu entwickelnden ufosV2  
 Anmerkung: In Klammern wird jeweils die Nummer der inhaltlichen Hypothese angegeben.

Neben den Hypothesen, die die Usability betreffen, werden im konzeptuellen Erfolgsmodell weitere Annahmen berücksichtigt, die sich auf Zusammenhänge zwischen den weiteren Erfolgsgrößen beziehen. In Tabelle 3-2 werden diese Hypothesen H10 bis H21 unter Verweis auf den jeweiligen Abschnitt in Kapitel zwei zusammengefasst. Die Untersuchung dieser Annahmen ist im Vergleich zu den Hypothesen zur Skalengüte zwar von untergeordneter Relevanz, erfolgt jedoch, um die Gültigkeit des konzeptuellen Erfolgsmodells im Ganzen beurteilen zu können.

Tab. 3-2: Inhaltliche Hypothesen des konzeptuellen Erfolgsmodells zu den Zusammenhängen der weiteren Konstrukte (H10 bis H21)

<i>Nr.</i>	<i>Annahme</i>	<i>Kapitel 2</i>
H10	Je ästhetischer Kunden die Gestaltung eines Online-Shops empfinden, umso höher ist ihre Intention, einen Einkauf bei dem Anbieter zu tätigen.	2.6.1
H11	Je stärker Kunden einem Online-Anbieter vertrauen, umso höher ist ihre Intention, einen Einkauf im jeweiligen Shop zu tätigen.	2.6.2
H12	Je höher aus Sicht der Kunden die Anzahl der in einem Online-Shop angebotenen Produkte ausfällt, umso höher ist ihre Intention, einen Einkauf bei dem Anbieter zu tätigen.	2.6.3
H13	Je höher das Einkaufsvergnügen auf Seiten der Kunden ausfällt, umso höher ist ihre Intention, bei dem Anbieter zu kaufen.	2.6.4
H14	Ein subjektiv als niedrig wahrgenommener Preis wirkt sich positiv auf den Erfolg eines Online-Anbieters aus.	2.6.5
H15	Das Bereitstellen verschiedener Service-Leistungen seitens eines Online-Anbieters (Möglichkeiten zur Abwicklung der Bezahlung, Möglichkeiten der persönlichen Kontaktaufnahme, Lieferfristen der Produkte, Garantien, Rückgaberegelungen sowie die Kosten für Porto und Verpackung) wirkt sich positiv auf den Erfolg des Shops aus.	2.6.6
H16	Je ästhetischer ein Online-Shop aus Sicht der Kunden gestaltet ist, umso mehr Vergnügen empfinden sie bei der Nutzung des Shops.	2.7.4
H17	Je mehr Produkte aus Sicht der Online-Kunden in einem Shop erworben werden können, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.	2.7.5
H18	Je bekannter ein Shop aus Sicht von Online-Kunden ist, umso mehr Vertrauen bringen sie dem jeweiligen Anbieter entgegen.	2.7.6
H19	Die Wirkung des Einkaufsvergnügens auf die Kaufintention wird durch die Art des Produktes moderiert. Im Fall von Entertainmentprodukten bzw. Produkten, die primär Experience-Eigenschaften besitzen, erweist sich die Variable als einflussreich. Bei Verbrauchsgütern bzw. Produkten, die primär Sucheigenschaften aufweisen, besitzt das Einkaufsvergnügen keine besonders hohe Relevanz.	2.7.7.1
H20	Die Wirkung der Ästhetik auf die Online-Kaufintention wird durch die Art des Produktes moderiert. Im Fall von Entertainmentprodukten bzw. Produkten, die primär Experience-Eigenschaften besitzen, erweist sich die Variable als einflussreich. Bei Verbrauchsgütern bzw. Produkten, die primär Sucheigenschaften aufweisen, ist es weniger relevant, ob der Online-Shop als ästhetisch wahrgenommen wird oder nicht.	2.7.7.1
H21	Je höher die Kaufintention von Online-Kunden ausfällt, umso höher ist die Wahrscheinlichkeit, dass sie eine positive Kaufentscheidung treffen.	2.1.6.2

Anmerkung: Der Spalte „Kapitel 2“ kann entnommen werden, in welchen Teil des zweiten Kapitels die Befunde dargestellt werden, die die jeweilige Annahme des konzeptuellen Modells untermauern.





## **4. Entwicklung des Itempools der ufosV2-Skalen**

Das Vorgehen zur Entwicklung der formativen Skala ufosV2f folgte im Wesentlichen der in Abschnitt 2.8.3 vorgestellten Prozedur zur Konstruktion formativer Maße. Da es sich jedoch um eine *Weiterentwicklung* und nicht eine *Neuentwicklung* von Skalen handelte, bestanden im Detail Abweichungen. Im ersten Schritt der genannten Prozedur gilt es, eine Definition des jeweiligen Konstrukts sowie eine Festlegung des Messmodells vorzunehmen. Wie in Abschnitt 4.1 erläutert wird, wurde dies im Rahmen der vorliegenden Arbeit bereits geleistet. Der wesentliche Schwerpunkt dieses Kapitels liegt in der Festlegung der Items für die formative Skala ufosV2f und die reflektive Skala ufosV2r zur Erfassung der Usability von Online-Shops aus Nutzersicht. In Abschnitt 4.2 wird dargestellt, welche verschiedenen Schritte durchlaufen wurden, um die Items beider Skalen zu bestimmen.

### **4.1 Definition des Konstrukts und Festlegung des Messmodells**

Der in Abschnitt 2.8.3 vorgestellten Prozedur folgend, muss im ersten Schritt der Skalenentwicklung eine angemessene Definition des Konstrukts zugrunde gelegt werden. Eine ausführliche Definition des Konstrukts Usability wurde bereits in Abschnitt 2.2.1 durch die Darstellung der Teile 10 und 11 der DIN EN ISO 9241 angeführt. Im vorliegenden Fall sollte sowohl eine formative als auch eine reflektive Skala entwickelt werden, die sich auf die *Usability aus Nutzersicht von Online-Shops* beziehen. In Abschnitt 3.1.1 wurde diskutiert, worin die Unterschiede zwischen einer reflektiven und einer formativen Usability-Skala gesehen werden. Die reflektive Skala ufosV2r sollte das Gesamturteil zur Gebrauchstauglichkeit eines Online-Shops von Seiten der Nutzer abfragen bzw. die Folgen der Usability-Ausprägung betreffen. Die formative Skala ufosV2f sollte sich aus den verschiedenen Einzelaspekten der Usability von Online-Shops aus Nutzersicht zusammensetzen.

### **4.2 Bestimmung der ufosV2-Items**

Im zweiten Schritt der in Abschnitt 2.8.3 vorgestellten Skalenentwicklungsprozedur wurden die Items zur Erfassung der jeweiligen latenten Variable bestimmt. Im Fall der formativen Skala ufosV2f waren Vorüberlegungen anzustellen, welche formative Usability-Aspekte Berücksichtigung finden sollten (siehe dazu Abschnitt 4.2.1). Ausgangspunkt für die Bestimmung der Items beider neuer Skalen stellte der ufosV1-Fragebogen dar (siehe Abschnitt 2.4.1.6). In Abschnitt 4.2.2 wird dargestellt, anhand welcher Kriterien die Eignung der ufosV1-Items geprüft wurde bzw. welche Modifikationen an den Items vorgenommen wurden. Anschließend wurden neue Indikatoren auf Grundlage einer umfangreichen Literaturrecherche abgeleitet (Abschnitt 4.2.3). In Abschnitt 4.2.4 werden Erwägungen und Maßnahmen zu potenzielle Störeffekten geschildert. Zudem galt es, das Antwortformat der Items festzulegen (Abschnitt 4.2.5). Um eine angemessene Verständlichkeit der Items zu gewähr-

leisten, fand eine Voruntersuchung statt (siehe Abschnitt 4.2.6). In Abschnitt 4.2.7 werden schließlich die ufosV2-Items vorgestellt, die das Resultat dieses Arbeitsschrittes darstellen.

#### 4.2.1 Vorüberlegungen zur Itemauswahl der Skala ufosV2f

Um eine valide formative Skala zu entwickeln, gilt es, alle wesentlichen formativen Aspekte des jeweiligen Konstrukts in Form einzelner Indikatoren zu berücksichtigen (vgl. Diamantopoulos & Winklhofer, 2001). Im Fall der Skala ufosV2f waren entsprechend prinzipiell alle wichtigen Usability-Einzelaspekte von Online-Shops aus Nutzersicht heranzuziehen. Eine Berücksichtigung aller *möglichen* formativen Usability-Facetten von Online-Shops erscheint jedoch unangemessen und auch gar nicht möglich. Es können verschiedene Gründe für diese Auffassung angeführt werden:

- (1) Bei der Usability handelt es sich angesichts ihrer starken Kontextabhängigkeit um ein sehr komplexes Konstrukt. Zudem stellen Online-Shops vielfach sehr umfangreiche Bewertungsgegenstände dar. Angesichts dieser Komplexität erschien es nicht realisierbar, alle möglichen Einzelfacetten der Usability von Online-Shops zu antizipieren und in Items umzusetzen.
- (2) Ein Nutzer kann ein Item zu einem konkreten Usability-Aspekt nur dann bewerten, wenn er überhaupt Erfahrungen mit dem Online-Shop gesammelt hat, da ihm andernfalls die nötige Bewertungsgrundlage fehlt. Folglich sollten nur solche Items in der Skala berücksichtigt werden, die sich auf Grundlage zentraler Benutzungsvorgänge bewerten lassen. Entsprechend war z.B. auf ein Item zu verzichten, das sich auf den Vorgang des Anlegens eines Passwortes bezieht, da es voraussichtlich von vielen Nutzern nicht bewertet werden könnte.
- (3) Viele Usability-Aspekte sind als sehr spezifisch und somit eher unbedeutend einzustufen, so dass eine wesentliche Auswirkung auf die Gesamtbeurteilung bzw. weitere relevante Variablen wie etwa die Kaufintention unwahrscheinlich ist. Zwar können auch sehr spezifische Usability-Probleme im Einzelfall auf Seiten des Nutzers als sehr gewichtig wahrgenommen werden, gemittelt über alle Nutzer besitzen sie jedoch eine geringe Relevanz. Die Entscheidung, derartig unbedeutende Items nicht in die Skala aufzunehmen, steht im Einklang mit Empfehlungen von Rossiter (2002, S. 314 f.) zur Identifikation der formativen Indikatoren. Er argumentiert, eine Aufnahme wenig relevanter Items würde nicht zu einer deutlichen Verbesserung der Messgüte einer formativen Skala führen. Im Fall der Skala ufosV2f wäre eine unnötige Verlängerung die Folge.
- (4) Es konnten nur Aspekte bzw. Items aufgenommen werden, die sich auf alle Arten von Shops übertragen lassen. Angenommen, die Suche nach einem Buch mit Hilfe der ISBN-Nummer in einem Online-Shop zum Verkauf von Büchern ist nicht gut gestaltet. Eine Aufnahme dieses Aspektes in Form eines Indikators erschien nicht sinnvoll, da er bei Shops zum Verkauf von Veranstaltungstickets nicht bewertet werden kann.

(5) Schließlich war auch davon abzusehen, Items aufzunehmen, die das objektive Vorhandensein spezifischer Shop-Features betreffen, da sie lediglich eine ja / nein-Antwort erfordern. Ein solches Urteil kann leicht von Expertenseite getroffen werden. So wäre es sicher unsinnig, jeden Nutzer zu befragen, ob er nach erfolgter Bestellung eine Bestätigungsmail von Seiten des Anbieters erhalten habe. Es galt daher, ausschließlich Items einzubeziehen, die die Zufriedenheit mit einer konkreten Facette des Bewertungsgegenstands abfragen.

Die angeführten Erwägungen grenzten die Art und Zahl der Usability-Aspekte, die für eine Aufnahme in die ufosV2f-Skala in Frage kommen, ein. Ziel war es also, ein Set an Indikatoren zu identifizieren, das einen ausreichend großen Teil des Konstrukts Usability von Online-Shops abdeckt und damit eine valide Abbildung desselben ermöglicht.

#### 4.2.2 Eignung und Modifikation der Items des ufosV1

Als Grundlage für die Entwicklung der formativen Skala ufosV2f und der reflektiven Skala ufosV2r diente die in Abschnitt 2.4.1.6 vorgestellte Untersuchung zur Entwicklung von ufosV1 (Balazs, 2002; Christophersen, 2002; Konradt et al., 2003). Angesichts des systematischen Vorgehens bei der Zusammenstellung des Itempools von ufosV1 erschien es angemessen, die 80 Usability Items, die in der damaligen Datenerhebung zum Einsatz kamen, als Ausgangspunkt zu wählen. Der Tabelle B-1 im Anhang ist zu entnehmen, wie mit den ufosV1-Items verfahren wurde. Im Folgenden wird geschildert, welche Items von der weiteren Verwendung ausgeschlossen bzw. modifiziert wurden.

##### 4.2.2.1 Analyse der Anteile fehlender Werte

Gemäß des in Abschnitt 2.8.3.2 geschilderten Vorgehens zur Bestimmung der Indikatoren wurde im ersten Teilschritt der *Anteil fehlender Werte* der 80 ufosV1 Items betrachtet, um auf Grundlage dieses Kriteriums Rückschlüsse auf die Eignung zur weiteren Verwendung zu ziehen. Die Daten, die zu dieser Prüfung herangezogen wurden, entstammten der Untersuchung von Konradt et al. (2003).

Sieben der 80 Items wiesen in der Untersuchung von Konradt et al. (2003) mit mehr als 10% einen hohen Anteil fehlender Werte auf (u17, u19, u52, u53, u57, u75 und u76). Im Fall der Items u17 und u75 könnte eine mangelnde Verständlichkeit die Ursache des hohen Anteils fehlender Werte darstellen. Zudem betreffen die sieben identifizierten Items sehr spezifische Bewertungsaspekte. Entsprechend ist zu vermuten, dass einem Großteil der Teilnehmer eine Beurteilungsgrundlage fehlte, um eine Einschätzung vorzunehmen. Folglich wurden die Items u17, u19, u52, u53 und u75 aus dem Itempool ausgeschlossen. Die konkreten Erwägungen, die bei der Elimination der Indikatoren angestellt wurden, können Tabelle B-1 im Anhang B entnommen werden.

Im Fall der Items u57 und u76 könnte der große Anteil fehlender Werte darin begründet sein, dass das Versuchsszenario der Untersuchung von Konradt et al. (2003) keine entsprechende

Interaktion mit den Online-Shops erforderte. Folglich war für viele Teilnehmer keine Beantwortung dieser Items möglich. Da es sich jedoch um wichtige Aspekte der Usability von Online-Shops handelt, wurde von einer Elimination beider Items abgesehen. Als Ergebnis dieses Vorgehensschrittes reduzierte sich der ufosV1-Itempool auf 75 Items.

#### 4.2.2.2 Treffen von Messmodellannahmen

Wie bereits dargestellt, wurde in der Untersuchung von Konradt et al. (2003) versäumt, zwischen formativen und reflektiven Indikatoren zu unterscheiden. Unter den 80 Items des finalen ufosV1-Itempools finden sich entsprechend sowohl formative als reflektive Indikatoren. Es galt somit, für jedes der 75 verbleibenden Items zu entscheiden, ob es sich um ein reflektives oder ein formatives Item handelt.

Die Entscheidungen für das eine oder andere Messmodell wurden unter Zuhilfenahme der in Tabelle 2-6 vorgestellten Fragen und Kriterien getroffen. Der Autor der vorliegenden Arbeit sowie drei weitere Usability-Experten gaben diesbezüglich getrennt voneinander Urteile ab. Im Fall von elf Items ergaben sich zunächst Unklarheiten bei der Messmodellannahme. Bei vier Items konnte eine Festlegung auf Grundlage einer Diskussion der beteiligten Experten erreicht werden. Im Fall von sieben Items erwies es sich nicht als möglich, die Messmodellzugehörigkeit eindeutig zu benennen. Um eine Fehlspezifikation der Messmodelle für die Skalen ufosV2r und ufosV2f zu vermeiden, wurden sechs Indikatoren aus dem Itempool ausgeschlossen (u4, u5, u10, u11, u12, u15). Ein Item wurde modifiziert, um eine klare Messmodellannahme zu ermöglichen (u38). Als Ergebnis dieses Untersuchungsschrittes wurde für 65 der ufosV1 Items ein formatives Messmodell angenommen. Acht Items waren nach Auffassung der beteiligten Experten als reflektiv zu spezifizieren (siehe Tabelle B-1 im Anhang B).

#### 4.2.2.3 Optimierung der Itemschwierigkeiten

Im nächsten Teilschritt wurden die Schwierigkeiten der Items analysiert. Sieben der verbleibenden 75 ufosV1-Items wurden mit dem Ziel umformuliert, die Aussagen zu ent- bzw. verschärfen. In Folge dessen sollten die Schwierigkeiten in den mittleren Bereich rücken. Der Tabelle B-1 im Anhang B sind die Schwierigkeiten der 80 ufosV1-Items zu entnehmen. Eine Änderung der Formulierung erfolgte für die Items u26, u28, u37, u43, u47, u62 und u73.

#### 4.2.2.4 Umformulierungen von Items

In Abschnitt 2.8.3.2 werden verschiedene Aspekte angeführt, die es bei der Formulierung von Items zu berücksichtigen gilt. Es wurde eine erneute Sichtung der verbleibenden Items aus ufosV1 vorgenommen, um zu prüfen, inwiefern die Items die genannten Anforderungen erfüllen. Diese Analyse legte in 25 Fällen eine Optimierung der Formulierungen nahe. Oftmals wurden nur einzelne Worte geändert; bei anderen Items waren die Umformulierungen weitreichender. Teilweise erwies es sich zunächst als problematisch, die Umformulierungen vorzunehmen. In diesen Fällen wurden mehrere alternative Formulierungsvorschläge erstellt.

Ein Verbesserungsbedarf wurde für 15 Items darin gesehen, dass sie anderen Items in ihrer Formulierung zu stark ähnelten. Einige Untergruppen von ufosV1-Items beginnen jeweils mit „Ich kann...“ und „Ich erhalte nützliche Informationen...“. Es wurden die Items u13, u22, u23, u32, u33, u45, u57, u62, u64, u65, u66, u67, u68, u69 und u73 umformuliert, um derartige Ähnlichkeiten zu reduzieren.

Weiterhin wurde festgestellt, dass bei einigen ufosV1-Items kein konkreter Bezug zum Bewertungsgegenstand bestand. Im Fall der Items u8, u9, u18, u22, u25, u35, u36, u41, u71 und u72 wurde der konkrete Bezug auf Online-Shops durch eine Umformulierung hergestellt. Dabei wurden bewusst unterschiedliche Formulierungen verwendet („in diesem Shop“ und „im Shop“ bzw. „dieser Shop“ und „der Shop“), um eine zu starke Eintönigkeit des gesamten Fragebogens zu vermeiden.

#### 4.2.2.5 Prüfung auf inhaltliche Redundanz unter den Items

Die Daten der Untersuchung von Konradt et al. (2003) wurden im nächsten Teilschritt dahingehend geprüft, inwiefern Paare oder gar Gruppen formativer ufosV1-Items derart ähnlich beantwortet wurden, dass von einer inhaltlichen Redundanz ausgegangen werden musste. Als Kriterium zur Prüfung diente die bivariate Korrelation zwischen den jeweiligen Indikatoren, wobei mindestens ein  $r \geq 0,6$  vorliegen sollte, damit eine Elimination in Frage kam. Ein Ausschluss erfolgte im Fall der Items u21, u27, u31, u34, u48, u56 und u58. Der Tabelle B-1 im Anhang B ist zu entnehmen, welche inhaltlichen Redundanzen als Gründe für diese Entscheidungen gesehen wurden.

### 4.2.3 Gewährleistung der Vollständigkeit des Itempools

Um zu gewährleisten, dass alle wichtigen formativen Aspekte der Usability in der ufosV2f-Skala Berücksichtigung finden würden bzw. dass genügend reflektive Items für die ufosV2r-Skala zur Verfügung stünden, wurde eine Sichtung neuerer Quellen vorgenommen (Abschnitt 4.2.3.1). Auf Grundlage dessen wurden einige neue Items abgeleitet (4.2.3.2).

#### 4.2.3.1 Sichtung relevanter Literatur

Da seit der Zusammenstellung des Itempools von ufosV1 viele Beiträge erschienen sind, anhand derer formative und reflektive Items zur Usability von Online-Shops abgeleitet werden konnten, erfolgte eine erneute Sichtung relevanter Studien. Unter den diversen Veröffentlichungen wurden 35 identifiziert, die entweder potenziell verwertbare Items, relevante Guidelines oder deskriptive Darstellungen konkreter Usability-Aspekte enthielten. Es handelte sich um die Beiträge von Barnes und Vidgen (2002), Bucher und Jäckel (2002), Charitou et al. (2003), Choi, Lee, Lee und Subramani (2004), Danielson (2003), Devaraj et al. (2002), Gefen (2002), Gefen und Straub (2003), Gehrke und Turban (1999), Gladstone, Rundle und Alexander (2002), Goonetilleke, Duffy und Jacques (2001), Hassenzahl et al. (2003), Jee und Lee (2002), Kim und Lee (2002), Kim et al. (2002), Kohli, Devaraj und Mahmood (2004), Koufaris (2002), Lavie und Tractinsky (2004), Lin et al. (1997), Lowengart und Tractinsky (2001), McKnight, Choudhury und Kacmar (2002), Nielsen et al. (2001), Ozok und Salvendy (2001), Lewis (2002), Sing (2004), Späth (2002), Stefani und Xenos (2001), Tilson und Dong (1998), van der Heijden, Verhagen und Creemers (2001), van der Heijden und Verhagen (2002), van Dyk und Renaud (2002), Wolfenbarger und Gilly (2003), Yom (2003) sowie Zhang und von Dran (2001). Diese Studien entstammen recht unterschiedlichen Forschungsdisziplinen, wie der Software-Ergonomie, der betriebswirtschaftlich orientierten Erfolgsfaktorenforschung bzw. dem Marketing. Es wurden alle Items, Guidelines und deskriptiven Usability-Aspekte aus den Studien zusammengetragen. Anschließend erfolgte ein Abgleich mit dem überarbeiteten Itempool von ufosV1.

#### 4.2.3.2 Ableitung neuer Usability-Items

Es konnten sieben formative Facetten identifiziert werden, die nicht in dieser Ausdrücklichkeit im Itempool von ufosV1 enthalten waren. Sie entstammen ursprünglich den Veröffentlichungen von Nielsen et al. (2001) sowie Bucher und Jäckel (2002). In Tabelle B-2 im Anhang B werden die sieben Originalitems aufgelistet. Es wurden jeweils mehrere alternative Formulierungen erstellt. Die Codes der resultierenden neuen Items lauten uf50 bis uf56c. Ihr Wortlaut ist der Tabelle 4-1 in Abschnitt 4.2.7 zu entnehmen; in diesem Abschnitt werden alle Items des ufosV2 zusammengefasst vorgestellt. Zusätzlich wurde ein weiteres neues Item aufgenommen (Itemcode uf58c), dessen Formulierung sich an Vorgaben von Kim et al. (2002) orientiert. Es betrifft die Ladezeiten von Shopseiten. Dieser inhaltliche Aspekt wurde bereits durch das ufosV1-Item u80 (neuer Itemcode uf49c) abgedeckt. Eine Aufnahme des neuen Indikators erfolgte trotzdem, um im Fall dieses sehr wichtigen Usability-Aspektes zwischen zwei Items wählen zu können.

Ein weiteres neues formatives Item (Itemcode uf57c) wurde vom Autor mit dem Ziel entwickelt, durch diesen Indikator verschiedene Items zusammenzufassen, die die Verfügbarkeit einkaufsrelevanter Informationen betreffen (ufosV1-Items u64 bis u69: Infos zu den Lieferbedingungen, der Rückgabe von Produkten, den Zahlungsmöglichkeiten, den Lieferzeiten der einzelnen Produkte, den AGBs sowie zum Umgang mit den persönlichen Daten). Auch für dieses neue Item wurden zunächst mehrere alternative Formulierungen erstellt.

Im reflektiven Fall wurde auf Grundlage der Literaturstudie lediglich ein neues Item identifiziert. Es handelt sich dabei um eine Übersetzung eines Items aus dem PSSUQ-Fragebogen (Lewis, 2002). Das Item trug im neuen Fragebogen den Item-Code ur7 (siehe Tabelle 4-2).

#### 4.2.4 Berücksichtigung von Störeffekten durch die Itemformulierung

Eine methodische Maßnahme zur Reduktion verzerrender Antworttendenzen wird darin gesehen, dass in einen Fragebogen sowohl positiv als auch negativ formulierte Items aufgenommen werden (Nunnally, 1978). Im Fall negativ gepolter Items kann es sich jedoch als problematisch erweisen, dass die Beantwortung auf Seiten des Antwortenden ein zu hohes Maß an kognitiver Beteiligung erfordert (vgl. Motl & DiStefano, 2002), da bei einer Negativantwort eine doppelte Verneinung und damit eine gedankliche Umpolung des Items stattfinden muss. In Folge dessen können erhöhte Messfehler auftreten (vgl. Horan, DiStefano & Motl, 2003; Schriesheim & Eisenbach, 1995). Entsprechend sollten im neu entwickelten ufosV2-Fragebogen nur solche negativ formulierten Items verwendet werden, deren Beantwortung *keine* starke kognitive Beanspruchung erfordert. Eine negative Polung bot sich aus Sicht der Entwickler bei zwölf der 58 formativen Usability-Items an. Im Falle der reflektiven Skala wurde lediglich ein negativ formuliertes Item verwendet. Negativ formulierte Items werden im Folgenden dadurch gekennzeichnet, dass dem Itemcode ein „c“ angehängt wird, also z.B. uf5c für formative und ur2c für reflektive Indikatoren.

Bei einer Skalenentwicklung sollte sich der Entwickler über das potenzielle Problem einer Beantwortung im Sinne der *sozialen Erwünschtheit* bewusst sein. Falls diese mögliche Störquelle als relevant eingestuft wird, sind Maßnahmen zur Reduktion zu ergreifen (vgl. Nunnally & Bernstein, 1994). Im Fall der fragebogenbasierten Erfassung der Usability stellt es jedoch gemeinhin kein Problem dar, dass gemäß einer sozialen Erwünschtheit geantwortet wird, da keine Bewertung bzgl. der eigenen Person gefordert wird (vgl. Lewis, 2002).

#### 4.2.5 Antwortformat der Usability-Items

In Abschnitt 2.4.2.6 wurden einige Unterschiede zwischen den vorhandenen Usability-Fragebögen bzgl. des Antwortformates angeführt. Es stellte sich die Frage, welches Antwortformat im Fragebogen ufosV2 zur Anwendung kommen sollte. Einerseits galt es, die Anzahl der Antwortstufen festzulegen (Abschnitt 4.2.5.1), andererseits sollte eine Antwortoption bestehen, bei den Items „keine Angabe“ zu machen (Abschnitt 4.2.5.2).

#### 4.2.5.1 Verwendung einer siebenstufigen Likert-Skala

Zur Beantwortung der Items des ufosV1 wurde eine fünfstufige Likert-Skala genutzt (vgl. Konradt et al., 2003). Wie in Abschnitt 2.4.2.6 dargestellt, sollte sich die Anzahl der Antwortstufen stets nach der Differenzierungsfähigkeit der Antwortenden richten (vgl. Bühner, 2004). Gemäß den Erfahrungen der Untersuchung von Konradt et al. (2003) scheinen Online-Nutzer durchaus in der Lage zu sein, sehr differenzierte Usability-Bewertungen abzugeben. Es wurde in Abschnitt 2.4.2.6 zudem angeführt, dass in einer Reihe von Usability-Fragebögen siebenstufige Antwortskalen Verwendung finden. Weiterhin besagen die Ergebnisse von Lewis (1993), dass eine sieben-stufige Skala in einem Usability-Fragebogen zu besseren Ergebnissen führt als eine fünfstufige. Aufgrund dieser Erwägungen und um eventuellen Varianzeinschränkungen vorzubeugen, wurde die Anzahl der Stufen auf sieben festgelegt. Als Pole wurden wie im ufosV1 die Bezeichnungen „stimme gar nicht zu“ (Kodierung mit dem Wert 0) bis „stimme voll zu“ (Kodierung 6) genutzt.

#### 4.2.5.2 Zusätzliche Antwortoption „keine Angabe“

In der Untersuchung von Konradt et al. (2003) wurde den Teilnehmern neben der fünfstufigen Antwortskala die Option „kann ich nicht beantworten“ angeboten, um einen potenziellen *Uniformed Response Error* zu minimieren (siehe Abschnitt 2.4.2.6). Hawkins und Coney (1981) zufolge erweist sich dies als ein sinnvolles Mittel, um den Messfehler des Response Errors zu mindern. Anstelle der Formulierung „kann ich nicht beantworten“ wurde für den neuen ufosV2-Fragebogen die Bezeichnung „keine Angabe“ festgelegt. Letztere erscheint besser geeignet, da ein Ankreuzen der Option „kann ich nicht beantworten“ eventuell von den Teilnehmer als ein Eingeständnis mangelnder Beurteilungsgabe missverstanden werden könnte (also „kann nicht“ im Sinne von „ich bin nicht fähig dazu“).

### 4.2.6 Voruntersuchung zur Verständlichkeit

Um zu gewährleisten, dass die Items für ufosV2 verständlich formuliert waren, wurde eine Voruntersuchung durchgeführt, an der neun Usability-Experten beteiligt waren. Sie wurden instruiert, die Usability-Items bzgl. ihrer Verständlichkeit und allgemeinen Eignung zu prüfen sowie gegebenenfalls Kommentare oder Formulierungsvorschläge einzubringen. Im Fall der Items, für die mehrere alternative Formulierungen vorlagen, sollten die Experten zudem eine Entscheidung treffen, welche Alternative aus ihrer Sicht geeignet erschien. Weiterhin wurden die Teilnehmer der Vorstudie gebeten, für jedes Item zu prüfen, inwiefern die getroffenen Annahmen zum Messmodell (formativ vs. reflektiv) angemessen seien. Die Durchführung der Voruntersuchung erfolgte mit Hilfe einer Excel-Datei, die die Items enthielt und den Experten per E-mail zugesandt wurde. Sie wurden gebeten, die favorisierten Alternativformulierungen zu kennzeichnen, Kommentare einzubringen und die Datei nach Bearbeitung wieder an den Autoren zurückzusenden. Auf Grundlage der Vorbefragungsergebnisse wurden die Items für die formative (ufosV2f) und die reflektive Skala (ufosV2r) festgelegt.



#### 4.2.7 Verwendete formative und reflektive Items

Den Tabellen 4.1 und 4.2 kann entnommen werden, welche formativen bzw. reflektiven Indikatoren in der Hauptuntersuchung zur Anwendung kamen, um die *Usability von Online-Shops aus Nutzersicht* zu erfassen. Die Codes der formativen Items beginnen mit „uf“, die reflektiven werden mit „ur“ kodiert. In beiden Tabellen wird für jedes Item angegeben, auf welchem ufosV1 Item es beruht bzw. ob es sich um ein neues Item handelt. 16 der 80 ufosV1 Items wurden unverändert übernommen.

Tab. 4-1.1: Formative Indikatoren zur Erfassung der Usability

Neuer Code	Alter Code ufosV1	Formatives Item
uf1	u9	Beim Einkauf in diesem Shop kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.
uf2	u13	Es sind nützliche Informationen zur Benutzung des Shops abrufbar.
uf3	u14	Bei auftretenden Problemen bietet der Shop von sich aus nützliche Hilfe an.
uf4	u16	Es ist in diesem Shop leicht, versehentlich durchgeführte Aktionen zu korrigieren.
uf5c	u18	Ich muss mir viel merken, wenn ich in dem Shop einkaufe.
uf6	u20	Es ist leicht, sich innerhalb des Shops zu bewegen.
uf7	u22	Es ist einfach, zwischen den verschiedenen Seiten des Shops hin- und herzuspringen.
uf8	u23	Zentrale Seiten des Shops (z.B. die Startseite) sind jederzeit schnell zu erreichen.
uf9	u24	Ich kann jede Aktion wieder rückgängig machen.
uf10	u25	Mir ist beim Einkauf in dem Shop stets klar, welche Handlungen ich als nächstes durchführen kann.
uf11c	u26	Ich weiß manchmal nicht, wo ich mich innerhalb des Shops befinde.
uf12	u29	Dieser Shop ist gut strukturiert.
uf13c	u30	Es sind zu viele Schritte nötig, um zu einem Produkt zu gelangen.
uf14	u32	Ich finde mich auf den einzelnen Seiten des Shops gut zurecht.
uf15	u33	Ich finde in diesem Shop schnell die Informationen, die für meinen Einkauf wichtig sind.
uf16c	u35	Es werden in diesem Online-Shop häufig zu viele Informationen dargestellt.
uf17	u36	Ich kann selbst bestimmen, welche Informationen vom Shop angezeigt werden und welche nicht.
uf18c	u37	Die Texte sind manchmal nur schwer lesbar.
uf19c	u38	Das optische Design des Shops stört mich bei meinem Einkauf.
uf20	u41	Wenn ich in diesem Shop auf etwas klicke, passiert genau das, was ich erwarte.
uf21c	u43	Manchmal finde ich die im Shop verwendeten Texte inhaltlich unverständlich.
uf22	u44	Die Benennungen sind so gewählt, dass ich den Shop gut benutzen kann.
uf23	u45	Es ist schnell erkennbar, welche Produkte ich in dem Online-Shop kaufen kann und welche nicht.
uf24	u46	Es ist leicht, gesuchte Produkte zu finden.
uf25	u47	Ich kann sehr gut auf die Suchfunktionen des Shops zugreifen.
uf26	u49	Die Suchfunktionen ermöglichen mir, schnell bestimmte Produkte zu finden.
uf27	u51	Die Suchfunktionen liefern mir sinnvolle Ergebnisse.
uf28	u54	Die Produktlisten im Shop sind übersichtlich.
uf29	u55	Die Produkte sind in sinnvolle Kategorien eingeteilt.
uf30	u57	Ein Vergleich verschiedener Produkte ist leicht.
uf31	u59	Ich bin mit den Beschreibungen der Produkte zufrieden.
uf32	u60	Ich finde die verwendeten Bilder und Grafiken hilfreich.
uf33	u61	Der Shop empfiehlt mir interessante Produkte, nach denen ich ursprünglich nicht gesucht habe.
uf34	u62	Die Preise der einzelnen Produkte sind sofort ersichtlich.

Tab. 4-1.2: Formative Indikatoren zur Erfassung der Usability (Fortsetzung)

<i>Neuer Code</i>	<i>Alter Code ufosV1</i>	<i>Formatives Item</i>
uf35	u63	Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.
uf36	u64	Der Shop bietet mir nützliche Informationen zu den Lieferbedingungen.
uf37	u65	Die Informationen zur Rückgabe von Produkten sind verständlich.
uf38	u66	Ich werde gut über die verschiedenen Zahlungsmöglichkeiten informiert.
uf39c	u67	Man muss lange nach Angaben zu den Lieferzeiten der Produkte suchen.
uf40	u68	Es sind nützliche Informationen zu den allgemeinen Geschäftsbedingungen verfügbar.
uf41	u69	Ich werde ausreichend darüber informiert, wie mit meinen persönlichen Daten umgegangen wird.
uf42	u70	Der Warenkorb ist übersichtlich.
uf43	u71	Ich weiß in diesem Shop stets, welche Produkte sich gerade im Warenkorb befinden.
uf44	u72	Es ist in diesem Shop ganz einfach, Produkte in den Warenkorb zu legen.
uf45	u73	Der Inhalt des Warenkorbes kann ganz einfach verändert werden.
uf46	u74	Der Bestellvorgang kann problemlos durchlaufen werden.
uf47	u76	Die Eingabe meiner persönlichen Daten (z.B. Anschrift, Kreditkartennummer) während des Bestellvorgangs ist unkompliziert.
uf48	u78	Während des Bestellvorgangs weiß ich bei einigen persönlichen Angaben nicht, wofür sie benötigt werden.
uf49c	u80	Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.
uf50	neu	Der Ablauf des Bestellvorgangs gefällt mir.
uf51	neu	Die Auflistungen der Produkte lassen sich bequem nach meinen Bedürfnissen ordnen.
uf52	neu	Ich erhalte nützliche Informationen zum Unternehmen, das den Shop betreibt.
uf53	neu	Ich bin mit den Informationen zur Sicherheit des Datentransfers zufrieden.
uf54c	neu	Die Werbung auf den Seiten des Shops lenkt mich vom Einkaufen ab.
uf55	neu	Die Startseite dieses Shops stellt einen guten Ausgangspunkt für meinen Einkauf dar.
uf56c	neu	Bei der Benutzung des Shops treten störende technische Probleme auf.
uf57c	neu	Ich vermisse wichtige Informationen, die ich für meine Kaufentscheidung benötige.
uf58c	neu	Die Ladezeiten der Seiten sind bei diesem Online-Shop zu lang.

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

<i>Neuer Code</i>	<i>Alter Code ufosV1</i>	<i>Reflektives Item</i>
ur1	u1	Dieser Shop macht das Einkaufen im Netz leicht.
ur2c	u2	Es ist mir zu kompliziert, diesen Shop zu benutzen.
ur3	u3	Man kann sich schnell einen Überblick über den Shop verschaffen.
ur4	u6	Der Umgang mit dem Shop ist leicht zu erlernen.
ur5	u7	Dieser Shop regt mich an, ihn weiter zu erkunden.
ur6	u8	In diesem Shop kann ein Einkauf zügig erledigt werden.
ur7	neu	Insgesamt bin ich mit der Benutzerfreundlichkeit dieses Shops zufrieden.
ur8	u40	Der Shop bietet alle Möglichkeiten, die ich mir von einem Online-Shop wünsche.
ur9	u42	Ich kann den Shop so benutzen, wie ich es erwarte.

## 5. Untersuchungsplanung

In diesem Kapitel wird die Planung der quasi-experimentellen Untersuchung zur Überprüfung der Fragestellungen dargestellt. In den ersten Abschnitten dieses Kapitels (5.1 bis 5.4) wird erläutert, wie die unabhängige Variable (UV), die Erfolgsfaktoren, die Erfolgsgrößen Kaufintention und Kaufentscheidung (abhängige Variablen, AVn) sowie einige Variablen operationalisiert wurden. Die Entwicklung des verwendeten Versuchsszenarios ist Gegenstand von Abschnitt 5.5. Im Folgeabschnitt 5.6 wird darauf eingegangen, welche Formen von Reihenfolgeeffekten bei der Fragebogenbeantwortung zu beachten waren, um ihren Einfluss minimieren bzw. kontrollieren zu können. Letztlich stellt Abschnitt 5.7 den Versuchsplan vor.

### 5.1 Unabhängige Variable Online-Shop

Das wesentliche Ziel der vorliegenden Untersuchung lag in der Entwicklung von Skalen zur Erfassung der *Usability von Online-Shops aus Nutzersicht*. Entsprechend stellte sich die Frage nach der Auswahl der im Quasi-Experiment zugrunde gelegten Shops. Es war davon abzu-sehen, bei der Entwicklung eines Usability-Fragebogens nur eine einzelne Benutzerschnittstelle zur Validierung des Instruments heranzuziehen. Gemäß der Argumentation von Finn und Kayande (1997) ist ein solches Vorgehen insofern zu kritisieren, als dass die Ergebnisse von Validierungsuntersuchungen nur spezifisch für die eine gewählte Schnittstelle interpretiert werden können. Aussagen zur Skalengüte bei der Anwendung anhand eines weiteren Interfaces sind nicht möglich. Entsprechend wurden im Quasi-Experiment mehrere Shops berücksichtigt. Um den Einfluss der Produktart kontrollieren zu können, wurden die Shops nach der Art der angebotenen Produkte in Untergruppen eingeteilt. Die verschiedenen Online-Shops stellten die Ausprägungen der systematisch variierten UV dar.

#### 5.1.1 Kriterien zur Auswahl der Produktarten

Wie in Abschnitt 2.7.7.1 dargestellt wurde, kann die Ausprägung der Erfolgsfaktoren von Online-Shops für verschiedene Arten von Produkten unterschiedlich ausfallen. Folglich lag es nahe, die Produktart zur Untergruppenbildung heranzuziehen. Da ein branchenübergreifender Usability-Fragebogen konstruiert werden sollte, wurden Online-Shops zum Verkauf verschiedener Produktarten in die Untersuchung einbezogen. Es handelte sich um Shops zum Verkauf von Büchern, DVDs, CDs, Veranstaltungstickets und Druckerpatronen. Der Auswahl dieser Produktarten lagen fünf verschiedene Kriterien zugrunde. Die Kriterien (1) bis (3) wurden vor dem Hintergrund der Operationalisierung der Kaufentscheidung festgelegt (siehe Abschnitt 5.3.2):

*1. Kriterium:* Es sollten nur Shops mit Produkten herangezogen werden, die sich genau spezifizieren lassen, z.B. durch Angabe von Hersteller, Interpret und Titel.<sup>15</sup> Entsprechend

---

<sup>15</sup> Die Teilnehmer sollten sich im Vorfeld des Experimentes auf ein konkretes Produkt festlegen, das sie im Rahmen des Versuchs bestellen würden (siehe Abschnitt 5.3.2).

schieden Online-Shops zum Verkauf von Handys und Kleidungsstücken aus, da sich diese Produkte über verschiedene Shops hinweg oft deutlich voneinander unterscheiden. Der Verkauf von Handys ist zumeist an Verträge gekoppelt; eine Vergleichbarkeit zwischen diesen Angeboten ist selten gegeben. Im Fall von Kleidungsstücken werden nur selten Produkte desselben Herstellers in mehreren Shops angeboten, so dass Unterschiede in vielen Details wie der Farbe, dem Schnitt etc. bestehen.

2. *Kriterium*: Aufgrund des Umstandes, dass die Entlohnung der Teilnehmer im Versuch an die Kaufentscheidung gekoppelt wurde (Abschnitt 5.3.2), konnten nur solche Shops berücksichtigt werden, die Produkte mit einem Verkaufspreis unterhalb bzw. leicht oberhalb der Versuchspersonenvergütung (15 Euro) anbieten. Andernfalls wäre nicht gewährleistet gewesen, dass die Probanden den Rechnungsbetrag durch die ausgezahlte Versuchsentlohnung hätten begleichen können. Daher musste von einer Einbeziehung von Shops zum Verkauf von teuren Produkten wie Gebrauchtwagen, Urlaubsreisen, Unterhaltungselektronik etc. abgesehen werden.

3. *Kriterium*: Das dritte Kriterium betraf die Anzahl der in den Shops angebotenen Produkte. Da es den Teilnehmern des Quasi-Experimentes freigestellt wurde, welches Produkt sie im Rahmen des Versuchs erwerben (siehe Abschnitt 5.3.2), mussten die einbezogenen Shops über eine sehr breite Produktpalette verfügen. Folglich war beispielsweise von der Einbeziehung von Shops zum Verkauf von Computerspielen und Software abzusehen, da die meisten nur über ein begrenztes Sortiment verfügen.

4. *Kriterium*: Um genügend Teilnehmer für die Untersuchung akquirieren zu können, erschien es sinnvoll, nur solche Produktarten zu wählen, die unter Internetnutzern eher beliebt sind (siehe Abschnitt 2.1.3). Entsprechend schieden Online-Shops zum Verkauf von Musikinstrumenten, Filmplakaten, Comics, Parfums, Kosmetika, Arzneimitteln etc. aus. Dies hat einen Selektionseffekt zufolge, der im Rahmen der Diskussion berücksichtigt wird.

5. *Kriterium*: Zwei der in Abschnitt 3.3 angeführten Hypothesen beziehen sich einen moderierenden Einfluss der Produktart (H19 und H20, siehe auch Abschnitt 2.7.7.1). Um diese Hypothesen untersuchen zu können, galt es einerseits Verbrauchsgüter bzw. Produkten, die primär Sucheigenschaften aufweisen, und andererseits Entertainmentprodukten bzw. Produkten, die primär Experience-Eigenschaften besitzen, heranzuziehen.

Die Produktarten CDs, DVDs, Druckerpatronen, Bücher und Veranstaltungstickets erfüllten alle genannten Kriterien. Sie ließen sich gemäß dem Kriterium (1) konkret spezifizieren (CDs durch die Angabe von Titel und Interpret etc.). Auch das Kriterium (2) wurde erfüllt, da innerhalb der fünf Produktarten Produkte erworben werden konnten, deren Verkaufspreis nicht wesentlich höher lag als 15 Euro. Das Kriterium (3) war ebenfalls als erfüllt anzusehen, da sich für alle fünf Produktarten ausreichend viele Online-Shops im Internet finden ließen, die über eine große Produktpalette verfügten. Das Kriterium (4) wurde erfüllt, da die fünf gewählten Produktarten unter deutschen Internetnutzern beliebt waren (siehe Abschnitt 2.1.3).

Kriterium (5) folgend wurden mit Druckerpatronen Verbrauchsgüter berücksichtigt, die primär Sucheigenschaften aufweisen. Die übrigen vier Produktarten können als Entertainmentprodukte angesehen werden, die primär Experience-Eigenschaften besitzen.

### 5.1.2 Festlegung der Online-Shops

Im Anschluss an die Bestimmung der Produktarten wurde die Auswahl an Shops festgelegt. Pro Produktart wurden jeweils sieben Online-Shops herangezogen. Diese Anzahl ergab sich aufgrund versuchsplanerischer Erwägungen bzw. angesichts der Mindeststichprobengrößen der gewählten statistischen Analyseverfahren (siehe Kapitel 6). Die Auswahl der Shops erfolgte anhand dreier Kriterien:

1. *Kriterium*: Um prüfen zu können, wie gut die entwickelten Skalen des ufosV2 Fragebogens zwischen unterschiedlichen Online-Shops differenzieren, sollten sich die Shops bzgl. ihrer Usability unterscheiden. Da es eines hohen Aufwandes bedarf, umfassende Shop-Evaluationen durchzuführen, musste die Einschätzung zur Ausprägung der Benutzerfreundlichkeit heuristisch von Seiten des Autors vorgenommen werden. Hierzu wurden die Heuristiken von Nielsen (1994) herangezogen. Auf eine detaillierte Darstellung dieser Prüfung muss aus Platzgründen verzichtet werden.

2. *Kriterium*: Um eine möglichst große Unterschiedlichkeit in der Bewertung der Online-Shops zu erreichen (vgl. Finn & Kayande, 1997), sollten sich diese hinsichtlich ihrer *Bekanntheit* unterscheiden. Es ist davon auszugehen, dass bekanntere, umsatzstarke Anbieter in Bezug auf die Variablen *Usability*, *Vertrauen*, *Ästhetik* etc. ein besseres graphisches Benutzerinterface zur Verfügung stellen als unbekanntere Shops, da bekannteren Unternehmen mehr finanzielle Mittel zur angemessenen Gestaltung ihres Online-Auftritts zur Verfügung stehen. Entsprechend sollten sowohl Anbieter herangezogen werden, die unter Internetnutzern (z.B. Amazon.de, Alternate.de) und teilweise auch im traditionellen Handel (z.B. Karstadt.de, WOM.de) bekannt sind, als auch unbekanntere (z.B. Abella.de, PrintPower.net).

3. *Kriterium*: Wie bereits verdeutlicht, sollten die Shops über eine sehr breite Produktpalette verfügen, um die Operationalisierung der *Kaufentscheidung* zu gewährleisten (vgl. Kriterium 3 in Abschnitt 5.3.2).

Es erwies sich als problematisch, pro Produktart sieben unterschiedliche Shops zu identifizieren, die jeweils über eine ausreichend große Produktpalette verfügten. Entsprechend mussten einige Shops herangezogen werden, die jeweils von demselben Anbieter betrieben wurden. Diese Shops wiesen in zentralen Punkten große Ähnlichkeiten auf (z.B. bzgl. der Menüstruktur, des Bestellvorgangs etc.), unterschieden sich jedoch im graphischen Design. Folgenden Untergruppen an Shops wurden jeweils vom selben Unternehmen betrieben:

(1) drei Ticket-Shops „Provided by Get-go“: Get-Go.de und Eventim.de der CTS EVENTIM AG, Kartenkaufen.de der TKS Ticket-Service und Veranstaltungen GmbH,

- (2) vier Shops der JPC-Schallplatten Versandhandelsgesellschaft mbH: WOM.de und Abc.de (beide zur Produktart Musik-CDs), JPC.de (Bücher), Incinea.de (DVDs),
- (3) zwei Shops der Buecher.de GmbH & Co. KG: Booxtra.de (Bücher), Bücher.de (DVDs),
- (4) zwei Shops der buch.de AG: Buch.de (Produktart CDs), Bol.de (Produktart Bücher).

Tabelle 5-1 listet die 35 im Rahmen der Untersuchung berücksichtigten Shops geordnet nach Produktart auf.<sup>16</sup>

Tab. 5-1: Herangezogene Online-Shops geordnet nach Produktart

Nummer	Produktart	Code	Name des Shops	URL
01	Bücher	b1	Karstadt.de	www.karstadt.de
02		b2	JPC.de <sup>a</sup>	www.jpc.de
03		b3	Libri.de	www.libri.de
04		b4	Booxtra.de <sup>b</sup>	www.booxtra.de
05		b5	BOL.de <sup>c</sup>	www.bol.de
06		b6	Buchkatalog.de	www.buchkatalog.de
07		b7	Buch24.de	www.buch24.de
08	DVDs	d1	DVD-Laden-24.de	www.dvdladen-24.de
09		d2	Fröschl.de	www.froeschl.de
10		d3	inCinea.de <sup>a</sup>	www.incinea.de
11		d4	Bücher.de <sup>b</sup>	www.buecher.de
12		d5	Avides.com	www.vides.com
13		d6	Delta-Shopping.de	www.delta-shopping.de
14		d7	DVD-Galaxis.de	www.dvd-galaxis.de
15	CDs	m1	WOM.de <sup>a</sup>	www.wom.de
16		m2	CyberCD.de	www.cybercd.de
17		m3	AlphaMusic.de	www.alphamusic.de
18		m4	Abella.de	www.abella.de
19		m5	Buch.de <sup>c</sup>	www.buch.de
20		m6	Amazon.de	www.amazon.de
21		m7	abc-CD.de <sup>a</sup>	www.abc-cd.de
22	Druckerpatronen	p1	Druckerpatronen.de	www.druckerpatronen.de
23		p2	PrintPower.net	www.printpower.net
24		p3	Express-Toner.de	www.express-toner.de
25		p4	Alphaco.de	www.alphaco.de
26		p5	ACI-Druckerpatronen.de	www.aci-druckerpatronen.de
27		p6	Planet-Shop.de	www.planet-shop.de
28		p7	Alternate.de	www.alternate.de
29	Tickets	t1	Eventim.de <sup>d</sup>	www.eventim.de
30		t2	Getgo.de <sup>d</sup>	www.getgo.de
31		t3	Kartenkaufen.de	www.kartenkaufen.de
32		t4	Kartenhaus.de <sup>d</sup>	www.kartenhaus.de
33		t5	Karten-Online.de	www.karten-online.de
34		t6	Ticket-Online.de	www.ticketonline.de
35		t7	TicketCenter.de	www.ticketcenter.de

Anmerkung: <sup>a</sup> Shops der jpc-schallplatten Versandhandelsgesellschaft mbH, <sup>b</sup> Shops der buecher.de GmbH & Co. KG, <sup>c</sup> Shops der buch.de internetstores AG, <sup>d</sup> Shops „Provided by Get-go“.

<sup>16</sup> Es bestand keine Form der Zusammenarbeit mit den Betreibern dieser Online-Shops im Rahmen der Untersuchung – etwa durch ein Sponsoring. Die Shopbetreiber wurden nicht über die Studie informiert.

## 5.2 Operationalisierung der Variablen auf Prädiktorebene

Im Folgenden wird dargestellt, wie die verschiedenen Konstrukte auf Prädiktorebene im Rahmen der Untersuchung operationalisiert wurden. Die Originalitems, auf die die verwendeten Indikatoren zurückgehen, sind dem Anhang C zu entnehmen. Wie im Fall der Usability Items wurden alle negativ formulierten mit einem zusätzlichen „c“ im Itemcode gekennzeichnet (z.B. se2c; siehe Abschnitt 5.2.6). Wenn es nicht ausdrücklich anders erwähnt wird, kam bei allen Skalen dasselbe Antwortformat wie im Fall der Usability-Items zur Anwendung (siebenstufige Likert-Skala mit den Polen „stimme gar nicht zu“ und „stimme voll zu“ und zusätzlicher Antwortoption „keine Angabe“; siehe Abschnitt 4.2.5).

### 5.2.1 Reflektive Erfassung der Ästhetik

Zur Erfassung des Konstruktes Ästhetik (siehe Abschnitt 2.6.1) wurde auf die *Visual Appeal*-Skala von Mathwick et al. (2001) zurückgegriffen. Diese drei englischsprachige Items umfassende Skala weist nach Angaben ihrer Entwickler eine interne Konsistenz von  $\alpha=0,92$  auf (vgl. auch Hong, Thong & Tam, 2004; Lowengart & Tractinsky, 2001). Um die Anzahl der Items im Fragebogen nicht zu groß werden zu lassen, kamen in der Hauptuntersuchung lediglich zwei der Items zur Anwendung. Die beiden verwendeten Items lauteten:

*Item a1:* Ich finde es optisch ansprechend, wie dieser Online-Shop seine Produkte anbietet.

*Item a2:* Der Internetauftritt dieses Online-Shops ist ästhetisch gestaltet.

### 5.2.2 Reflektive Erfassung des Vertrauens

Zur Erfassung des reflektiv spezifizierten Konstruktes Vertrauen (siehe Abschnitt 2.6.2) wurde auf Items aus der Untersuchung von Konradt et al. (2003) zurückgegriffen, in der drei Aussage-Items mit Hilfe einer fünfstufigen Likert-Skala abgefragt wurden. Bei diesen Items handelte es sich um Übersetzungen englischsprachiger Items von Jarvenpaa et al. (2000), die wiederum auf einer Skala von Doney und Cannon (1997) basieren.

Bei Jarvenpaa et al. (2000) findet sich keine konkrete Angabe zur Skalenreliabilität. Es wird jedoch angegeben, dass die Reliabilität gut ausfällt. In der Untersuchung von Konradt et al. (2003) wurden mehrere Alternativübersetzungen der drei Originalitems angefertigt. Die Festlegung auf die deutsche Fassung erfolgte auf Grundlage einer Vorbefragung mit einer studentischen Stichprobe. In der Hauptuntersuchung ergab sich ein wenig zufrieden stellendes Cronbachs Alpha von  $\alpha=0,74$ . Es zeigte sich für eines der drei Items ein hoher Anteil fehlender Werte von 26,7%. Scheinbar ist die Formulierung dieses Items nicht optimal ausgefallen („Es ist dem Online-Shop wichtig, dass er Versprechen und Verpflichtungen einhält“). Durch eine Umformulierung wurde versucht, die Verständlichkeit zu erhöhen (siehe Wortlaut von Item t3 unten). Die Schwierigkeiten der drei Items erreichten in der Untersuchung von Konradt et al. (2003) sehr gute Werte zwischen  $p_m=0,51$  und  $p_m=0,53$ .

Um eine höhere interne Konsistenz der Skala zu erreichen, wurde zusätzlich zu den drei Items ein weiteres herangezogen, das eine Abwandlung eines Items von McKnight et al. (2002) darstellt und mit dem Item-Code t4 bezeichnet wurde. Die vier verwendeten Items lauteten:

Item t1: *Dieser Online-Shop ist vertrauenswürdig.*

Item t2: *Ich bin mir sicher, dass der Online-Shop seine Versprechen und Verpflichtungen einhalten wird.*

Item t3: *Ich vertraue darauf, dass dieser Shop meine Interessen berücksichtigt.*

Item t4: *Ich habe das Gefühl, dieser Online-Shop ist ehrlich.*

### 5.2.3 Reflektive Erfassung der wahrgenommenen Produktanzahl

Auch für das als reflektiv angenommene Konstrukt wahrgenommene Produktanzahl (siehe Abschnitt 2.6.3) wurde auf die Vorgaben der Untersuchung von Konradt et al. (2003) zurückgegriffen. In der damaligen Studie fanden drei Items Verwendung. Zwei davon wurden von Seiten der Autoren selbst entwickelt und in einem Pretest mit einer studentischen Stichprobe auf Verständlichkeit geprüft. Das dritte Item stellte eine Übersetzung eines Originalitems von Jarvenpaa et al. (2000) dar („Der Online-Shop ist ein sehr großer Anbieter im Internet“). Für dieses Item ergab sich im Quasi-Experiment von Konradt et al. (2003) ein sehr hoher Anteil fehlender Werte von 17,5%. Es wurde aus der Skala eliminiert, da sich auf diese Weise eine höhere interne Konsistenz ergab, die mit einem Cronbachs Alpha Wert von  $\alpha=0,81$  (Pearson  $r=0,69$ ) jedoch trotzdem als wenig zufrieden stellend anzusehen ist.

Auf Grundlage dieser Vorergebnisse wurde eine Umformulierung des besagten Items von Jarvenpaa et al. (2000) vorgenommen (Item-Code wa1). Zusätzlich wurde eines der beiden von Konradt et al. (2003) entwickelten Items verwendet (Item-Code wa2). Das dritte Item wa3 stellte eine Abwandlung eines Items von Danielson (2003) dar, bei dem die Anzahl von Einzelseiten einer Internetsite eingeschätzt werden sollte. Analog dazu wurde mit Hilfe von wa3 in der aktuellen Untersuchung abgefragt, wie viele Produkte der jeweils bewertete Online-Shop zum Verkauf anbietet. Das Antwortformat dieses Items unterschied sich entsprechend von dem der übrigen. Anstelle einer siebenstufigen Antwortskala wurden sechs mögliche Größenordnungen der Produktanzahl vorgegeben. Die drei Items, die in das Quasi-Experiment eingebunden wurden, sowie die Antwortvorgaben des Items wa3 lauteten:

Item wa1: *Ich habe den Eindruck, es handelt sich um einen vergleichsweise großen Online-Shop.*

Item wa2: *Der Online-Shop führt sehr viele verschiedene Produkte.*

Item wa3: *Bitte schätzen Sie, wie viele verschiedene Produkte dieser Online-Shop zum Verkauf anbietet. Antwortoptionen: weniger als 1.000, 1.000-10.000, 10.000-500.000, 500.000-1Mio, 1Mio-2Mio, mehr als 2 Mio.*



### 5.2.4 Reflektive Erfassung des Einkaufsvergnügens

Die Erfassung des Konstruktes Einkaufsvergnügens (siehe Abschnitt 2.6.4) erfolgte mit Hilfe zweier reflektiver Indikatoren. Zum einen wurde ein Item von Wolfinbarger und Gilly (2003) in einer vom Autor vorgenommenen Übersetzung aus dem Englischen verwendet (se1)<sup>17</sup>. Zusätzlich wurde ein weiteres Item herangezogen (se2), das abfragte, als wie frustrierend die Benutzung des Online-Shops erlebt wurde (Koufaris, 2002; Ghani, Supnick & Rooney, 1991). Bei Koufaris (2002) wurde das Item zur Erhebung des Konstruktes *Perceived Control* genutzt. Wie in Abschnitt 2.6.4 erläutert, wurde Frustration auf Seiten der Nutzer im Rahmen der vorliegenden Arbeit als eine negative Form des Einkaufsvergnügens angesehen. Aufgrund seiner negativen Formulierung, wurde dieses Item zur weiteren Auswertung umkodiert. Die Items, die für das Quasi-Experiment herangezogen wurden, lauteten:

Item se1: *Es macht wirklich Spaß, bei diesem Online-Shop einzukaufen.*

Item se2c: *Die Benutzung dieses Online-Shops frustriert mich.*

### 5.2.5 Erfassung des wahrgenommenen Preises

Die subjektiv erlebte Höhe des Preisniveaus eines Shops, also der wahrgenommene Preis (siehe Abschnitt 2.6.5), wurde mit Hilfe eines einzelnen, vom Autor entwickelten Items erfasst. Der Argumentation von Rossiter (2002) folgend, wird zur Erhebung eines Konstruktes, das nicht über unterschiedliche Facetten verfügt und sich auf einen konkreten Bewertungsgegenstand bezieht, lediglich ein Indikator benötigt. Nach Drolet und Morrison (2001) bringt die Verwendung weiterer Items nicht unbedingt eine Erhöhung der Messgüte bzw. einen Zuwachs an Information mit sich. Verschiedene Studien belegen, dass mit Hilfe von Single-Item-Skalen eine reliable und valide Messung möglich ist (vgl. Gardener, Cummings, Dunham & Pierce, 1998; Nagy, 2002; Wanous, Reichers & Hudy, 1997). Das Item zur Erfassung des wahrgenommenen Preises lautete:

Item subpc: *Die Preise in diesem Shop sind mir zu hoch.*

### 5.2.6 Formative Erfassung des Service

Wie in Abschnitt 2.6.6 dargestellt, wurden die sechs inhaltlichen Aspekte Zahlungsmöglichkeiten, Garantiegewährleistungen, Rückgaberegulung, Lieferzeiten, Kosten für Porto und Verpackung sowie die Möglichkeiten der persönlichen Kontaktaufnahme im Rahmen der vorliegenden Untersuchung als die relevanten formativen Facetten des Service von Online-Shops angesehen. Die Formulierungen der entsprechenden Items sv1 bis sv6 orientierten sich an englischsprachigen Vorgaben von Wolfinbarger und Gilly (2003), die dem Anhang A entnommen werden können. Die sechs Items, die zur Erfassung des Service genutzt wurden, lauteten im Einzelnen:

---

<sup>17</sup> Die Abkürzung „se“ steht für die englische Konstruktbezeichnung „Shopping Enjoyment“.

Item sv1: *Ich bin mit den Möglichkeiten der Bezahlung, die mir der Online-Shop bietet, zufrieden.*

Item sv2: *Die Garantieleistungen des Online-Shops finde ich angemessen.*

Item sv3: *Ich finde, der Online-Shop hat eine gute Regelung zur Rückgabe von Produkten.*

Item sv4c: *Die Lieferzeiten der Produkte empfinde ich bei diesem Online-Shop als zu lang.*

Item sv5c: *Die Kosten, die der Online-Shop für Porto und Verpackung berechnet, sind mir zu hoch.*

Item sv6: *Ich bin mit den Möglichkeiten der persönlichen Kontaktaufnahme zum Verkaufspersonal dieses Shops zufrieden.*

### **5.2.7 Erfassung der Bekanntheit der Shops**

Um die Bekanntheit der Shops zu erfassen, wurde jedem Teilnehmer ein einzelnes Item präsentiert (Item bek). Es fragte ab, inwiefern die Versuchsperson der Meinung ist, der Shop sei *allgemein* bekannt. Die Beantwortung erfolgte mit Hilfe der bereits vorgestellten siebenstufigen Likert-Skala mit den Polen „stimme gar nicht zu“ bis „stimme voll zu“ mit der zusätzlichen Option „keine Angabe“. Das Item lautete konkret:

Item bek: *Der Shop [Name des Shops], den ich im Rahmen des Versuchs besucht habe, ist allgemein sehr bekannt unter Internetnutzern.*

### **5.3 Operationalisierung der abhängigen Variablen**

Die Abschnitte 5.3.1 und 5.3.2 stellen vor, wie die abhängigen Variablen (AVn) Kaufintention und Kaufentscheidung operationalisiert wurden. Allgemein ist zu beachten, dass Untersuchungsergebnisse durch einen *Common Method Bias* verzerrt sein können, wenn die AVn und UVn auf die gleiche Art und Weise erfasst werden (vgl. Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Entsprechend wurde im Rahmen der Untersuchung mit der Kaufentscheidung ein Verhalten auf Seiten der Teilnehmer erfasst. Da die verschiedenen Erfolgsfaktoren mit Hilfe eines Fragebogens erhoben wurden, unterschied sich die Operationalisierung auf unabhängiger und abhängiger Variablenebene somit deutlich.<sup>18</sup> Ein Common Method Bias sollte folglich minimiert worden sein. Die Kaufintention wurde dagegen wie die kundennahen Erfolgsfaktoren in Form von Fragebogenitems abgefragt. In diesem Fall kann ein verzerrender Einfluss des Common Method Bias nicht ausgeschlossen werden. Allerdings wies dieses Maß den Vorteil auf, dass es aufgrund eines Intervallskalenniveaus die Anwendung multivariater statistischer Auswertungsverfahren erlaubte, die im Fall einer kategorialen AV wie der Kaufentscheidung nicht in Frage kommen (siehe Kapitel 6).

---

<sup>18</sup> Systematisch variiert wurde im Experiment nur die UV Online-Shop. Allerdings stellen die Erfolgsfaktoren in den berechneten Analysen die Prädiktoren bzw. die UVn dar.

### 5.3.1 Reflektive Erfassung der Kaufintention

Wie bereits bei diversen Items auf Prädiktorenebene wurde auch bei der Erfassung der AV *Kaufintention* auf Indikatoren zurückgegriffen, die schon in der Voruntersuchung von Konradt et al. (2003) zum Einsatz kamen. Die Autoren nutzten in der damaligen Studie drei Items, die nicht wie die übrigen in einem Aussageformat gestellt wurden, sondern Fragen darstellten. Es wurde damals eine interne Konsistenz von  $\alpha=0,90$  für die Skala erreicht. Die Annahme eines reflektiven Zusammenhanges zwischen der latenten Variable *Kaufintention* und den manifesten Variablen wird auch von Jarvis et al. (2003) getroffen, die das Konstrukt als ein „typisch“ reflektives bezeichnen. Die drei Indikatoren von Konradt et al. (2003) stellen Übersetzungen englischsprachiger Items von Jarvenpaa et al. (2000) dar, die ihrerseits auf Basis der Empfehlungen von Fishbein und Ajzen (1975) zur Erfassung der Verhaltensintention formuliert wurden. Inhaltlich bezieht sich einer der Indikatoren nicht direkt auf eine *Intention zu Kaufen* sondern vielmehr auf die Intention, den Online-Shop *wieder aufzusuchen*. Eine Aufnahme dieses Items erfolgte zum einen vor dem Hintergrund, dass es eine Besuchsbereitschaft abfragt, die als eine Voraussetzung für ein erneutes Kaufverhalten anzusehen ist – schließlich kann ein Kauf im Anschluss an den Versuch nur dann erfolgen, wenn eine Bereitschaft besteht, den Shop wieder aufzusuchen. Zudem wurde hierdurch berücksichtigt, dass vermutlich nicht alle Teilnehmer der Untersuchung in der Lage wären zu antizipieren, ob sie wieder im jeweiligen Online-Shop bestellen werden, ohne vorher zu wissen, welches Produkt sie kaufen. Die weiteren beiden Items, die in der Voruntersuchung zum Einsatz kamen, unterscheiden sich in Hinblick auf den Zeithorizont: das eine fragt die Intention ab, innerhalb von drei Monaten in dem jeweiligen Shop zu bestellen; das andere betrifft die Intention, im laufenden Versuch zu bestellen.

In der aktuellen Untersuchung wurde davon abgesehen, die Items wie in der Voruntersuchung in Frageform zu stellen. Stattdessen wurde auch hier eine Aussageform gewählt. Als Antwortformat wurde die siebenstufige Likert-Skala mit den Polen „stimme gar nicht zu“ und „stimme voll zu“ mit zusätzlicher Antwortoption „keine Angabe“ herangezogen, die auch bei nahezu allen weiteren Items zur Anwendung kam (siehe Abschnitt 4.2.5). Hintergrund dieser Änderung war, dass die Formulierungen der drei Items von Konradt et al. (2003) ein zu hohes Maß an Ähnlichkeit aufweisen; sie beginnen stets mit den Worten „Wie wahrscheinlich ist es...?“. Es ließe sich vermuten, dass sehr ähnlich formulierte Items unabhängig von ihrer semantischen Bedeutung tendenziell ähnlich beantwortet werden. Eine Unterschiedlichkeit der Itemformulierungen ließ sich aus Sicht des Autors jedoch leichter bei Verwendung eines Aussageformates als im Falle eines Frageformates erreichen. Zusätzlich zu den drei umformulierten Indikatoren wurde ein neues, vom Autor entwickeltes Item zur Erfassung der *Kaufintention* verwendet (Item-Code ki4). Mit dieser „Testverlängerung“ wurde das Ziel verfolgt, eine höhere Messgenauigkeit zu erreichen. Die vier in der Untersuchung verwendeten Items lauteten:

- Item ki1: *Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen.*
- Item ki2: *Ich denke, dass ich innerhalb der nächsten drei Monate in diesem Online-Shop bestellen werde.*
- Item ki3: *Es ist recht wahrscheinlich, dass ich im Rahmen des laufenden Versuchs in diesem Shop bestelle.*
- Item ki4: *Wenn ich wieder einmal [Bücher, DVDs, CDs, Druckerpatronen, Veranstaltungstickets] im Internet bestellen möchte, dann tue ich dies wahrscheinlich in diesem Shop.*

### 5.3.2 Operationalisierung der Kaufentscheidung

Als zentrale Erfolgsgröße der Untersuchung diene die *Kaufentscheidung* der Teilnehmer. Die Kaufentscheidung wurde im durchgeführten Versuch an die Entlohnung der Probanden gekoppelt. Jede Vp sollte ein Produkt in einem von zwei Online-Shops erwerben, die ihr im Quasi-Experiment präsentiert wurden. Alternativ stand ihr die Wahl eines Gutscheins über 15 Euro für die Warenhauskette Karstadt offen, dessen Ausgabe jedoch erst nach Abschluss der gesamten Datenerhebung erfolgte – also nicht direkt im Anschluss an die Teilnahme des Probanden. Für den Fall, dass sich die Vp für einen Kauf in dem einen der beiden Shops entschied, wurden ihr 15 Euro direkt im Anschluss an seine Teilnahme am Versuch ausgezahlt. Jeder Teilnehmer wurde im Vorfeld der Untersuchung gebeten, sich auf ein konkretes Produkt festzulegen. Dadurch sollte gewährleistet werden, dass auf Teilnehmerseite eine hohe Kaufmotivation bestand und somit von einer realistischen Kaufsituation gesprochen werden könnte. Es konnte sich bei dem Produkt um ein Buch, eine CD, eine DVD, eine Druckerpatrone oder ein Veranstaltungsticket handeln (siehe Abschnitt 5.1.1). Sowohl im Vorfeld als auch während des Quasi-Experiments wurden die Teilnehmer darauf hingewiesen, dass die Bestellung in einem der beiden Shops auf ihren *eigenen Namen* erfolgen müsse.

Alternativ zu diesem Vorgehen wäre es möglich gewesen, ein konkretes Produkt der jeweiligen Kategorie für alle Teilnehmer als Belohnung festzulegen. Dies hätte jedoch dazu führen können, dass die Motivation zur Teilnahme unter den Vpn sehr unterschiedlich ausgefallen wäre. Andererseits wäre es auch denkbar gewesen, keine Festlegung auf ein bestimmtes Produkt zu fordern. Dies hätte jedoch den deutlichen Nachteil gehabt, dass viele Teilnehmer im Versuch nicht gewusst hätten, welches Produkt sie kaufen sollten, was wiederum mit einer niedrigen Kaufmotivation verbunden gewesen wäre. Die Namen der beiden Shops, die der Proband im Versuch kennen lernte, wurden im Vorfeld nicht genannt.

Die Wahlmöglichkeit eines Gutscheins wurde aufgrund unterschiedlicher Erwägungen eingeräumt. Zum einen wären die Teilnehmer gewissermaßen zu einem Kauf in einem der Shops gedrängt worden, wenn es diese Option nicht gegeben hätte. Dies erschien jedoch aus forschungsethischen Gesichtspunkten unangemessen. Des Weiteren wäre andernfalls eine Wahl zwischen nur zwei Alternativen getroffen worden (Kauf in Shop 1 oder 2). Von einer

*Kaufentscheidung* kann jedoch erst gesprochen werden, wenn mehr als nur zwei Alternativen zur Verfügung stehen und wenn die Entscheidung ein gewisses Maß an kognitiver Beteiligung seitens des Kunden erfordert (siehe Abschnitt 2.1.6.1; Wiswede, 1995). Außerdem besteht für einen Kunden in einer realen Kaufsituation die Möglichkeit, sich *gegen* einen Kauf zu entscheiden. Hätte man den Teilnehmern im Rahmen des Versuchs alternativ zum Gutschein Geld ausgezahlt, so wäre die Folge gewesen, dass sie Geld „verdient“ hätten, wenn sie sich gegen den Kauf entschieden. Zudem hätte eine niedrige Bereitschaft resultiert, im Versuch einen Kauf zu tätigen, so dass keine realistische Kaufsituation entstanden wäre. Über die Alternative, anstelle einer Bestellung in einem der beiden Shops einen Gutschein zu erhalten, wurde der Proband erst informiert, nachdem er beide Shops im Versuch kennen gelernt hatte. Hätte der Teilnehmer bereits im Vorfeld von der Option des Gutscheins erfahren, so wäre eine mögliche Konsequenz gewesen, dass er sich gar nicht erst auf ein konkretes Produkt festlegt hätte.

Eine realistischere Kaufentscheidung wäre dann entstanden, wenn die Teilnehmer die Wahl zwischen mehr als nur zwei Online-Shops hätten treffen könnten. Die Einbeziehung von mehr als zwei Shops pro Teilnehmer konnte jedoch nicht erfolgen, da dies den zeitlichen Rahmen der Versuchsdurchführung gesprengt hätte. Schließlich ist eine differenzierte Bewertung eines Online-Shops nur dann möglich, wenn sie auf einem gewissen Erfahrungsschatz, also einer Mindestdauer der Interaktion mit dem Shop, beruht. Wie sich in der Vorgängeruntersuchung zeigte, erfordert die Kennlernphase und Bewertung *eines* Shops im Mittel ca. 25 Minuten (Konradt et al., 2003). Bei einer Einbindung von drei Shops hätte die Untersuchungsdauer pro Teilnehmer somit häufig bei mehr als zwei Stunden gelegen, was nicht als zumutbar anzusehen ist.

Zusammenfassend wurde die Kaufentscheidung als Wahl zwischen den drei Alternativen *Kauf in Shop 1*, *Kauf in Shop 2* oder *Gutschein* operationalisiert. Da der Bestellvorgang in vielen Online-Shops eine sehr kritische Phase des Einkaufs darstellt (siehe Abschnitt 2.1.6.2 und 2.5.1), stand es den Teilnehmern im Versuch offen, eine zunächst getroffene Kaufentscheidung zu revidieren und die Wahl erneut zu treffen.

## 5.4 Erhebung weiterer Variablen

Neben den zentralen Variablen der Untersuchung – also den potenziellen Erfolgsfaktoren auf Prädiktorenebene und den Erfolgsgrößen *Kaufintention* und *Kaufentscheidung* auf Ebene der AVs – wurden einige weitere Variablen erhoben. Es erfolgte die Erfassung verschiedener demographischer Merkmale der Probanden (Abschnitt 5.4.1). Mit derselben Zielsetzung wurden einige Indikatoren erfasst, die die Vorerfahrung in den Bereichen Computer- und Internetnutzung betreffen (Abschnitt 5.4.2). Des Weiteren wurden einige Kontrollvariablen herangezogen (Abschnitt 5.4.3).

### 5.4.1 Demographische Merkmale der Teilnehmer

Es wurden die demographischen Merkmale *Alter*, *Geschlecht*, *Einkommen*, *Haushaltsgröße*, *Berufliche Tätigkeit* und *Bildungsstand* der Teilnehmer abgefragt – einerseits um modellierende Einflüsse zu untersuchen, andererseits um die Teilnehmerstichprobe angemessen beschreiben und Vergleiche zur Population deutscher Internetnutzer ziehen zu können. Im Falle der rangskalierten Variable *Einkommen* wurden die folgenden Antwortoptionen 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=*mehr als 2500 Euro*. Die Stufen der Variable *Bildungsstand* (im Sinne des höchsten Abschlusses) lauten: 0=*Abitur*, 1=*Mittlere Reife*, 2=*Hauptschule*, 3=*keiner* sowie 4=*Sonstiges*. Als Antwortkategorien für die *Berufliche Tätigkeit* dienten die Kategorien: 0=*Angestellter/Arbeiter*, 1=*Selbständig*, 2=*Beamter*, 3=*Student*, 4=*Schüler*, 5=*keine* sowie 6=*Sonstiges*.

### 5.4.2 Vorerfahrung im Bereich PC und Online-Shopping

Im Rahmen des Versuchs wurde die Vorerfahrung in Bereichen *PC-Nutzung*, *WWW-Nutzung* und *Online-Shopping* auf Seiten der Teilnehmer abgefragt. Die Vorerfahrung der *PC-* und *Internetnutzung* wurde jeweils mit Hilfe von drei Items erhoben, die die Nutzungserfahrung in Jahren, in Tagen im Monat und Stunden am Tag betreffen. Es wurden drei weitere Items genutzt, um zu erheben, über welches Maß an Erfahrung die Teilnehmer im Online-Einkaufen verfügen. Die Items lauteten im Einzelnen:

Items zur Vorerfahrung im Bereich PC-Nutzung:

Item pc\_seit: *Seit wie vielen Jahren nutzen Sie einen PC?* (offenes Eingabefeld) \_\_\_\_\_

Item pc\_t: *Wie häufig nutzen Sie einen PC momentan?* \_\_\_\_\_ *Tage im Monat*

Item pc\_s: (zusätzliches Eingabefeld zum letztgenannten Item) \_\_\_\_\_ *Stunden pro Tag*

Items zur Vorerfahrung im Bereich Internet-Nutzung:

Item www\_st: *Seit wie vielen Jahren nutzen Sie das Internet?* (offenes Eingabefeld) \_\_\_\_\_

Item www\_t: *Wie häufig nutzen Sie das Internet momentan?* \_\_\_\_\_ *Tage im Monat*

Item www\_s: (zusätzliches Eingabefeld zum letztgenannten Item) \_\_\_\_\_ *Stunden pro Tag*

Items zur Vorerfahrung im Bereich Online-Shopping:

Item os\_a: *Wie oft haben Sie bisher [Bücher, DVDs, CDs, Druckerpatronen, Veranstaltungstickets] im Internet eingekauft?* (offenes Eingabefeld) \_\_\_\_\_

Item os\_p: *Haben Sie schon einmal vor diesem Versuch [Bücher, DVDs, CDs, Druckerpatronen, Veranstaltungstickets] im Internet bestellt?* ja, nein

### 5.4.3 Kontrollvariablen

Im Rahmen des Quasi-Experiments galt es, einige Kontrollvariablen zu erheben, die Einfluss auf die Ergebnisse der Studie hätten nehmen können. Zum einen wurde das *Produkt-Involvement* erfasst, das sich darauf bezog, wie wichtig den Vpn das Produkt war, auf das sie sich festgelegt hatten (Abschnitt 5.4.3.1). Zum anderen konnte ein Einfluss durch die Versuchsumgebung nicht ausgeschlossen werden. In Abschnitt 5.4.3.2 werden die Kontrollvariablen dargestellt, die die experimentellen Rahmenbedingungen betreffen.

#### 5.4.3.1 Produkt-Involvement

Es war in Betracht zu ziehen, dass auf Seiten der Teilnehmer Unterschiede dahingehend bestanden, wie wichtig ihnen die Produkte waren, auf die sie sich im Vorfeld der Untersuchung als Belohnung festgelegt hatten. Die Wahrscheinlichkeit eines Kaufs im Rahmen des Quasi-Experiments dürfte für eine Person höher gewesen sein, der das gewählte Produkt sehr wichtig war, als für einen Probanden, der dem Produkt keine hohe Wichtigkeit beimaß. Letzterer hätte sich vermutlich eher für einen Gutschein entscheiden, als im Versuch einen Einkauf zu tätigen. Um Unterschiede bzgl. der Wichtigkeit des Produktes und ihren Einfluss auf die abhängigen Variablen kontrollieren zu können, wurde das Produkt-Involvement erhoben (vgl. Koufaris, 2002). Das *Involvement* stellt nach Mittal (1989) (a) einen auf ein Objekt gerichteten motivationalen Zustand (z.B. Arousal, Interesse, Antrieb) dar, wobei (b) der motivationale Zustand durch die subjektiv erlebte Relevanz des Objektes angeregt wird (Blackwell, Miniard & Engel, 2001). Das Involvement im Rahmen des Einkaufs kann auf unterschiedliche Aspekte bezogen werden, z.B. auf Werbung (Andrews & Durvasula, 1991), auf den Einkaufsprozess (Slama & Tashchian, 1985) oder das zu erwerbende Produkt (Park & Moon, 2003). Das Produkt-Involvement ist für ein bestimmtes Produkt keineswegs immer gleich hoch, sondern kann von Kunde zu Kunde erheblich variieren (Zaichkowsky, 1985). In einigen empirischen Studien finden sich Hinweise auf den Einfluss des Produkt-Involvements im Rahmen des Online-Shoppings (Balabanis & Reynolds, 2001; Gordon, McKeage & Fox, 1998; Hoffman & Novak, 1996; Lee, Kim & Moon, 2000; McGaughey & Mason, 1998; Ranaweera et al., 2005).

Zur Erfassung der Variable *Produkt-Involvement* haben McQuarrie und Munson (1992) eine Skala entwickelt, die auf einem semantischen Differential beantwortet wird. Die Unterskala *Importance* deckt dabei jene inhaltliche Dimension ab, die im Rahmen der vorliegenden Studie relevant war (vgl. Koufaris, 2002; Novak, Hoffman & Yung, 2000; Zaichkowsky, 1985, 1994). Von einer Nutzung eines semantischen Differentials wurde in der vorliegenden Untersuchung abgesehen, da die Teilnehmer nicht mit verschiedenen Antwortformaten konfrontiert werden sollten. Entsprechend wurden, aufbauend auf den Items von McQuarrie und Munson (1992), Items entwickelt, die wie die übrigen in Aussageform mit einer siebenstufigen Likert-Skala gestellt wurden. Es handelte sich um ein positiv (pi1) sowie ein negativ formuliertes Item (pi2c). Zusätzlich wurde ein selbst entwickeltes Item pi3 genutzt, das nicht

auf die Skala von McQuarrie und Munson (1992) zurückgeht. Die drei in der Studie verwendeten Items lauteten:

Item pi1: *Das Produkt ist mir sehr wichtig.*

Item pi2c: *Eigentlich ist das Produkt gänzlich irrelevant für mich.*

Item pi3: *Ich will das Produkt wirklich sehr gerne haben.*

#### 5.4.3.2 Kontrollvariablen zum experimentellen Kontext

Um eventuelle Einflüsse der Erhebungssituation zu identifizieren und statistisch zu kontrollieren, galt es, einige weitere Kontrollvariablen zu erheben. Die Art der *Bildschirmauflösung* sowie die genutzte *Browsersoftware* wurden automatisch erfasst. Eine automatisierte Dokumentation erfolgte weiterhin für die Kontrollvariable *Durchführungsdauer*. Die Art der *Internetverbindung* ließ sich nicht automatisiert erfassen, daher wurden die Teilnehmer gebeten, diese anzugeben (Item modem: *Mit was für einer Verbindung gehen Sie während dieses Versuchs ins Internet?* Antwortoptionen: 0=33,6K-Modem oder langsamer; 1=56K-Modem; 2=ISDN; 3=schneller als ISDN).

### 5.5 Versuchsszenario

Es galt, ein Versuchsszenario zu entwerfen, durch das gewährleistet wurde, dass die Teilnehmer die Online-Shops in ausreichendem Maße kennen lernen konnten, um eine Bewertung bzgl. der verschiedenen Bewertungsaspekte vornehmen zu können. Entsprechend sollte die Interaktion des Probanden mit dem Shop nicht zu kurz ausfallen. Zudem galt es, die grundlegenden Schritte eines typischen Einkaufsvorganges abzubilden, um eine hohe externe Validität der Untersuchung zu erreichen und dadurch eine Übertragbarkeit der Ergebnisse zu rechtfertigen.

Um sicherzustellen, dass die Teilnehmer die typischen Schritte eines Online-Einkaufs durchliefen, wurden ihnen verschiedene Aufgaben gestellt. Nutzungsintentionen von Online-Kunden können sehr unterschiedlich ausfallen (Abschnitt 2.5.3). Folglich lassen sich sehr viele verschiedene Aufgaben und Versuchsszenarien ableiten. Aus software-ergonomischer Sicht wäre zu fordern, dass ein Online-Shop alle wesentlichen Nutzungsintentionen unterstützt und der Kunde seine Ziele effizient, effektiv und zufrieden stellend erreichen kann. Sicher wäre wenig sinnvoll gewesen, im Rahmen des Versuchs alle möglichen Nutzungsszenarien in Form einzelner Aufgaben zu berücksichtigen. Dies hätte den zeitlichen Rahmen des Quasi-Experimentes gesprengt. Vielmehr galt es, besonders *zentrale* Benutzungsvorgänge eines Online-Einkaufs in Aufgaben umzusetzen. Wie in Abschnitt 2.5.3 dargestellt wurde, sind die gezielte Suche sowie das Stöbern als zentrale Nutzungsformen anzusehen. Da jeder Online-Einkauf zudem eine Interaktion mit dem Warenkorb und das Durchlaufen des Bestellvorgangs erfordert, sollten diese Phasen des Einkaufs ebenfalls Berücksichtigung finden.



Die erste Aufgabe forderte den Teilnehmer auf, gezielt nach dem Produkt zu suchen, auf das er sich im Vorfeld der Untersuchung als Belohnung festgelegt hatte (Aufgabe *Gezielte Suche*). Von diesem Produkt sollten drei Exemplare in den Warenkorb gelegt werden. Bei der zweiten Aufgabe sollte der Proband im Shop nach Produkten suchen bzw. „stöbern“, die ihn interessieren, und drei verschiedene davon in den Warenkorb legen (Aufgabe *Stöbern*). Zusätzlich wurde die Instruktion gegeben, dass die Vp sich über alle Rahmenbedingungen des Einkaufs in dem Shop informieren sollte, die für sie im Fall einer Bestellung von Wichtigkeit sei. Dazu gehörte, den Bestellvorgang probeweise zu beginnen, um zu erfahren, wie der Bestellprozess des Shops gestaltet sei. Im Rahmen der dritten Aufgabe wurde der Teilnehmer schließlich dazu aufgefordert, den Inhalt des Warenkorbs gezielt so zu verändern, dass sich darin nur noch ein Exemplar des Produkts befände, auf das er sich im Vorfeld festgelegt hatte (Aufgabe *Veränderung des Warenkorbs*). Im Anhang A findet sich der konkrete Wortlaut der Teilnehmerinstruktionen. In Abschnitt 7.4 wird der Versuchsablauf ausführlich vorgestellt.

Durch die Vorgabe konkreter Aufgaben im Quasi-Experiment wurde der Shopbesuch zu einem gewissen Grad standardisiert. Denkbar wäre gewesen, den Teilnehmern lediglich eine generelle Anweisung zu geben, sich den jeweiligen Online-Shop über einen gewissen Zeitraum anzusehen. Die Erfahrungen der Nutzer hätten in Folge dessen sehr stark voneinander abgewichen. Bei einer solchen Unterschiedlichkeit der Bewertungsgrundlagen wäre eine Vergleichbarkeit der Versuchsergebnisse zwischen den Teilnehmern jedoch kaum gewährleistet gewesen.

## 5.6 Berücksichtigung möglicher Reihenfolgeeffekte

In diversen Studien wurden störende Einflüsse der Präsentationsreihenfolge von Items bei der Abfrage der Zufriedenheit auf die Ergebnisse fragebogenbasierter Erhebungen untersucht (vgl. Konradt, Christophersen & Schäffer-Külz, 2006; Oishi, Schimmack & Colcombe, 2003; Schwarz, Strack & Mai, 1991; Strack & Werth, 2002). Danach können Reihenfolgeeffekte die Ergebnisse in teilweise erheblichem Maße in dem Sinne verzerren, dass vorausgehende Items einen Einfluss auf die Beantwortung von später präsentierten Items nehmen. In den meisten Fällen lassen sich diese Effekte dadurch erklären, dass die kognitive Beschäftigung mit den Bewertungsaspekten die Verfügbarkeit bestimmter Gedächtnisinhalte beeinflusst (vgl. Sudman, Bradburn & Schwarz, 1996, S. 58). Bei Zufriedenheitsabfragen gilt es daher stets zu berücksichtigen, dass eine allgemeine Zufriedenheitseinschätzung dadurch beeinflusst werden kann, inwiefern und in welchem Ausmaß zum Zeitpunkt der Abfrage positive bzw. negative Informationen zum Beurteilungsgegenstand verfügbar sind. In der vorliegenden Untersuchung hätten Reihenfolgeeffekte auf unterschiedlichen Ebenen auftreten können.

### 5.6.1 Reihenfolge der formativen Items zur Usability

Im Fall der formativen Items der *Usability* erschien eine Berücksichtigung von Reihenfolgeeffekten besonders relevant, da vom Teilnehmer eine differenzierte Bewertung von vielen Facetten der Benutzerfreundlichkeit vorgenommen werden sollte. Entsprechend wurden die formativen *Usability* Items im Rahmen der computerbasierten Versuchsumgebung in zufälliger Reihenfolge dargeboten. Zwar können Reihenfolgeeffekte auf Ebene des einzelnen Teilnehmers auch bei einer Randomisierung der Items auftreten. Allerdings ergibt sich der Vorteil, dass die Ergebnisse auf Ebene der gesamten Untersuchung nicht durch eine spezifische Reihenfolge der abgefragten Aspekte reflektiert werden, da mögliche Effekte sich der Gesamtstichprobe gleichverteilen (vgl. Sudman et al., 1996, S. 163).

Es wäre denkbar gewesen, sämtliche in der Studie verwendeten Items in einer zufälligen Reihenfolge abzufragen. Von einem solchen Vorgehen wurde aus zweierlei Gründen abgesehen. Zum einen hätte es auf Seiten der Respondenten ein sehr viel höheres Maß an kognitiver Beschäftigung erfordert, da sich die Teilnehmer bei (nahezu) jedem Item mit einem gänzlich anderen Bewertungsaspekt hätten beschäftigen müssen. Die sehr hohe Anzahl an Items und die insgesamt lange Durchführungsdauer des Quasi-Experiments erforderte bereits ein hohes Maß an Konzentrationsleistung, so dass eine zusätzliche Belastung vermieden werden sollte. Zum anderen reagieren Versuchspersonen oftmals verärgert, wenn sie meinen, eine Frage bereits zu einem früheren Zeitpunkt im Fragebogen beantwortet zu haben. Zwar wurde bei der Zusammenstellung sämtlicher Items darauf geachtet, dass sich die Formulierungen möglichst deutlich voneinander unterscheiden, um einer derartigen negativen Affektreaktion entgegen zu wirken. Bei einer randomisierten Darbietung aller Indikatoren hätte schneller der Eindruck entstehen können, dass ein Item bereits beantwortet wurde, als im Falle einer geblockten Präsentation der Skalen. Entsprechend wurde von einer randomisierten Darbietung aller Items abgesehen. Die Präsentation der Skalen erfolgte jeweils geblockt, wobei lediglich innerhalb des Blocks der formativen Usability Items eine zufällige Reihenfolge gewählt wurde.

### 5.6.2 Reihenfolge der Usabilityabfrage reflektiv vs. formativ

Wie in Abschnitt 4.1 dargestellt, wurde durch die reflektive Usability-Skala ein Globalurteil zur Benutzerfreundlichkeit eines Online-Shops abgefragt. Dagegen diente das formative Usability-Maß der Erfassung einzelner Facetten des Konstruktes. Diese formativen Facetten können je nach Bewertungsgegenstand und Nutzer durchaus differenziert bewertet werden. Angesichts der oben dargestellten möglichen Reihenfolgeeffekte bei der Beantwortung von Fragebögen erscheint es denkbar, dass formative Einzelaspekte zur Usability eines Shops anders bewertet werden je nachdem, ob zuvor ein Globalurteil abgegeben wurde oder nicht. Andererseits könnte auch ein globales Urteil anders ausfallen, wenn sich der Antwortende zuvor mit der Beurteilung einzelner Facetten der Benutzerfreundlichkeit beschäftigt hat. Diese

Erwägungen legten nahe, eine systematische Variation der Präsentationsreihenfolge der formativen und reflektiven Usability-Skala im Sinne einer Ausbalancierung vorzunehmen.

### 5.6.3 Reihenfolge der Shopdarbietung

Da jeder Teilnehmer im Quasi-Experiment zwei Online-Shops bewerten sollte, ist ein Einfluss der Präsentationsreihenfolge beider Shops auf die Bewertung der Konstrukte nicht auszuschließen. Dieser könnte beispielsweise darin begründet sein, dass bei der Bewertung des zweiten Shops im Versuchsablauf auf Seiten der Nutzer kognitive Vergleichsprozesse zum jeweils vorhergehenden Shop abliefen. Die Bewertung des ersten Shops fungierte unter Umständen als ein Maßstab zur Beurteilung des zweiten. Entsprechend galt es eine Ausbalancierung vorzunehmen (vgl. Rack & Christophersen, 2006, S. 25). Zum einen wurde innerhalb der einzelnen Produktarten jede mögliche Paarung der sieben Shops berücksichtigt und zum anderen die Reihenfolge der Präsentation für jedes Paar variiert.

### 5.6.4 Reihenfolge der Konstruktabfrage

In Abschnitt 5.6.1 wurde bereits dargelegt, dass die Präsentation der Skalen innerhalb des Fragebogens geblockt erfolgte. Es stellte sich zudem die Frage, in welcher Reihenfolge die verschiedenen Itemblocks am sinnvollsten dargeboten werden sollten. Abbildung 5-1 gibt einen schematischen Überblick über den Ablauf der Konstruktabfrage.

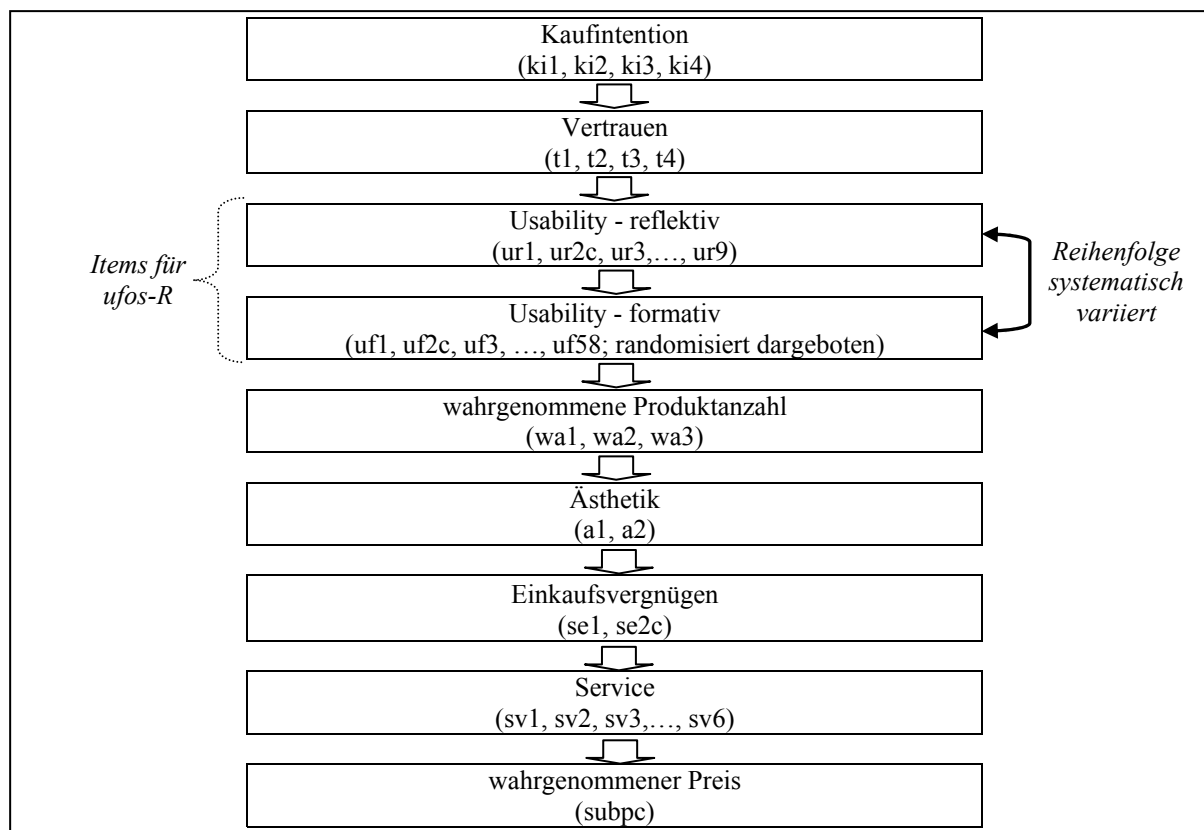


Abb. 5-1: Schematischer Aufbau des Fragebogens zur Bewertung der Online-Shops

Die gewählte Reihenfolge wurde primär dadurch bestimmt, welche Annahmen bzgl. der Zusammenhänge zwischen den Konstrukten zugrunde liegen (siehe Abschnitt 3.3). Dies soll beispielhaft anhand der Skala zur Kaufintention erläutert werden, die als erstes präsentiert wurde. Die Kaufintention stellte die zentrale AV dar, die im Fragebogen erhoben wurde. Es wird also angenommen, diese latente Variable werde von den übrigen beeinflusst – und nicht andersherum. Eine Abfrage der Kaufintention im Anschluss an die der übrigen Konstrukte hätte möglicherweise eine Verzerrung der Ergebnisse zur Folge gehabt. Der Einfluss der Prädiktoren auf die AV hätte sich aufgrund der oben angeführten Reihenfolgeeffekte durch eine kognitive Auseinandersetzung mit diesen Bewertungsaspekten verstärken können. Entsprechend hätte die Gefahr einer Überschätzung der Zusammenhänge zwischen den Variablen bestanden. Analog zu dieser Argumentation wurde die Skala zum Vertrauen den übrigen Skalen vorangestellt, da sie gemäß den Annahmen von den anderen beeinflusst wird. Die Positionierung der weiteren Skalen wurde ebenfalls durch die Zusammenhangshypothesen bestimmt. Eine Ausnahme wurde im Fall der beiden Usability-Skalen gemacht, deren Abfrage im Anschluss an die Vertrauen-Items erfolgte. Zwar wurde theoretisch begründet angenommen, dass die Usability eher das Einkaufsvergnügen beeinflusst als umgekehrt das Einkaufsvergnügen die Usability (siehe dazu Abschnitt 3.3). Allerdings kommt der Usability im Rahmen der Studie die höchste Priorität zu. Es kann nicht ausgeschlossen werden, dass die Genauigkeit von Versuchspersonen bei der Beantwortung von Items sinkt, je mehr Items sie bereits beantwortet haben. Daher wurden die Skalen wahrgenommene Produktanzahl, Ästhetik, Einkaufsvergnügen, Service sowie das Item zum wahrgenommenen Preis nach den beiden Usability-Skalen präsentiert.

## 5.7 Versuchsplan

Wie in Abschnitt 5.1 erläutert, stellten die verschiedenen Online-Shops die Ausprägungen der ersten systematisch variierten UV dar. Pro Produktart wurden sieben Online-Shops herangezogen. Da jeder Teilnehmer zwei Online-Shops derselben Produktart kennen lernte und bewertete, ergaben sich innerhalb der fünf einzelnen Produktarten jeweils 21 mögliche Paarungen von Shops. Um eventuelle Reihenfolgeeffekte bei der Präsentation der Shops kontrollieren zu können, wurde zusätzlich die Präsentationsreihenfolge der jeweils zwei Shops im Sinne einer Ausbalancierung variiert, so dass sich die Anzahl der Paarungen auf 42 verdoppelte (siehe Abschnitt 5.6.3). Zudem wurde die Abfragereihenfolge der reflektiven und formativen Usability ausbalanciert (siehe Abschnitt 5.6.2). Dadurch verdoppelte sich die Anzahl der Versuchsbedingungen pro Produktart ein weiteres Mal auf 84. Bei fünf Produktarten ergaben sich somit insgesamt 420 verschiedene Versuchsbedingungen.

Jede der 420 Versuchsbedingungen sollte durch die Erhebung eines einzelnen Teilnehmers berücksichtigt werden. Da jeder Proband zwei Shops bewertete, sollte sich eine Anzahl von insgesamt 840 Shopbewertungen ergeben mit 24 Bewertungen pro Shop. Die Teilnehmer

wurden den Versuchsbedingungen in Bezug auf die demographischen Kriterien *Alter* und *Geschlecht* so weit wie möglich parallelisiert zugeordnet (vgl. Rack & Christophersen, 2006, S. 29). Da nicht jeder Teilnehmer jeden Shop bewertete, handelt es sich um ein unvollständiges Versuchsdesign, so dass nur von einer *quasi-experimentellen* Untersuchung die Rede sein kann (vgl. Rack & Christophersen, 2006, S. 25).

In Tabelle C-8 im Anhang C werden beispielhaft die 84 Versuchsbedingungen der Produktart *Bücher* unter Angabe der Teilnehmercodes aufgelistet. Abbildung 5-2 stellt dar, wie sich der Teilnehmercode zusammensetzt. Zur Veranschaulichung dient der Teilnehmercode *d6\_4\_2*. Hier wurde der Versuchsperson zunächst der sechste von sieben Online-Shops (1. Zahl=6) zum Verkauf von DVDs (Buchstabe=d) präsentiert und im Anschluss daran der vierte DVD-Shop (2. Zahl=4). Die Items zur Erfassung der *formativen Usability* wurden bei diesem Teilnehmer vor den Items zur *reflektiven Usability* abgefragt (3. Zahl=2).

Buchstabe	1. Zahl	2. Zahl	3. Zahl
↓ <u>Produktart</u>	↓ <u>Nummer des ersten präsentierten Shops</u>	↓ <u>Nummer des zweiten präsentierten Shops</u>	↓ <u>Reihenfolge der Usabilityabfrage</u>
<i>b</i> =Bücher <i>d</i> =DVDs <i>m</i> =CDs <i>p</i> =Druckerpatronen <i>t</i> =Veranstaltungstickets	<i>1</i> =Shop 1 <i>2</i> =Shop 2 <i>3</i> =Shop 3 <i>4</i> =Shop 4 <i>5</i> =Shop 5 <i>6</i> =Shop 6 <i>7</i> =Shop 7	<i>1</i> =Shop 1 <i>2</i> =Shop 2 <i>3</i> =Shop 3 <i>4</i> =Shop 4 <i>5</i> =Shop 5 <i>6</i> =Shop 6 <i>7</i> =Shop 7	<i>1</i> =zunächst reflektiv, dann formativ <i>2</i> =zunächst formativ, dann reflektiv

Abb. 5-2: Zusammensetzung des Teilnehmercodes



## 6. Statistische Auswertung der Daten

In diesem Kapitel wird dargestellt, welche statistischen Verfahren zur Analyse der Fragestellungen herangezogen werden und welche Voraussetzungen bzw. Gütekriterien diesen Verfahren zugrunde liegen. Zudem werden die verschiedenen methodischen Schritte erläutert, die zur Aufbereitung des Datensatzes und der Analyse der Messgüte dienen, bevor die Auswertung der Daten stattfinden kann.

In einem ersten Schritt erfolgt eine Analyse und Behandlung fehlender Werte. Es wird erläutert, wie mit hohen Missing-Anteilen auf Fallebene umgegangen wird, wie die Analyse des Fehlendmechanismus erfolgt und welche Ersetzungsmethoden für fehlende Werte zum Einsatz kommen (Abschnitt 6.1). Anschließend werden in Abschnitt 6.2 die Kriterien zur Analyse der Verteilungen von Indikatoren vorgestellt. Um zu prüfen, inwiefern ein Common Method Bias (vgl. Podsakoff et al., 2003) im Datensatz vorliegt, kommt mit dem Harman's Single Factor Test ein Testverfahren zum Einsatz, das in Abschnitt 6.3 erläutert wird. Zur Identifikation potenziell problematischer Teilnehmer dienen drei Kriterien, die den Abschnitten 6.4.1 bis 6.4.3 zu entnehmen sind. Die Skalengüte reflektiver Maße auf Prädiktorebene wird anhand einer Prozedur von Homburg und Giering (1996) geprüft. Sie verbindet eine klassische Itemanalyse (Abschnitt 6.5.2) mit der Durchführung einer explorativen sowie einer konfirmatorische Faktorenanalyse, deren Voraussetzungen und Kriterien in den Abschnitten 6.5.1 bzw. 6.5.3 dargelegt werden.

Abschnitt 6.6 stellt die Gütekriterien des Strukturgleichungsverfahrens *Partial Least Squares* (PLS) vor, mit Hilfe dessen zwei der wesentlichen Fragestellungen dieser Arbeit nachgegangen wird. Einerseits wird mit Hilfe dieses Verfahrens ein Zwei-Konstrukt-Modell zur Festlegung der formativen Usability-Skala geschätzt, andererseits erfolgt die Untersuchung des konzeptuellen Erfolgsmodells zur Prädiktion der Kaufintention durch PLS-Modelle. Das zweite zentrale Verfahren, das der Beantwortung einer Hauptfragestellung dient, ist die *multinomiale logistische Regression* (MLR), durch die eine Vorhersage der Kaufentscheidung vorgenommen wird. Die Voraussetzungen und Gütekriterien dieser Methode sind Abschnitt 6.7 zu entnehmen.

Zentrale Voraussetzung zur Anwendung diverser Verfahren ist, dass zwischen den einbezogenen Variablen keine Multikollinearität besteht. Im Rahmen der vorliegenden Untersuchung werden drei Ansätze zur Identifikation von Multikollinearität verfolgt (Abschnitt 6.8). In den Abschnitten 6.9.1 bis 6.9.4 werden die Voraussetzungen und Kriterien der einfaktoriellen Varianzanalyse, des t-Tests sowie des Levene-Tests vorgestellt, die im Rahmen der Datenanalyse ebenfalls zur Anwendung kommen.

Um dem Leser die Bezüge zwischen den verschiedenen Abschnitten dieser Arbeit vor Augen zu führen, wird in Tabelle 6-1 eine Übersicht gegeben, welches Verfahren zur Untersuchung welcher Hypothese diene. Für jede einzelne Hypothese werden die korrespondierenden

Kapitelabschnitte zum theoretischen bzw. empirischen Hintergrund (Kapitel 2), zur Operationalisierung (Kapitel 4 und 5), zu den angewandten statistischen Verfahren (aktuelles Kapitel 6) und den Ergebnissen (Kapitel 8) aufgeführt.

Tab. 6-1: Übersicht zur Lokalisierung korrespondierender Abschnitte zu den verschiedenen Hypothesen

Nr.	Hypothese (siehe Abschnitt 3.3)	Hintergrund	Operationalisierung		Statistisches Verfahren	Ergebnis		
H1	Interne Konsistenz der Skala ufosV2r	2.4.4	4.		6.5.2 ( $\alpha$ )	8.2.2		
					6.6.2.1 (Composite Reliability)	8.4.3		
H2	Faktorielle Validität der Skala ufosV2r	2.4.4	4.		6.5.1 (EFA)	8.2.1		
					6.5.3 (CFA)	8.2.3		
					6.6.2.1 (PLS: Komplexes Modell)	8.5.1		
H3	Differenzierungsfähigkeit der Skala ufosV2r zwischen Online-Shops	2.4.4	4.		6.9.1 (ANOVA)	8.2.4		
H4	ufosV2f → ufosV2r	2.8.3.4	4.		6.6 (PLS: Zwei-Konstrukt-Modell)	8.4.3		
H5a	ufosV2r → Kaufintention	2.5.2	4.	5.3.1	6.6.2.1 (PLS: Komplexes Modell)	8.5.2		
H5b	ufosV2f → Kaufintention	2.5.2	4.	5.3.1				
H6a	ufosV2r → Kaufentscheidung	2.5.2	4.	5.3.2	6.7 (MLR)	8.6.2		
H6b	ufosV2f → Kaufentscheidung							
H7a	ufosV2r → Vertrauen	2.7.1	4.	5.2.2	6.6.2.1 (PLS: Komplexes Modell)	8.5.2		
H7b	ufosV2f → Vertrauen							
H8a	ufosV2r → Einkaufsvergnügen	2.7.2	4.	5.2.4				
H8b	ufosV2f → Einkaufsvergnügen							
H9a	Ästhetik → ufosV2r	2.7.3	5.2.1	4.				
H9b	Ästhetik → ufosV2f							
H10	Ästhetik → Kaufintention	2.6.1	5.2.1	5.3.1				
H11	Vertrauen → Kaufintention	2.6.2	5.2.2	5.3.1				
H12	Wahrge. Produktanzahl → Kaufintention	2.6.3	5.2.3	5.3.1				
H13	Einkaufsvergnügen → Kaufintention	2.6.4	5.2.4	5.3.1				
H14	Wahrge. Preis → Kaufintention	2.6.5	5.3.5.	5.3.1				
H15	Service → Kaufintention	2.6.6	5.2.6	5.3.1				
H16	Ästhetik → Einkaufsvergnügen	2.7.4	5.2.1	5.2.4				
H17	Wahrge. Produktanzahl → Vertrauen	2.7.5	5.2.3	5.2.2				
H18	Bekanntheit → Vertrauen	2.7.6	5.2.7	5.2.2				
H19	Einkaufsvergnügen x Drucker → Kaufintention	2.7.7.1	5.2.4	5.1			5.3.1	
H20	Ästhetik x Drucker → Kaufintention	2.7.7.1	5.2.1	5.1			5.3.1	
H21	Kaufintention → Kaufentscheidung	2.1.6.2	5.3.1	5.3.2			6.7 (MLR)	8.6.2

Anmerkung: Hypothese (H), Partial Least Squares (PLS), multinomiale logistische Regression (MLR), einfaktorielle Varianzanalyse (ANOVA), explorative Faktorenanalyse (EFA), Cronbachs Alpha ( $\alpha$ ), konfirmatorische Faktorenanalyse (CFA).



## 6.1 Analyse und Behandlung fehlender Werte

Wie in nahezu jeder Befragung sind auch im geplanten Experiment fehlende Werte in Form von Item-Nonresponses aufgetreten (vgl. Schafer & Graham, 2002). Dabei kann zwischen zwei unterschiedlichen Formen der Nicht-Beantwortung unterschieden werden. Einerseits erhielten die Probanden bei der Beantwortung der Items die Option, „keine Angabe“ zu machen (siehe Abschnitt 4.2.5.2, Kodierung mit dem Wert 7). Andererseits kann ein Missing-Value dadurch entstanden sein, dass gar keine Beantwortung eines Items vorgenommen wurde (Kodierung mit Wert -1). Es lassen sich keine Schlüsse aus dem Auftreten des einen oder anderen Fehlertypus ziehen, da die kognitiven Prozesse auf Seiten der Teilnehmer bei einer Nicht-Antwort unbekannt bleiben. Daher erfolgt keine getrennte Analyse beider Fehlertypen. Im Folgenden wird dargestellt, wie methodisch mit fehlenden Werten umgegangen wird.

### 6.1.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte

Zeigt sich bei einem Teilnehmer ein sehr hoher Anteil fehlender Werte, so kann dies ein Indiz auf ein zu niedriges Maß an Teilnahmebereitschaft darstellen. Entsprechend können die vorhandenen Daten eines solchen Teilnehmers das nötige Maß an Genauigkeit vermissen lassen, also einen hohen Messfehler aufweisen. Des Weiteren beruhen die angewandten Methoden zur Ersetzung fehlender Werte darauf, dass auf Grundlage vorhandener Daten eines Falles auf die fehlenden Werte desselben Falls geschlossen wird (siehe Abschnitte 6.1.3 und 6.1.4). Ein solches Vorgehen erscheint jedoch nur dann angemessen, wenn fallweise ein Mindestmaß an Daten vorliegt. Es wird folglich ein Cut-Off von höchstens 20% fehlender Werte pro Fall angesetzt; hat ein Teilnehmer also mehr als 20% der Items nicht beantwortet, wird er aus dem Datensatz ausgeschlossen (vgl. Cohen, Cohen, West & Aiken, 2003, S. 433).

### 6.1.2 Analyse des Fehlendmechanismus

In einem Datensatz können drei verschiedene Formen des Auftretens fehlender Werte (auch *Fehlendmechanismen*) vorliegen:

*Missing Not At Random* (MNAR, auch bezeichnet als *Missing Non-Ignorable At Random*): Bei diesem Mechanismus ist die Ursache für das Fehlen von Werten nicht im Rahmen der Gesamterhebung gemessen worden (Graham, Cumsille & Elek-Fisk, 2002, S. 89). Es besteht somit keine Möglichkeit, Ursacheinformationen bei der Ersetzung der Missings heranzuziehen. Diese Form der Nichtantwort kann entsprechend nicht ignoriert werden. Als weiteres Kriterium für diese Form von Missing Data nennen Tabachnik und Fidell (2006, S. 63), dass das Auftreten fehlender Werte in einem Zusammenhang zur abhängigen Variable steht.

*Missing At Random* (MAR): Beim MAR-Mechanismus wird unterstellt, dass die Wahrscheinlichkeit des Auftretens fehlender Werte von den erhobenen Variablen abhängt. Es besteht keinerlei Abhängigkeit dieser Wahrscheinlichkeit zu Variablen bzw. Bedingungen, die nicht auch im Rahmen der Untersuchung erfasst wurden (vgl. Rubin, 1976).

*Missing Completely At Random* (MCAR): Dieser unterstellt ein gänzlich zufälliges Auftreten fehlender Werte im Datensatz; es besteht dabei kein Zusammenhang zwischen den erhobenen Variablen und dem Missingmechanismus (vgl. Göthlich, 2006, S. 135).

Zur Prüfung, inwiefern die abhängigen Variablen in einem Zusammenhang zum Fehlendmechanismus stehen, wird jeweils für beide Shops, die die Teilnehmer im Experiment kennen lernen und bewerten, die bivariate Korrelation zwischen dem fallweisen Prozentteil an Missing Data und dem Skalenwert der Kaufintention berechnet. Ein hoher, signifikanter Korrelationskoeffizient stellt ein Indiz für einen MNAR-Mechanismus dar. Ein kleiner, insignifikanter Koeffizient ist somit erwünscht.

Die Software SPSS bietet den *MCAR-Test nach Little* an, mit Hilfe dessen geprüft werden kann, inwiefern der Fehlendmechanismus eines Datensatzes als MCAR gelten kann (vgl. Brosius, 2004). Bei einem insignifikanten Ergebnis kann ein zufälliges Auftreten fehlender Werte unterstellt werden (vgl. Tabachnik & Fidell, 2006, S. 63). Es werden alle Indikatoren einbezogen, die die Teilnehmer im Experiment beantworten.

### **6.1.3 Mittelwertbasierte Ersetzung bei reflektiven Indikatoren**

Gemäß einer Empfehlung von Graham et al. (2002, S. 108) kommt für die reflektiven Variablen eine besondere Form der Mittelwertersetzung fehlender Werte zur Anwendung. Im Fall reflektiver Variablen sind hohe Korrelationen zwischen den Indikatoren zu erwarten, da sie Folgen derselben Ursache darstellen (vgl. Abschnitt 2.8.1; Churchill, 1979). Entsprechend liegt es nahe, einen fehlenden Wert für einen reflektiven Indikator fallweise durch den Mittelwert der übrigen Indikatoren derselben Skala zu ersetzen. Diese Methode wird in der vorliegenden Untersuchung für die reflektiven Skalen Usability, Vertrauen, Kaufintention, wahrgenommene Größe, Einkaufsvergnügen sowie Ästhetik verfolgt. Das geschilderte Vorgehen kann dann keine Anwendung finden, wenn ein Teilnehmer bei einer Skala keinen einzigen Indikator beantwortet hat. Für diese wenigen Fälle werden fehlende Werte mit Hilfe der im Folgeabsatz beschriebenen Methode der *Multiplen Imputation* ersetzt.

### **6.1.4 Multiple Imputation bei formativen Indikatoren**

Eine Anwendung der oben geschilderten Ersetzungsprozedur auf Basis eines Skalenmittelwertes kann für formative Indikatoren nicht erfolgen, da hohe Korrelationen zwischen den Indikatoren zwar vorliegen können, aber nicht müssen (vgl. Abschnitt 2.8.2). Als Alternative ließe sich eine Ersetzung fehlender Werte durch die Mittelwerte auf Indikatorebene vornehmen, da der Mittelwert als ein guter Schätzer für den erwarteten Wert angesehen werden kann, insofern keine weiteren Informationen vorliegen (vgl. Bortz & Döring, 1999, S. 310). Zwar wird der Mittelwert der jeweiligen Variable bei diesem Vorgehen nicht verändert, allerdings führt die Methode zu einer Varianzeinschränkung der Indikatoreausprägung. Auch ein listenweiser bzw. paarweiser Ausschluss von Fällen mit fehlenden Werten erscheint unangemessen, da dieses Vorgehen zu einer erheblichen Reduzierung des zugrunde gelegten

Datensatzes und damit zu einer Nichtberücksichtigung vorhandener Daten führt (vgl. Göthlich, 2006, S. 137 ff.).

Als die beste Methode zur Ersetzung fehlender Werte wird vielfach die Technik der *Multiplen Imputation* angesehen (vgl. Tabachnik & Fidell, 2006). Bei diesem Verfahren werden für jeden fehlenden Wert  $m$  mögliche Ersetzungswerte auf Basis von Regressionsanalysen geschätzt, die alle einer vorhergesagten Verteilung entstammen (Schafer & Olsen, 1998). Es ergeben sich somit  $m$  Datensätze, die in einem nächsten Schritt auf Basis einfacher statistischer Regeln (vgl. Rubin, 1976) miteinander kombiniert werden, wobei die geschätzten Werte und ihre Standardfehler die mit der Ersetzung verbundene Unsicherheit berücksichtigen (Graham et al., 2002). Die Durchführung dieser Prozedur ist an Voraussetzungen geknüpft. Zwar beruht die Vorhersage auf der multivariaten Normalverteilung, allerdings erweist sich das Verfahren in der Praxis als sehr robust gegenüber der Verletzung von Verteilungsannahmen (Schafer & Graham, 2002). Des Weiteren wird ein Missing at Random (MAR) Fehlendmechanismus vorausgesetzt.

Die Multiple Imputation der fehlenden Werte wird mit Hilfe der Software *Norm* Version 2.03 vorgenommen (Graham et al., 2002; Schafer, 1999). Im Imputationsdatensatz werden die Bewertungen von Shop 1 und Shop 2 zusammengefasst, so dass sich die Fallzahl verdoppelt. Im ersten Schritt der Multiplen Imputation wird ein EM (*Expectation Maximation*) Algorithmus berechnet. Dieser stellt eine allgemeine Technik dar, um Modelle an unvollständige Datensätze anzupassen, wobei Beziehungen zwischen fehlenden Werten und den unbekanntem Parametern des Datenmodells hergestellt werden. Die geschätzten Parameter werden für den zweiten Schritt, die *Daten Augmentation*, herangezogen, bei der  $m=50$  verschiedene imputierte Datensätze erzeugt werden. Die Schätzungen der fehlenden Werte erfolgt gemäß einer Empfehlung von Graham et al. (2002) bis auf zwei Stellen hinter dem Komma gerundet. Eine Anzahl von 50 Datensätzen erscheint ausreichend, da die Schätzgenauigkeit ab  $m=5$  imputierten Datensätzen nur noch geringfügig steigt (Schafer & Olsen, 1998). Im letzten Schritt werden die 50 imputierten Datensätze zu einem verdichtet.

## 6.2 Verteilungsprüfung

Als Voraussetzung zur Berechnung vieler statistischer Verfahren muss eine Normalverteiltheit der einbezogenen Variablen gegeben sein (Tabachnick & Fidell, 2006, S. 78 ff.). Um zu prüfen, inwiefern die in der Untersuchung erhobenen Indikatoren univariat normalverteilt sind, soll für jede Variable die Berechnung eines *Kolmogorov-Smirnov*-Tests erfolgen sowie die Schiefe und der Exzess bestimmt werden.

Mit Hilfe eines Kolmogorov-Smirnov-Tests wird 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.). Um einen Beta-Fehler zu vermeiden, wird das Signifikanzniveau mit  $p < 0,1$  entsprechend hoch angesetzt. Trotz eines signifikanten Ergebnisses ist es allerdings möglich, dass die Werte der Variable in der Grundgesamtheit annähernd normalverteilt sind, was für viele statistische Verfahren als Voraussetzung genügt (Brosius, 2004, S. 755).

Zur genaueren Beschreibung der Verteilungsform sollen mit der *Schiefe* ( $Sch$ ) und dem *Exzess* ( $Ex$ , auch *Kurtosis*) zwei Komponenten der Normalverteilung bestimmt werden (vgl. Bortz & Döring, 2005, S. 46). Bei einem Wert  $Sch < 0$  für die Schiefe ist der Verteilungsverlauf rechtssteil (linke Seite der Verteilung ist zu lang); ein Wert  $Sch > 0$  weist auf eine Linkssteilheit der Verteilung hin (rechte Seite ist zu lang). Der Exzess-Wert gibt an, wie stark die Wölbung der Verteilung ausfällt (breit- vs. spitzgipflig). Der Exzess-Wert einer Normalverteilung liegt bei  $Ex = 0,263$ ; in der verbreiteten Statistiksoftware wird der Exzess-Wert jedoch transformiert, so dass der erwartete Wert bei  $Ex = 0$  liegt. Liegt der Exzess oberhalb dieses Wertes, so indiziert dies im Vergleich zur Normalverteilung einen zu steilen Höhepunkt, wohingegen ein Wert  $Ex < 0$  eine zu flache Verteilung anzeigt. Die z-Werte der Schiefe und des Exzess ergeben sich dadurch, dass die Werte durch ihren jeweiligen Standardfehler geteilt werden. Auf Basis der z-Werte lässt sich berechnen, inwiefern sich diese Parameter signifikant von denen einer Normalverteilung unterscheiden ( $p < 0,01$ ). Dabei ist eine Abhängigkeit von der Stichprobengröße zu berücksichtigen. Je höher das  $N$ , umso eher zeigen die z-Werte für Kurtosis und Schiefe an, die Nullhypothese solle verworfen werden. Folglich können signifikante Ergebnisse auftreten, obwohl die Verteilung einer Variable nicht in relevantem Maße von der Normalverteilung abweicht. Daher empfehlen Tabachnik und Fidell (2006, S. 80) bei großem  $N$  eine Betrachtung der Verteilungsformen, anstelle inferenzstatistische Tests anzuwenden. Auch sei eine Analyse der Höhe des Schiefe-Wertes bei großen Stichproben relevanter als dessen Signifikanz.

### 6.3 Prüfung auf Vorliegen eines Common Method Bias

Zur Untersuchung, inwiefern ein *Common Method Bias* in einem Datensatz eine Rolle spielt, soll der *Harman's Single-Factor Test* zur Anwendung kommen (vgl. Podsakoff et al., 2003; Spector, 2006). Dabei werden alle Indikatoren der Untersuchung in eine explorative Faktorenanalyse ohne Rotation der Faktoren einbezogen. Als Indiz für einen Common Method Bias gilt, wenn zum einen lediglich ein wesentlicher Faktor extrahiert wird und dieser zum anderen 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 wahre (unverzerrte) Kovarianz der Indikatoren in aller Gänze überschattet. Dies sei jedoch eher unwahrscheinlich (Podsakoff et al., 2003). In Ermangelung eines sensitiveren Analyseansatzes wird der besagte Harman's Single-Factor Test trotz der genannten Einschränkungen angewandt.

### 6.4 Vorgehen zur Identifikation problematischer Fälle

Es ist stets zu befürchten, dass auf Seiten einzelner Probanden des Experimentes eine geringere Motivation zur Teilnahme vorherrschte, was sich negativ auf die Güte der Beantwortung auswirken und in höheren Messfehlern der erhobenen Variablen niederschlagen kann. Hohe Messfehler bergen das Problem, dass sich vorhandene Zusammenhänge mit Hilfe der angewandten statistischen Verfahren nicht bestätigen lassen (vgl. Cote & Buckley, 1988). Entsprechend gilt es, diejenigen Teilnehmer zu identifizieren, die die Beantwortung der präsentierten Items nicht sorgfältig vornahmen.

Es werden drei Kriterien betrachtet, die als Hinweise auf problematische Fälle dienen sollen. Erweisen sich die anhand dieser Kriterien identifizierten Fälle im Rahmen der weiteren Auswertungsschritte als Ausreißer, so erscheint ein Ausschluss aus den jeweiligen Analysen gerechtfertigt. Es kann argumentiert werden, dass der jeweilige Ausreißer nicht der zugrunde gelegten Population angehört (vgl. Tabachnick & Fidell, 2006, S. 73). Zum einen wird geprüft, welche Teilnehmer eine auffällig lange bzw. kurze Bearbeitungsdauer benötigt haben (Abschnitt 6.4.1). Des Weiteren erfolgt eine Untersuchung der Varianz im Antwortverhalten der Teilnehmer; ein besonders gleichförmiges Antwortprofil wird als Indiz für eine unreflektierte Bewertung gewertet (Abschnitt 6.4.2). Außerdem wird analysiert, welche Teilnehmer negativ kodierte Indikatoren gegenüber inhaltlich nahe stehenden Items erwartungskonträr bzw. inkonsistent beantworteten (Abschnitt 6.4.3).

#### 6.4.1 Teilnehmer mit auffällig langer bzw. kurzer Bearbeitungsdauer

Angesichts der hohen Anzahl an Indikatoren (219 im Maximalfall) stellt sich die Frage, inwiefern bei einer sehr kurzen Bearbeitungsdauer auf Seiten von Probanden eine angemessene kognitive Auseinandersetzung mit den abgefragten Aspekten stattgefunden haben kann. Schließlich dürften Teilnehmer, die das Shopinterface nicht in ausreichendem Maße

kennen gelernt haben, kaum in der Lage gewesen, eine differenzierte Bewertung eines Shops abzugeben. Bei einer Durchführungsdauer von 45 Minuten müsste alle 12 Sekunden ein Item beantwortet worden sein, vorausgesetzt die Vp würde die gesamte Zeit mit der Beantwortung der Fragen verbringen. Führt man sich jedoch vor Augen, dass die Versuchsdurchführung auch das Lesen sämtlicher Versuchspersoneninstruktionen, die Interaktion mit beiden bewerteten Shops sowie die Kaufentscheidung umfasst, so erscheint eine derartige Durchführungsdauer sehr kurz, und es drängen sich Zweifel bzgl. der Sorgfalt auf. Es ist anzunehmen, dass erlebter Stress oder Eile auf Seiten der Vpn einen Einfluss auf den Prozess ihrer Informationssuche genommen haben könnte (vgl. Kulviwat, Guo & Engchanil, 2004; Weening & Maarleveld, 2002). Auch eine besonders lange Durchführungsdauer ließe sich als Hinweis auf ein problematisches Antwortverhalten deuten. So erscheint fraglich, ob eine Vp über einen Zeitraum von mehr als zwei Stunden konzentriert arbeiten kann. Andererseits kann eine vergleichsweise lange Bearbeitungsdauer auch darauf hinweisen, dass eine Vp ein hohes Maß an Genauigkeit an den Tag gelegt und seine Antworten genau abwägt hat.

Auf Grundlage der angeführten Argumentation sollen alle Teilnehmer mit auffällig langer bzw. kurzer Bearbeitungsdauer als potenziell problematische Fälle gewertet werden. Als besonders kurz wird dabei eine Dauer von unter 45 Minuten angesehen; als auffällig lang soll eine Dauer von über zwei Stunden gelten. Es ließe sich kritisieren, dass die gewählten kritischen Zeiten von 45 bzw. 120 Minuten keine statistische Grundlage besitzen. Zwei Gründe sprechen gegen eine Festlegung im Sinne einer Abweichung in Standardabweichungen. Einerseits kann nicht dieselbe Abweichung vom Mittelwert nach oben und unten gewählt werden, da zu erwarten ist, dass die Verteilung der Teilnahmedauer nicht einer Normalverteilung folgen; eher ist eine Rechtsschiefe zu vermuten, da mehr Ausreißer in Form besonders langer als besonders kurzer Teilnahmedauern auftreten. Zum zweiten sind längere Bearbeitungsdauern nicht grundsätzlich unerwünscht, da sie, wie bereits angesprochen, auch auf eine große Sorgfalt der Teilnehmer schließen lassen. Daher sollen lediglich diejenigen Probanden als auffällig gewertet werden, die besonders viel Zeit benötigt haben.

#### **6.4.2 Fallweise Analyse der Varianz im Antwortverhalten**

Eine niedrige Teilnahmemotivation bzw. eine mangelnde Sorgfalt eines Probanden äußert sich in Befragungen häufig darin, dass er kein differenziertes Antwortverhalten zeigt, sich also durch den Fragebogen klickt, ohne sich kognitiv mit den einzelnen Items zu beschäftigen. Um derartige „Durchklicker“ zu identifizieren, soll eine fallweise Analyse der Antwortvarianz vorgenommen werden. Es werden diejenigen Teilnehmer als problematisch gewertet, die ein besonders gleichförmiges Bewertungsprofil aufweisen. Dazu werden für jeden Teilnehmer alle Indikatoren betrachtet, die sich auf die Bewertung des ersten bzw. des zweiten Shops beziehen; die Angaben zur Demographie und Vorerfahrung werden nicht mit berücksichtigt. Herangezogen werden die vier Items zur Kaufintention (ki1, ki2, ki3, ki4), die vier Items zum Vertrauen (t1, t2, t3, t4), die neun Items zur reflektiven Usability (ufosV2r;

ur1, ur2c, ur3, ..., ur9), die 58 Items zur formativen Usability (ufosV2f; uf1, uf2, uf3, ..., uf58), die drei Items zur wahrgenommenen Größe (wg1, wg2, wg3), die beiden Items zum Ästhetik (a1, a2), die beiden Items zum Einkaufsvergnügen (se1, se2), die sechs Items zum Service (sv1, sv2, sv3, ..., s6) sowie das Item zum wahrgenommenen Preis (subp). Als Kriterium für die Gleichförmigkeit der Beantwortung wird die Streuung über alle genannten Indikatoren berechnet. Als kritischer Wert wird eine Streuung von  $SD \leq 1$  festgelegt. Liegt die Streuung unterhalb dieses Wertes, so fällt die Beantwortung des Fragebogens derart undifferenziert aus, dass ein „Durchklickverhalten“ unterstellt werden kann.

### 6.4.3 Konsistenz des Antwortverhaltens

Als weiterer Ansatz zur Identifikation eventuell problematischer Fälle wird geprüft, inwiefern die Beantwortung negativ gepolter Items inhaltlich konsistent zur Beantwortung der übrigen positiv formulierten Items derselben reflektiven Skala erfolgte. Der größte Teil der im Rahmen des Fragebogens gestellten Items war positiv formuliert. Eine negative Formulierung von Items wurde nur selten verwendet, da ihre Beantwortung ein höheres Maß an kognitiver Anstrengung erfordert und in einem höherer Messfehler resultieren kann (Abschnitt 4.2.4). Es werden die beiden reflektiven Skalen zur Usability und zum Einkaufsvergnügen für diese Analyse herangezogen. Dabei werden ausschließlich die Bewertungsdaten zum zweiten Shop verwendet; es ist zu vermuten, dass sich ein inkonsistentes Antwortverhalten aufgrund mangelnder Motivation eher zu einem späteren Zeitpunkt im Versuch gezeigt hat. Als Hinweis auf eine inkonsistente Bewertung soll eine Abweichung des negativ gepolten Items ur2c von mehr als 4 vom Mittelwert der übrigen Indikatoren der reflektiven Usability Skala gelten. Im Fall der zwei Items umfassenden reflektiven Skala zum Einkaufsvergnügen des zweiten Shops wird ebenfalls eine Abweichung von 4 zwischen den beiden Indikatoren se1 und se2c als Hinweis auf ein Durchklick-Verhalten gewertet. Schließlich erscheint es wenig einsichtig, dass die Nutzung eines Shops einerseits Spaß macht (se1) und gleichzeitig auch frustrierend erlebt wird (se2c, siehe Abschnitt 2.6.4).

Weitere umgepolte Indikatoren finden sich neben den beiden bereits herangezogenen ur2c und se2c lediglich in den formativen Skalen des Fragebogens. Es erscheint jedoch unangemessen, diese einer entsprechenden Analyse zu unterziehen, da im Fall von formativen Skalen nicht von einer hohen Interkorrelation zwischen den Indikatoren ausgegangen werden kann (siehe Abschnitt 2.8.2). Folglich lässt sich ein differenziertes Antwortverhalten nicht von einem Durchklicken unterscheiden.

Kritisch ist in Bezug auf das angewandte Vorgehen anzumerken, dass sich auf diese Weise lediglich diejenigen „Durchklicker“ identifizieren lassen, die insgesamt eine besonders positive bzw. negative Bewertung des zweiten Shops abgeben. Im Fall von Teilnehmern, die einen Shop als mittelmäßig bewerteten, erweist sich dieses Verfahren als ungeeignet, um ein unreflektiertes Antwortverhalten aufzudecken, da sich in den negativ gepolten Indikatoren keine Abweichungen gegenüber den positiv formulierten Items zeigen werden.

## 6.5 Analyse der Güte reflektiver Maße auf Prädiktorebene

Zur Güteprüfung der reflektiven Skalen, die in der Untersuchung auf Prädiktorebene herangezogen werden, wird ein drei-stufiges Vorgehen verfolgt, das sich an Empfehlungen von Homburg und Giering (1996) orientiert. Die von diesen Autoren vorgeschlagene Validierungsprozedur hat sich im Bereich der Marketing-Forschung bereits weitgehend etabliert und lässt sich auf die im Rahmen der vorliegenden Studie erhobenen reflektiven Maße anwenden (vgl. Panten, 2005). Die drei Untersuchungsschritte sind die Folgenden:

*Schritt 1:* Zunächst werden sämtliche reflektiven Indikatoren in eine *explorative Faktorenanalyse* (EFA) einbezogen. Ziel dieses Schrittes ist es zu untersuchen, inwiefern sich die theoretisch angenommene Faktoren- bzw. Skalenstruktur bestätigt. Dazu wird eine Hauptachsenanalyse (HAA) mit schiefwinkliger Rotation berechnet; die nicht-orthogonale Rotation erleichtert eine Interpretation der extrahierten Komponenten (vgl. Backhaus et al., 2003, S. 300). Zeigt sich eine erwartungskonträre Faktorenstruktur, so gilt es, eine theoriegeleitete Anpassung der Skalen vorzunehmen. Die Voraussetzungen zur Berechnung einer EFA sowie die zugrunde gelegten Gütekriterien dieses Verfahrens werden in Abschnitt 6.5.1 dargestellt.

*Schritt 2:* Im zweiten Untersuchungsschritt wird jede einzelne reflektive Skala nach Kriterien der klassischen Testtheorie untersucht bzw. optimiert. Zunächst wird jeweils eine weitere *explorative Faktorenanalyse* berechnet, um die *Eindimensionalität* der Komponente zu prüfen. Weiterhin werden die Reliabilitäten der Einzelskalen in Form des internen Konsistenzmaßes *Cronbachs Alpha* bestimmt. Als Gütekriterien auf Einzelitemebene dienen die *korrigierte Trennschärfe*, die *Homogenität*, der *Anteil fehlender Werte* sowie die *Itemschwierigkeit*. Abschnitt 6.5.2 stellt die zugrunde gelegten Kriterien im Einzelnen dar.

*Schritt 3:* Schließlich erfolgt in einem dritten Schritt mit Hilfe einer *konfirmatorischen Faktorenanalyse* (CFA) eine weitere Prüfung der unterstellten Skalenstruktur. Diese wird durch die Berechnung eines mehrfaktoriellen kovarianzbasierten Strukturgleichungsmodells durchgeführt. In Abschnitt 6.5.3 werden die Voraussetzungen und Gütekriterien der CFA vorgestellt.

Ein Vorteil des von Homburg und Giering (1996) vorgeschlagenen Vorgehens besteht darin, dass es Güteprüfungsverfahren der so genannten ersten und zweiten Generation kombiniert (vgl. Panten, 2005, S. 228). Als Methoden der ersten Generation kommen die explorative Faktorenanalyse sowie die weiteren Kriterien der klassischen Testtheorie zur Anwendung; im Bereich der Marketing-Forschung wurden diese Ansätze der Reliabilitätsbestimmung bzw. Validierung vor allem von Churchill (1979) geprägt. Die konfirmatorische Faktorenanalyse weist gegenüber diesen klassischen Kriterien insofern Vorteile auf, als dass vorhandene Erkenntnisse und theoretische Erwägungen gezielt in die Analyse mit einfließen können (vgl. Homburg & Pflesser, 2000, S. 417); daher wird dieses Verfahren als eines der zweiten Generation angesehen.



### 6.5.1 Explorative Faktorenanalyse

Mit Hilfe einer *explorativen Faktorenanalyse* (EFA) wird untersucht, inwiefern sich die theoretisch angenommene Struktur der reflektiven Skalen auf Prädiktorebene bestätigt. Es wird eine *Hauptachsenanalyse* (HAA) berechnet unter Einbeziehung der Indikatoren zu den Konstrukten Usability (reflektiv; ufosV2r-Skala), Einkaufsvergnügen, Produkt-Involvement, Vertrauen, wahrgenommene Größe und Ästhetik. Widaman (1993) zeigte in einer Monte-Carlo-Studie eine Überlegenheit der HAA gegenüber der Hauptkomponentenanalyse (HKA). Es erfolgt eine schiefwinklige Rotation mit der Technik *Direct Oblimin*, die Korrelationen zwischen den Faktoren zulässt. Dies erscheint angemessen, da Zusammenhänge zwischen den prädiktierenden Konstrukten angenommen werden (siehe Abschnitt 2.7). Durch die Festlegung eines Delta-Wertes kann a priori festgelegt werden, welcher Grad an Korrelation zwischen den Faktor zugelassen werden soll. Bei einem Delta-Wert  $\delta=0$  können die Faktoren maximal korrelieren, ein Wert von  $\delta=-4$  entspricht einer minimalen Korrelation. Da Korrelationen mittleren Ausmaßes zwischen den betrachteten Erfolgsfaktoren aus theoretischer Sicht durchaus plausibel erscheinen (siehe Abschnitt 2.7), wird ein Wert von  $\delta=-2$  gewählt. Als Kriterium zur Bestimmung der Faktorenanzahl wird das Kaiser-Eigenwert-Kriterium herangezogen, wonach alle Faktoren extrahiert werden, die einen Eigenwert  $\lambda > 1$  aufweisen (vgl. Heck, 1998, S. 189). Dieses Vorgehen erscheint gegenüber einer a priori Festlegung der Faktorenanzahl auf Grundlage der Anzahl einbezogener Skalen angemessener, da es eher dem explorativen Rational dieses Untersuchungsschrittes entspricht.

#### 6.5.1.1 Voraussetzungen und Kriterien der Dateneignung

Im Folgenden werden zunächst die Voraussetzungen zur Berechnung einer explorativen Faktorenanalyse und sowie Kriterien zur Eignung der Daten vorgestellt:

*Stichprobengröße und Anzahl der Indikatoren pro Faktor:* Für jeden theoretisch unterstellten Faktor sollten nach Bühner (2004, S. 157) mindestens drei, besser vier oder mehr Items vorliegen. Dabei gilt es eine Abhängigkeit von den psychometrischen Eigenschaften der einzelnen Items zu berücksichtigen, die anhand der Kommunalitäten beurteilt wird. Liegen geringe Kommunalitäten ( $h^2 < 0,5$ ) und gering besetzte Faktoren (nur drei oder vier Indikatoren pro Faktor) vor, so ist ein  $N=500$  als *sehr gut* anzusehen, ein  $N=1000$  ist *exzellent*. Je größer die Stichprobe ist, umso stabiler fällt die Faktorenlösung aus, da der Stichprobenfehler abnimmt (MacCallum, Widaman, Zhang & Hong, 1999).

*Linearität:* Die Berechnung einer explorativen Faktorenanalyse kann nur dann erfolgen, wenn zwischen den einbezogenen Variablen ausreichend hohe *lineare* Zusammenhänge bestehen.

*Ausreißer und Verteilungen:* Die Faktorenanalyse ist nicht direkt an Verteilungsannahmen geknüpft (Backhaus et al., 2003, S. 273). Allerdings sollten die einzelnen Variablen keine Ausreißer aufweisen. Die Verteilungen der Indikatoren sollten sich zudem ähneln, da die Höhe der Korrelationen der Variablen durch ihre Verteilungsformen beeinflusst wird.

*Eignung der Korrelationsmatrix:* Die Indikatoren müssen untereinander substantielle Korrelationen aufweisen, damit eine Faktorenanalyse sinnvoll berechnet werden kann. Zur Prüfung der Eignung der Korrelationsmatrix stehen diverse Kriterien zur Verfügung:

Einerseits wird der *Kaiser-Meyer-Olkin-Koeffizient* (KMO) bestimmt, der die Höhe der Korrelationen zwischen den Faktoren und den Variablen widerspiegelt. Die Bewertung des KMO-Koeffizienten wird folgendermaßen vorgenommen (vgl. Kaiser & Rice, 1974, S. 111): 0,9 bis 1,0 *fabelhaft*; 0,8 bis unter 0,9 *recht gut*; 0,7 bis unter 0,8 *mittelprächtigt*; 0,6 bis unter 0,7 *mäßig*; 0,5 bis unter 0,6 *schlecht*; unter 0,5 *inakzeptabel*.

Als Entsprechung zum KMO-Maß werden auf Indikatorebene die *MSA-Werte* (Measure of Sampling Adequacy) betrachtet, die der Diagonalen der Anti-Image-Korrelationsmatrix zu entnehmen sind. Die Bewertung der Wertausprägung erfolgt dabei analog zum KMO-Koeffizienten, wie oben beschrieben. Bei einem  $MSA < 0,5$  sollte der entsprechende Indikator eliminiert werden (Backhaus et al., 2003, S. 310).

Mit Hilfe des *Bartlett-Tests auf Sphärizität* wird die Nullhypothese getestet, dass alle Korrelationen null sind. Dabei besteht eine Abhängigkeit von der Stichprobengröße: die Nullhypothese, alle Korrelationen seien gleich null, wird bei steigendem N eher verworfen, da die Prüfgröße annähernd  $\chi^2$ -verteilt ist. Entsprechend muss als Voraussetzung zur Berechnung dieses Tests eine Normalverteilung der Variablen in der Erhebungsgesamtheit gegeben sein (vgl. Backhaus et al., 2003, S. 274). Die in der *Anti-Image-Korrelationsmatrix* dargestellten negativen Werte der partiellen Korrelationskoeffizienten sollten nahe null liegen.

#### 6.5.1.2 Beurteilung der Faktorlösung

Bei obliquen Rotation entsprechen die *Ladungsgewichte*  $a_{ij}$  aus der Mustermatrix den partiellen Regressionsgewichten der Indikatoren. Sie geben Aufschluss über die Wichtigkeit der einzelnen Indikatoren für den jeweiligen Faktor unter Herauspartialisierung des Einflusses der übrigen Indikatoren. Es wird erwartet, dass sich eine *Einfachstruktur* ergibt, d.h. dass sich für jede Skala ein Faktor ergibt, auf dem die Items einer Skala hoch laden. Die Ladungen auf den übrigen Faktoren sollten gering ausfallen. Treten hohe Nebenladungen auf oder ergibt sich eine Skalenstruktur, die nicht der theoretisch angenommenen entspricht, so muss eine theoriegeleitete Anpassung vorgenommen werden. Unter Umständen sind Zusammenfassungen von Skalen oder Eliminationen einzelner Items aus theoretischer Sicht gerechtfertigt. Es gilt die *Kommunalitäten*  $h^2$  der Indikatoren zu prüfen, die 50% nicht wesentlich unterschreiten sollten (*Fürntratt-Kriterium*; vgl. Fürntratt, 1969; Bühner, 2004, S. 167); die Kommunalität gibt an, wie viel Streuung in den Indikatoren durch alle Faktoren zusammengenommen erklärt wird. Der *Eigenwert*  $\lambda_j$  gibt an, wie viel Varianz eines Faktors  $j$  durch alle Indikatoren aufgeklärt wird. Anhand seiner Ausprägung kann die Wichtigkeit des Faktors eingeschätzt werden. Nach Heck (1998, S. 187) ist in der Regel zu erwarten, dass ein Anteil von 50% bis 80% der Gesamtvarianz aufgeklärt wird.

### 6.5.2 Güteprüfung nach Kriterien der klassischen Testtheorie

Im Folgenden werden die Kriterien des zweiten Prüfschrittes zur Messgüte der reflektiven Skalen nach Homburg und Giering (1996) vorgestellt; ergänzend werden einige weitere Kriterien herangezogen. In der psychologischen Test- bzw.- Fragebogenkonstruktion werden einige der Kriterien im Rahmen der *Itemanalyse* verwendet (vgl. Bortz & Döring, 2006, S. 217 ff.).

Als Anspruchsniveau für das interne Konsistenzmaß *Cronbachs Alpha* ( $\alpha$ ; Cronbach, 1951), das zur Einschätzung der Reliabilität dient, wird vielfach ein Wert von  $\alpha > 0,7$  als angemessen angesehen, wobei auf die Studien von Nunnally (1978, S. 245) Bezug genommen wird. Lance, Butts und Michels (2006) weisen jedoch darauf hin, ein Mindestwert für Alpha sei stets unter Berücksichtigung des jeweiligen Untersuchungssettings anzusetzen. Danach kann es mitunter sinnvoll sein, einen Standard von  $\alpha > 0,9$  bzw.  $\alpha > 0,95$  festzulegen – vor allem dann, wenn auf Grundlage von Versuchsergebnissen folgenreiche Entscheidungen getroffen werden. Es gilt zu beachten, dass die Höhe des  $\alpha$ -Koeffizienten von der Anzahl der Indikatoren abhängt (vgl. Homburg & Giering, 1996). Im Einzelfall kann das Cronbachs Alpha durch Elimination von Indikatoren gesteigert werden (vgl. Churchill, 1979).

Die *Itemschwierigkeit* ( $p_m$ ) eines Indikators kann Werte zwischen  $0 < p_m < 1$  annehmen und sollte im Idealfall im mittleren Bereich von  $p_m = 0,5$  liegen, da bei einem solchen Wert am ehesten sichergestellt ist, dass eine hohe Varianz in der Variablenausprägung erreicht werden kann (vgl. Fisseni, 2004, S. 43 ff). Ein Wert von  $p_m = 0$  zeigt besonders „schwere“ Items an, bei einer  $p_m = 1$  handelt es sich um ein besonders leichtes Item. Eine gute Itemschwierigkeit liegt gemäß Bortz und Döring (2005, S. 218) bei 0,2 bis 0,8.

Homburg und Giering (1996) empfehlen des Weiteren ein Heranziehen des Kriteriums der *Item to Total*-Korrelation, die im Bereich der Psychometrie unter der Bezeichnung *Trennschärfe* bekannt ist (Bortz & Döring, 2005, S. 219). Sie stellt die Korrelation eines Indikators mit der Summe aller Indikatoren seiner Skala dar. Anstelle dieses Maßes wird in der vorliegenden Studie die *korrigierte Trennschärfe* ( $r_{it}$ ) bestimmt, die als Korrelation eines Indikators mit der Summe aller *übrigen* Indikatoren derselben Skala bestimmt wird und entsprechend theoretisch einen Maximalwert von  $r_{it} = 1$  annehmen kann. Durch die Mittelung der Werte für alle Indikatoren einer Skala wird außerdem die *mittlere korrigierte Trennschärfe* bestimmt. Hohe Trennschärfeausprägungen sind erstrebenswert. Bei niedrigem Cronbachs Alpha kann gegebenenfalls eine Itemelimination auf Grundlage des Kriteriums der korrigierten Trennschärfe vorgenommen werden, um die Reliabilität der Messung zu steigern.

Die Bestimmung des *Anteils fehlender Werte* auf Variablenebene dient der Identifikation schwer zu beantwortender Items. Weist ein Indikator einen Anteil von mehr als 10% Missings auf, so wird von einer weiteren Verwendung im Rahmen der Auswertung abgesehen.

Die *Homogenität* einer reflektiven Skala gibt an, wie hoch die Indikatoren im Mittel miteinander korrelieren. Bei hoher Homogenität erfassen die Items eines Tests ähnliche Informationen (Bortz & Döring, 2005, S. 219).

Zusätzlich wird für jede Skala eine einzelne *explorative Faktorenanalyse* berechnet, um zu kontrollieren, dass jeweils nur ein Faktor extrahiert wird. Dies kann als ein Hinweis auf eine angemessene konvergente Validität der jeweiligen Skala gewertet werden (vgl. Robinson, Shaver & Wrightsman, 1991). Zudem sollte der einzelne Faktor stets mehr als 50% der Varianz der herangezogenen Indikatoren aufklären (Homburg & Gierung, 1996). Ist dies nicht der Fall, so kommt für diejenigen Items eine Elimination in Frage, die eine niedrige *Faktorladung* auf weisen.

### 6.5.3 Konfirmatorische Faktorenanalyse zur Prüfung der Skalenstruktur

Im dritten Schritt der Messgüteprüfung nach Homburg und Giering (1996) werden die reflektiven Skalen zur Erfassung der Prädiktoren einer konfirmatorischen Faktorenanalyse (CFA) unterzogen. Abschnitt 6.5.3.1 stellt die Voraussetzungen zur Berechnung einer CFA vor; teilweise unterscheiden sich diese je nach gewählter Schätzmethode. In Abschnitt 6.5.3.2 werden die verschiedenen Kriterien dargestellt, anhand derer die Passung zwischen dem geschätzten Modell und den zugrunde gelegten Daten beurteilt wird.

#### 6.5.3.1 Voraussetzungen zur Berechnung einer CFA

*Verteilungen:* Die Wahl der Schätzmethode muss vor dem Hintergrund der Voraussetzungen der verschiedenen CFA-Schätzmethoden getroffen werden. Sowohl die *Maximum Likelihood* (ML) als auch die *Generalized Least Square* (GLS) Methode setzen eine multivariate Normalverteilung sowie ein Intervallskalenniveau der einbezogenen Variablen voraus. Verletzungen der Verteilungsannahme können sich nach Byrne (2001) negativ auf die Schätzungen des  $\chi^2$ -Wertes sowie der Standardfehler auswirken. Allerdings besteht nach wie vor eine gewisse Unklarheit, als wie robust sich die verschiedenen Schätzmethoden der CFA gegenüber einer Verletzung der Normalverteilungsannahme erweisen. Curran, West und Finch (1996) zeigen, dass univariate Schiefe- und Exzesswerte über  $Sch=2,0$  bzw.  $Ex=7,0$  im Falle der gängigen *Maximum Likelihood* (ML) sowie der *Generalized Least Square* (GLS) Methode als suspekt angesehen werden müssen. Liegt eine Verletzung der Verteilungsannahmen vor, so kann die *asymptotically distribution-free* (ADF) Schätzmethode angewandt werden, die bei großen Stichproben ( $N>500$ ) und Modellen mit einem geringen Komplexitätsgrad zu genauen Ergebnissen führt (Hu & Bentler, 1998).

*Linearität:* Die Interpretation einer CFA ist nur dann sinnvoll möglich, wenn zwischen den Variablen lineare Zusammenhänge bestehen und diese nicht durch Ausreißer verzerrt sind (vgl. Bühner, 2004, S. 208).

*Stichprobengröße:* Kline (1998, S. 112) empfiehlt ein Verhältnis von 5:1 bzw. im besseren Fall 10:1 zwischen der Anzahl der einbezogenen Variablen und der Stichprobengröße. Allerdings wird die Einbeziehung von mindestens drei Indikatoren pro Faktor empfohlen, da sich dadurch die Identifizierbarkeit des Modells erhöht und die Parameterschätzungen stabiler ausfallen (vgl. Marsh, Hau, Balla & Grayson, 1998, S. 188).

*Multikollinearität:* Extrem hohe Korrelationen zwischen den Indikatoren können nach Bühner (2004, S. 208) besonders bei der ML-Schätzmethode zu Schätzproblemen führen. Es sollen entsprechend keine Indikatoren in die Analyse eingehen, die untereinander bivariate Korrelationen von  $r > 0,85$  aufweisen (siehe auch Abschnitt 6.8).

*Identifikation des Modells:* Ein potenzielles Problem bei der Berechnung einer CFA kann in der Identifikation des Modells liegen. Diese wird danach bestimmt, wie die Anzahl der zu schätzenden Parameter im Verhältnis zur Anzahl der Beobachtungen ausfällt (Bühner, 2004, S. 209). Eine Modellschätzung ist dann nicht möglich, wenn das Modell *unteridentifiziert* ist, also mehr Parameter geschätzt werden müssen als Beobachtungen vorliegen. Ist die Anzahl der zu schätzenden Parameter mit der der Beobachtungen identisch, das Modell also *gerade identifiziert*, lässt sich kein Modelltest berechnen, da die Anzahl der Freiheitsgrade null ist.

#### 6.5.3.2 Gütekriterien der CFA

Es werden die im Folgenden dargestellten Kriterien zur Gütebewertung des CFA-Modells herangezogen (vgl. Backhaus et al., 2003, S. 371). Eine Zusammenfassung der zugrunde gelegten Gütekriterien der CFA und der jeweils geforderten Anspruchsniveaus findet sich in Tabelle 6-2.

*Zuverlässigkeit der Schätzung:* Lokale Anpassungsmaße werden einerseits auf Indikatorebene und andererseits auf Faktorebene herangezogen. Die *Indikatorreliabilität* erlaubt eine Aussage, wie viel Varianz des einzelnen Indikators durch den jeweiligen Faktor aufgeklärt wird. Die Bestimmung der erfolgt nach der folgenden Formel (vgl. Homburg & Giering, 1996):

$$(3) \quad rel(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}}$$

wobei:  $\lambda_{ij}$  = die geschätzte Faktorladung zwischen dem Indikator  $x_i$  und dem zugrunde liegenden Faktor  $\xi_j$ ,

$\phi_{jj}$  = die geschätzte Varianz von  $\xi_j$  und

$\theta_{ii}$  = die geschätzte Varianz des Messfehlers  $\delta_i$ .

Relevanter als die *Indikatorreliabilität* ist die Faktorreliabilität, die anhand der folgenden Formel bestimmt wird:

$$(4) \quad rel(\xi_j) = \frac{\left( \sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj}}{\left( \sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}}$$

wobei eine Summation über alle k Indikatoren des jeweiligen Faktors vorgenommen wird.

Als strenges Kriterium zur Einschätzung der diskriminanten Validität wird das *Fornell-Larcker*-Kriterium herangezogen (Fornell & Larcker, 1981), wonach die durchschnittlich erfasste Varianz  $DEV(\xi_j)$  eines jeden Faktors größer sein muss als die quadrierten Korrelationskoeffizienten  $\phi_{ij}^2$  des jeweiligen Faktors mit allen anderen Faktoren. Wird dieses Kriterium erfüllt, so kann von einer angemessenen Diskriminanzvalidität ausgegangen werden. Die Formel zur Berechnung der durchschnittlich erfassten Varianz eines Faktors  $\xi_1$  lautet (Fornell & Larcker, 1981, S. 46):

$$(5) \quad DEV(\xi_j) = \frac{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj}}{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}}.$$

*Beurteilung der Gesamtstruktur:* Die Beurteilung des Gesamt-Fits erfolgt vielfach auf Grundlage eines  $\chi^2$ -Tests. Ein solches Vorgehen ist nach Ansicht von Homburg und Baumgartner (1995) jedoch nicht zu empfehlen, da mit der Anwendung dieses Tests einige Probleme verbunden sind. So beeinflusst beispielsweise die Stichprobengröße das Testergebnis derart, dass der  $\chi^2$ -Test bei größeren Samples schneller signifikant wird, was eine Verwerfung der Modell-Fit-Annahme zur Folge hat. Stattdessen wird ein Quotient aus dem  $\chi^2$ -Wert und der Anzahl der Freiheitsgrade herangezogen, wobei ein niedriger (hoher) Wert dieses Quotienten auf eine gute (schlechte) Modellpassung hinweist (vgl. Gerbing & Anderson, 1993, S. 41). Es werden unterschiedliche Faustregeln zur Beurteilung dieses Kriteriums verwendet; in der vorliegenden Arbeit wird gemäß einer Empfehlung von Homburg und Pflesser (2000, S. 430) ein Wert kleiner/gleich 3 angesetzt.

Neben diesem Test stehen verschiedene Fit-Indizes zur Verfügung, anhand derer die globale Modellanpassung bewertet wird. Sowohl für den GFI- als auch den AGFI-Index wird gemeinhin ein Wert von über 0,9 gefordert (Homburg & Pflesser, 2000, S. 430). Nach Hu und Bentler (1998) ist die Anwendung beider Fit-Maße im Falle der ADF- und ML-Schätzmethoden nicht zu empfehlen. Bühner (2004, S. 205) empfiehlt stattdessen den SRMR Index zur globalen Beurteilung der Modellpassung, der einen Wert von 0,11 oder geringer annehmen sollte. Der RMSEA Index sollte einen Wert von unter 0,05 annehmen; jedoch wird von seiner Anwendung im Fall der ADF Methode abgeraten.

Tab. 6-2: Anspruchsniveaus der herangezogenen Gütemaße zur Beurteilung des CFA-Modells (Homburg & Pflesser, 2000, S. 430; Bühner, 2004, S. 205)

Bezeichnung des Gütekriteriums	Anspruchsniveau
RMSEA*	$\leq 0,05$ (bzw. 0,08)
GFI**	$\geq 0,9$
$\chi^2/df$	$\leq 3$
AGFI**	$\geq 0,9$
SRMR	$\leq 0,11$
NFI	$\geq 0,9$
CFI	$\geq 0,9$
Indikatorreliabilität	$\geq 0,4$
Faktorreliabilität $rel(\xi_i)$	$\geq 0,6$
Durchschnittlich erfasste Varianz eines Faktors	$\geq 0,5$
Signifikanztest der Faktorladung (einseitiger Test auf $p < 0,05$ -Niveau)	$t \geq 1,645$
$\chi^2$ -Differenztest (5%-Niveau)	$\chi^2$ -Differenz $\geq 3,841$
Fornell-Larcker-Kriterium	$DEV(\xi_i) > \text{quadr. Korrelation } (\xi_i, \xi_j)$ , für alle $i \neq j$

Anmerkung: \* Bei der ADF-Schätzmethode ist von einer Anwendung des RMSEA abzusehen (Bühner, 2004, S. 205). \*\* Bei den Schätzmethoden ADF und ML wird von einer Anwendung abgeraten (Hu & Bentler, 1998, S. 446).

## 6.6 Strukturgleichungsmodelle auf Basis von Partial Least Squares (PLS)

Mit Hilfe des Strukturgleichungsansatzes auf Basis des *Partial Least Squares* (PLS) Algorithmus werden zwei der Fragestellungen empirisch untersucht, die der vorliegenden Arbeit zugrunde liegen. Zum einen wird ein PLS-Modell zur Entwicklung des formativen Usability-Maßes berechnet. Dabei wird die formative latente Variable Usability von Online-Shops aus Nutzersicht (ufosV2f-Skala) in einem Zwei-Konstrukt-Modell als Prädiktor in Beziehung zur reflektiven Variante derselben latenten Variable (ufosV2r-Skala) gesetzt; eine ausführliche Darstellung dieses Validierungsansatzes findet sich in Abschnitt 2.8.3.4. Zum anderen wird das konzeptuelle Erfolgsmodell zur Prädiktion der Kaufintention mit Hilfe eines PLS-Modells untersucht (siehe Abschnitte 2.6 und 2.7). Im Folgenden werden zunächst die Voraussetzungen zur Anwendung des Verfahrens und anschließend die Gütekriterien vorgestellt.

### 6.6.1 Voraussetzungen zur Berechnung von PLS-Modellen

Die Berechnung eines PLS-Modells ist nicht an konkrete Verteilungsvoraussetzungen der einbezogenen manifesten Variablen gebunden (vgl. Panten & Thies, 2006, S. 319). Eine Einbindung kategorialer Daten ist allerdings nur auf exogener Variablenebene in Form von Dummy-Variablen möglich. Es zeigte sich in Simulationsstudien, dass die PLS-Schätzmethode sehr robust auf Verletzungen des allgemeinen linearen Modells reagiert (vgl. Cassel, Hackl & Westlund, 1999). Die Reproduktion unterstellter Modellbeziehungen bleibt selbst bei schief verteilten Variablen und dem Auslassen bedeutender Regressoren trotz theoretisch inkonsistenter Schätzer stabil.

Als Mindeststichprobengröße wird in PLS das Fünf- bis Zehnfache der Prädiktorenanzahl der größten Regressionsgleichung benötigt (Scholderer & Balderjahn, 2006). Es können verschiedene Schätzverfahren gewählt werden, wobei das Path Weighting Scheme den Standard darstellt (Lohmöller, 1989, S. 42). Es ergeben sich lediglich marginale Unterschiede zu den konkurrierenden Schätzverfahren Factor bzw. Centroid Weighting Scheme (vgl. Panten & Thies, 2006, S. 316).

### 6.6.2 Gütebeurteilung von PLS-Modellen

Die PLS-Modellgütebeurteilung erfolgt sowohl für Zwei-Konstrukt-Modelle (siehe Abbildung 2.8.3.4) als auch komplexere Modell auf Grundlage derselben Parameter und in zwei Schritten. Zunächst gilt es, in einem ersten Schritt das Messmodell zu prüfen; im zweiten Schritt wird die Güte des Strukturmodells untersucht.

Aufgrund der Tatsache, dass PLS keine Verteilungsvoraussetzungen stellt, müssen Signifikanzen der Mess- und Strukturmodellparameter mit Hilfe nicht parametrischer Verfahren bestimmt werden. Zu diesem Zweck stehen verschiedene Resamplingprozeduren zur Verfügung (vgl. Reimer, 2006), die keine theoretischen Verteilungsannahmen erfordern. Dazu gehören die *Jackknife*-Methode sowie das *Bootstrapping*, wobei das erstgenannte Verfahren eine spezielle Form des zweitgenannten darstellt. Das Bootstrapping kann als der flexibelste Resamplingsansatz angesehen werden und ist standardmäßig in die Software PLS-Graph bzw. Smart-PLS integriert (Hansmann & Ringle, 2004). Bei dem von Efron (1979) entwickelten Bootstrapping Verfahren werden Teilstichproben aus dem Gesamtdatensatz gezogen und wieder „zurückgelegt“. Durch den Vergleich der Teilstichproben lassen sich t-Werte für die Modellparameter bestimmen, die annähernd einer Student-t-Verteilung folgen (Lunneborg, 2000, S. 125 ff.). Die Anzahl der Freiheitsgrade wird durch die Anzahl der Teilstichprobe, also die Größe des Bootstrap-Samples, determiniert. Da das Ziehen der Teilstichproben zufällig erfolgt, können sich die geschätzten t-Werte mehrerer durchgeführter Bootstrapping-Prozeduren voneinander unterscheiden.

#### 6.6.2.1 Beurteilung des Messmodells

Bei der Betrachtung des Messmodells (auch als äußeres Modell bezeichnet) erfolgt die Gütebeurteilung für reflektive und formative Indikatoren auf Basis unterschiedlicher Kriterien (vgl. Kraft et al., 2005, S. 72 ff.).

Reflektive Indikatoren sollten signifikante Ladungen ( $\lambda$ ) über 0,707 aufweisen (Barclay, Thompson & Higgins, 1995); Indikatoren mit Ladungen unter 0,4 bzw. 0,5 sind aus dem reflektiven Messmodell zu entfernen. Weiterhin lassen sich die Ladungsquadrate (Kommunalitäten) heranziehen, die den Prozentsatz der erklärten Indikatorvarianz angeben (Fornell & Cha, 1994, S. 68 f.). Die Signifikanz der Indikatorladungen wird anhand der t-Werte aus der Bootstrapping-Prozedur geprüft. Bei einem t-Wert von über 2 erweist sich der



Einfluss des jeweiligen Indikators als signifikant ( $p < 0,05$ ); ist dies nicht der Fall, besteht die Option, den nicht-signifikanten Indikator aus dem Messmodell zu eliminieren, da sein Erklärungsgehalt an der LV zu gering ausfällt. Liegt das Ziel des Gesamtmodells wie im Fall des konzeptuellen Erfolgsmodells in einer Prognose, so sollte ein formativer Indikator beibehalten werden, sobald sein t-Wert größer als 1 ist, da der Erklärungsgehalt des Indikators in diesem Fall bereits größer ausfällt als das Rauschen (Hansen, 1987). Erweisen sich mehrere Indikatoren als insignifikant, so gilt es zunächst, denjenigen mit dem niedrigsten t-Wert zu eliminieren. Anschließend ist das gesamte Modell neu zu berechnen bzw. einer Resampling-prozedur wie Bootstrapping zu unterziehen. Der Optimierungsprozess des Messmodells erfolgt also in einem iterativen Prozess.

Zur Bestimmung der Reliabilität eines reflektiven Messmodells wird die *Composite Reliability* bestimmt; diese sollte über 0,7 liegen. Zusätzlich misst das von Fornell und Larcker (1981) entwickelte AVE-Maß (Average Variance Extracted) denjenigen Varianzanteil der Indikatoren, der durch die latente Variable im Verhältnis zum Anteil des Messfehlers erklärt wird (Panten & Thies, 2006, S. 323). Der mit Hilfe der Bootstrap-Prozedur bestimmte AVE-Wert sollte über 0,5 liegen. Das AVE-Maß gibt darüber hinaus auch Aufschluss über die diskriminante Validität eines reflektiven Messmodells. Die Quadrat-Wurzeln der AVE-Werte sollten substantiell größer sein als die Korrelationen zwischen der latenten Variable mit den übrigen LVn (Chin, 1998, S. 327).

Zur Gütebeurteilung formativer Messmodelle gilt es zunächst sicherzustellen, dass kein kritisches Maß an Multikollinearität zwischen den formativen Indikatoren vorliegt, da dies im Rahmen von Strukturgleichungsmodellen wie PLS Verzerrungen in den Parameterschätzungen zur Folge haben kann (vgl. Grewal, Cote & Baumgartner, 2004). Die Gewichte der formativen Indikatoren sollten Werte  $\gamma_n > 0,1$  und ein erwartungskonformes Vorzeichen aufweisen, andernfalls kommt eine Elimination des jeweiligen Indikators in Frage, die jedoch stets theoretisch gerechtfertigt sein muss (vgl. Abschnitt 2.8.3.4). Die Signifikanz der Gewichte wird wiederum auf Basis der t-Werte aus der Bootstrap-Resampling-Prozedur bestimmt.

Zur Beurteilung der Diskriminanzvalidität formativer Variablen lassen sich nach Herrmann, Huber und Kressmann (2005) die Korrelationen zwischen den latenten Variablen des Gesamtmodells betrachten, wobei Werte von  $r < 0,9$  auf eine hinreichende Diskriminanz schließen lassen. Eine Anwendung dieses Gütekriteriums ist im Falle eines Zwei-Konstrukt-Modells jedoch nicht sinnvoll. Da hohe Korrelationen zwischen formativen Indikatoren zwar auftreten können, aber nicht müssen, sind die zur Beurteilung reflektiver Messmodelle genutzten Gütekriterien (AVE, Composite Reliability) im formativen Fall nicht geeignet.

### 6.6.2.2 Beurteilung des Strukturmodells

Auf Strukturmodellebene gilt es beim Zwei-Konstrukt-Modell, zwei Parameter in Augenschein zu nehmen. Zum einen ist der Pfadkoeffizient  $\beta$  zu betrachten, der die Stärke des Zusammenhangs zwischen den LVn bezeichnet. Die Signifikanz des Pfadkoeffizienten wird wiederum anhand eines t-Wertes auf Basis eines Bootstrapping Resamplings bestimmt. Bei einem  $\beta > 0,80$  kann nach Chin (1998, S. 324) von einer angemessenen Inhaltsvalidität des formativen Konstruktes ausgegangen werden.<sup>19</sup> In komplexeren Modellen mit mehr als zwei LVn sollten die Pfadkoeffizienten über 0,1 liegen (Sellin & Keeves, 1984).

Angaben zu einer angemessenen Höhe des Determinationskoeffizienten  $R^2$  für Zwei-Konstrukt-Modelle lassen sich in der Literatur nicht finden; in komplexen Modellen (mit mehr als zwei LVn) ist ein  $R^2 > 0,3$  anzustreben (Hermann et al., 2006). Im Falle eines Zwei-Konstrukt-Modells ist sicherlich ein weitaus höherer  $R^2$ -Koeffizient zu erwarten.

Um die Einflussstärken prädiktierender Variablen zu bestimmen, kann im Fall von komplexen PLS-Modelle die Effektgrößen  $f^2$  herangezogen werden kann (Panten & Thies, 2006, S. 325). Dieses Maß beschreibt das Verhältnis von zusätzlicher erklärter Varianz und vormals nicht erklärter Varianz (vgl. Gefen, Straub & Boudreau, 2000, S. 64). Von einem geringen, mittleren bzw. großen Effekt ist bei Werten von über 0,02, 0,15 bzw. 0,35 auszugehen (Chin, 1998, S. 317). Die Effektgröße berechnet sich wie folgt:

$$(6) \quad f_j^2 = \frac{R_{\text{eingeschlossen}}^2 - R_{\text{ausgeschlossen}}^2}{1 - R_{\text{eingeschlossen}}^2}$$

wobei:  $f^2$  = die Effektstärke der exogenen latenten Variable  $\eta_j$ ,

$R_{\text{eingeschlossen}}^2$  = Bestimmtheitsmaß der abhängigen latenten Variable im PLS-Modell, das das fragliche Konstrukt  $\eta_j$  einschließt,

$R_{\text{ausgeschlossen}}^2$  = Bestimmtheitsmaß der abhängigen latenten Variable im PLS-Modell, das das fragliche Konstrukt  $\eta_j$  nicht mit einschließt.

Weiterhin empfiehlt sich im Falle formativer Indikatoren die Berechnung von Totaleffekten, die sich durch eine der Multiplikation der Gewichte mit den Pfadkoeffizienten ergeben, wobei alle Pfade aufzuaddieren sind. Im Falle einer reflektiven Kriteriumsvariable kann das auf den Redundanzen basierende Stone-Geisser-Kriterium  $Q^2$  zur Einschätzung der gemeinsamen Vorhersagequalität von Struktur- und Messmodell dienen (Gefen, Straub & Boudreau, 2000). Nimmt es einen Wert über 0 an, so besitzt das Modell eine Vorhersagerelevanz. Im Gegensatz zu kovarianzbasierten Strukturgleichungsansätzen wie LISREL stehen keine weiteren globalen Fit-Maße zur Gütebeurteilung des gesamten Modells zur Verfügung.

<sup>19</sup> Ein Pfadkoeffizient von  $\beta=0,87$  in einer Beispieluntersuchung weist nach Angaben von Chin (1998, S. 324) auf eine starke Konvergenz zwischen dem formativen und dem reflektiven Maß hin und impliziert eine adäquate Abdeckung der Wahrnehmungen im formativen Indikatorenset.

## 6.7 Multinomiale logistische Regression (MLR)

Mit Hilfe einer multinomialen logistischen Regression (MLR), die eine Verallgemeinerung der binären logistischen Regression darstellt, soll das nominalskalierte Kriterium Kaufentscheidung mit den  $j=3$  Ausprägungen *Kauf in Shop 1*, *Kauf in Shop 2* und *Wahl des Gutscheins* vorhergesagt werden (vgl. Hosmer & Lemeshow, 2000, S. 31 ff.). Einen ähnlichen Ansatz zur Identifikation von Erfolgsfaktoren im Bereich des Online-Shoppings verfolgen Lowengart und Tractinsky (2001). Dieses Verfahren besitzt gegenüber dem PLS Strukturgleichungsansatz den Vorteil, dass die verhaltensbasierte abhängige Variable *Kaufentscheidung* einbezogen werden kann. Da die prädiktierenden Variablen mit Hilfe von Rating-Skalen erhoben wurden, unterscheidet sich die Art der Daten auf unabhängiger und abhängiger Variablenseite deutlich voneinander, was die Gefahr eines Common Method Bias anerkanntermaßen deutlich mindert (vgl. Podsakoff et al., 2003).

Wie aus Tabelle 6-1 am Anfang dieses Kapitels hervorgeht, sollen mit Hilfe der MLR im Rahmen der vorliegenden Untersuchung mehrere inhaltliche Hypothesen untersucht werden. Zum Zwecke der Bestimmung der prädiktiven Validität wird analysiert, inwieweit auf Grundlage der ufosV2-Skalen eine Vorhersage der verhaltensbasierten Variable Kaufentscheidung möglich ist (Hypothesen H6a und H6b). Weiterhin erfolgt eine Überprüfung, ob sich der im konzeptuellen Modell angenommene Zusammenhang zwischen der Kaufintention als direkter Prädiktor für die Kaufentscheidung bestätigt (Hypothese 21). Letztlich wird das MLR Verfahren auch dazu genutzt, um einen potenziellen Störeinfluss einiger Kontrollvariablen auf die Versuchsergebnisse zu untersuchen.

### 6.7.1 Voraussetzungen der MLR

Die Berechnung einer MLR ist an eine Reihe von Voraussetzungen gebunden. Die Fallzahl sollte pro Kategorie der AV mindestens 25 betragen, wobei bei einer größeren Zahl an Prädiktoren eine höhere Anzahl an Fällen erforderlich wird (Backhaus et al., 2003, S. 470). Genaue Angaben zur benötigten Fallanzahl im Verhältnis zur Anzahl der Prädiktoren stehen nicht zur Verfügung. Nach Tabachnik und Fidell (2006, S. 442) können besonders hohe Parameterschätzungen und Standardschätzfehler auf ein zu geringes N hinweisen. Ein hohes Maß an Multikollinearität zwischen den Prädiktoren stellt eine potenzielle Gefahr dar, da wie auch im Fall der linearen Regression ungenaue Schätzungen der Regressionskoeffizienten und Verzerrungen der Teststatistiken resultieren können (vgl. Bortz, 1999, S. 438; Tabachnik & Fidell, 2006, S. 443). Die Prädiktoren einer MLR können dichotome, diskrete oder kontinuierliche Werte annehmen. Weiterhin werden eine lineare Beziehung zwischen den kontinuierlichen Variablen und der logit-transformierten AV sowie eine fallweise Unabhängigkeit der Fehlerterme vorausgesetzt (vgl. Rohrlack, 2006, S. 205 f.).

### 6.7.2 Ausreißeranalyse bei der MLR

Um zu prüfen, ob einzelne Fälle den im MLR-Modell angenommenen Zusammenhang nicht aufweisen und den Gesamtfit des Modells in hohem Maße verzerren, wird für jede Beobachtung bestimmt, wie groß die Diskrepanz zwischen den empirischen Werten und den durch das Modell geschätzten Wahrscheinlichkeiten ausfällt (vgl. Backhaus et al., 2003, S. 448 ff.). Die Bestimmung der fallbezogenen Residuen kann dabei nicht auf Gesamtmodellebene stattfinden, sondern muss für beide Logit-Modelle getrennt erfolgen, die im Rahmen der MLR berechnet werden. Als problematisch erweist sich dabei, dass keine allgemeinen statistischen Maße zur Verfügung stehen, auf Grundlage derer beurteilt werden könnte, ob und welche Residuen zu Verzerrungen führen. Nach Backhaus et al. (2003, S. 449) weist im binär logistischen Regressionsmodell ein Residuum mit einem absoluten Wert über 0,5 darauf hin, dass ein Ausreißer im Sinne eines Klassifikationsfehlers vorliegt. Zur leichteren Identifikation von Ausreißern wird darüber hinaus eine Gewichtung durch die Bestimmung standardisierter Residuen (auch Pearson-Residuen genannt) empfohlen (Tabachnick & Fidell, 2006, S. 448).

### 6.7.3 Gütekriterien der MLR

Zum Zwecke einer Beurteilung der globalen Anpassungsgüte wird mit Hilfe eines Likelihood-Quotiententests die globale Nullhypothese getestet, alle Regressionsparameter im untersuchten Modell seien Null (vgl. Rohrlack, 2006, S. 209). Dabei wird das vollständige Modell (mit allen einbezogenen Prädiktoren und der Konstante) mit einem Modell verglichen, in das lediglich die Konstante aufgenommen wird.

Um bei einer sequentiell berechneten MLR zu untersuchen, inwiefern hinzugefügte Prädiktoren signifikant zur Vorhersage beitragen, kann ein Vergleich zwischen dem veränderten und dem ursprünglichen Modell gezogen werden, indem eine Differenz zwischen den beiden globalen Log-Likelihood-Werten (jeweils multipliziert mit -2) berechnet wird. Dabei wird der Wert für das kleinere Modell von dem des größeren abgezogen. Es ergibt sich ein  $\chi^2$ -Wert, dessen Freiheitsgrade sich ebenfalls aus der Differenz der Freiheitsgrade beider Modelle berechnen lassen (vgl. Tabachnick & Fidell, 2006, S. 447).

Zur weiteren Gesamtgütebeurteilung werden Modellgültigkeitstests durchgeführt, die auf der Pearson-Chi<sup>2</sup>- bzw. der Devianz-Statistik basieren. Mit Hilfe dieser Goodness-of-fit-Tests wird ein Vergleich zwischen erwarteten und beobachteten Häufigkeiten der AV-Gruppen gezogen (vgl. Tabachnick & Fidell, 2006, S. 486). Als Voraussetzung zur Berechnung dieser Tests muss gelten, dass die Matrix zur Gegenüberstellung der Kovariatenmuster mit den Kategorien der AV mehrfach besetzte Zellen aufweist, wobei jede Zelle durch eine Antwortkombination und eine Kategorie der AV bestimmt ist. Die Zellbesetzung sollte ausreichend groß sein, da ansonsten eine niedrige Power der Analyse resultiert.

Zur Beurteilung der Gesamtmodellanpassung stehen außerdem drei deskriptive Maßzahlen – so genannte Pseudo- $R^2$ -Koeffizienten – zur Verfügung (Backhaus et al., 2003)<sup>20</sup>. Das  $R^2$  nach Cox und Snell kann nur Werte kleiner 1 annehmen, wobei Werte über 0,2 als *akzeptabel* und ab 0,4 als *gut* gelten können. Für das McFadden- $R^2$  kann sich maximal ein Wert von 1 ergeben; dabei entspricht ein Wert ab 0,2 bzw. 0,4 einer guten Modellanpassung. Im Fall des  $R^2$  nach Nagelkerke kann ein Wert über 0,5 als *sehr gut* interpretiert werden; diesem Gütemaß ist nach Backhaus et al. (2003, S. 441) der Vorzug zu geben. Adjustierte Formen der Pseudo- $R^2$ -Koeffizienten, die die Anzahl der Prädiktoren mit berücksichtigen, existieren für die MRL nicht (Tabachnik & Fidell, 2006, S. 442).

Zur Gütebeurteilung auf Variablenebene werden Likelihood-Quotienten-Tests berechnet, die beide Logit-Gleichungen simultan berücksichtigen und die einzelnen Prädiktoren bzgl. ihres Einflusses auf die abhängige Variable untersuchen (vgl. Rohrlack, 2006, S. 213). Ein großer 2LL-Wert (LL für LogLikelihood) bzw.  $\chi^2$ -Koeffizient weist auf einen starken Unterschied zwischen dem vollständigen Modell und dem um den jeweiligen Prädiktor reduzierten Modell hin. Dabei wird geprüft, inwiefern sich die Differenz zwischen vollständigem und reduziertem Modell und damit der Einfluss des einzelnen Regressoren als signifikant erweist; die Nullhypothese besagt hier also, dass die jeweilige Variable keinen Einfluss auf die Gruppentrennung nimmt.

Auf Variablenebene werden des Weiteren die Regressionskoeffizienten  $B$  und deren Standardfehler  $SE$  der beiden Logit-Modelle herangezogen. Eine inhaltliche Interpretation dieser Schätzer ist jedoch problematisch, da sie u.a. von der Skalierung bzw. Maßeinheit der Regressoren abhängen. Auf Basis der Wald-Statistik, welche bei hinreichend großer Stichprobe annähernd  $\chi^2$ -verteilt ist, werden die Signifikanzen der Regressionskoeffizienten beider Logit-Modelle berechnet. Es wird ein  $p < 0,05$ -Signifikanzniveau angesetzt. Zur Interpretation der Einflussstärke des einzelnen Prädiktoren wird der Exponentialfunktionswert  $\text{Exp}(B)$  herangezogen. Dieser besagt, wie stark sich der  $j$ -te Wahrscheinlichkeitsquotient ändert, wenn sich die Prädiktorausprägung um eine Einheit erhöht und alle anderen Prädiktoren unverändert bleiben. Entsprechend kann dieser Parameter als eine Art Effekt-Koeffizient (auch als „odds-ratio“ bezeichnet) der exogenen Variablen in den beiden Logit-Modellen angesehen werden (Backhaus et al., 2003, S. 436). Für die  $\text{Exp}(B)$ -Werte stehen des Weiteren Konfidenzintervalle zur Verfügung.

Anhand der Klassifikationsergebnisse lässt sich wiederum die Gesamtgüte des Modells beurteilen. Die Rate korrekter Klassifikationen (Correct Classification Rate) gibt an, wie hoch der Anteil der AV-Werte ist, die durch das Modell korrekt bzw. inkorrekt hervorgesagt werden. Die Trefferraten sind mit der Leistung eines Basismodells ohne Prädiktoren zu vergleichen, bei dem jeder Fall zufällig einer der Gruppen zugeordnet werden würde. Ein

---

<sup>20</sup> Die Angaben zur Güteeinschätzung werden von Backhaus et al. (2003, S. 440 ff.) übernommen.

Vergleich mit der proportionalen Zufallswahrscheinlichkeit (PZW) wird wie folgt berechnet (Morrison, 1969):

$$(7) \quad \text{PZW} = \sum_{g=1}^G \left( \frac{n_g}{n} \right)^2$$

wobei:  $n_g$  = Anzahl beobachteter Elemente in der Kategorie  $g$  der AV,  
 $G$  = Anzahl der Kategorien der AV.

Bei der Bestimmung von Klassifikationsraten anhand einer einzigen Stichprobe gilt es zu bedenken, dass es zu einer Überschätzung kommen kann. Um einen realistischen Eindruck von der prädiktiven Güte zu erhalten, kann eine Kreuzvalidierung des Modells durch Aufteilung in eine Teilstichprobe zur Modellgewinnung und eine zweite Substichprobe zur Modellvalidierung vorgenommen werden.

## 6.8 Multikollinearitätsprüfung

Zum Nachweis der Multikollinearität werden in der Literatur drei verschiedene Ansätze angeführt (vgl. Schneider, 2006, S. 190 ff.):

- a) Es können die bivariaten Korrelationen zwischen den Prädiktoren betrachtet werden. Hohe Interkorrelationen stellen einen Hinweis auf das Vorliegen von Multikollinearität dar.
- b) Bivariate Korrelationskoeffizienten betreffen nur paarweise Abhängigkeiten zwischen den Indikatoren. Multikollinearität kann jedoch auch durch den Zusammenhang mehrerer Regressoren bedingt sein (Backhaus et al., 2003, S. 90). Entsprechend werden auf Basis einer Hilfsregression Toleranz- bzw. Variance-Inflation-Factor (VIF)-Werte bestimmt. Der VIF ergibt sich aus dem Kehrwert der Toleranz. Klare Angaben zu kritischen Werten lassen sich in der Literatur nicht finden. Nach Fickel (2001, S. 41) ist bei Toleranz-Werten unter 0,4 eine mittlere Multikollinearität zu unterstellen. VIF-Werte über 10 sind nach Chatterjee und Price (1977, S. 182) als Hinweis auf bestehende Multikollinearität zu werten. Nach Schneider (2006, S. 192) können bereits VIF-Werte über 2 problematisch sein.
- c) Eine weitere Möglichkeit der Multikollinearitätsdiagnose besteht in der Bestimmung der Konditionsindizes. Mäßige bis starke Multikollinearität kann bei Werten über 30 bestehen (Belsley, Kuh und Welsch, 1980, S. 105).

## 6.9 Weitere Testverfahren

### 6.9.1 t-Tests

Als Voraussetzung eines t-Test sollten die Messwerte in der Population normalverteilt sein; allerdings erweist sich der Test bei größeren Stichproben als robust gegenüber einer Verletzung der Normalverteilungsannahme. Um einen schwachen Effekt mit einer Teststärke von 80% ( $1-\beta=0,8$ ) und einer Irrtumswahrscheinlichkeit von  $\alpha=0,05$  identifizieren zu können, wird mindestens eine Stichprobengröße von  $N=310$  benötigt (vgl. Bortz & Döring, 2005, S. 139).

Jeder Teilnehmer bewertete im Rahmen des Experimentes zwei Online-Shops. Um zu prüfen, inwiefern es sich als einflussreich erweist, ob die Bewertung eines Shops an der ersten oder zweiten Stelle im Versuchsablauf erfolgte, soll ein t-Test für unabhängige Stichproben berechnet werden (vgl. Bortz & Döring, 2005, S. 140). Es wird geprüft, ob sich signifikante Mittelwertunterschiede in der AV *Kaufintention* für die Erhebung an erster und an zweiter Stelle ergeben; das Signifikanzniveau wird bei  $\alpha=0,05$  festgelegt. Als Voraussetzung dieses t-Tests muss wiederum gelten, dass die untersuchte Variable in beiden Gruppen normalverteilt ist. Zusätzlich dürfen keine Varianzunterschiede zwischen den Gruppen bestehen (Annahme der Varianzhomogenität). Falls diese Voraussetzung verletzt wird, steht eine alternative Statistik zur Verfügung, anhand derer die Signifikanz des Mittelwertunterschiedes geprüft werden kann.

### 6.9.2 Einfaktorielle Varianzanalyse

Mit Hilfe einer mehrdimensionalen einfaktoriellen Varianzanalyse (ANOVA) soll die Diskriminationsfähigkeit der reflektiven Skalen überprüft werden. Als Voraussetzung zur Berechnung einer Varianzanalyse sollte eine Normalverteilung der Werte in der Grundgesamtheit gegeben sein (vgl. Backhaus et al., 2003, S. 150). Außerdem wird eine Gleichheit der Varianzen in den verschiedenen Gruppen vorausgesetzt. Bei Verletzungen dieser Annahmen erweist sich das Verfahren im Falle gleicher Zellbesetzungen als einigermaßen robust. Zur Testung der Varianzhomogenität werden Levene-Tests berechnet.

### 6.9.3 Levene-Test

Um zu prüfen, inwiefern sich die Varianzen zweier Stichproben unterscheiden, steht in der Software SPSS der Levene-Test zur Verfügung (vgl. Dayton, 1970, S. 34 f.). Dieser Varianzhomogenitätstest erweist sich als relativ unempfindlich gegenüber Verletzungen der Normalverteilungsvoraussetzung (Bortz & Döring, 2005, S. 286). Als statistisches Prüfmaß wird dabei ein F-Wert berechnet, der die Wahrscheinlichkeit angibt, mit der eine Gleichheit der Varianzen in der Grundgesamtheit vorliegt (Brosius, 2004, S. 385). Im Fall eines insignifikanten Ergebnisses im Test wird die Nullhypothese beibehalten, dass sich die Varianzen nicht voneinander unterscheiden.





## **7. Durchführung**

In diesem Kapitel wird die Durchführung des Quasi-Experiments beschrieben, die im Zeitraum zwischen April 2005 und Februar 2006 stattfand. Im ersten Abschnitt 7.1 wird berichtet, wie die Teilnehmerstichprobe rekrutiert wurde, welche Probleme bei der Probandenakquise auftraten und wie zentrale Merkmale der Stichprobe ausgeprägt sind. In Abschnitt 7.2 folgt eine Darstellung der physikalischen Versuchsumgebung im Labor. Abschnitt 7.3 stellt die Online-Versuchsumgebung vor, mit Hilfe derer das Experiment computergestützt geleitet wurde. Der genaue Ablauf des Versuchs wird in Abschnitt 7.4 geschildert. Schließlich kann dem letzten Unterabschnitt 7.5 entnommen werden, welche Probleme bei der Versuchsdurchführung auftraten und wie mit ihnen umgegangen wurde.

### **7.1 Teilnehmerstichprobe**

Um eine möglichst gute Übertragbarkeit der Untersuchungsergebnisse zu gewährleisten, wurde das Ziel verfolgt, eine repräsentative Stichprobe für die Population deutscher Internetnutzer zu ziehen. Als Segmentierungskriterien dienten zunächst die demographischen Merkmale *Alter* und *Geschlecht* der Teilnehmer.

#### **7.1.1 Rekrutierung der Teilnehmer**

Wie in Abschnitt 7.1.1.1 geschildert wird, wurden bei der Rekrutierung der Teilnehmer mehrere unterschiedliche Wege der Akquise verfolgt, um bei der Ziehung der Stichprobe eine möglichst gute Abbildung der Population deutscher Internetnutzer zu erreichen. In Abschnitt 7.1.1.2 wird dargestellt, wie die Rekrutierung erfolgte.

##### **7.1.1.1 Wege der Rekrutierung**

Die Versuchsleiter versendeten Werbe-E-mails, in denen die Rahmenbedingungen des Quasi-Experiments geschildert wurden, an potenziell Interessierte (siehe Mail in Anhang D). Die Empfänger der Mails wurden gebeten, die Nachricht ihrerseits weiterzuleiten. Mit Hilfe von Handzetteln wurde in den Job-Centern des Kieler Arbeitsamtes, in der Mensa der Kieler Uni, Einrichtungen zur Fortbildung von Arbeitslosen und in Sportvereinen geworben (siehe Handzettel in Anhang D). Es wurden Aushänge in vielen öffentlichen Einrichtungen gemacht, z.B. an den „schwarzen Brettern“ der Uni, in Berufsschulen, in Geschäften des Einzelhandels und in Kliniken. Zudem wurde der Kontakt zu diversen Leitern von Computerkursen hergestellt, die in Seniorentreffs und in Einrichtungen der lokalen Kirchengemeinde durchgeführt werden. Ältere Teilnehmer über 60 Jahre wurden vor allem in bestimmten Lehrveranstaltungen der Kieler Uni angeworben, z.B. in Vorlesungen zur Geschichte Schleswig-Holsteins. Unter den Studenten der Uni wurde zudem in Veranstaltungen des hiesigen Marketing-Lehrstuhls sowie in einer Statistik Vorlesung im Psychologie-Grundstudium geworben.

Die Kieler Uni gab eine Pressemitteilung heraus; der genaue Wortlaut kann Anhang D entnommen werden. In regionalen Zeitungen wurden Artikel zur Untersuchung veröffentlicht. Zum Zwecke der Rekrutierung wurde eine Internet-Werbepage programmiert, auf der sich Interessierte über die wesentlichen Details zur Untersuchung informieren und sich zur Teilnahme melden konnten (siehe Screenshots im Anhang D). Auf der Werbepage war anhand einer Tabelle abzulesen, in welchen Versuchsbedingungen (Geschlecht, Alter, Produktart) noch Teilnehmer gesucht wurden bzw. in welchen nicht mehr. Auf den Homepages der Kieler Uni sowie des Graduiertenkollegs *Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business* wurde ein Werbetext zur Untersuchung veröffentlicht. Einige Probanden konnten über das Internetportal [kiel4kiel.de](http://kiel4kiel.de) geworben werden.

#### 7.1.1.2 Ablauf der Rekrutierung

Sobald sich ein Interessent oder eine Interessentin über die Werbepage oder eine Email bei der Versuchsleitung meldete, wurde zunächst geprüft, inwiefern noch Teilnehmer dieses Alters und Geschlechts in der jeweiligen Produktart benötigt wurden. War dies nicht der Fall, so erhielt der Interessent eine Absage per Email. Wurden noch Probanden in der jeweiligen Versuchsbedingung gesucht, erfolgte zunächst eine Prüfung, inwiefern das angegebene Produkt in den herangezogenen Shops überhaupt lieferbar war. Einige der angegebenen Produkte waren in den Shops nicht lieferbar. In diesen Fällen wurden die potenziellen Teilnehmer gebeten, ihre Wahl erneut zu treffen. Nachdem die Frage nach dem zu bestellenden Produkt geklärt wurde, erfolgte eine Absprache zur Festlegung des Untersuchungstermins per Telefon oder per Email.

#### 7.1.2 Probleme bei der Teilnehmerakquise

Allgemein erwiesen sich die Rahmenbedingungen der Untersuchung in schriftlicher Form als schwer vermittelbar. Bei einigen Probanden traten Verständnisprobleme bzgl. der Notwendigkeit auf, sich im Vorfeld der Untersuchung auf ein konkretes Produkt festzulegen. Einige Personen lasen die Werbetexte offensichtlich nicht aufmerksam genug und meinten, sich den Online-Shop selber aussuchen zu können oder den Versuch online durchführen zu können. Die verschiedenen Produktarten wurden aus Sicht der potenziellen Teilnehmer der Studie als deutlich unterschiedlich attraktiv wahrgenommen. Die insgesamt fünf Versuchsleiter erhielten den Eindruck, dass das Interesse an der Teilnahme in den fünf Produktkategorien in Abhängigkeit des Geschlechts und des Alters schwankte. Allgemein schwer zu akquirieren waren männliche Teilnehmer im Alter zwischen 40 und 60 Jahren.

Erhebliche Probleme bereitete die Erhebung der Substichprobe zu den Veranstaltungstickets. Einerseits fanden sich wenige Teilnehmer, die überhaupt Interesse äußerten, im Rahmen des Versuchs ein Ticket zu bestellen. Eine Rolle könnte dabei spielen, dass Konzerttickets oft weitaus mehr als 15 Euro kosten, die die Vpn als Entlohnung erhielten. Viele populäre

Künstler treten zudem nicht in Kiel, dem Ort der Versuchsdurchführung, sondern nur in größeren Städten auf, so dass der Besuch diverser Veranstaltungen mit einer längeren Anfahrt und weiteren Kosten verbunden ist. Zum anderen ergab sich ein Problem bzgl. der Verfügbarkeit vieler Tickets in den einzelnen Shops. Diverse Tickets, die sich die Teilnehmer im Vorfeld auswählten, waren jeweils nur in einem der sieben Shops verfügbar bzw. nur in einer bestimmten Untergruppe. So wurden beispielsweise Karten für Theaterveranstaltungen in Kiel oder auch für bestimmte Sportveranstaltungen lediglich von zwei der sieben Ticket-Shops zum Verkauf angeboten. Da jedoch der Anspruch bestand, alle Versuchsbedingungen des Versuchsplans zu füllen, d.h. alle Ticket-Shops jeweils gleich häufig miteinander zu paaren, erwies es sich in vielen Fällen nicht als möglich, dem Produktwunsch der Teilnehmer zu entsprechen. Aufgrund der dargestellten Probleme wurde die Datenerhebung abgeschlossen, bevor alle Versuchsbedingungen innerhalb der Substichprobe zu den Veranstaltungstickets gefüllt waren, so dass für 43 Bedingungen keine Daten vorliegen (siehe Tabelle D-1 im Anhang D).

### 7.1.3 Beschreibung der Stichprobe

Bei Ziehung der Teilnehmerstichprobe sollte eine möglichst repräsentative Abbildung der Population deutscher Internetnutzer erreicht werden. Als Vergleichsdaten dienen einerseits die Ergebnisse der Studie internet facts 2005-III, in der Angaben zu verschiedenen Aspekten der Internet-Nutzung in Deutschland gemacht werden (AGOF, 2006). An dieser repräsentativen Befragung innerhalb der deutschen Bevölkerung nahmen mehr als 100.000 Personen teil. Der Berichtszeitraum bezieht sich auf die Monate Juli bis September 2005. Zum anderen werden die Ergebnisse des (N)ONLINER Atlas 2005 (TNS Infratest, 2005) herangezogen, die auf einer repräsentativen Befragung der deutschen Bevölkerung mit 48.500 Interviews im Januar bis Mai 2005 basieren (siehe Tabellen im Anhang D).

Tab. 7-1: Alter der Teilnehmer im Vergleich zur Population – Angaben in Prozent

Altersgruppe	Experiment						Population
	Bücher	DVDs	CDs	Druckerpatronen	Tickets	Gesamt	
unter 18	0,0	1,2	3,6	0,0	0,0	1,1	13,6
18-29	36,9	63,5	27,4	59,5	70,7	49,5	19
30-39	29,8	21,2	25,0	16,7	17,1	22,5	23,5
40-49	19,0	8,2	23,8	11,9	2,4	14,3	21,2
50-59	8,3	3,5	13,1	4,8	4,9	7,1	12,8
60 und älter	6,0	2,4	7,1	7,1	4,9	5,6	9,9

Anmerkung: Die Angaben zur Population entstammen der Studie Internet Facts 2006-III (AGOF, 2006).

Der Tabelle 7-1 kann die Altersverteilung der Teilnehmer in den einzelnen Produktgruppen im Vergleich zu den Angaben zur Population entnommen werden. Der Probandenanteil unter 18 Jahren fiel deutlich geringer aus als in der Population. Es entstammten lediglich 1,1% der Teilnehmer dieser Altersgruppe. Aufgrund potenzieller rechtlicher Probleme wurde davon abgesehen, entsprechend viele minderjährige Teilnehmer heranzuziehen. Der Anteil an 18 bis

29-jährigen Versuchspersonen (Vpn) war mit 49,5% im Verhältnis zu 19% in der Population offensichtlich deutlich überrepräsentiert, wobei auch innerhalb der Stichprobe zwischen den Produktarten große Unterschiede dahingehend bestanden, wie viele Vpn in diese Altersgruppe fallen. Den geringsten Anteil nahm diese Altersgruppe bei der Produktart CD mit 27,4% ein, bei den Veranstaltungstickets waren es dagegen 70,7%. Die Gruppe der 30 bis 39 Jahre alten Teilnehmer war mit 22,5% ähnlich stark vertreten wie in der Bezugspopulation mit 23,5%, allerdings schwankte auch hier der Anteil zwischen den Produktarten – wenngleich nicht in einem ähnlich starken Maße wie bei den 20 bis 29-Jährigen. 40 bis 49-jährige Internetnutzer waren mit 14,6% in der Stichprobe gegenüber der Population mit 21,2% unterrepräsentiert; gleiches gilt für die Altersgruppen 50 bis 59 Jahre (Stichprobe: 7,1%; Population: 12,8%) und 60 Jahre und älter (Stichprobe: 5,6%; Population: 9,9%). Auch für diese Alterssegmente schwankten die Anteile zwischen den verschiedenen Produktarten. Auffällig gering sind die Anteile der 40 bis 49-Jährigen bei den Veranstaltungstickets sowie der Gruppe 60 Jahre und älter bei den DVDs (jeweils 2,4%). Für die verschiedenen Produktarten kann festgehalten werden, dass die Altersverteilung bei den CDs und Büchern am ehesten mit denen der Population vergleichbar waren, wenngleich auch hier der Anteil der 18 bis 29-Jährigen über- bzw. der Anteil der Minderjährigen unterrepräsentiert war.

Tab. 7-2: Geschlecht der Teilnehmer im Vergleich zur Population – Angaben in Prozent

	<i>Experiment</i>						<i>Population</i>
	<i>Bücher</i>	<i>DVDs</i>	<i>CDs</i>	<i>Druckerpatronen</i>	<i>Tickets</i>	<i>Gesamt</i>	
männlich	48,8	51,8	54,8	56,0	51,2	52,6	55,7
weiblich	51,2	48,2	45,2	44,0	48,8	47,4	44,3

Anmerkung: Die Angaben zur Population entstammen der Studie Internet Facts 2006-III (AGOF, 2006).

Wie Tabelle 7-2 entnommen werden kann, fielen die Unterschiede zwischen Population und Stichprobe im Fall der Geschlechterverteilung deutlich geringer aus. In den Tabellen D-2 bis D-12 im Anhang werden weitere Charakteristika der Population und der Stichprobe einander gegenüber gestellt. Mit steigendem Alter waren in sowohl in der Population als auch in der Stichprobe mehr Männer als Frauen online (Tabellen D-2 und D-3). Der Anteil an Personen mit niedrigem Einkommen war in der Stichprobe im Vergleich zur Population deutlich überrepräsentiert (Tabellen D-4 und D-5). Der Anteil an Internetnutzern mit Abitur fiel in der Stichprobe höher und der Anteil an Personen mit Mittlerer Reife und Hauptschulabschluss niedriger aus als in der Population (Tabellen D-6 und D-7). In der Stichprobe fanden sich anteilig deutlich mehr Studenten als in der Population (Tabellen D-8 und D-9). Offenbar verfügten die Versuchsteilnehmer im Mittel über mehr Online-Erfahrung, als es in der Population der Fall war (Tabellen D-10 und D-11).

Die Tabellen D-12 bis D-16 im Anhang D widmen sich der Beschreibung der Versuchspersonenstichprobe hinsichtlich verschiedener Aspekte der PC- bzw. Internetnutzung. Die Erfahrung der Probanden im Bereich Online-Shopping im Sinne der Anzahl vorheriger Käufe im Internet fiel in der Stichprobe sehr unterschiedlich aus (D-17). Mit

10,3% hatten einige der Teilnehmer noch gar keinen Online-Einkauf getätigt und 19,1% kauften zuvor nur ein- bis fünfmal im Internet ein. Dagegen kaufte ein durchaus großer Anteil von 28,6% immerhin 21 bis 50mal im Netz ein. Deutlich geringer fielen die Erfahrungen der Teilnehmer mit dem Kauf von Produkten aus, die sie im Quasi-Experiment bestellen konnten (Item os\_a: „Wie oft haben Sie bisher [CDs, DVDs, Bücher, Druckerpatronen bzw. Ticket] im Internet eingekauft?“). Mit 50,53% hatte etwa die Hälfte der Vpn noch kein Produkt der jeweiligen Art bestellt, 26,72% kauften bereits ein- bis fünfmal ein ähnliches Produkt (D-18).

Von den  $N=378$  Teilnehmern der Untersuchung gaben 17 (4,5%) an, keine Muttersprachler zu sein. Alle Nicht-Muttersprachler wurden gebeten, auf einer siebenstufigen Likert-Skala einzuschätzen, wie gut ihre Deutschkenntnisse seien. Von ihnen schätzten 15 Teilnehmer ihr Deutsch als maximal gut ein (Wert 6); nur zwei Probanden bewerteten es mit einem Wert von 5, also der zweithöchsten Ausprägung. Entsprechend ist nicht zu befürchten, dass Verständnisprobleme bzgl. der Versuchsinstruktionen aufgrund mangelnder Deutschkenntnisse aufgetreten sind.

## 7.2 Beschreibung der Versuchsumgebung

Die Durchführung des Quasi-Experimentes erfolgte an unterschiedlichen Orten. Ausschlaggebend für den konkreten Ort war die Präferenz der Teilnehmer. In den meisten Fällen wurde der Versuch im Computerlabor des Psychologischen Instituts an der Kieler Uni durchgeführt. Gelegentlich fand die Untersuchung auch in den privaten Räumlichkeiten der Teilnehmer oder der Versuchsleiter statt. Eine Erhebungssitzung wurde in einem Computerraum einer Fortbildungsinstitution für Arbeitslose abgehalten. Die konkrete Testumgebung variierte entsprechend deutlich. Es wurden stets die jeweils vorhandenen Computer genutzt, die sich recht stark bzgl. ihrer Hardware- und Softwareausstattung unterscheiden konnten. Die Tabelle 7-3 stellt dar, welche Arten von Internetverbindung bei der Durchführung des Experiments genutzt wurden. Da ein Großteil der Erhebungssitzungen an der Uni stattfand, konnte zumeist eine sehr schnelle Verbindung zum Internet genutzt werden. Zumeist fand die Erhebung gruppenweise statt, wobei im Höchstfall fünf Probanden zur selben Zeit anwesend waren. Einige Erhebungssitzungen wurden mit nur einem Teilnehmer durchgeführt. Fünf verschiedene Versuchsleiter bestritten die gesamte Erhebung. Der Tabelle 7-4 kann entnommen werden, wie viele Teilnehmer die einzelnen Versuchsleiter erhoben haben.

Tab. 7-3: Während des Versuchs genutzte Verbindung ins Internet

Verbindung zum Netz	Häufigkeit	%
33,6K-Modem	1	0,3
56K-Modem	16	4,2
ISDN	31	8,2
schneller als ISDN	330	87,3
Gesamt	378	100,00

Tab. 7-4: Anteil der erhobenen Teilnehmer pro Versuchsleiter

Versuchsleiter	Anzahl der Teilnehmer	%
1	49	13
2	12	3,2
3	27	7,1
4	78	20,6
5	212	56,1
Gesamt	378	100

### 7.3 Computergestützte Durchführung des Versuchs

Zur Durchführung des Quasi-Experimentes wurde eine computergestützte Versuchsumgebung erstellt. Es wurde ein PHP-Skript verwendet unter Nutzung eines Servers, der von Seiten des Rechenzentrums der Kieler Uni zur Verfügung gestellt wurde. Abbildung 7-1 zeigt beispielhafte eine Seite der Online-Versuchsumgebung. Im Anhang D finden sich weitere Screenshots zentraler Einzelseiten dieses sequentiell aufgebauten Hypertext-Dokuments. Der Zugriff erfolgte mit Hilfe der allgemein gängigen Browsersoftware. Durch ausführliche Testungen im Vorfeld wurde gewährleistet, dass sowohl die Seitendarstellung als auch die Speicherung der Datenfiles im Falle der gängigen Browser-Typen fehlerlos funktionierte. Der Zugang wurde durch eine Abfrage des Vpn-Codes gesichert (siehe Abschnitt 5.7). Durch diesen Code erfolgt die Zuordnung der Shops bzw. Versuchsbedingungen zu den einzelnen Teilnehmern.

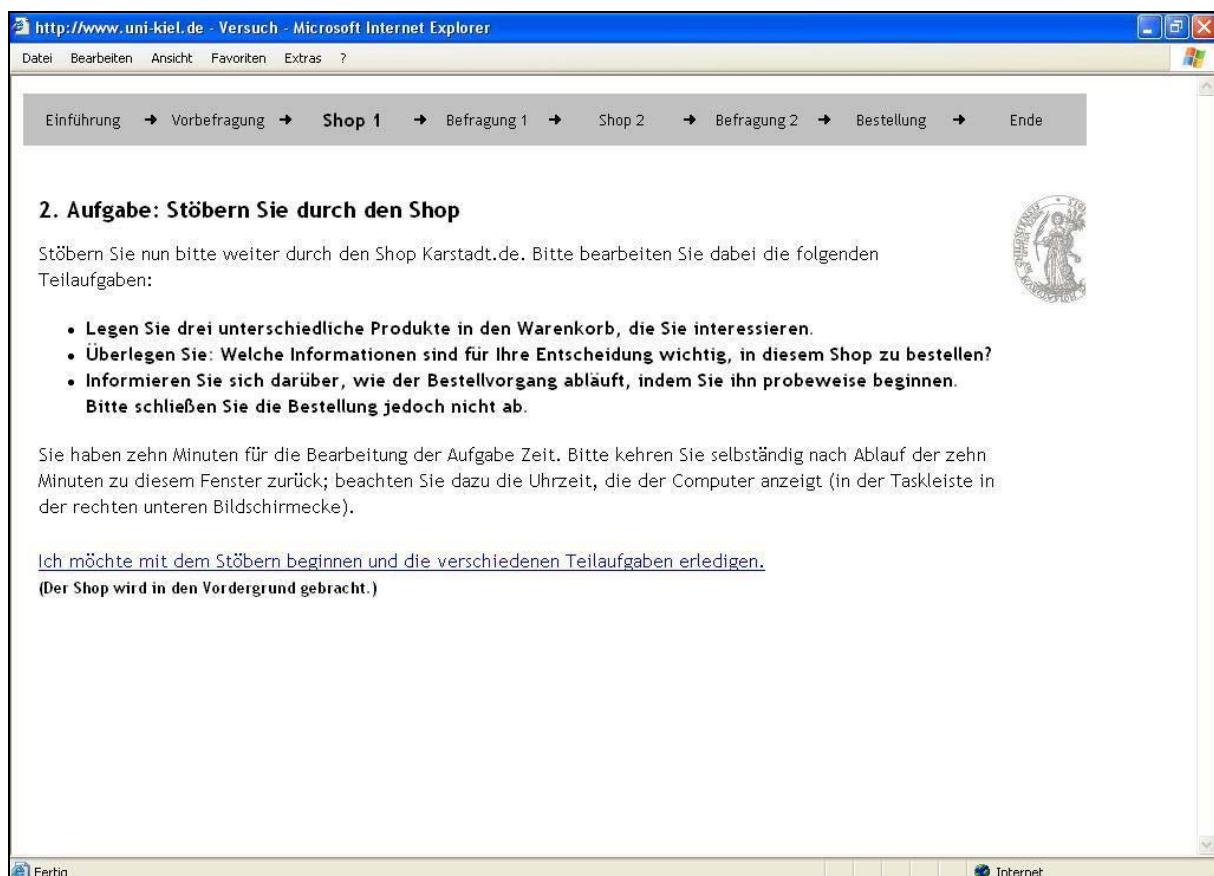


Abb. 7-1: Screenshot einer Seite der Hypertext-Versuchsumgebung

Auf den Seiten der Online-Versuchsumgebung wurden die Instruktionen zur Durchführung des Quasi-Experimentes präsentiert. Es wurde darauf verzichtet, die Teilnahmeinstruktionen während des Versuchs zusätzlich in Papierform vorzulegen, da dies in der Vorstudie zu Problemen führte (Christophersen, 2002, S. 68). Einige Seiten der Versuchsumgebung waren derart umfangreich, dass den Teilnehmern abverlangt wurde, die Scrollfunktionen der Browser-Software zu nutzen, um die Instruktionen im Ganzen lesen zu können. Die

Probanden wurden zu Beginn des Versuchs instruiert, nicht die History-Funktionen des Browsers (*Vor* und *Zurück*) zu nutzen, da dies zu Problemen beim Abruf der Seiten bzw. der Speicherung der Daten führen konnte. Die Vorwärts-Navigation sollte mit Hilfe der Hyperlinks erfolgen, die jeweils in blau gekennzeichnet am Ende der Einzelseiten stehen. Durch diese Einschränkungen wurde ein weitgehend einheitlicher Ablauf des Versuchs gewährleistet. Am oberen Rand jeder Einzelseite wurde eine Orientierungshilfe gegeben, anhand derer die Probanden entnehmen konnten, wie weit sie bereits im Versuchsablauf fortgeschritten waren bzw. in welcher der verschiedenen Versuchsphasen sie sich gerade befanden.

Die Präsentation des Fragebogens zur Erfassung der verschiedenen Konstrukte erfolgte ebenfalls mittels der Hypertext-Versuchsumgebung. Die computergestützte Abfrage der Items erwies sich einer vorangegangenen Untersuchung als unproblematisch (Konradt et al., 2003). Richter (1999) erzielte beim Vergleich von Pencil/Paper- und online-basierter Erhebung des Usability Fragebogens ISONorm weitgehend dieselben Ergebnisse. Verschiedene Studien weisen daraufhin, dass die Ergebnisse eines präsentierten Fragebogens jedoch durchaus davon abhängen können, welche Bildschirmdarstellungsform gewählt wird (vgl. Andrews, Nonnecke & Preece, 2003; Couper, Traugott & Lamias, 2001; Norman, Friedman, Norman & Stevenson, 2001). Die Ergebnisse einer Online-Befragungsstudie von Smyth, Dillman, Christian und Stern (2006) weisen auf einen deutlichen Einfluss der räumlichen Gestaltung von Antwortoptionen auf das Antwortverhalten von Respondenten hin. Es gilt, das Gestalt-Gesetz der Nähe zu berücksichtigen, wonach nahe beieinander stehende Bestandteile einer Webseite als zusammengehörig wahrgenommen werden (Metzger, 1953).

The image shows two identical Likert scale items presented in a web environment. Each item consists of a title, three response options, and seven radio buttons. The first item is titled "Die Preise der einzelnen Produkte sind sofort ersichtlich." and the second is titled "Der Warenkorb ist übersichtlich." Both items have the response options "stimme gar nicht zu", "stimme voll zu", and "keine Angabe". The radio buttons are arranged in a horizontal line below the response options, with the first radio button corresponding to "stimme gar nicht zu", the last to "stimme voll zu", and the middle one to "keine Angabe".

Abb. 7-2: Art der Itemabfrage in der verwendeten Hypertext-Versuchsumgebung

Vor dem Hintergrund der genannten Befunde wurde ein eher geringer Abstand zwischen der einzelnen Itemformulierung und den dazu gehörigen Radio-Buttons zur Beantwortung gewählt. Die Items zur Erfassung der Konstrukte und Kontrollvariablen verteilten sich insgesamt über drei Seiten der Versuchsumgebung. Der Abstand zwischen den Items wurde ausreichend groß gewählt, um die Gefahr zu minimieren, dass die Teilnehmer beim Beantworten des Fragebogens in die falsche Zeile rutschen. Abbildung 7-2 zeigt einen

Screenshot mit einem Ausschnitt von zwei Items aus dem verwendeten Fragebogen. Die Entscheidung, die Items nicht jeweils auf einzelnen Seiten darzubieten, sondern auf drei Seiten aufzuteilen, wurde angesichts von Studienergebnissen getroffen, wonach die Beantwortung von Single-Item-Pages deutlich mehr Zeit erfordert als im Fall von Multi-Item-Pages (Couper et al., 2001; Norman et al., 2001).

Für jeden Teilnehmer wurde automatisch ein Ergebnisfile in einem .txt-Dateiformat erstellt, das auf dem Server des Rechenzentrums abgespeichert wurde und alle im Rahmen des Versuchs erhobenen Daten dokumentierte. Mit Hilfe einer zusätzlichen Applikation ließ sich jedes Einzelfile in ein Dateiformat konvertieren, in dem die Daten derart aufbereitet enthalten waren, dass sie problemlos in einen Datensatz der Statistik-Software SPSS werden konnten.

## 7.4 Versuchsablauf

Im Folgenden wird der Ablauf des Experiments dargestellt. Nach mündlicher Begrüßung der Vp durch den Versuchsleiter wurde die Online-Versuchsumgebung am jeweiligen Rechnerarbeitsplatz durch Eingabe des im Vorfeld zugewiesenen Vpn-Codes gestartet. Außerdem wurde der Proband darum gebeten, einen Zettel mit einer Einverständniserklärung zur Teilnahme zu unterzeichnen (siehe Anhang D). Diese Erklärung enthielt einige Formalien, z.B. dass die Vpn mit der anonymisierten Auswertung der erhobenen Daten einverstanden seien. Im Folgenden werden die zentralen Instruktionen, die den Teilnehmern durch die Versuchsumgebung präsentiert wurden, wiedergegeben. Der Anhang D enthält den gesamten Wortlaut der Anweisungen. Der Versuchsablauf gliedert sich in acht Phasen; Abbildung 7-3 visualisiert den Ablauf schematisch.

### *Phase 1: Einführung*

Die Einführungsphase umfasste drei Seiten der Online-Versuchsumgebung. Auf der ersten wurden die Teilnehmer begrüßt. Es wurde darauf hingewiesen, dass jegliche im Rahmen des Experiments erhobenen Daten vertraulich behandelt werden bzw. dass der Versuch nicht auf Leistungen des Probanden sondern vielmehr auf die Testung der präsentierten Online-Shops fokussiere. Auf der zweiten Seite wurden die Teilnehmer gebeten, das Produkt anzugeben, auf das sie sich im Vorfeld der Datenerhebung festgelegt hatten. Des Weiteren sollten die drei Items pi1, pi2c und pi3 zum *Produkt Involvement* beantwortet werden (siehe Abschnitt 5.2.7). Die dritte Seite gab einen kurzen Überblick über den Ablauf des Versuchs. Es wurde außerdem darauf aufmerksam gemacht, dass im Laufe des Versuchs mehrere Browserfenster zur selben Zeit geöffnet sein werden, von denen keines eigenständig geschlossen werden sollte. Bei Fragen stand stets der Versuchsleiter zur Verfügung.



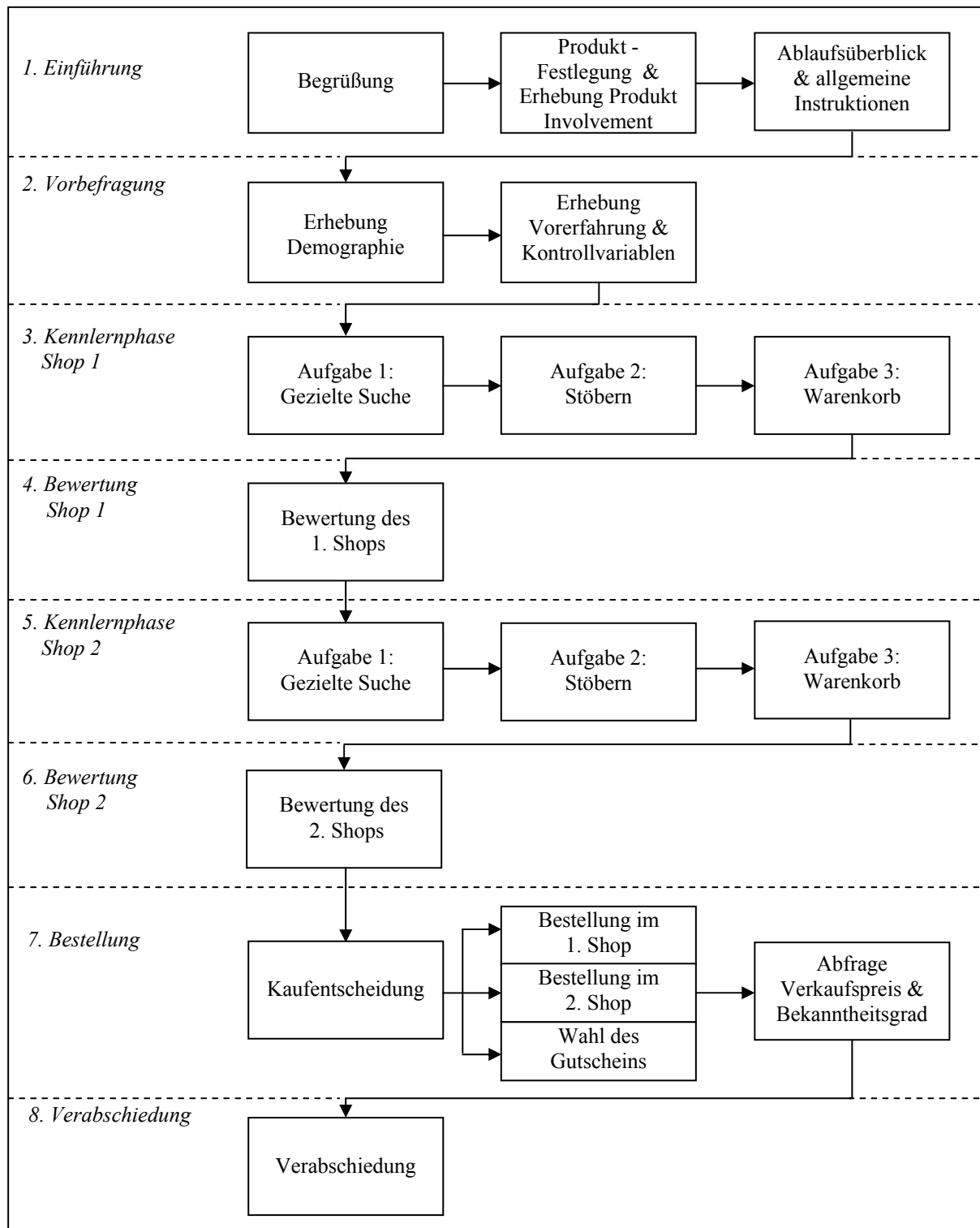


Abb. 7-3: Schematische Darstellung des Versuchsablaufs

*Phase 2: Vorbefragung*

In diesem zweiten Abschnitt des Quasi-Experiments, der lediglich zwei Hypertextseiten umfasste, wurden zunächst die verschiedenen Variablen zur Demographie (Indikatoren: alter, sex, einko, haus, beruf, schul; siehe Abschnitt 5.4.1) erhoben. Die zweite der beiden Seiten enthielt die Items zur Vorerfahrung in den Bereichen PC-Nutzung (pc\_seit, pc\_t, pc\_s), Internet-Nutzung (www\_st, www\_t, www\_s) und Online-Shopping (os\_q, os\_p; siehe Abschnitt 5.4.2). Zusätzlich wurde die Art der Internetverbindung abgefragt (modem; siehe Abschnitt 5.4.3).

*Phase 3: Kennen lernen von Shop 1*

In der dritten Phase des Versuchs sollten die Teilnehmer den ersten der beiden Online-Shops ausführlich kennen lernen. Dazu wurden ihnen drei Aufgaben gestellt, die einen breiten Freiraum zur Nutzung des Shop-Interfaces einräumen. Die *erste Aufgabe* bezog sich auf eine gezielte Suche nach einem Produkt. Die Instruktion lautete im Fall der Produktkategorie CD:

*Suchen Sie in diesem Shop genau die CD heraus, die Sie sich im Vorfeld des Versuchs ausgesucht haben (Eingabe des Nutzers aus Phase 1 des Versuchs). Legen Sie drei Exemplare dieser CD in den Warenkorb (also dreimal genau dasselbe Produkt!), bestellen Sie aber bitte noch nichts.*

Je nachdem, welcher Produktart die Vp zuzuordnen war, variierte bei dieser Instruktion der Wortlaut. In der ersten Klammer stand dabei jene Produktbezeichnung, die der Proband auf der zweiten Seite der Versuchsumgebung selber angegeben hatte. Durch Klicken auf einen Link mit der Bezeichnung „Ich möchte jetzt mit der Suche nach der CD beginnen“ (wiederum je nach Produktart anders lautend) öffnete sich zusätzlich zu dem Fenster der Versuchsumgebung ein zweites Browserfenster. In diesem wurde automatisch die Startseite des ersten Online-Shops aufgerufen. Die Vp wurde angewiesen, wieder zur Versuchsumgebung zurück zu kommen, nachdem sie diese erste Aufgabe erledigt hatte. Dazu sollte sie auf das entsprechende Kästchen mit der Bezeichnung *Versuch* in der Taskleiste klicken. Die Versuchsumgebung zeigte nun eine Seite an, auf der nochmals die Instruktionen der ersten Aufgabe paraphrasiert wurden. Insofern die Teilnehmer die Aufgabe nicht vollständig erledigt hatten, sollten sie dies nachholen. Nach vollständiger Bearbeitung der Aufgabe wurden sie gebeten, mit dem Versuch fort zu fahren.

Die *zweite Aufgabe (Stöbern)* in Phase 3 gliederte sich in drei Teilaufgaben. Der Wortlaut der Instruktionen war wie folgt:

*Legen Sie drei unterschiedliche Produkte in den Warenkorb, die Sie interessieren.*

*Überlegen Sie: Welche Informationen sind für Ihre Entscheidung wichtig, in diesem Shop zu bestellen?*

*Informieren Sie sich darüber, wie der Bestellvorgang abläuft, indem Sie ihn probeweise beginnen. Bitte schließen Sie die Bestellung jedoch nicht ab.*

Im Rahmen dieser Aufgabe sollten die Teilnehmer also durch den Online-Shop stöbern, um weiterreichende Erfahrungen in der Nutzung der Schnittstelle zu sammeln. Durch den zweiten Teil der dreigeteilten Aufgabe sollte gewährleistet werden, dass die Probanden alle aus ihrer Sicht relevanten Informationen eingesehen hatten, wenn sie in dem Shop eine Bestellung abgaben. Für die Bearbeitung der gesamten Aufgabe 2 wurden zehn Minuten eingeräumt. Nach Ablauf dieses Zeitraums sollten die Probanden selbstständig zum Versuchsfenster zurückkehren, um den Versuch mit der nächsten Aufgabe fort zu setzen. Falls die Teilnehmer weniger als zehn Minuten für die Bearbeitung benötigten, konnten sie auch mit dem Versuch fortfahren, bevor dieser Zeitraum verstrichen war. Auf der Folgeseite der Versuchsumgebung wurden die Teilnehmerinstruktionen wiederholt, damit die Probanden prüfen konnten, inwiefern sie die Bearbeitung der zweiten Aufgabe erfolgreich abgeschlossen hatten. War dies nicht der Fall, wurden sie dazu aufgefordert, die Aufgabe vollständig beenden, bevor sie mit dem Versuch fort fuhren.

Die *dritte Aufgabe* betraf den Umgang mit dem *Warenkorb* des jeweiligen Online-Shops. Die Instruktion lautete wie folgt - wobei je nach Produktart wiederum Abwandlungen dieses Textes verwendet wurden:

*Verändern Sie den Warenkorb nun so, dass sich darin nur noch ein Exemplar der CD befindet, auf die Sie sich im Vorfeld der Untersuchung festgelegt haben (Eingabe des Teilnehmers). Falls Sie bestellen möchten, tun sie dies bitte noch nicht.*

Eine Zeitbegrenzung wurde in diesem Fall nicht gegeben. Die Teilnehmer sollten wie zuvor eigenständig zum Versuchsfenster des Browsers zurückkehren, sobald sie die dritte Aufgabe erledigt hatten. Dort fanden sie eine Seite vor, auf der die Aufgabe nochmals paraphrasiert wurde. Falls die Probanden diese noch nicht vollständig erledigt hatten, sollten sie dies nachholen und ansonsten mit dem Versuch durch Klicken auf einen entsprechenden Hypertextlink fortfahren.

*Phase 4: Bewertung von Shop 1*

Im Anschluss an die Bearbeitung der drei Aufgaben folgte die Phase der Bewertung des Online-Shops, den die Teilnehmer gerade kennen gelernt hatten. Die Präsentationsreihenfolge der gestellten Items wurde dabei systematisch variiert, um eventuelle Reihenfolgeeffekte der Items in der Beantwortung aufdecken zu können. In den Abschnitten 5.6.1 und 5.6.2 wurden diese Variationen dargestellt. Gemäß einer einleitenden Instruktion wurde es den Teilnehmern offen gestellt, auch während der Beantwortung zwischenzeitlich zum Shop-Fenster wechseln, um sich ein besseres Bild bzgl. der abgefragten Bewertungsaspekte machen zu können. Zur Bewertung dienten die Indikatoren, die die Konstrukte Kaufintention (ki1, ki2, ki3, ki4), Vertrauen (t1, t2, t3, t4), Usability reflektiv (ufosV2r: ur1, ur2c, ur3,..., ur9), Usability formativ (ufosV2f: uf1, uf2, uf3, ..., uf58), wahrgenommene Produktanzahl (wg1, wg2, wg3), Ästhetik (a1, a2), Einkaufsvergnügen (se1, se2), Service (sv1, sv2, sv3,..., sv6) sowie wahrgenommener Preis (subjc) erfassen sollten. Da sich diese Variablen auf die Bewertungen auf den ersten der beiden Online-Shops bezogen, wurden sie bei der Speicherung der Daten mit der zusätzlichen Endung „\_s1“ kodiert (z.B. ki1\_s1).

Aufgrund der recht hohen Anzahl an Items wurde der gesamte Online-Fragebogen über drei Seiten der Versuchsumgebung gestreckt. Beantworteten die Probanden mehr als fünf Items auf einer drei Fragebogenseiten nicht, erhielten sie eine in roter Schrift dargestellte Fehlermeldung, wenn sie auf den Hyperlink zur jeweils nächsten Seite klickten. Die Teilnehmer wurden durch diese Meldung gebeten, die nicht getroffenen Bewertungen nachzuholen. Die nicht beantworteten Items wurden ebenfalls mit roter Schrift gekennzeichnet, damit schnell ersichtlich war, welche Beantwortungen nachträglich vorzunehmen waren.

*Phase 5: Kennen lernen von Shop 2*

Nach abgeschlossener Bewertung des ersten Online-Shops lernten die Teilnehmer in Phase 5 den zweiten Shop ihrer Versuchsbedingung kennen. Es wurden ihnen dieselben Aufgaben wie in Phase 3 zum ersten Shop gestellt. Zur Verbesserung des Leseflusses fanden sich marginale Variationen im Wortlaut der Instruktionen, ansonsten bestehen keine Unterschiede zur dritten Phase des Versuchs. In der dritten Aufgabe sollte wie bereits beim ersten Shop jenes Produkt im Warenkorb des Shops verbleiben, auf das sich die Teilnehmer im Vorfeld der Untersuchung festgelegt hatten.

*Phase 6: Bewertung von Shop 2*

Die Bewertung des zweiten Shops erfolgte auf identische Art und Weise wie zuvor. Bei der Speicherung der Daten wurde jeder Indikator mit der zusätzlichen Endung „\_s2“ kodiert (z.B. ki1\_s2), um im Datensatz zwischen den Beantwortungen des Fragebogens für Shop 1 und Shop 2 unterscheiden zu können.

*Phase 7: Bestellung*

Nachdem die Teilnehmer die beiden Online-Shops in den Phasen 3 bis 6 kennen gelernt und bewertet hatten, wurde in der siebten Phase des Versuchs die Kaufentscheidung operationalisiert bzw. erhoben; diese Phase war in der Statusleiste des Versuchsfensters mit *Bestellung* betitelt. Die Probanden wurden vor die Wahl gestellt, das zuvor herausgesuchte Produkt entweder im ersten oder zweiten Shop zu bestellen oder alternativ einen Gutschein zu erhalten. Im Falle einer Bestellung wurden ihnen im Anschluss an das Quasi-Experiment 15 Euro in bar ausgezahlt. Von diesem Geld konnten sie die anfallenden Kosten für den Kauf des Produktes begleichen. Sollten die Kosten mehr als 15 Euro betragen, so mussten sie den Restbetrag selber tragen. Es erfolgte in jedem Fall die Auszahlung der kompletten 15 Euro, auch dann, wenn die Kosten für die Bestellung niedriger ausfallen würden. Der 15 Euro Gutschein, der als dritte Wahlmöglichkeit zur Verfügung stand, konnte bei einem Einkauf in einer Filiale der Warenhauskette Karstadt eingelöst werden. Die Teilnehmer erhielten die Information, dass der Gutschein jedoch nicht direkt im Anschluss an das Experiment ausgehändigt werde, sondern erst nach Abschluss der kompletten Datenerhebung.

Entschied sich eine Vp für eine Bestellung in einem der beiden Shops, so wurde sie angewiesen, diese durchzuführen, den Versuchsleiter jedoch kurz vor dem endgültigen Abschicken zu sich zu rufen. Auf diese Weise sollte verhindert werden, dass die Vp unwillentlich eine fehlerhafte Bestellung abgab. In direkter Anwesenheit des Versuchsleiters schloss sie den Kauf ab. Es wurde dabei die Möglichkeit eingeräumt, sich für einen Kauf im jeweils anderen Shop oder den Erhalt des Gutscheins um zu entscheiden, solange die Bestellung nicht endgültig abgeschickt wurde. Dadurch wurde berücksichtigt, dass sich Online-Kunden gerade während des Bestellvorgangs häufig umentscheiden und eine zuvor intendierte Bestellung doch nicht abgeben. Vielfach sind diese Abbrüche darin begründet, dass in dieser Phase des Einkaufs Benutzungsprobleme auftreten oder bestimmte negative Rahmenbedingungen wie etwa zu hohe Kosten für den Versand oder eingeschränkte Zahlungsoptionen erst zu diesem Zeitpunkt vom Kunden wahrgenommen werden (Abschnitt 2.5.1).

Entschied sich die Vp für den Erhalt des Gutscheins, so wurde sie gebeten, ein Codewort einzugeben, das sie bei Abholung des Gutscheins zum späteren Zeitpunkt angeben solle. Sowohl die Kaufentscheidung, also die Wahl zwischen *Kauf in Shop 1*, *Kauf in Shop 2* oder *Gutschein*, als auch eine eventuelle Revidierung dieser Entscheidung wurden automatisch im Ergebnisfile der Vp protokolliert.

Im Anschluss an die Erhebung dieser abhängigen Variable sollten die Teilnehmer angeben, wie hoch der Verkaufspreis des Produkts im ersten und zweiten Shop ausfiel. Bei einer Entscheidung für einen Kauf in einem der beiden Shops, wurde zusätzlich erfragt, ob die Vp über das eine Produkt hinaus noch weitere Artikel bestellt hatte bzw. wie hoch die Bestellkosten insgesamt waren. Die Phase 7 endete mit einer Seite, auf der der Bekanntheitsgrad beider Shops aus Sicht der Teilnehmer erhoben wurde (*bek\_s1*, *bek\_s2*; siehe Abschnitt 5.2.7).

*Phase 8: Ende*

Der Versuch endete mit der achten Phase, die zwei Seiten der Versuchsumgebung umfasste. Auf der ersten wurden die Vp gebeten, freiwillig ihre Email-Adresse anzugeben für den Fall, dass eine Nachbefragung stattfinden würde. Die letzte Seite verabschiedete die Probanden.

Im Anschluss an den gesamten Versuch wurden der Vp die 15 Euro Teilnahmeentschädigung von Seiten des Versuchsleiters ausgezahlt, insofern sie sich für einen Kauf in einem der beiden Shops entschieden hatte. Abschließend wurde die Vp persönlich verabschiedet.

Für den Fall, dass während der Versuchsdurchführung Probleme auftraten, stand der Versuchsleiter stets als Ansprechpartner zur Verfügung. Allerdings griff er im Wesentlichen nur bei Schwierigkeiten im Versuchsablauf oder der Handhabung der Versuchsumgebung ein. Traten Verständnisschwierigkeiten bzgl. der Versuchsinstruktionen auf, so paraphrasierte der Versuchsleiter die Anweisungen, die durch die Online-Versuchsumgebung gestellt wurden. Es wurden von Seiten des Versuchsleiters keine bewertenden Kommentare zu den Shops abgegeben, die die Teilnehmer in ihrem Urteil hätten beeinflussen können. Hilfestellungen zur Benutzung der Online-Shops erhielten die Probanden erst dann, wenn sie mit dem Versuch ohne weitere Unterstützung nicht fortfahren konnten.

## **7.5 Probleme der Erhebung**

Bei der Durchführung des Quasi-Experiments traten insgesamt selten Probleme auf. In den meisten Fällen erwiesen sich sowohl der Versuchsablauf als auch die Teilnehmerinstruktionen für die Probanden als verständlich. Es zeigte sich, dass einige der Teilnehmer mit einer besonders langen Bearbeitungsdauer (teils mehr als 120 Minuten, siehe dazu Abschnitt 8.5 im Ergebnisteil) erhebliche Verständnisprobleme bzw. Probleme in der Interaktion mit dem Computer hatten. Es war augenfällig, dass gerade ältere Probanden oft mehr Zeit benötigten als jüngere. Die Erhebung älterer Teilnehmer erwies sich allgemein oft als problematisch, da diese über weniger Medienkompetenz im Bereich der PC- bzw. Internetnutzung verfügten. Die Untersuchung an sich wurde von den meisten Vpn im persönlichen Gespräch im Nachhinein als interessant und nicht überfordernd bewertet. Allerdings empfanden einige Probanden die Anzahl der abgefragten Items als recht hoch.

Die zum Zwecke der Erhebung programmierte Versuchsumgebung lief weitgehend stabil. Meist konnte der Versuch auch dann regulär fortgesetzt werden, wenn das Versuchsfenster im Laufe des Experiments plötzlich nicht mehr vorhanden war, z.B. weil ein Proband es unentdeckt geschlossen hatte. Dazu musste ein neues Browserfenster geöffnet und die URL zum Versuch eingegeben werden. Die Versuchsumgebung öffnete sich dann auf der letzten durch den Teilnehmer aufgerufenen Seite. Dieses Vorgehen führte jedoch nur dann zum Erfolg, wenn mindestens ein Browserfenster noch immer vorhanden war. In Einzelfällen war dies nicht der Fall, und der Versuchsablauf musste erneut gestartet werden, wobei die betroffenen Teilnehmer bereits beantwortete Items nicht erneut bearbeiten mussten. Die bereits

vorhandenen Daten der Probanden wurden nicht gelöscht, sondern im Nachhinein händisch mit denen der jeweils zweiten Session zusammen gefügt. Dieses Vorgehen wurde angewandt im Falle eines Computerabsturzes kurz vor Abgabe der Bestellung (m5\_1\_1) sowie bei einer Teilnehmerin, die versehentlich alle Browser-Fenster schloss (m6\_2\_2). Zu derartigen Problemen führte auch die Funktion *Merkliste* einer bestimmten Gruppe von Online-Shops (JPC.de, WOM.de, Incinea.de, ABC-CD.de). Bei einem Klick auf den Link zur Merkliste dieser Shops öffnete sich diese im Versuchsfenster. Normalerweise sollte sich die Merkliste in einem zusätzlichen Fenster öffnen, dies funktionierte jedoch im Rahmen des Versuchs nicht problemlos.

Häufig äußerten die Teilnehmer negative Kommentare zur Gestaltung der Online-Shops, nur in wenigen Fällen wurden spezifische Bestandteile der Benutzerschnittstellen explizit positiv bewertet. In vielen Fällen führte die mangelnde Benutzerfreundlichkeit dazu, dass die gestellten Aufgaben nicht ohne Hilfe des VL bewältigt werden konnten. Zwei Vpn schlossen versehentlich in einem der Shops eine Bestellung ab, da der Bestellvorgang dieses Anbieters nicht transparent genug gestaltet war. Diese Bestellungen konnten jedoch problemlos storniert werden. Einige Teilnehmer zeigten sich angesichts der mangelhaften Gestaltung einzelner Shops emotional erregt. Weiterhin beobachteten die Versuchsleiter in einigen Fällen, dass die dritte Subaufgabe der Aufgabe 2 nicht korrekt bearbeitet wurde. Statt den Bestellvorgang probeweise durch die Eingabe von Kundendaten zu beginnen, wurde nur die erste Seite der Bestellphase in Augenschein genommen. Diese Personen wurden von Seiten des VL gebeten, den Bestellvorgang eingehender zu betrachten.





## 8. Ergebnisse

Die Ergebnisdarstellung folgt nicht immer exakt dem Ablauf der vorgenommenen Analysen. In Einzelfällen werden Ergebnisse später dargestellter Auswertungsschritte unter Verweis auf den jeweiligen Kapitelabschnitt vorweggenommen. Um gezielt einen Abschnitt innerhalb dieses Kapitels zu finden, der sich auf eine konkrete Hypothese aus Kapitel drei bezieht, sei auf Tabelle 6-1 verwiesen.

### 8.1 Vorbereitende Datenanalysen

Vor Überprüfung der Hypothesen wurden einige vorbereitende Analysen durchgeführt, die in den Abschnitten 8.1.1 bis 8.1.5 dargestellt werden.

#### 8.1.1 Aufbereitung der Daten: Zusammenfassung der Bewertungen

Für die  $N=378$  Teilnehmer liegen jeweils zwei Shopbewertungen vor. In dem entsprechenden Datensatz mit *beiden Bewertungen pro Fall* ist somit die Information enthalten, wie der eine Shop im Verhältnis zum jeweils anderen bewertet wird. Jeder Indikator zur Erfassung der Erfolgsfaktoren bzw. -größen findet sich somit zweimal in diesem Datensatz. Für viele der Auswertungsschritte wird die Information der verhältnismäßigen Bewertung beider Shops nicht benötigt. Es erscheint daher plausibel, jeden Fall aufzuteilen. Auf diese Weise verdoppelt sich die Anzahl der Datenreihen auf  $N=756$ . In diesem Datensatz mit *einer Bewertung pro Fall* ist jeder Teilnehmer somit zweimal vertreten, und jeder Bewertungsaspekt wird durch einen Indikator abgedeckt. Die Zusammenfassung zu einem großen Datensatz erscheint gerechtfertigt, da sich eventuelle Reihenfolgeeffekte der Shoppräsentation aufgrund der Ausbalancierung der verschiedenen Versuchsbedingungen über den gesamten Datensatz hinweg ausgleichen dürften (siehe Abschnitte 5.6.3 und 5.7). Da jedoch nicht alle Versuchsbedingungen zur Produktart Veranstaltungstickets gefüllt werden konnten (siehe Abschnitt 7.1.2), soll statistisch geprüft werden, inwiefern ein Einfluss der Shopreihenfolge besteht.

Entsprechend wird ein t-Test für unabhängige Stichproben berechnet (siehe Abschnitt 6.9.1; Datensatz mit zwei Bewertungen pro Fall). Dabei wird untersucht, ob sich signifikante Mittelwertsunterschiede in der AV Kaufintention<sup>21</sup> für die Erhebung an erster und an zweiter Stelle ergeben. Die Mittelwerte dieser Variable in beiden Bedingungen werden in Tabelle E-1 im Anhang E wiedergegeben. Die Voraussetzung des t-Tests, dass die untersuchte Variable in beiden Gruppen der Normalverteilung folgt, wird verletzt (siehe Abschnitt 8.1.3). Da der t-Test jedoch robust auf Verletzungen dieser Voraussetzung reagiert, wird trotzdem eine Berechnung durchgeführt. Die Annahme der Varianzhomogenität kann aufgrund eines signifikanten Ergebnisses des Levene-Tests ebenfalls nicht aufrechterhalten werden (siehe Tabelle E-2 im Anhang E). Für den Fall, dass diese Voraussetzung zur Berechnung des

---

<sup>21</sup> An dieser Stelle wird auf die Ergebnisse eines später dargestellten Analyseschrittes vorgegriffen: Als Wert für die Kaufintention dient der Mittelwert über die Items ki1, ki2 und ki4 ( $\alpha=0,91$ ; siehe Abschnitt 8.3.1).

t-Tests verletzt wird, steht eine alternative Statistik für ungleiche Varianzen zur Verfügung, anhand derer die Signifikanz des Mittelwertunterschiedes geprüft werden kann. Diese weist ein insignifikantes Ergebnis auf ( $n_1=n_2=368$ ;  $t(716, 9)=0,374$ , n.s.). Somit kann gefolgert werden, dass es in Bezug auf die AV Kaufintention keinen Einfluss hat, ob die Bewertung an erster oder zweiter Stelle des Versuchsablaufs erfolgte. Entsprechend erscheint es gerechtfertigt, den Datensatz mit zwei Bewertungen pro Fall aufzuteilen und mit dem verdoppelten Datensatz mit  $N=736$  Fällen bei einer Bewertung pro Fall zu rechnen. Für jeden im Folgenden dargestellten Analyseschritt wird angegeben, welcher Datensatz bei der Berechnung zugrunde liegt (*Datensatz mit zwei Bewertungen pro Fall vs. Datensatz mit einer Bewertung pro Fall*).

### 8.1.2 Analyse und Behandlung fehlender Werte

In den folgenden drei Abschnitten wird die Analyse und Behandlung fehlender Werte dargestellt. Zunächst werden in Abschnitt 8.1.2.1 diejenigen Fälle identifiziert, die einen zu hohen Missing-Anteil aufweisen und aus der weiteren Auswertung ausgeschlossen werden. Abschnitt 8.1.2.2 stellt die Analyse des Fehlendmechanismus vor, und in Abschnitt 8.1.2.3 wird über die Ersetzung fehlender Werte für formative und reflektive Indikatoren berichtet.

#### 8.1.2.1 Elimination von Fällen mit zu hohem Anteil fehlender Werte

Der fallweise Anteil fehlender Werte wird anhand des Datensatzes mit zwei Bewertungen pro Fall bestimmt. Der Anteil fällt unter den Teilnehmern erwartungsgemäß sehr unterschiedlich aus. In Tabelle 8-1 wird eine Übersicht gegeben, wie viele Fälle jeweils wie viel Prozent fehlender Werte aufweisen. Mit 16,67% ist der Anteil an Teilnehmern mit weniger als einem Prozent Missings recht hoch. Ein großer Teil von 88,62% der Probanden hat mehr als 90% aller 202 Items, die im Rahmen des Versuchs gestellt wurden, beantwortet. Zehn Fälle weisen mehr als 20% fehlende Werte auf und werden entsprechend von der weiteren Analyse ausgeschlossen (siehe Abschnitt 6.1.1); in Tabelle E-3 im Anhang werden die Vpn-Codes dieser Probanden aufgelistet. Nach diesen Eliminationen verbleiben  $N=368$  Fälle im Datensatz mit jeweils zwei Bewertungen pro Fall bzw.  $N=736$  Fälle im Datensatz mit einer Shopbewertung pro Fall.

Tab. 8-1: 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>
<i>0 bis 1%</i>	63	63	16,67	16,67
<i>1 bis 3%</i>	86	149	22,75	39,42
<i>3 bis 5%</i>	78	227	20,63	60,05
<i>5 bis 10%</i>	108	335	28,57	88,62
<i>10 bis 15%</i>	20	355	5,29	93,92
<i>15 bis 20%</i>	13	368	3,44	97,35
<i>20 bis 25%</i>	6	374	1,59	98,94
<i>25 bis 28%</i>	4	378	1,06	100

Anmerkung: Anzahl Teilnehmer als absolute Häufigkeit (n). Lies erste Zeile: Bei 63 Teilnehmern (16,67% der Stichprobe) liegt der Anteil fehlender Werte pro Fall zwischen 0 und 1%.

### 8.1.2.2 Analyse des Fehlendmechanismus

Es werden zwei methodische Ansätze verfolgt, um die Frage nach dem zugrunde liegenden Fehlendmechanismus zu untersuchen (siehe Abschnitt 6.1.2). Zur Prüfung, inwiefern die abhängigen Variablen in einem Zusammenhang zum Fehlendmechanismus stehen, werden jeweils für beide Shops die bivariaten Korrelationen zwischen dem fallweisen Prozentteil an Missing Data und dem Skalenwert der Kaufintention ( $ki_1$ ,  $ki_2$ ,  $ki_4$ ; zur Skalenanalyse siehe Abschnitt 8.3.1) berechnet. Grundlage dieser Analyse ist somit der Datensatz mit zwei Bewertungen pro Fall. Keine der beiden Korrelationen weist einen Koeffizienten über  $r=0,09$  bzw. eine Signifikanz auf einem  $\alpha=0,05$ -Niveau auf. Die Berechnung des *MCAR-Test nach Little* ergibt eine Signifikanz von  $p=0,366$  ( $\chi^2=53720,29$ ;  $df=53609$ ), so dass ein zufälliges Auftreten fehlender Werte unterstellt werden kann (vgl. Tabachnik & Fidell, 2006, S. 63). Auf Grundlage der angeführten Überlegungen zu den Ursachen fehlender Werte im Datensatz sowie der beiden rechnerischen Analysen kann davon ausgegangen werden, dass entweder ein MCAR- oder ein MAR-Mechanismus vorliegt (siehe Abschnitt 6.1.2).

### 8.1.2.3 Ersetzung fehlender Werte

Wie in Abschnitt 6.1.3 dargestellt, kommt im Falle der reflektiven Indikatoren zur Erfassung der Konstrukte Usability (ufosV2r-Skala), Vertrauen, Kaufintention, wahrgenommene Produktanzahl, Einkaufsvergnügen sowie Ästhetik eine besondere Form der Mittelwertersetzung fehlender Werte zur Anwendung. Nach Durchlaufen der Ersetzungsprozedur für die reflektiven Indikatoren verbleiben insgesamt 60 manifeste Variablen im Datensatz, die fehlende Werte aufweisen. Es ergeben sich bei einem  $N=368$  somit 1212 Zellen fehlender Werte von insgesamt 44160 Zellen (Datensatz mit zwei Bewertungen pro Fall). Der Anteil an Missing Data liegt damit bei 2,74%. Die fallweise Ersetzung durch den Mittelwert der übrigen Indikatoren derselben Skala kann dann nicht erfolgen, wenn ein Teilnehmer bei einer Skala keinen einzigen Indikator beantwortet hat. Für diese wenigen Fälle werden die fehlenden Werte mit Hilfe der Multiplen Imputation ersetzt (siehe Tabelle E-4 im Anhang E zur Anzahl der Fälle für jede der reflektiven Skalen). Die Ersetzung der verbleibenden fehlenden Werte erfolgt mit Hilfe der Multiple Imputation unter Nutzung der Software Norm. Da ein Missing at Random (*MAR*) Fehlendmechanismus unterstellt wird und sich das Verfahren darüber hinaus als sehr robust gegenüber Verletzungen von Verteilungsannahmen erweist (Schafer & Graham, 2002), sind die Voraussetzungen zur Berechnung gegeben. In Tabelle E-5 im Anhang E werden alle Indikatoren unter Angabe ihres Anteils fehlender Werte aufgelistet, die in den Imputationsdatensatz aufgenommen werden. Grundlage der Berechnung ist der Datensatz mit einer Bewertung pro Fall ( $N=736$ ).<sup>22</sup>

---

<sup>22</sup> Der im ersten Schritt der Multiplen Imputation berechnete EM (Expectation Maximation) Algorithmus konvergiert bei einem standardmäßigen Konvergenzkriterium von 0,0001 nach 18 Iterationen.

### 8.1.3 Verteilungen der erhobenen Indikatoren

Die Verteilungen der Indikatoren werden anhand des Datensatzes mit einer Bewertung pro Fall untersucht. Bei allen reflektiven manifesten Variablen zeigen sich signifikante Ergebnisse im Kolmogorov-Smirnov-Test (sämtlich  $p < 0,001$ ), so dass von einer Abweichung von der Normalverteilung ausgegangen werden müsste. Allerdings weist Brosius (2004) darauf hin, dass es aufgrund einer Signifikanz des Kolmogorov-Smirnov-Tests nicht unbedingt gerechtfertigt ist, eine Variable als nicht normalverteilt anzusehen. Daher werden zusätzlich Schiefe (Sch) und Exzess (Ex) sowie die entsprechenden z-Werte der Verteilungen bestimmt (Tabelle E-7 im Anhang E). Es zeigt sich, dass für keinen der Indikatoren eine Normalverteilungsannahme aufrechterhalten werden kann. Eine mögliche methodische Maßnahme zur Herstellung von Normalverteilungen bestünde nun in einer Transformation der Variablen. Dies brächte zwar einerseits den Vorteil, dass eine wesentliche Voraussetzung zur Berechnung vieler statistischer Verfahren erfüllt wäre. Andererseits ergäbe sich das Problem, dass transformierte Daten vielfach schlechter interpretiert werden können als nicht transformierte (vgl. Tabachnik & Fidell, 2006, S. 86 f.). Da die wesentlichen Auswertungsschritte (PLS und MLR) nicht zwingend normalverteilte Variablen voraussetzen, erscheint die Notwendigkeit einer Transformation nicht gegeben zu sein.

### 8.1.4 Prüfung auf Vorliegen eines Common Method Bias

Zur Untersuchung, inwiefern ein Common Method Bias im Datensatz eine Rolle spielt, wird der Harman's Single-Factor Test berechnet (siehe Abschnitt 6.3). Grundlage ist der Datensatz mit einer Bewertung pro Fall ( $N=736$ ). Es wird eine Hauptkomponentenanalyse (HKA) ohne Rotation durchgeführt unter Einbeziehung der Indikatoren zu den Konstrukten Usability (formative und reflektive Skalen ufosV2f bzw. ufosV2r), Vertrauen, wahrgenommene Produktanzahl, Produkt-Involvement, Ästhetik, Service, Einkaufsvergnügen, wahrgenommener Preis, Kaufintention und einiger weiterer Einzelindikatoren. Insgesamt werden bei Berechnung der HKA 19 Komponenten mit einem Eigenwert  $\lambda > 1$  extrahiert. Wie in Tabelle E-6 im Anhang E entnommen werden kann, klärt die erste Komponente einen Anteil von 29,02% der Gesamtvarianz auf, die zweite Komponente noch 4,97%. In Tabelle E-8 im Anhang E wird die Komponentenmatrix wiedergegeben. Ein Großteil der 80 Indikatoren weist eine Ladung  $a_{il} > 0,3$  auf dem ersten Faktor auf; lediglich bei zwölf Items nimmt die Ladung einen kleineren Wert an. Insgesamt tragen die Komponenten 2 bis 19  $\Sigma=41,53\%$  zur Erklärung der Modellvarianz bei. Anhand des eher schwachen Kriteriums des Harman's Single-Factor Tests kann somit davon ausgegangen werden, dass kein kritischer Common Method Bias in den Daten vorherrscht.

### 8.1.5 Identifikation problematischer Fälle

Wie in Abschnitt 6.4 ausführlich dargestellt wurde, dienen drei Analysen der Identifikation problematischer Fälle (Datensatz mit zwei Bewertungen pro Fall). Es werden a) Teilnehmer mit auffällig langer bzw. kurzer Bearbeitungsdauer (weniger als 45 Minuten und mehr als 120 Minuten) identifiziert. Weiterhin erfolgt b) eine fallweise Analyse der Varianz im Antwortverhalten sowie c) eine Prüfung der Konsistenz des Antwortverhaltens bei der Beantwortung negativ gepolter Items.

Die mittlere Bearbeitungsdauer liegt bei 65,5 Minuten. Eine Streuung von  $SD=17,9$  weist darauf hin, dass die Dauer der Untersuchungsdurchführung stark variiert. Das Minimum liegt bei 27 Minuten (Vp b6\_3\_1); die längste Bearbeitungsdauer beträgt 186 Minuten (Vp m6\_4\_2\_1). Im Fall der Probandin b6\_3\_1 bedeutet eine Durchführungsdauer von 27 Minuten, dass alle 7,4 Sekunden ein Item beantwortet wurde – vorausgesetzt die Teilnehmerin hätte die gesamte Zeit mit der Beantwortung der Fragen verbracht. Die genannten Bearbeitungsdauern deuten auf eine rechtsschiefe Verteilung hin, was sich in einem Schiefeffizientwert  $Sch=1,42$  ( $SE=0,13$ ) widerspiegelt. Ein Exzess von  $Ex=6,11$  ( $SE=0,26$ ) zeigt einen eher spitzen Gipfel der Verteilung an. Zwischen dem Alter und der Dauer der Bearbeitung liegt eine signifikante Korrelation von  $r=0,34$  ( $p<0,001$ ) vor. Ältere Teilnehmer benötigen somit tendenziell länger für die Untersuchung. Elf Teilnehmer werden als auffällig identifiziert, da sie weniger als 45 Minuten für die Durchführung des Quasi-Experiments benötigen. Bei fünf Teilnehmern liegt die Bearbeitungsdauer bei über 120 Minuten.

Insgesamt werden in der Analyse der Varianz im Antwortverhalten  $n=20$  Teilnehmer auffällig. Bei der Bewertung des ersten Shops ergibt sich bei fünf Teilnehmern über alle Items eine Streuung von  $SD\leq 1$ . Beim zweiten Shop fällt die Anzahl der als „Durchklicker“ identifizierten Probanden mit  $n=15$  deutlich höher aus. Keiner der Teilnehmer zeigt eine auffällig niedrige Varianz in der Beurteilung *beider* Shops.

Zur Prüfung der Konsistenz des Antwortverhaltens wird die Abweichung zwischen dem Mittelwert der positiv gepolten Usability-Indikatoren und dem negativ gepolten Item ur2c sowie die Abweichung zwischen se1 und se2c berechnet. Zehn Teilnehmer weisen im Fall der reflektiven Usability-Skala ein inkonsistentes Antwortmuster auf, zwölf im Fall der Skala Einkaufsvergnügen. Es wird kein Proband identifiziert, der sowohl bei der reflektiven Usability-Skala als auch der Skala zum Einkaufsvergnügen eine inkonsistente Bewertung des zweiten Shops zeigt.

Eine Übersicht über die insgesamt  $n=54$  Probanden, die in mindestens einer der drei durchgeführten Analysen auffällig werden, gibt Tabelle E-9 im Anhang E; dabei wird durch eine graue Färbung der Zellen verdeutlicht, in welcher Analyse der jeweilige Teilnehmer auffällig ist.

## 8.2 Erfassung der Variablen auf Prädiktorebene

Im Folgenden werden die Ergebnisse der drei Analyseschritte der in Abschnitt 6.5 vorgestellten Prozedur zur Gütebeurteilung reflektiver Maße nach Homburg und Giering (1996) wiedergegeben (Abschnitte 8.2.1 bis 8.2.3). Sämtliche Berechnungen werden auf Basis des Datensatzes mit einer Bewertung pro Fall ( $N=736$ ) angestellt. Um zu prüfen, inwiefern die reflektiven Skalen in der Lage sind, Merkmalsunterschiede zwischen den herangezogenen Online-Shops zu identifizieren, wird zusätzlich eine Varianzanalyse berechnet (siehe Abschnitt 8.2.4).

### 8.2.1 Explorative Faktorenanalyse zur Prüfung der Skalenstruktur

Im ersten Schritt der Güteanalyse der reflektiven Skalen auf Prädiktorebene wird mit Hilfe der Methode der explorativen Faktorenanalyse (EFA) untersucht, inwiefern sich die theoretisch angenommene Struktur bestätigt. Es werden mehrere Hauptachsenanalysen (HAA) unter Verwendung der Rotationstechnik *Direct Oblimin* berechnet. Der Delta-Wert wird mit  $\delta=-2$  festgelegt – es wird also ein mittleres Ausmaß der Korrelation zwischen den Komponenten zugelassen (siehe Abschnitt 6.5.1), da Zusammenhänge zwischen den latenten Variablen aus theoretischer Sicht durchaus plausibel erscheinen. Einbezogen werden die reflektiven Skalen zur Erfassung der Konstrukte Usability (ufosV2r; Items ur1 bis ur9), Einkaufsvergnügen (se1, se2c), Produkt-Involvement (pi1, pi2, pi3), Vertrauen (t1 bis t4), wahrgenommene Produktanzahl (wa1 bis wa3) und Ästhetik (ae1, ae2). Die grundsätzlichen Voraussetzungen zur Berechnung einer EFA sind gegeben, da angesichts der Vorgabe des siebenstufigen Likert-Antwortformates keine univariaten Ausreißer auftreten und die Stichprobengröße mit  $N=736$  als sehr gut zu bewerten ist (vgl. Abschnitt 6.5.1.1). Kritisch ist jedoch zu werten, dass für zwei der angenommenen Konstrukte jeweils nur zwei Items erhoben wurden (Einkaufsvergnügen, Ästhetik), obwohl gemäß Bühner (2004, S. 157) mehr Indikatoren pro Konstrukt in die EFA eingehen sollten (siehe Abschnitt 6.5.1.1).

Es werden mehrere HAA berechnet, da sich die erwartete Faktorenlösung zunächst nicht zeigt. Durch die iterative Elimination einzelner Items wird versucht, die angenommene Skalenstruktur und damit eine angemessene faktorielle Validität der Skalen zu erreichen. Tabelle 8-2 gibt einen Überblick über die insgesamt fünf berechneten EFAs. Lediglich die Mustermatrizen der vierten und fünften Analyse werden in diesem Kapitel wiedergegeben; die Matrizen der ersten drei HAA finden sich im Anhang E (Tabellen E-10, E-11 sowie E-12). Wie der Tabelle 8-2 entnommen werden kann, sind die Voraussetzungen zur Berechnung der HAA (bis auf die oben genannte Einschränkung der Indikatorenanzahl pro Konstrukt) für alle fünf Analysen gegeben, da sich ausreichend hohe KMO- bzw. MSA-Werte zeigen und die Bartlett-Tests auf Sphärizität signifikante Werte aufweisen.

Tab. 8-2: 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=0,931; alle MSA-Werte über 0,5; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis ( <i>ungefähr</i> $\chi^2=12635,11$ ; $df=253$ ; $p<0,001$ ) → <b>Voraussetzungen zur Berechnung erfüllt*</b>	Extraktion von fünf Faktoren gemäß der angenommenen Anzahl; Kommunalitäten der Indikatoren wa3 (wahrge. Produktanzahl) und pi2 (Produkt-Involvement) weisen Ausprägungen von $h^2<0,5$ auf → Elimination von wa2 und pi2	Tab. E-10 (Anhang E)
2	KMO=0,931; alle MSA-Werte über 0,5; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis ( <i>ungefähr</i> $\chi^2=12151,35$ ; $df=210$ ; $p<0,001$ ) → <b>Voraussetzungen zur Berechnung erfüllt*</b>	Extraktion von fünf Faktoren gemäß der angenommenen Anzahl; alle Kommunalitäten liegen bei $h_i^2>0,5$ ; auf dem ersten Faktor laden die neun Usability-Items; Usability-Item ur5 weist jedoch eine Nebenladung $a>0,3$ auf → Elimination von ur5	Tab. E-11 (Anhang E)
3	KMO=0,925; alle MSA-Werte über 0,5; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis ( <i>ungefähr</i> $\chi^2=11427,67$ ; $df=190$ ; $p<0,001$ ) → <b>Voraussetzungen zur Berechnung erfüllt*</b>	Extraktion von fünf Faktoren gemäß der angenommenen Anzahl; alle Kommunalitäten liegen bei $h_i^2>0,5$ ; die beiden Items zum Einkaufsvergnügen laden nicht erwartungskonform: se2c lädt gemeinsam mit den acht Usability Items auf dem ersten Faktor; se1 lädt mit den beiden Items zur Ästhetik auf dem fünften Faktor → Elimination von se1 und se2c	Tab. E-12 (Anhang E)
4	KMO=0,905; alle MSA-Werte über 0,5; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis ( <i>ungefähr</i> $\chi^2=9967,86$ ; $df=153$ ; $p<0,001$ ) → <b>Voraussetzungen zur Berechnung erfüllt*</b>	Extraktion von vier statt fünf erwarteten Faktoren; die Items ae1 und ae2 zur Ästhetik laden auf dem ersten Faktor, auf dem ansonsten die acht Items zur Usability laden → Berechnung einer fünften HAA, in die nur die Items zur Usability und Ästhetik eingehen, um zu prüfen, inwiefern sich Eindimensionalität der Usability und Ästhetik bestätigt	Tab. 8-3
5	KMO=0,920; alle MSA-Werte über 0,5; Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis ( <i>ungefähr</i> $\chi^2=5992,58$ ; $df=45$ ; $p<0,001$ ) → <b>Voraussetzungen zur Berechnung erfüllt*</b>	Eindimensionalität bestätigt sich nicht; es werden zwei Faktoren extrahiert: erster Faktor mit den acht Items zur Usability, zweiter Faktor mit den zwei Items zur Ästhetik	Tab. 8-4

**Anmerkung:** \* Als Einschränkung der Voraussetzungserfüllung ist die teilweise geringe Anzahl an Indikatoren pro Konstrukt zu sehen. Kaiser-Meyer-Olkin-Maß (KMO), Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_i$ ), Measure of Sampling Adequacy (MSA), Freiheitsgrade (df), Signifikanzniveau (p), Chi-Quadrat-Wert ( $\chi^2$ ).

Eine erste HAA wird nicht als zufrieden stellend angesehen, da sich für die Kommunalitäten der Indikatoren wa2 und pi2 zu geringe Werte ergeben ( $h^2<0,5$ ; siehe Mustermatrix in Tabelle E-10). Die Items werden zur Berechnung der zweiten HAA eliminiert. In dieser Analyse ergibt sich eine Nebenladung des Usability-Items ur5 (siehe Mustermatrix in Tabelle E-11), das entsprechend ebenfalls eliminiert wird. In der dritten HAA laden die beiden Items zum Einkaufsvergnügen nicht erwartungskonform (siehe Mustermatrix in Tabelle E-12). Nach

Ausschluss dieser Items werden in der vierten HAA anstelle der erwarteten fünf Faktoren lediglich vier extrahiert (siehe Tabelle 8-3). Die beiden Indikatoren zur Ästhetik laden gemeinsam mit den acht Items zur Usability auf dem ersten Faktor, der mit 44,25% einen Großteil der Gesamtvarianz ( $\Sigma=67,44\%$ ) erklärt. Ein eigenständiger Ästhetik-Faktor zeigt sich somit entgegen der Erwartungen nicht.

Tab. 8-3: Mustermatrix der vierten berechneten Hauptachsenanalyse

Faktor	1	2	3	4	$h_i^2$	MSA-Wert
Usability ur7	<b>0,835</b>	0,089	0,037	0,082	0,956	0,818
Usability ur6	<b>0,821</b>	0,054	0,031	-0,075	0,948	0,679
Usability ur1	<b>0,797</b>	0,128	0,023	0,036	0,957	0,749
Usability ur4	<b>0,796</b>	0,074	0,061	-0,089	0,947	0,654
Usability ur3	<b>0,741</b>	0,045	0,049	0,040	0,962	0,604
Usability ur9	<b>0,741</b>	0,093	0,010	0,132	0,943	0,691
Usability ur2	<b>0,719</b>	-0,004	0,042	-0,010	0,950	0,517
Usability ur8	<b>0,588</b>	0,137	-0,025	0,282	0,940	0,620
Ästhetik ae1	<b>0,565</b>	0,038	-0,041	0,298	0,864	0,532
Ästhetik ae2	<b>0,509</b>	0,057	-0,025	0,254	0,850	0,433
Vertrauen t4	0,107	<b>0,865</b>	0,019	0,035	0,894	0,846
Vertrauen t2	0,096	<b>0,805</b>	0,041	0,079	0,902	0,763
Vertrauen t1	0,201	<b>0,712</b>	0,001	0,141	0,933	0,734
Vertrauen t3	0,165	<b>0,670</b>	0,086	0,067	0,932	0,608
Produkt-Involvement pi3	0,046	0,004	<b>0,884</b>	-0,015	0,557	0,791
Produkt-Involvement pi1	-0,022	0,009	<b>0,841</b>	0,031	0,536	0,707
Wahrge. Produktanzahl wa1	0,028	0,037	-0,010	<b>0,899</b>	0,812	0,844
Wahrge. Produktanzahl wa2	0,041	0,050	0,022	<b>0,711</b>	0,782	0,548
$\Sigma=67,44$	44,25	8,87	8,37	5,94		
$\lambda_{j=5}=0,963$	8,292	1,877	1,772	1,358		

Anmerkung: Hauptachsenanalyse (HAA), Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_j$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

Um die Eindimensionalität des ersten Faktors zu prüfen, wird eine weitere HAA berechnet, in die als Variablen lediglich die acht Usability- sowie die beiden Ästhetik-Indikatoren eingehen. Eine Zusammenfassung der beiden Skalen erscheint dann angemessen, wenn als Ergebnis der Faktorenanalyse nur ein Faktor extrahiert wird. Das Ergebnis dieser fünften HAA wird in Tabelle 8-4 dargestellt. Es werden zwei Faktoren extrahiert. Auf dem ersten laden die acht Indikatoren zur Usability. Dieser Faktor klärt 61,44% der Gesamtdatenvarianz auf, welche bei  $\Sigma=69,76$  liegt. Der zweite Faktor umfasst die beiden Indikatoren zur Ästhetik und erklärt einen Anteil von 8,32% der Varianz. Die Analyse bestätigt die Eindimensionalität somit nicht. Es wird entsprechend die Entscheidung getroffen, beide Skalen getrennt voneinander in die weitere Analyse einzubinden. Auf Basis der berechneten explorativen Faktorenanalysen kann die **Hypothese H2**, in der angenommen wurde, die Skala ufosV2r erweise sich gegenüber Skalen zur Erfassung weiterer kundennaher Erfolgsfaktoren als faktoriell valide, insgesamt als bestätigt angesehen werden.



Tab. 8-4: Mustermatrix der fünften berechneten Hauptachsenanalyse

Faktor	1	2	$h_i^2$	MSA-Wert
Usability ur7	<b>0,905</b>	-0,001	0,818	0,948
Usability ur1	<b>0,880</b>	-0,061	0,756	0,954
Usability ur6	<b>0,852</b>	-0,149	0,697	0,95
Usability ur9	<b>0,841</b>	-0,026	0,699	0,93
Usability ur4	<b>0,829</b>	-0,143	0,659	0,937
Usability ur3	<b>0,765</b>	0,027	0,594	0,954
Usability ur2	<b>0,736</b>	-0,094	0,522	0,947
Usability ur8	<b>0,720</b>	0,113	0,564	0,912
Ästhetik ae2	0,479	<b>0,665</b>	0,803	0,820
Ästhetik ae1	0,547	<b>0,647</b>	0,863	0,83
$\Sigma=69,76$	61,44	8,32		
$\lambda_{j=3}=0,618$	6,443	1,041		

Anmerkung: Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_j$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

### 8.2.2 Prüfung der Skalengüte auf Einzelskalenebene

Im Folgenden werden die Ergebnisse des zweiten Prüfschrittes zur Messgüte der reflektiven Skalen nach Homburg und Giering (1996) vorgestellt. In Tabelle 8-5 werden die Kriterienausprägungen der reflektiven Skalen und ihrer Indikatoren zusammengefasst wiedergegeben.

Nachdem im ersten Schritt der Skalengüteprüfung der Indikator ur5 eliminiert wurde (siehe Abschnitt 8.2.1), stehen noch acht reflektive Usability-Items zur Verfügung, die gemeinsam die Skala ufosV2r bilden. Das Cronbachs Alpha dieser Skala liegt bei einem sehr hohen Wert von  $\alpha=0,94$ . Es könnte vermutet werden, dass ein Grund für diese hohe Ausprägung in der vergleichsweise großen Indikatorenanzahl der Skala liegt; dagegen spricht jedoch die Tatsache, dass auch bei der Verwendung von lediglich drei der Usability-Indikatoren (u1, ur7, ur9) noch ein  $\alpha=0,90$  erreicht wird, was ebenfalls als zufrieden stellend hoch anzusehen ist. Somit kann die Reliabilität der ufosV2r-Skala zur Erfassung dieses Kernkonstruktes als sehr gut gewertet werden. Es bestätigt sich damit die **Hypothese 1**, in der angenommen wurde, die reflektive Usability-Skala ufosV2r sei reliabel (siehe Abschnitt 3.3). In Tabelle E-13 im Anhang E werden Cronbachs Alpha-Werte für die insgesamt 35 Benutzerschnittstellen angegeben, die in die Untersuchung einbezogen wurden. Die Varianz der Werte ist groß; der niedrigste Wert liegt bei  $\alpha=0,83$  ( $n=11$  Bewertungen) für den Shop Getgo.de, der höchste Wert wird mit  $\alpha=0,96$  ( $n=8$  Bewertungen) für den Shop Karten-Online.de erreicht.

Die interne Konsistenz der Vertrauensskala ist mit einem  $\alpha=0,91$  als sehr gut zu beurteilen. Da zur Messung der Konstrukte wahrgenommene Produktanzahl, Ästhetik und Produkt-Involvement 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 bestimmt wird. Diese liegt für die wahrgenommene Produktanzahl ( $r=0,74$ ;  $\alpha=0,85$ ) und das Produkt-Involvement ( $r=0,76$ ;  $\alpha=0,85$ ) im noch zufrieden stellenden Bereich und kann im Fall der Ästhetik ( $r=0,84$ ;  $\alpha=0,91$ ) als gut angesehen werden.

Tab. 8-5: Skalen- und Itemkennwerte der reflektiven Skalen auf Prädiktorenebene

Konstrukt/ Skalenkennwerte	Item- Code	M	SD	$a_{ij}$	% Missing	$r_{it}$	$p_m$
Usability (ufosV2r) $\alpha=0,936$ $M=3,63$ $SD=1,54$ Mittlere korr. Trennschärfe=0,779 Homogenität=0,652 Erklärte Varianz: 65,65% Reduzierung auf 3 Items (ur1, ur7, u9): $\alpha=0,902$	ur1	3,58	1,88	0,869	2,38	0,839	0,597
	ur2	4,12	1,96	0,718	1,19	0,691	0,687
	ur3	3,50	1,88	0,764	1,19	0,735	0,583
	ur4	4,31	1,58	0,802	1,19	0,774	0,718
	ur6	4,07	1,82	0,825	1,19	0,793	0,678
	ur7	3,45	1,90	0,905	3,97	0,872	0,575
	ur8	2,80	1,91	0,737	3,31	0,711	0,467
	ur9	3,25	1,91	0,843	2,51	0,816	0,542
	Ästhetik $\alpha=0,908$ Spearman-Brown Korr. $r=0,837$ $M=2,82$ $SD=1,81$ Mittlere korr. Trennschärfe=0,837 Homogenität=0,837 Erklärte Varianz: 83,02 %	a1	2,89	1,93	0,911	1,59	0,837
a2		2,74	1,85	0,911	4,37	0,837	0,457
Vertrauen $\alpha=0,913$ $M=3,91$ $SD=1,42$ Mittlere korr. Trennschärfe=0,804 Homogenität=0,727 Erklärte Varianz: 73,07%	t1	3,84	1,70	0,847	7,94	0,798	0,639
	t2	4,16	1,45	0,876	7,80	0,824	0,693
	t3	3,63	1,63	0,772	8,33	0,736	0,606
	t4	4,01	1,54	0,917	6,75	0,857	0,669
Wahrgenommene Produktanzahl $\alpha=0,846$ Spearman-Brown Korr. $r=0,744$ $M=3,41$ $SD=1,90$ Mittlere korr. Trennschärfe=0,733 Homogenität=0,733 Erklärte Varianz: 73,25%	wa1	3,28	2,07	0,856	5,69	0,733	0,547
	wa2	3,55	2,00	0,856	4,23	0,733	0,592
Produkt-Involvement $\alpha=0,850$ Spearman-Brown Korr. $r=0,756$ $M=4,74$ $SD=1,26$ Mittlere korr. Trennschärfe=0,747 Homogenität=0,747 Erklärte Varianz: 74,61 %	pi1	4,52	1,44	0,864	0	0,747	0,753
	pi3	4,96	1,25	0,864	0	0,747	0,827

Anmerkung: Cronbachs Alpha ( $\alpha$ ), Mittelwert (M), Standardabweichung (SD), Faktorladung ( $a_{ij}$ ), Anteil fehlender Werte in Prozent (% Missing), korrigierte Trennschärfe ( $r_{it}$ ), Itemschwierigkeit ( $p_m$ ).

Für den Indikator ur4 ergibt sich eine vergleichsweise hohe Itemschwierigkeit von  $p_m=0,72$ . Da die Schwierigkeiten der übrigen Indikatoren der ufosV2r-Skala jedoch im mittleren Bereich liegen, wird dies nicht als problematisch angesehen. Neben dem Indikator ur4 zeigen auch die beiden Indikatoren pi1 und pi3 zum Produkt-Involvement hohe Itemschwierigkeiten. Die korrigierten Trennschärfen  $r_{it}$  sämtlicher Indikatoren sind ausreichend hoch. Auch die Ergebnisse der explorativen Faktorenanalysen fallen positiv aus. Bei allen reflektiven Skalen wird lediglich ein Faktor extrahiert, der jeweils mehr als 50% der Datenvarianz erklärt. Die Faktorladungen der Einzelindikatoren liegen in allen Fällen über  $a_{ij}=0,7$  und weisen damit

eine angemessene Höhe auf. Insgesamt kann auf Grundlage der Ergebnisse des zweiten Prüfungsschrittes von einer guten Messgüte der reflektiven Maße gesprochen werden.

### 8.2.3 Konfirmatorische Faktorenanalyse zur Prüfung der Skalenstruktur

Im dritten Schritt der Messgüteprüfung nach Homburg und Giering (1996) werden die reflektiven Skalen zur Usability (ufosV2r; ur1, ur2, ur3, ur4, ur6, ur7, ur8, ur9), Ästhetik (a1, a2), wahrgenommenen Produktanzahl (wa1, wa2) und zum Vertrauen (t1 t2, t3, t4) einer konfirmatorischen Faktorenanalyse (CFA) unterzogen (siehe Abschnitt 6.5.3). Zwar übersteigen die in die CFA einbezogenen Indikatoren nicht die kritischen Werte für Schiefe und Exzess von  $Sch > 2,0$  bzw.  $Ex > 7,0$  (siehe Tabelle E-7 im Anhang E), allerdings zeigt sich für die Variablenauswahl ein multivariater Exzesswert von  $Ex = 68,11$ . Es liegt somit eine Voraussetzungsverletzung der Schätzmethode Maximum Likelihood (ML) und der Generalized Least Square (GLS) vor. Entsprechend erscheint die Anwendung der Asymptotically Distribution-Free Methode (ADF) angemessen. Da 16 Indikatorvariablen in die CFA eingehen, ist ein  $N = 736$  als sehr gut anzusehen. Allerdings liegt die Anzahl der Indikatoren bei zwei der latenten Variablen (wahrgenommene Produktanzahl und Ästhetik) unterhalb der empfohlenen Anzahl von mindestens drei manifesten Variablen pro Faktor. Eine Einbeziehung der exogenen Variable Produkt-Involvement in die CFA kann nicht erfolgen, da dies eine Unteridentifikation des Modells zur Folge hätte. Es ergibt sich das in Abbildung 8-1 dargestellte CFA-Modell. Da korrelative Zusammenhänge zwischen den einbezogenen latenten Variablen nicht auszuschließen sind, werden diese im zu testenden Modell in Form von Kovarianzen berücksichtigt. Die Schätzung des Modells erfolgt mit Hilfe der Software AMOS in der Programm-Version 5.0.1 (Arbuckle & Wothke, 1999).

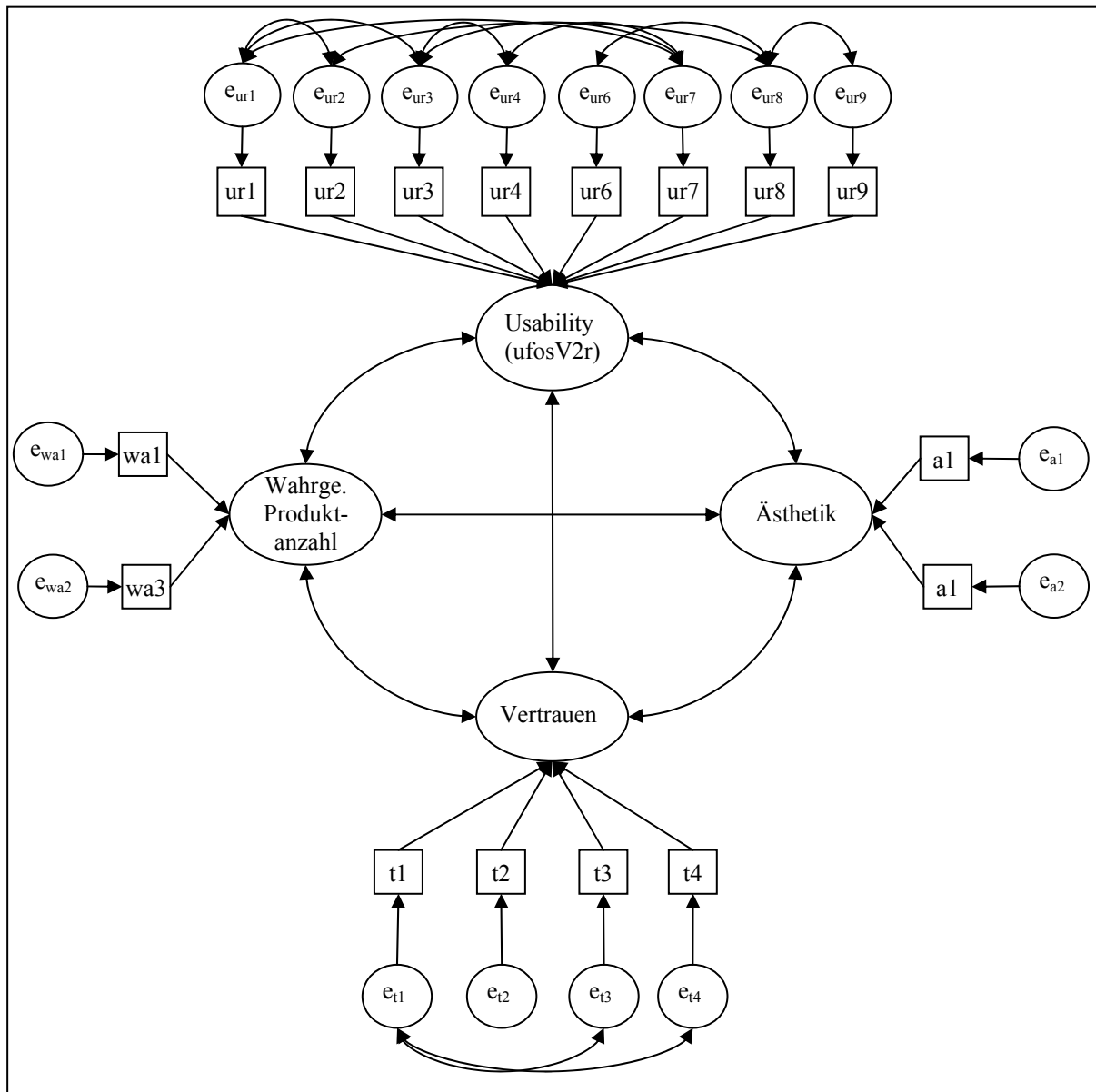


Abb. 8-1: CFA-Modell zur Überprüfung der Skalenstruktur

Anmerkung: Messfehler des Indikators (e).

Ein angemessener Fit des Modells wird erst dann erreicht, wenn Korrelationen zwischen den Fehlertermen innerhalb der latenten Variablen zugelassen werden. Derartige Modelle werden auch von anderen Autoren herangezogen, z.B. bei Bollen (1989, S. 235). Allerdings gibt Fornell (1983) zu bedenken, dass Korrelationen zwischen Fehlertermen nur dann zugelassen werden sollten, wenn (1) es aus theoretischer oder methodischer Sicht gerechtfertigt ist oder (2) keine wesentlichen Veränderungen der Strukturparameterschätzungen daraus folgen. Eine methodische Rechtfertigung ist z.B. gegeben, wenn sich ein Teil der Fehlervarianz auf die Form der Datenerhebung zurückführen lässt. Bagozzi (1983) fügt hinzu, dass (3) sich die Parameter des Messmodells nicht in signifikantem Maße ändern sollten. Gerbing und Anderson (1984) weisen darauf hin, dass die Einbeziehung von Kovarianzen zwischen Fehlertermen eine zugrunde liegende Struktur maskieren kann, auch wenn die Bedingungen (2) und (3) gegeben sind, was die Bedeutung der Voraussetzung (1) unterstreicht.

Im Fall der durchgeführten CFA werden keine wesentlichen Veränderungen von Parametern auf Struktur- bzw. Messmodellebene beobachtet, nachdem die Korrelationen zwischen Fehlertermen der Indikatoren im Modell zugelassen werden. Eine methodische Begründung könnte darin gesehen werden, dass die Darbietung der Items eines Konstruktes im Fragebogen stets geblockt erfolgte. Aufgrund dieses Umstandes könnten sich potenzielle Störvariablen bei der Beantwortung, wie z.B. eine Tendenz zum konsistenten Antworten, auf alle Indikatoren einer latenten Variable im gleichen Maße ausgewirkt haben, was eine Korrelation der Fehlerterme zur Folge hätte. Die Identifikation der zu korrelierenden Fehlerterme erfolgt anhand der Modifikations-Indizes, die durch AMOS berechnet werden.

Tab. 8-6: Standardisierte Faktorladungen und Messfehlervarianzen der CFA zur Analyse der Skalenstruktur

$\xi_i$	Indikator	$\lambda_{ij}$	$\theta$
Usability (ufosV2r) $\phi_1=2,656$	ur1	0,895	0,659
	ur2	0,757	1,469
	ur3	0,824	1,056
	ur4	0,809	0,782
	ur6	0,837	0,931
	ur7	0,947	0,350
	ur8	0,814	1,193
	ur9	0,847	0,986
Vertrauen $\phi_2=2,6$	t1	0,954	0,245
	t2	0,879	0,483
	t3	0,846	0,705
	t4	0,931	0,301
Wahrge. Produktanzahl $\phi_3=3,628$	wa1	0,945	0,439
	wa2	0,783	1,429
Ästhetik $\phi_4=3,641$	a1	0,996	0,030
	a2	0,870	0,842

Anmerkung: Standardisierte Faktorladung zwischen Indikator  $x_i$  und Faktor  $j$  ( $\lambda_{ij}$ ), Messfehlervarianz ( $\theta$ ), Faktor  $j$  ( $\xi_j$ ), geschätzte Varianz von  $\xi_j$  ( $\phi_j$ ).

Alle nicht standardisierten partiellen Regressionsgewichte sind auf einem Niveau von  $p < 0,001$  signifikant; die standardisierten Faktorladungen der Indikatoren finden sich in Tabelle 8-6. In Tabelle 8-7 werden darüber hinaus die Ausprägungen der relevanten Gütemaße zur Beurteilung des CFA-Modells aufgeführt. Der Koeffizient aus  $\chi^2$ -Wert und Freiheitsgraden liegt mit  $\chi^2=2,94$  wie gefordert unterhalb eines Werts von 3 (siehe Abschnitt 6.5.3). Der SRMR zeigt einen Wert von 0,067 und bleibt damit deutlich unterhalb des kritischen Maximalwerts von 0,11. Die Fit-Indizes NFI und CFI entsprechen ebenfalls dem geforderten Mindestwert von 0,9. Sämtliche Indikatorreliabilitäten weisen Werte über  $rel(x_i)=0,6$  auf; ferner sind die Faktorreliabilitäten deutlich größer als der geforderte Wert von  $rel(\xi_j)=0,6$ . Das Fornell-Larcker-Kriterium, wonach die durchschnittlich erfasste Varianz eines Faktors  $DEV(\xi_1)$  größer sein sollte als die quadrierte Korrelation mit den weiteren LVn, wird ebenfalls für jeden Faktor erfüllt.

Tab. 8-7: Evaluation des Messmodells der CFA

1. Globalkriterien:						
$\chi^2$	= 256,27	GFI	= 0,95	SRMR	= 0,067	
df	= 87	AGFI	= 0,92	RMSEA	= 0,05	
$\chi^2/df$	= 2,95	CFI	= 0,93	(LO 90 = 0,4; HI 90 = 0,6)		
NFI	= 0,90					
2. Detailkriterien:						
Konstrukt $\xi_j$	Indikator	Indikator- reliabilität $rel(x_i)$	Konvergenzvalidität		Diskriminanzvalidität	
			DEV( $\xi_j$ )	$rel(\xi_j)$	DEV( $\xi_j$ )	$\varphi_{ij}^2$
$\xi_1$ 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					
$\xi_2$ Vertrauen	t1	0,901	0,822	0,949	0,822	> 0,39 (Usability) > 0,16 (W. Produktanz.) > 0,22 (Ästhetik)
	t2	0,797				
	t3	0,714				
	t4	0,876				
$\xi_3$ Wahrge. Produkt- anzahl	wa1	0,881	0,745	0,853	0,745	> 0,27 (Usability) > 0,16 (Vertrauen) > 0,24 (Ästhetik)
	wa2	0,609				
$\xi_4$ Ästhetik	a1	0,992	0,880	0,936	0,880	> 0,50 (Usability) > 0,22 (Vertrauen) > 0,24 (W. Produktanz.)
	a2	0,766				

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

In Tabelle 8-8 werden die von AMOS berechneten Korrelationen zwischen den latenten Variablen wiedergegeben. Auffällig sind die hohen Korrelationen zwischen den Variablen Usability und Vertrauen ( $r=0,63$ ) sowie Usability und Ästhetik ( $r=0,71$ ). Insgesamt ist festzuhalten, dass sowohl die globalen Fit-Maße als auch die Detailkriterien auf eine angemessene Passung des Modells an die Daten hinweisen. Auf Grundlage dieser Analyse findet die **Hypothese 2**, wonach sich die reflektive Usability-Skala ufosV2r gegenüber anderen Skalen zu kundennahen Erfolgsfaktoren als faktoriell valide erweist, folglich eine zweite empirische Bestätigung (siehe auch Abschnitt 8.2.1). Einschränkend ist dabei zu berücksichtigen, dass dieser Fit erst nach dem Zulassen von Korrelationen zwischen Fehlertermen erreicht wird.

Tab. 8-8: Konstruktinterkorrelationen auf Basis der CFA

Konstrukt	$\xi_1$ Usability (ufosV2r)	$\xi_2$ Vertrauen	$\xi_3$ Wahrge. Produktanzahl	$\xi_4$ Ästhetik
$\xi_1$ Usability (ufosV2r)	1			
$\xi_2$ Vertrauen	0,626	1		
$\xi_3$ Wahrge. Produktanzahl	0,516	0,401	1	
$\xi_4$ Ästhetik	0,706	0,468	0,487	1

Anmerkung: Faktor j ( $\xi_j$ ).

### 8.2.4 Diskriminationsfähigkeit der reflektiven Skalen

Um zu prüfen, inwiefern die reflektiven Maße in der Lage sind, Unterschiede in der jeweiligen Merkmalsausprägung der 35 untersuchten Online-Shops zu identifizieren, wird eine mehrdimensionale einfaktorielle Varianzanalyse (ANOVA) berechnet (siehe Abschnitt 6.9.2). Die Shops stellen dabei die unterschiedlichen Ausprägungen der unabhängigen Variable dar. Als abhängige Variablen werden die Mittelwerte der vier reflektiven Skalen Usability (ufosV2r), Vertrauen, wahrgenommene Produktanzahl sowie Ästhetik herangezogen. Es wird die Nullhypothese  $H_0$  geprüft, dass keine Mittelwertunterschiede zwischen den Shops bestehen. Die Alternativhypothese  $H_1$  nimmt an, dass sich wenigstens ein Shop in seinem Mittelwert von den übrigen unterscheidet. Den Tabellen E-14 bis E-17 im Anhang E können die Mittelwerte und Standardabweichungen der Shops auf den Skalen Usability (ufosV2r), Vertrauen, wahrgenommene Produktanzahl und Ästhetik entnommen werden. Dabei werden pro Konstrukt getrennt nach den fünf Produktarten (Buch, DVD, CD, Druckerpatronen und Tickets) Rangreihen der jeweils sieben Online-Shops dargestellt.

Tab. 8-9: 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	-0,410	-4,56	-0,844	-4,689
Ästhetik	0,000	0,036	0,40	-1,155	-6,417
Vertrauen	0,000	-0,542	-6,02	-0,285	-1,583
Wahrgenommene Produktanzahl	0,000	-0,168	-1,87	-1,177	-6,539

Anmerkung: Exzess (Ex), Schiefe (Sch).

Da die Normalverteilung der einbezogenen Variablen eine Voraussetzung zur Berechnung der ANOVA darstellt, wird dies mit Hilfe von Kolmogorov-Smirnov-Tests überprüft. Es zeigen sich signifikante Ergebnisse in den Tests ( $p < 0,001$ ), so dass keine Normalverteilungen der Variablen angenommen werden können. Da es trotz einer Signifikanz des Kolmogorov-Smirnov-Tests möglich ist, dass die Werte einer Variable in der Grundgesamtheit annähernd normalverteilt sind (Brosius, 2004, S. 755), werden zusätzlich die Schiefe und der Exzess der Skalenwerte bzw. deren z-Werte bestimmt (siehe Tabelle 8-9). Diese weisen ebenfalls auf eine Abweichung von der Normalverteilung hin, was eine Voraussetzungsverletzung der Varianzanalyse darstellt. Letztlich erfolgt eine Berechnung der ANOVA trotzdem, da sich das Verfahren gegen Verletzungen der Normalverteilungsannahme als recht robust erweist.

Tab. 8-10: Ergebnisse der Levene-Tests auf Varianzhomogenität

<i>Abhängige Variable</i>	<i>Levene-Statistik</i>	<i>df1</i>	<i>df2</i>	<i>Signifikanz</i>
Usability (ufosV2r)	1,409	34	701	0,063
Ästhetik	1,389	34	701	0,072
Vertrauen	1,046	34	701	0,399
Wahrgenommene Produktanzahl	2,012	34	701	0,001

Anmerkung: Freiheitsgrade (df).

Um zu prüfen, inwiefern die Voraussetzung der Varianzhomogenität erfüllt wird, erfolgt die Berechnung von Levene-Tests (siehe Abschnitt 6.9.3). Die Ergebnisse werden in Tabelle 8-10 wiedergegeben. Im Fall der wahrgenommenen Produktanzahl ergibt sich ein signifikantes Ergebnis im Levene-Test, so dass hier die Voraussetzung der Varianzhomogenität als verletzt angesehen werden muss. Daher sind die Ergebnisse der Varianzanalyse, die in Tabelle 8-11 dargestellt werden, für diese Skala nur eingeschränkt interpretierbar. Es zeigt sich bei allen Skalen ein signifikantes Ergebnis der Varianzanalyse, so dass davon ausgegangen werden kann, dass sich die Skalenwerte der verschiedenen Shops in den jeweiligen Branchen signifikant voneinander unterscheiden. Entsprechend wird die **Hypothese H3** bestätigt, welche besagt, die ufosV2r-Skala sei in der Lage, zwischen mehreren Online-Shops zu differenzieren.

Tab. 8-11: 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 (ufosV2r)	Zwischen den Shops	275,382	34	8,099	3,836	0,001
	Innerhalb der Shops	1480,091	701	2,111		
	Gesamt	1755,473	735			
Ästhetik	Zwischen den Shops	383,639	34	11,284	3,900	0,001
	Innerhalb der Shops	2027,913	701	2,893		
	Gesamt	2411,553	735			
Vertrauen	Zwischen den Shops	256,234	34	7,536	4,316	0,001
	Innerhalb der Shops	1223,953	701	1,746		
	Gesamt	1480,188	735			
Wahrgenommene Produktanzahl	Zwischen den Shops	685,179	34	20,152	7,226	0,001
	Innerhalb der Shops	1955,049	701	2,789		
	Gesamt	2640,228	735			

Anmerkung: Freiheitsgrade (df), F-Wert (F).



### 8.2.5 Erfassung der formativen Variable Service

Eine Analyse der Anteile fehlender Werte auf Ebene der Variablen<sup>23</sup> ergibt, dass vier der sechs Indikatoren zum Konstrukt Service mehr als 10% Missing Data aufweisen. Es handelt sich um die Items zu den Aspekten *Garantieleistungen* (sv2), *Rückgaberegelerung* (sv3), *Lieferzeiten* (sv4) sowie *Möglichkeiten der persönlichen Kontaktaufnahme* (sv6). Von einer Einbeziehung dieser Indikatoren in die weitere Analyse muss daher abgesehen werden. Es verbleiben sv1 sowie das negativ formulierte Item sv5c, die die inhaltlichen Aspekte der *Zufriedenheit mit den vorhandenen Zahlungsmöglichkeiten* bzw. die *Versandkosten* betreffen. Dies hat zur Folge, dass keine Abbildung des Konstruktes Service erfolgt, wie es in Abschnitt 2.6.6 vorgestellt wird. Stattdessen bleiben in den Folgeanalysen jegliche Aussagen zum Service auf den Einfluss der Aspekte Zahlungsoptionen und Versandkosten beschränkt. In Tabelle 8-12 werden die Kennwerte der beiden verbleibenden Indikatoren sv1 und sv5c aufgeführt. Wie in Abschnitt 8.1.3 bereits dargestellt, kann auf Grundlage von signifikanten Ergebnissen im Kolmogorov-Smirnov-Test und den ermittelten Schiefe- bzw. Exzesswerten nicht von einer Normalverteilung der beiden Indikatoren ausgegangen werden (siehe Tabelle E-7 im Anhang E). Die Korrelation zwischen den beiden Indikatoren liegt bei  $r=0,212$  ( $p<0,001$ ).

Tab. 8-12: Kennwerte der formativen Indikatoren zur Erfassung des Service

Item-Code	% Missing	M	SD	$p_m$
sv1	6,5	4,13	1,99	0,699
sv2*	30,6			
sv3*	24,3			
sv4*	17,9			
sv5c	9,0	3,51	2,11	0,585
sv6*	29,0			

Anmerkung: \* Indikator wird aufgrund eines zu hohen Anteils fehlender Werte nicht in die weitere Analyse einbezogen. Mittelwert (M), Standardabweichung (SD), Anteil fehlender Werte in Prozent (% Missing), Itemschwierigkeit ( $p_m$ ).

### 8.2.6 Erfassung des wahrgenommenen Preises

Im Fall des Single-Items zum wahrgenommenen Preis liegt der Anteil fehlender Werte bei 0 %, d.h. sämtliche Teilnehmer gaben für beide Shops, die sie kennen lernten, eine Bewertung des Items ab. Der Mittelwert liegt bei  $M=3,28$  ( $SD=1,96$ ). In der Tabelle E-18 im Anhang E werden die deskriptiven Ergebnisse zur Ausprägung dieses Items für die verschiedenen einbezogenen Online-Shops in Form von Rangreihen vorgestellt.

<sup>23</sup> Für jedes Item wird dabei der Anteil an Fällen bestimmt, bei denen keine Beantwortungen von Seiten der Teilnehmer vorliegen.

### 8.3 Erfassung der abhängigen Variablen

Nachdem in Abschnitt 8.2 die Messgüte der verschiedenen Erfolgsfaktoren (Prädiktoren) betrachtet wurde, widmet sich der folgende Kapitelabschnitt der Erfassung der beiden Erfolgsgrößen (Kriterien) Kaufintention und Kaufentscheidung.

#### 8.3.1 Reflektive Erfassung der Kaufintention

Die Güte der Skala zur Kaufintention wird anhand der in Abschnitt 6.5.2 dargestellten Prüfkriterien der klassischen Testtheorie bzw. der Itemanalyse vorgenommen. Grundlage ist der Datensatz mit einer Bewertung pro Fall. Die Schritte eins und drei (explorative bzw. konfirmatorische Faktorenanalyse) der Prozedur von Homburg und Giering (1996) können nicht zur Anwendung kommen, da keine weitere intervallskalierte abhängige Variable erhoben wurde. Bei Einbeziehung aller vier Indikatoren der Kaufintention (ki1, ki2, ki3, ki4) in die Skala ergibt sich ein Cronbachs Alpha von  $\alpha=0,89$ . Dieser Wert wird durch eine Elimination des Items ki3 auf  $\alpha=0,91$  gesteigert. Tabelle 8-13 gibt einen Überblick über die Kennwerte dieser Drei-Item-Skala. In einer explorativen Faktorenanalyse wird lediglich ein Faktor extrahiert, der wie gefordert mehr als 50% der Datenvarianz aufklärt. Die korrigierten Trennschärfe-Werte der drei verbleibenden Indikatoren sind ausreichend hoch, und die Itemschwierigkeiten liegen im angemessenen Bereich zwischen  $0,2 < p_m < 0,8$  (vgl. Bortz & Döring, 2005, S. 218). Die Messgüte der reflektiven Kaufintentionsskala ist somit positiv zu werten. Wie Abschnitt 8.1.3 zu entnehmen ist, kann nicht von einer Normalverteilung der drei manifesten Variablen ausgegangen werden, da sich in den Kolmogorov-Smirnov-Tests signifikante Ergebnisse zeigen. Auch im Fall des Skalenwertes der Kaufintention, der sich aus der gemittelten Summe der drei Indikatoren ergibt, liegt ein signifikantes Testergebnis vor ( $p < 0,001$ ). Schiefe und Exzess weisen ebenso auf eine Abweichung des Skalenwertes von der Normalverteilung hin ( $sch=-0,55$ ;  $z\text{-Wert } sch=-0,61$ ;  $ex=-0,29$ ;  $z\text{-Wert } ex=-1,58$ ). In der Tabelle E-19 im Anhang E werden die deskriptiven Ergebnisse zur Ausprägung der Kaufintention (Skalenwerte) für die 35 im Quasi-Experiment berücksichtigten Online-Shops durch eine Rangreihenbildung über die Skalenwerte dargestellt.

Tab. 8-13: Skalen- und Itemkennwerte der reflektiven Skala zur Kaufintention

Skalenkennwerte	Item-Code	M	SD	$a_{ij}$	% Missing	$r_{it}$	$p_m$
$\alpha=0,911$ $M=1,96$ ; $SD=1,77$ Mittlere korr. Trennschärfe=0,816 Homogenität=0,766 Erklärte Varianz: 67,81% $\alpha$ vor Elimination von ki3=0,894	ki1	2,25	2,03	0,944	1,59	0,843	0,375
	ki2	1,74	1,87	0,868	4,23	0,837	0,290
	ki4	1,91	1,86	0,944	2,51	0,789	0,318

Anmerkung: Cronbachs Alpha ( $\alpha$ ), Mittelwert (M), Standardabweichung (SD), Faktorladung ( $a_{ij}$ ), Anteil fehlender Werte in Prozent (% Missing), korrigierte Trennschärfe ( $r_{it}$ ), Itemschwierigkeit ( $p_m$ ).

### 8.3.2 Erfassung der Kaufentscheidung

Da es sich bei der Kaufentscheidung um eine kategoriale Variable handelt, ist es nicht möglich, die Güte dieser AV anhand derselben Kriterien wie im Fall der AV Kaufintention zu beurteilen. Entsprechend werden in diesem Abschnitt zunächst deskriptive Ergebnisse zur Ausprägung der Variable dargestellt. Tabelle 8-14 gibt eine Übersicht, wie viele der Teilnehmer sich für einen *Gutschein*, einen *Kauf in Shop 1* oder einen *Kauf in Shop 2* entschieden. Dabei wird nach den fünf berücksichtigten Produktarten differenziert. Nach der in Abschnitt 8.1.2.1 geschilderten Elimination von zehn Fällen mit zu hohen Anteilen fehlender Werte liegt die Anzahl der Online-Käufe bei insgesamt 284. Es haben somit 77,2% der  $N=368$  Probanden einen Einkauf vorgenommen.<sup>24</sup>

Die meisten Käufe wurden in den Shops der Produktart CDs getätigt; 90,1% der Teilnehmer in dieser Substichprobe bestellten eine CD. Hoch ist die Anzahl der Käufe im Verhältnis zur Wahl des Gutscheins auch in den Produktarten Bücher (84%) und DVDs (82,1%). Prozentual ist der Anteil der Käufer unter den Teilnehmern zur Bewertung der Ticketshops deutlich geringer; hier entschieden sich 68,3% für den Kauf eines Tickets. Mit 56,8% wurden die wenigsten Käufe im Fall der Druckerpatronen getätigt; 43,2% der Probanden zu dieser Produktart zogen die Wahl eines Gutscheins vor. Insgesamt kauften die Teilnehmer häufiger im ersten der beiden Shops, der ihnen präsentiert wurde. Von den 284 Bestellungen insgesamt wurden 54,6% in Shop 1 getätigt, in Shop 2 waren es 45,4%. Lediglich im Fall der Druckerpatronen tätigten mit jeweils  $n=23$  gleich viele Versuchspersonen eine Bestellung im ersten und im zweiten Shop. Deskriptive Ergebnisse zu den einzelnen Online-Shops werden in Form von Rangreihen in Tabelle E-20 im Anhang E vorgestellt.

Tab. 8-14: Wahlhäufigkeit der Gutschein, Kauf in Shop 1 bzw. Shop 2 nach Produktart

		<i>Gutschein</i>	<i>Shop 1</i>	<i>Shop 2</i>	<i>Käufe</i>	<i>Gesamt</i>
<i>Bücher</i>	Abs. Häufigkeit	13	40	28	68	81
	%	16,05	49,38	34,57	83,95	100
<i>DVDs</i>	Abs. Häufigkeit	15	36	33	69	84
	%	17,86	42,86	39,29	82,14	100
<i>CDs</i>	Abs. Häufigkeit	8	41	32	73	81
	%	9,88	50,62	39,51	90,12	100
<i>Druckerpatr.</i>	Abs. Häufigkeit	35	23	23	46	81
	%	43,21	28,40	28,40	56,79	100
<i>Tickets</i>	Abs. Häufigkeit	13	15	13	28	41
	%	31,71	36,59	31,71	68,29	100
<i>Gesamt</i>	Abs. Häufigkeit	84	155	129	284	368
	%	22,83	42,12	35,05	77,17	100

Anmerkung: Die Werte in der Spalte „Käufe“ ergeben sich aus der Summe der Spalten „Shop 1“ und „Shop 2“.

<sup>24</sup> Von den insgesamt  $N=378$  Teilnehmern (vor der Elimination von Fällen aufgrund zu hoher Anteile fehlender Werte) bestellten 291 im Rahmen des Experiments.

## 8.4 Festlegung der formativen Skala ufosV2f

Im Folgenden wird die Prozedur zur Entwicklung formativer Maße für die formative Usability-Skala ufosV2f fortgesetzt (vgl. Abschnitt 2.8.3). Nachdem die Zusammenstellung des Itempools im vierten Kapitel erläutert wurde, erfolgt auf Basis der erhobenen Daten zunächst eine Festlegung der Items, die in die ufosV2f-Skala eingehen (siehe Abschnitt 8.4.1). Da die formativen ufosV2f-Items in PLS-Modelle integriert werden sollen, ist zu prüfen, inwiefern sich die Indikatoren multikollinear zueinander verhalten. Liegt ein hohes Maß an Multikollinearität vor, so sind Unterindizes zu bilden (vgl. Abschnitt 8.4.2). Schließlich wird im letzten Schritt der Entwicklungsprozedur eine Validierung des formativen Usability-Maßes ufosV2f mit Hilfe eines PLS-Zwei-Konstrukt-Modells vorgenommen (Abschnitt 8.4.3). Grundlage sämtlicher angeführter Berechnungen zur Festlegung der ufosV2f-Skala ist der Datensatz mit einer Bewertung pro Fall.

### 8.4.1 Endgültige Bestimmung der ufosV2f-Indikatoren

Zur endgültigen Bestimmung der ufosV2f erfolgt zunächst eine Analyse, inwiefern einzelne Indikatoren aufgrund zu hoher Anteile fehlender Werte eliminiert werden müssen (Abschnitt 8.4.1.1). Anschließend wird auf Grundlage der bivariaten Korrelationen zwischen den Indikatoren geprüft, ob inhaltliche Redundanzen innerhalb des Itempools vorliegen, die eine Elimination weiterer Items nahe legen (Abschnitt 8.4.1.2).

#### 8.4.1.1 Elimination von Indikatoren aufgrund zu vieler fehlender Werte

Da die Indikatoren zur Usability zukünftig die Bestandteile des ufosV2-Fragebogens darstellen werden, besteht der Anspruch, dass die Items allgemein zur Bewertung von Online-Shops anwendbar sind. Es sollen ausschließlich Items in die ufosV2f-Skala eingehen, die von einem Großteil der Population der Online-Kunden beantwortet werden können und aus Expertensicht als relevant eingeschätzt werden (siehe Abschnitt 4.2.1). Als Cut-Off-Kriterium wurde festgelegt, dass der Anteil fehlender Werte für jeden Indikator 10% nicht überschreiten soll. Es finden sich neun Indikatoren, die dieses Kriterium verletzen: uf2, uf3, uf17, uf30, uf37, uf47, uf51, uf52 und uf53. In Tabelle 8-15 werden der Wortlaut sowie die jeweiligen Missing-Anteile dieser Items angeführt. Es gilt zu prüfen, inwiefern eine Elimination der neun identifizierten Indikatoren aus theoretischer Sicht gerechtfertigt ist. Dazu wird eine Sichtung der inhaltlichen Aspekte dieser Items vorgenommen.

Die vier Items uf5, uf30, uf3 und uf17 betreffen sehr spezifische Aspekte der Usability von Online-Shops (Funktionen zur Individualisierung von Produktlisten, Vergleichsmöglichkeiten für Produkte, automatisierte Benutzungshilfen, Individualisierbarkeit der Informationsdichte). Diese können im Einzelfall durchaus relevant sein, jedoch ist nicht zu erwarten, dass sie den kundenseitigen Usability-Gesamteindruck in wesentlichem Maße mitbestimmen (siehe Abschnitt 4.2.1). Entsprechend erscheint eine Elimination der vier Items gerechtfertigt.

Tab. 8-15: Formative Usability Indikatoren mit mehr als 10 % fehlender Werte

Item-Code	% Missing	Wortlaut des formativen Indikators
uf2	14,2	Es sind nützliche Informationen zur Benutzung des Shops abrufbar.
uf3	38,5	Bei auftretenden Problemen bietet der Shop von sich aus nützliche Hilfe an.
uf17	14,4	Ich kann selbst bestimmen, welche Informationen vom Shop angezeigt werden und welche nicht.
uf30	13,5	Ein Vergleich verschiedener Produkte ist leicht.
uf37	16,4	Die Informationen zur Rückgabe von Produkten sind verständlich.
uf47	10,6	Die Eingabe meiner persönlichen Daten (z.B. Anschrift, Kreditkartennummer) während des Bestellvorgangs ist unkompliziert.
uf51	14,0	Die Auflistungen der Produkte lassen sich bequem nach meinen Bedürfnissen ordnen.
uf52	16,4	Ich erhalte nützliche Informationen zum Unternehmen, das den Shop betreibt.
uf53	16,4	Ich bin mit den Informationen zur Sicherheit des Datentransfers zufrieden.

Anmerkung: Anteil fehlender Werte (% Missing).

Die vier Items uf2, uf37, uf52 und uf53 betreffen die Verfügbarkeit von Benutzungshilfen bzw. spezifischer einkaufsbezogener Informationen. Angesichts der hohen Anteile fehlender Werte kann davon ausgegangen werden, dass diese Aspekte für viele Kunden keine derart hohe Relevanz besitzen, als dass diese Informationen gezielt auf den Seiten eines Shops gesucht werden. Eine Elimination der vier Items wird daher als gerechtfertigt angesehen. Der Indikator uf47 überschreitet mit einem Missing-Anteil von 10,6% den gesetzten Cut-Off-Wert nur leicht. Dieses Item deckt mit der Eingabe von Daten im Bestellprozess einen inhaltlichen Aspekt ab, der aus Expertensicht hoch relevant ist (Nielsen et al., 2001). Es erscheint plausibel, dass der tendenziell hohe Anteil fehlender Werte dieses Items im Benutzungsszenario begründet ist, denn die Beantwortung des ufosV2-Fragebogens erfolgte im Versuch zu einem Zeitpunkt, an dem die Teilnehmer noch keine Bestellung durchgeführt hatten (siehe Abschnitt 5.5). Folglich fehlte den Probanden eine angemessene Bewertungsgrundlage zur Beantwortung des Items u47. Daher wird von einer Elimination des Items uf47 abgesehen. Zusammenfassend werden die acht Items uf2, uf3, uf5, uf17, uf30, uf37, uf52 und uf53 aufgrund zu hoher Anteile fehlender Werte nicht in die ufosV2f-Skala aufgenommen.

#### 8.4.1.2 Elimination inhaltlich redundanter Indikatoren

Damit der Item-Umfang der formativen Skala ufosV2f nicht unnötig groß ausfällt, sollen inhaltliche redundante Indikatoren eliminiert werden. Um derartige Items zu identifizieren, werden die bivariaten Korrelationen zwischen den Indikatoren betrachtet (siehe Tabelle E-21 im Anhang E). Alle Indikatorpaarungen mit einem Koeffizienten  $r \geq 0,5$  werden näher betrachtet. Die Tabelle 8-16 führt die vier Indikatorpaarungen mit hoher bivariater Korrelation auf, bei denen aus Untersuchersicht eine derart hohe inhaltliche Redundanz gegeben ist, dass eine Elimination gerechtfertigt erscheint. Als Kriterien, welches der jeweils zwei Items eliminiert wird, dienen die Itemschwierigkeit, der Anteil fehlender Werte sowie Korrelationen mit anderen Indikatoren. Es werden die Indikatoren uf4, uf6, uf26 sowie uf58c von der weiteren Verwendung in der formativen Skala ufosV2f ausgeschlossen, um inhaltliche Redundanzen im Fragebogen zu verhindern. Es verbleiben somit 46 Items in der ufosV2f-Skala.

Tab. 8-16: Indikatorpaarungen mit hoher bivariater Korrelation und inhaltlicher Redundanz

Item-Code	VIF	$p_m$	% Missing	Item	Elimination	$r$
uf24	3,23	0,538	0,01	Es ist leicht, gesuchte Produkte zu finden.		0,749
uf26	3,53	0,589	0,01	Die Suchfunktionen ermöglichen mir, schnell bestimmte Produkte zu finden.	×	
uf49c	2,17	0,361	0,03	Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.		0,673
uf58c	2,07	0,345	0,04	Die Ladezeiten der Seiten sind bei diesem Online-Shop zu lang.	×	
uf4	2,63	0,439	0,04	Es ist in diesem Shop leicht, versehentlich durchgeführte Aktionen zu korrigieren.	×	0,621
uf9	2,00	0,334	0,04	Ich kann jede Aktion wieder rückgängig machen.		
uf6	2,86	0,476	0,00	Es ist leicht, sich innerhalb des Shops zu bewegen.	×	0,604
uf7	2,03	0,338	0,02	Es ist einfach, zwischen den verschiedenen Seiten des Shops hin- und herzuspringen.		

Anmerkung: Variance Inflation Factor (VIF), Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte in Prozent (% Missing), bivariate Pearson-Korrelation ( $r$ ), Elimination des Items (X).

#### 8.4.2 Behandlung von Multikollinearität zwischen den ufosV2f-Items

Da die formativen ufosV2f-Items in ein PLS-Zwei-Konstrukt-Modell eingebunden werden sollen, stellt sich die Frage, inwiefern die Indikatoren untereinander multikollinear sind. Dies könnte zu Verzerrungen der Schätzparameter führen. Zur Prüfung der Multikollinearität werden die drei in Abschnitt 6.8 vorgestellten Ansätze verfolgt.

Zunächst werden die Interkorrelationen der Items herangezogen. Wie aus Tabelle E-21 im Anhang E hervorgeht, lassen sich nach der im Vorabschnitt dargestellten Elimination noch 44 bivariate Korrelationen  $r \geq 0,5$  identifizieren, von denen fünf Werte bei  $r \geq 0,6$  liegen. Angesichts dieser hohen Interkorrelationen muss von einer mäßigen bis starken Multikollinearität zwischen den formativen Indikatoren ausgegangen werden. Auf Basis einer Hilfsregression werden zur Einschätzung der Multikollinearität zusätzlich die Toleranz- und VIF-Werte der 46 formativen Indikatoren bestimmt (siehe Tabelle E-22 im Anhang E). Vier Items weisen eine *Toleranz*  $< 0,4$  auf; 14 VIF-Werte liegen oberhalb von 2. Somit ist auf Grundlage dieser Analyse zu befürchten, dass ein erhebliches Multikollinearitätsproblem besteht. Einen weiteren Hinweis auf ein zu hohes Maß an Multikollinearität erbringt die Betrachtung der Konditionsindizes. Für die 47. Dimension ergibt sich ein Konditionsindex von 50,96 (Eigenwert  $\lambda = 0,02$ ), was deutlich über dem kritischen Wert von 30 liegt.

Angesichts des hohen Maßes an Multikollinearität gilt es, diese im Zwei-Konstrukt-Modell methodisch angemessen zu berücksichtigen. Wie in Abschnitt 2.8.3.3 dargestellt wurde, empfiehlt es sich, die untereinander multikollinearen ufosV2f-Indikatoren zu Unterindizes (uf\_in) zusammenzufassen (Albers & Hildebrandt, 2006). Dazu wird ein methodischer Ansatz verfolgt, der dem Prinzip der *Hauptkomponentenregression* folgt. Dieses Verfahren kann bei hoher Multikollinearität zwischen den Prädiktoren eines Regressionsmodells herangezogen werden (vgl. Cohen et al., 2003, S. 428 f.; Moosmüller, 2004, S. 132; Schneider, 2006,

S. 199). Zunächst wird für die formativen Usability-Indikatoren eine HKA mit orthogonalen Hauptkomponenten berechnet.<sup>25</sup> Dabei wird eine orthogonale Rotation (Varimax-Methode) der Komponenten vorgenommen.<sup>26</sup> Die extrahierten Komponenten sollen als eigenständige formative Indikatoren in das PLS-Zwei-Konstrukt-Modell einbezogen werden (siehe Abschnitt 2.8.3.4 bzw. Folgeabschnitt). Eine inhaltliche Interpretation dieser Unterindizes ist nicht möglich, da aufgrund hoher Korrelationen zwischen formativen Indikatoren nicht auf eine inhaltliche Nähe geschlossen werden kann. Die Anzahl der extrahierten Komponenten wird auf Grundlage des Eigenwert-Kriteriums bestimmt (Eigenwert  $\lambda_j > 1$ ; vgl. Heck, 1998, S. 187 f.). Die teilweise schiefen Verteilungen der formativen Usability-Items erweisen sich als unkritisch, da die Berechnung einer HKA keine Normalverteilung der Indikatoren voraussetzt (vgl. Bühner, 2004, S. 159).

Es werden neun Komponenten extrahiert. Tabelle E-23 im Anhang E stellt die rotierte Komponentenmatrix dar. Das KMO-Maß liegt bei 0,95, so dass die Indikatoren als sehr gut zur Berechnung einer Faktorenanalyse geeignet angesehen werden können. Auf Indikatorebene liegt kein MSA-Wert unterhalb 0,5; der Bartlett-Test auf Sphärizität zeigt ein signifikantes Ergebnis (*ungefähres*  $\chi^2=13626,140$ ;  $df=1035$ ;  $p<0,001$ ). Für einige der Indikatoren ergeben sich Kommunalitäten, die einen Wert von  $h^2 < 0,5$  aufweisen (uf1, uf5, uf8, uf11, uf18, uf23, uf32, uf33, uf34, uf39, uf44 und uf47), was eine Elimination dieser Indikatoren nahe legen könnte. Da dies jedoch zum Verlust der dahinter stehenden inhaltlichen Usability-Aspekte führen würde, muss von einem solchen Vorgehen abgesehen werden. Der Gesamtanteil aufgeklärter Varianz liegt bei 55%, was zwar nicht als schlecht gewertet werden muss, jedoch darauf hinweist, dass ein nicht unerheblicher Anteil der Datenvarianz nicht erklärt wird. Eine nähere Interpretation der Komponentenstruktur kann nicht erfolgen, da eine HKA nicht geeignet erscheint, um eine *inhaltliche* Dimensionierung eines Sets formativer Indikatoren vorzunehmen. Zur Bildung der Unterindizes werden die Faktorwerte der HKA berechnet (vgl. Grice & Harris, 1998; Bühner, 2004, S. 169). Sie werden im Folgenden mit uf\_in1 bis uf\_in9 benannt.

Zusammenfassend ist festzuhalten, dass ein zu hohes Maß an Multikollinearität zwischen den formativen Usability-Items besteht, als dass sie als einzelne Indikatoren in eine PLS-Analyse eingehen könnten. Daher werden die Items mittels einer HKA zu Unterindizes zusammengefasst. Diese Unterindizes können als formative Indikatoren in die im Folgenden berechneten PLS-Analysen aufgenommen werden, da zwischen ihnen keine Multikollinearität vorliegt.

---

<sup>25</sup> Da das Ziel der Analyse darin liegt, möglichst viel Information der vorhandenen Daten beizubehalten, erweist sich die HKA in diesem Fall als die geeignete Form der explorativen Faktorenanalyse (Bühner, 2004, S. 159).

<sup>26</sup> Die Rotation erfolgt, um eine Verteilung der erklärten Varianz über mehrere Komponenten zu erreichen. Unrotierte Komponentenlösungen weisen häufig nur eine Hauptkomponente auf, die einen Großteil der Varianz erklärt. Die weiteren Komponenten tragen deutlich weniger zur Varianzaufklärung auf. Dies würde in einem PLS-Zwei-Konstrukt-Modell zu einem wenig zufrieden stellenden Ergebnis auf Ebene des Messmodells führen.

### 8.4.3 PLS-Zwei-Konstrukt-Modell zur Validierung

Gemäß dem in Abschnitt 2.8.3.4 vorgestellten Vorgehen wird die formative latente Variable Usability (Skala uofsV2f) in einem Zwei-Konstrukt-Modell als Prädiktor für das reflektive Maß der Usability (Skala ufosV2r) eingebunden. Die Schätzung dieses Modells erfolgt mit Hilfe des PLS-Algorithmus unter Nutzung des Path Weighting Scheme. Als Software wird das Programm Smart-PLS in der Version 2.0.M3 genutzt (Hansmann & Ringle, 2004). Die Stichprobengröße mit  $N=736$  Fällen erweist sich angesichts der Anzahl an Indikatoren als ausreichend groß. Zur Bestimmung der Modellgüte wird eine Bootstrapping Resampling Prozedur mit 200 Subsamples berechnet.

Tab. 8-17: 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 (ufosV2f-Skala)	uf_in1	0,720	0,063	11,512
	uf_in2	0,485	0,073	6,614
	uf_in3	0,224	0,054	4,158
	uf_in4	0,176	0,067	2,634
	uf_in5	0,265	0,076	3,511
	uf_in6	0,173	0,075	2,305
	uf_in7	0,202	0,064	3,169
	uf_in8	0,154	0,072	2,150
	uf_in9	0,032	0,062	0,514
Usability reflektiv (ufosV2r-Skala) <i>Composite Reliability</i> =0,949 <i>AVE</i> =0,698	ur1	0,881	0,032	27,333
	ur2	0,758	0,058	13,023
	ur3	0,805	0,041	19,864
	ur4	0,832	0,031	26,922
	ur6	0,847	0,033	26,075
	ur7	0,908	0,023	39,832
	ur8	0,779	0,042	18,365
	ur9	0,863	0,027	32,497

Anmerkung: \* Im Fall der formativen Indikatoren werden die Gewichte angegeben, bei den reflektiven Indikatoren handelt es sich um die Ladungen. Average Variance Extracted (AVE).

Es erfolgt zunächst eine Betrachtung der Messmodellgüte. Tabelle 8-17 enthält die Kennwerte. Acht der neun formativen Unterindizes zur Erfassung der formativen Usability, die mittels der HKA bestimmt wurden, weisen wie gefordert Gewichte  $\gamma > 0,1$  sowie signifikante t-Werte über 1,66 auf. Lediglich für die neunte Hauptkomponente uf\_in9 ergibt sich ein Gewicht von  $\gamma = 0,032$  mit einem insignifikanten t-Wert von  $t = 0,514$ . Die Ladungen der acht reflektiven Indikatoren liegen wie gefordert über  $\lambda > 0,707$ ; für alle Ladungen ergeben sich in der t-Test-Statistik signifikante Werte von  $t > 1,66$ . Für die Composite Reliability als Maß für die interne Konsistenz der reflektiven Usability-Skala ufosV2r wird ein Wert von 0,949 berechnet, was auf eine sehr reliable Messung des Konstruktes schließen lässt. Auf Basis dieses Wertes wird die **Hypothese 1**, wonach die Skala ufosV2r eine hohe interne Konsistenz aufweist, ein weiteres Mal bestätigt (siehe Abschnitt 8.2.2). Der AVE-Wert dieser latenten Variable liegt bei  $AVE = 0,698$  und damit deutlich oberhalb des geforderten Wertes von 0,5. Insgesamt ist die Güte des Messmodells als hervorragend anzusehen.



Das niedrige Gewicht von  $\gamma=0,032$  für den Unterindex uf\_in9 kann darauf zurückgeführt werden, dass diese Hauptkomponente in der berechneten HKA am wenigsten zur Varianzaufklärung beiträgt (2,61% aufgeklärte Varianz;  $\lambda=1,03$ ; siehe Tabelle E-23 im Anhang E). Von einer Elimination dieses Unterindizes muss jedoch abgesehen werden, da alle formativen ufosV2f-Indikatoren jeweils in alle neun Unterindizes eingehen. Somit würde ein Ausschluss von uf\_in9 keine Elimination im üblichen Sinne darstellen. Eine vergleichende Interpretation der Einflussstärken der Unterindizes ist nicht möglich, da keine inhaltliche Aussage bzgl. der aggregierten Indikatoren getroffen werden kann (siehe Abschnitt 8.4.2).

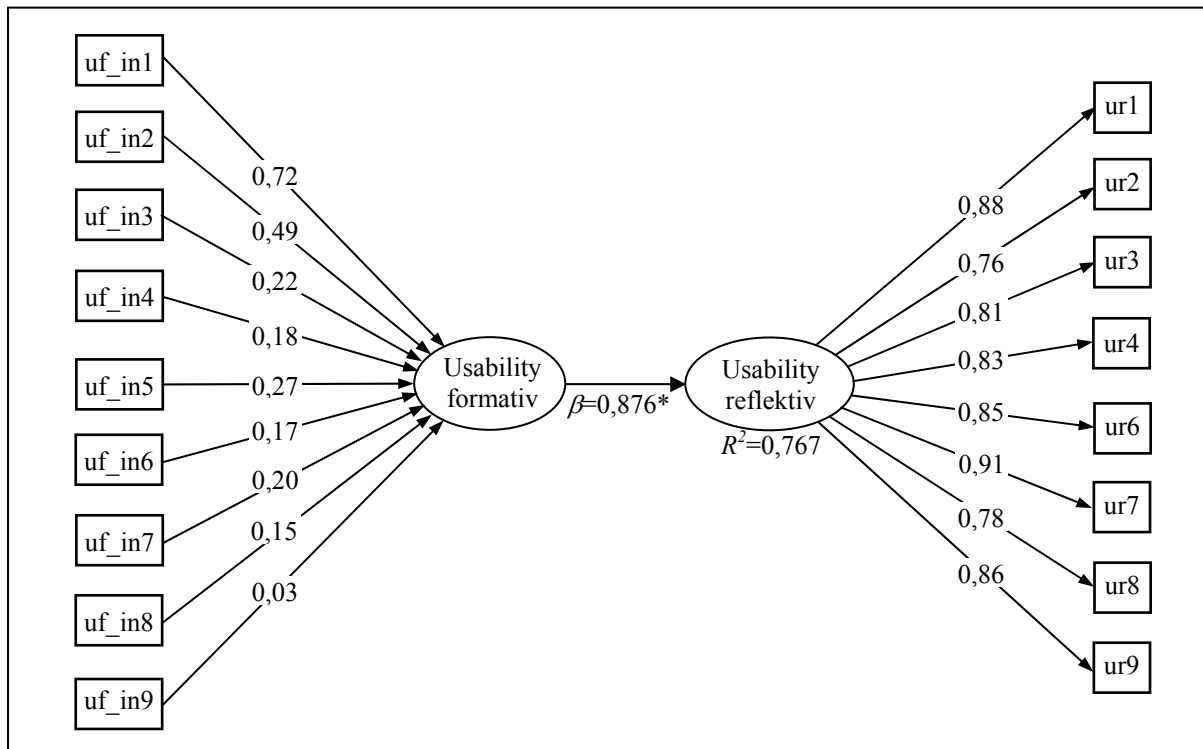


Abb. 8-2: Zwei-Konstrukt-Modell zur Validierung der formativen Usability-Skala ufosV2f anhand der reflektiven Usability-Skala ufosV2r

Anmerkung: \* Pfadkoeffizient ist signifikant ( $t > 1,66$ ). Für die formative Usability-Variable (ufosV2f, links) werden die Gewichte und für die reflektive Usability-Variable (ufosV2r, rechts) die Ladungen angegeben.

Nachdem das Messmodell als gut gewertet wurde, folgt die Begutachtung des Strukturmodells. In Abbildung 8-2 wird das geschätzte Modell graphisch dargestellt. Es ergibt sich ein sehr hoher Pfadkoeffizient von  $\beta=0,876$  sowie ein Determinationskoeffizient von  $R^2=0,767$ . Diese Werte lassen darauf schließen, dass ein sehr starker Zusammenhang zwischen der formativen und der reflektiven latenten Variable der Usability besteht; das Set der 46 formativen ufosV2f-Indikatoren deckt das Globalurteil der Usability, welches durch die reflektive Skala ufosV2r erfasst wird, offensichtlich sehr gut ab. Somit bestätigt sich auf Grundlage dieser Analysen die **Hypothese H4**, wonach ein starker Zusammenhang zwischen der formativen ufosV2f-Skala und der reflektiven ufosV2r-Skala besteht (siehe Abschnitt 2.8.3.4).

Es wird durch zwei zusätzliche Analysen geprüft, wie stabil sich das Ergebnis des PLS-Zwei-Konstrukt-Modells in Hinblick auf die Methode zur Bestimmung der formativen Unterindizes  $uf\_in$  erweist, die in Abschnitt 8.4.2 verfolgt wurde. (1) Zum einen werden anstelle der Faktorwerte aus der HKA mittelwertbasierte Skalenwerte gebildet, indem für jede Komponente der HKA die jeweils hoch ladenden Indikatoren zusammengefasst werden (siehe Abschnitt 8.4.2). (2) Zum zweiten wird anstelle einer HKA eine Hauptachsenanalyse berechnet, die ebenfalls zu einer Extraktion von neun Faktoren führt; als formative Unterindizes werden die regressionsbasierten Faktorwerte genutzt. Die Ergebnisse beider alternativer Ansätze zur Bestimmung der formativen Unterindizes führen im PLS-Zwei-Konstrukt-Modell im Vergleich zu den oben berichteten Werten zu nahezu identischen Strukturmodellparameter wie oben berichtet (Ansatz 1:  $\beta=0,877$ ,  $R^2=0,769$ ; Ansatz 2:  $\beta=0,881$ ,  $R^2=0,776$ ). Folglich erweist sich der Zusammenhang zwischen beiden  $ufosV2$ -Skalen als hoch – unabhängig davon, welche faktorenanalytische Methode zur Erstellung der formativen Unterindizes  $uf\_in$  gewählt wird.

## 8.5 Untersuchung des konzeptuellen Erfolgsmodells mit Hilfe von PLS

Im Folgenden wird die empirische Untersuchung des konzeptuellen Erfolgsmodells dargestellt, welches in Abschnitt 3.3 vorgestellt wurde. Die Überprüfung der Zusammenhangsannahmen erfolgt mit Hilfe des PLS-Strukturgleichungsansatzes. Das konzeptuelle Modell beinhaltet Zusammenhangsannahmen zum Erfolgsfaktor Einkaufsvergnügen. Da sich die Skala zur Erfassung dieses Konstruktes gegenüber den übrigen jedoch nicht als faktoriell valide erwies (siehe Abschnitt 8.2.1), muss von einer Berücksichtigung im konzeptuellen Modell abgesehen werden. Eine Überprüfung der Hypothesen H8, H13, H16 sowie H19 kann somit nicht erfolgen.

Es werden zwei PLS-Modelle berechnet. In einen wird die reflektive Skala ufosV2r zur Erfassung der LV Usability verwendet, im anderen wird die formative Skala ufosV2f eingebunden. Beide Modelle unterscheiden sich folglich nur dahingehend, welche der beiden ufos-Skalen herangezogen wird. Der Einfachheit halber wird im Folgenden vom ufosV2r-Modell bzw. ufosV2f-Modell gesprochen. Die Überprüfung der Modelle erfolgt parallel, d.h. alle Parameter der Modelle werden jeweils gemeinsam gesichtet. Auf diese Weise lassen sich potenzielle Unterschiede zwischen beiden Modellen leichter identifizieren. Zur Schätzung beider Modelle dient der Datensatz mit einer Bewertung pro Shop mit  $N=736$  Fällen. Als Schätz-Algorithmus wird das Path Weighting Scheme gewählt. Sämtliche t-Werte werden mit Hilfe eines Bootstrapping mit 200 Resamplings bestimmt.

### 8.5.1 Beurteilung der Messmodelle für beide PLS-Modelle

Zunächst erfolgt eine Gütebeurteilung der Messmodelle. In Tabelle 8-18 werden die Ladungen und Gewichte der Indikatoren für beide PLS-Modelle angegeben. Alle Ladungen der reflektiven Indikatoren weisen in beiden PLS-Modellen Werte  $\lambda > 0,707$  auf (Barclay et al, 1995). Die entsprechenden t-Werte zeigen eine Signifikanz sämtlicher Ladungen an. Die Tabelle 8-19 enthält die AVE-Maße sowie die Werte der Composite Reliability der reflektiven Skalen in beiden PLS-Modelle. Das AVE-Maß überschreitet in allen Fällen den geforderten Wert von 0,5. Die Ausprägungen der Composite Reliability erweisen sich für alle Skalen als hoch (Werte über 0,9); für keine der Skalen ergibt sich ein Wert unterhalb der kritischen Grenze von 0,7. Die sehr hohe Composite Reliability von 0,948 für die reflektive Usability-Skala ufosV2r wurde bereits in Abschnitt 8.4.3 zum Zwei-Konstrukt-Modell als Bestätigung der **Hypothese 1** gewürdigt.

Tab. 8-18: Kennwerte des PLS-Messmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: Ladungen und Gewichte

Latente Variable	Item	ufosV2r-Modell (reflektive Usability)			ufosV2f-Modell (formative Usability)		
		Gewicht bzw. Ladung	SE	t-Wert	Gewicht bzw. Ladung	SE	t-Wert
Usability - ufosV2r (reflektiv)	ur1	0,883	0,027	32,929			
	ur2	0,754	0,061	12,430			
	ur3	0,800	0,050	16,044			
	ur4	0,822	0,039	21,147			
	ur6	0,841	0,032	25,866			
	ur7	0,910	0,019	46,934			
	ur8	0,791	0,038	20,581			
Usability - ufosV2f (formativ)	ur9	0,868	0,026	33,090			
	uf_in1				0,804	0,070	11,413
	uf_in2				0,317	0,087	3,637
	uf_in3				0,368	0,083	4,435
	uf_in4				0,131	0,087	1,514
	uf_in5				0,268	0,090	2,985
	uf_in6				0,147	0,080	1,841
	uf_in7				0,063	0,080	0,785
Ästhetik (reflektiv)	uf_in8				0,056	0,087	0,644
	uf_in9				0,014	0,101	0,139
Vertrauen (reflektiv)	ae1	0,963	0,008	124,700	0,963	0,008	119,497
	ae2	0,953	0,013	74,542	0,953	0,013	75,791
Wahrgenommene Produktanzahl (reflektiv)	t1	0,900	0,023	39,226	0,898	0,022	40,604
	t2	0,901	0,032	27,984	0,899	0,033	26,831
	t3	0,846	0,041	20,520	0,849	0,039	21,523
	t4	0,920	0,021	43,457	0,921	0,022	41,053
Service (formativ)	wa1	0,939	0,019	48,382	0,939	0,018	52,812
	wa2	0,922	0,023	40,628	0,922	0,026	35,694
Wahrge. Preis (Single-Item)	sv1	0,716	0,185	3,875	0,716	0,198	3,617
	sv5c	0,562	0,218	2,582	0,562	0,229	2,458
Bekanntheitsgrad (Single-Item)	shop_ab	1			1		
Kaufintention (reflektiv)	subp1	1			1		
	ki1	0,933	0,019	50,400	0,934	0,017	53,452
	ki2	0,924	0,023	40,914	0,925	0,021	43,623
Produktart Drucker (Single-Item)	ki4	0,908	0,025	36,503	0,907	0,022	41,448
	druck	1			1		
Moderatoreffekt Ästhetik x Drucker (reflektiv)							
	ae1xdruck	0,939	0,221	4,492	0,995	0,202	4,917
	ae2xdruck	0,922	0,200	4,299	0,858	0,190	4,523

Anmerkung: Standardfehler (SE). Die Pfadkoeffizienten, Ladungen und Gewichte werden getrennt für die Verwendung der reflektiven Usability-Skala ufosV2r sowie für die formative Usability-Skala ufosV2f angegeben. Im Fall formativer Konstrukte sind der Spalte „Gewicht bzw. Ladung“ die Gewichte zu entnehmen; im Fall reflektiver Konstrukte handelt es sich um Ladungen.

Weiterhin ist die faktorielle Validität der reflektiven Skalen zu begutachten. Die Tabellen E-24 und E-25 im Anhang E stellen die Korrelationen zwischen allen latenten Variablen für das PLS-Modell dar (Tabelle E-24 zum ufosVr-Modell; Tabelle E-25 zum ufosV2f-Modell). Für alle reflektiven Skalen in beiden Modellen überschreitet die Wurzel des AVE-Wertes (in

den Diagonalen der Tabellen dargestellt) die Korrelationskoeffizienten der jeweiligen LV mit den weiteren LVn. Weiterhin können den Tabellen E-26 und E-27 im Anhang E die Korrelationen sämtlicher Indikatoren mit den LVn entnommen werden. Alle Indikatoren korrelieren mit den ihnen zugehörigen LVn höher als mit den anderen Variablen. Auf Basis dieser Analysen kann die faktorielle Validität der reflektiven Skalen für beide PLS-Modelle als gegeben angesehen werden. Dies stellt wiederum eine Bestätigung der **Hypothese 2** dar. Diese nimmt an, die reflektive Usability-Skala erweise sich gegenüber weiteren reflektiven Skalen, die andere kundennahe Erfolgsfaktoren erfassen, als faktoriell valide (siehe Abschnitt 3.3).

Nachdem die Messgüte der reflektiven Skalen begutachtet wurde, gilt es, die formativen Skalen zum Service in beiden Modellen sowie zur Usability (ufosV2f) im ufosV2f-Modell bzgl. ihrer Messgüte zu untersuchen. Alle Gewichte sämtlicher formativer Indikatoren weisen hypothesenkonform positive Vorzeichen auf. Die Gewichte der Service-Indikatoren sv1 und sv5c liegen in beiden Modellen über dem geforderten Wert von  $\gamma_n > 0,1$  und weisen signifikante t-Werte auf. Zwischen beiden Indikatoren herrscht keine Multikollinearität ( $VIF=1,047$ ;  $Toleranz=0,955$ ). Die Gewichte der formativen Usability-Unterindizes uf\_in1 bis uf\_in9 liegen nicht in allen Fällen über dem Wert  $\gamma_n > 0,1$ . Eine Elimination erscheint aus inhaltlicher Sicht jedoch nicht angemessen, da in jeden Unterindex sämtliche Indikatoren der ufosV2f-Skala eingehen.<sup>27</sup> Insgesamt wird das Messmodell für die Schätzung der Strukturparameter als gut geeignet angesehen.

Tab. 8-19: Kennwerte des PLS-Messmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: AVE-Werte und Composite Reliability

Latente Variable	ufosV2r-Modell (reflektive Usability)		ufosV2f-Modell (formative Usability)	
	AVE	Composite Reliability	AVE	Composite Reliability
Usability - ufosV2r	0,697	0,948		
Ästhetik	0,918	0,957	0,918	0,957
Vertrauen	0,796	0,940	0,796	0,940
Wahrgenommene Produktanzahl	0,866	0,928	0,866	0,928
Kaufintention	0,850	0,944	0,850	0,944
Ästhetik x Drucker	0,867	0,926	0,863	0,926

Anmerkung: Average Variance Extracted – Kommunalität (AVE). Die AVE-Werte sowie die Composite Reliability-Werte werden getrennt für die Verwendung der reflektiven Usability-Skala ufosV2r sowie für die formative Usability-Skala ufosV2f angegeben.

### 8.5.2 Beurteilung der Strukturmodelle für beide PLS-Modelle

Die Evaluation des PLS-Strukturmodells erfolgt zunächst anhand der Höhe und Signifikanz der Pfadkoeffizienten. Die Ergebnisse des Strukturmodells werden in Abbildung 8-3 zusammengefasst.

<sup>27</sup> Eine probeweise Elimination der Unterindizes uf\_in7, uf\_in8 und uf\_in9 führt im Übrigen zu keiner wesentlichen Änderung der Strukturparameter im ufosV2f-Modell.

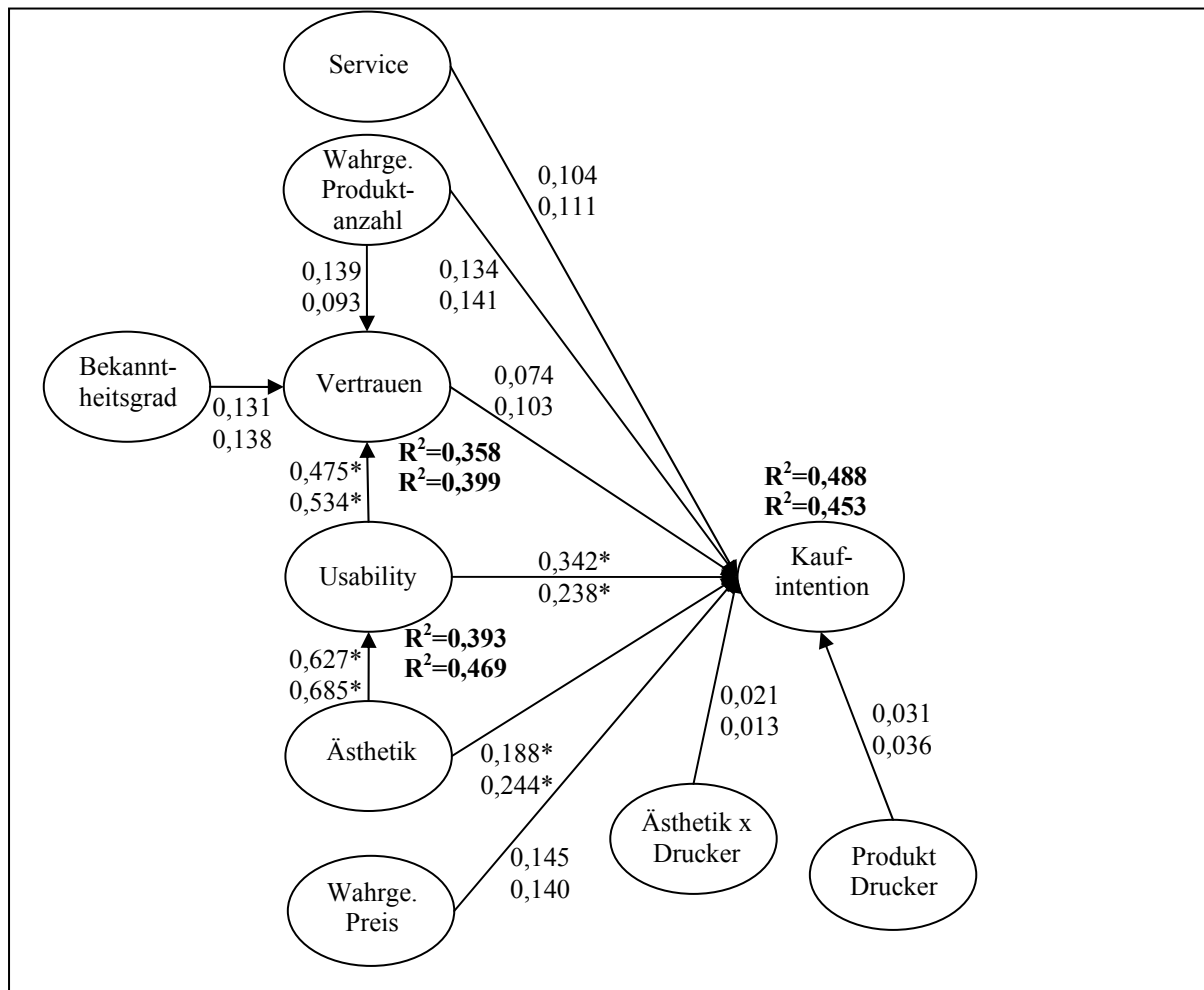


Abb. 8-3: Ergebnisse des PLS-Strukturmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen

Anmerkung: \* signifikanter Pfadkoeffizient, fettgedruckt Determinationskoeffizient ( $R^2$ ), Pfadkoeffizienten ( $\beta$ ). Der jeweils obere Wert bezieht sich auf das ufosV2r-Modell (Verwendung der reflektiven Usability-Skala ufosV2r) und der untere auf das ufosV2f-Modell (Verwendung der formativen Usability-Skala ufosV2f).

In Tabelle 8-20 werden die  $R^2$ -Werte der beiden PLS-Modelle wiedergegeben. Die abhängige Variable Kaufintention wird durch die Prädiktoren in substantiellem Maße erklärt, wobei nur geringfügige Unterschiede zwischen beiden PLS-Modellen festzustellen sind ( $R^2=0,488$  im ufosV2r-Modell bzw.  $R^2=0,453$  im ufosV2f-Modell). Es kann gefolgert werden, dass das konzeptuelle Modell zur Bestimmung der Kaufintention als gut geeignet anzusehen ist. Ebenfalls als hoch ist der Anteil aufgeklärter Varianz in den latenten Variablen Vertrauen und Usability zu bezeichnen. Das Vertrauen wird bei Verwendung der formativen Usability-Skala ufosV2f mit  $R^2=0,399$  leicht besser erklärt als im ufosV2r-Modell (dort  $R^2=0,358$ ). Die Ästhetik führt mit  $R^2=0,469$  zu einem höheren Maß an aufgeklärter Varianz in der LV ufosV2f-Usability als im Falle der LV ufosV2r-Usability, für die sich ein  $R^2=0,393$  ergibt. Die Vorhersagegüte der gesamten Modelle in Bezug auf das Kriterium Kaufintention wird auf Grundlage des Stone-Geisser-Q-Kriteriums beurteilt. Es nimmt im Fall des ufosV2r-Modells einen Wert von 0,383 an; für das ufosV2f-Modell liegt es bei 0,343. Folglich kann beiden Modellen eine substantielle Vorhersagekraft bescheinigt werden.

Tab. 8-20: Ergebnisse des PLS-Strukturmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: aufgeklärte Varianz ( $R^2$ )

Endogene Variable	Modell mit ufosV2r (reflektive Usability) $R^2$	Modell mit ufosV2f (formative Usability) $R^2$
Usability	0,393	0,469
Vertrauen	0,358	0,399
Kaufintention	0,488	0,453

Anmerkung: Determinationskoeffizient ( $R^2$ ).

In Tabelle 8-21 werden die für beide Modelle geschätzten Pfadkoeffizienten unter Angabe der t-Werte sowie deren Signifikanzen aufgelistet. Zusätzlich wurden die Effektgrößen  $f^2$  für alle exogenen LVn berechnet; diese Werte werden in Tabelle E-28 im Anhang E wiedergegeben.

Tab. 8-21: Ergebnisse des PLS-Strukturmodells für das konzeptuelle Modell zur Validierung der ufosV2-Skalen: Pfadkoeffizienten

Zusammenhang	Modell mit ufosV2r (reflektive Usability)		Modell mit ufosV2f (formative Usability)	
	$\beta$	t-Wert	$\beta$	t-Wert
Usability → Kaufintention	0,342	3,473	0,238	2,020
Usability → Vertrauen	0,475	5,886	0,534	6,490
Ästhetik → Kaufintention	0,188	2,041	0,224	2,149
Ästhetik → Usability	0,627	10,565	0,685	13,548
Vertrauen → Kaufintention	0,074	0,741	0,103	0,974
Wahrge. Produktanzahl → Vertrauen	0,139	1,409	0,093	0,975
Wahrge. Produktanzahl → Kaufintention	0,134	1,625	0,141	1,705
Bekanntheit → Vertrauen	0,131	1,421	0,138	1,691
Wahrge. Preis → Kaufintention	0,134	1,626	0,140	1,832
Service → Kaufintention	0,104	1,117	0,111	1,241
Produktart Drucker → Kaufintention	0,031	0,414	0,036	0,459
Ästhetik x Drucker → Kaufintention	0,021	0,249	0,013	0,152

Anmerkung: Pfadkoeffizient ( $\beta$ ). Der t-Wert der t-Statistik in PLS folgt nur annähernd der t-Verteilung. Die Pfadkoeffizienten werden getrennt für das ufosV2r-Modell (Verwendung der reflektiven Usability-Skala ufosV2r) sowie für das ufosV2f-Modell (Verwendung der formativen Usability-Skala ufosV2f) angegeben.

Von besonderem Interesse sind diejenigen Pfade, die die LV Usability betreffen. Die positive Wirkung der Usability auf die Kaufintention bestätigt sich in beiden Modellen. Sowohl die reflektive Skala ufosV2r als auch die formative Skala ufosV2f erweisen sich mit signifikanten Koeffizienten von  $\beta=0,342$  bzw.  $\beta=0,238$  als Prädiktoren dieser Erfolgsgröße; offensichtlich ist der Zusammenhang im Fall von ufosV2r höher als bei ufosV2f. Die Werte für die Effektstärke weisen den Einfluss auf die Kaufintention jedoch in beiden Fällen als eher gering aus ( $f^2=0,109$  im Fall von ufosV2r bzw.  $f^2=0,038$  für ufosV2f). Trotzdem bestätigen diese Ergebnisse die beiden **Hypothesen H5a** und **H5b** (siehe Tabelle 8-22) und sprechen für eine konkurrenente Kriteriumsvalidität der zwei neu entwickelten Skalen. Ein sehr hoher Zusammenhang zeigt sich in beiden PLS-Modellen für den Pfad zwischen den LVn Usability und Vertrauen. Dabei ist der Pfadkoeffizient zur Wirkung der ufosV2r-Variable auf das Vertrauen mit  $\beta=0,475$  niedriger als im Fall der ufosV2f-Variable  $\beta=0,534$ . Dies drückt sich auch in unterschiedlichen Effektstärken aus, welche für die ufosV2f-Skala hoch ( $f^2=0,381$ ) und bei der ufosV2r-Skala mittel hoch ( $f^2=0,293$ ) ausfällt. Es bestätigen sich also die

**Hypothesen H7a** sowie **H7b**, in denen diese Konstruktbeziehungen angenommen wurden. Ebenso bestätigen sich die Annahmen zum Zusammenhang zwischen der Ästhetik und der Usability, die in den **Hypothesen H9a** und **H9b** formuliert wurden. Der von der Ästhetik auf die Usability führende Pfad weist sehr hohe Koeffizienten auf. Bei der ufosV2r-Variable liegt er bei  $\beta=0,627$ ; für die ufosV2f-Variable ergibt sich ein Wert von  $\beta=0,685$ . In beiden Fällen ist von einer hohen Effektstärke zu sprechen ( $f^2=0,647$  im ufosV2r-Modell bzw.  $f^2=0,883$  im ufosV2f-Modell). Bis auf die Hypothese H8 zum Zusammenhang der Usability und dem Einkaufsvergnügen, die aus den genannten Gründen nicht überprüft werden kann, bestätigen sich damit in beiden PLS-Modellen alle Annahmen, die die latente Variable Usability bzw. die beiden neu entwickelten Skalen ufosV2r und ufosV2f betreffen. Die Ergebnisse stützen somit die nomologische Validität bzw. die konkurrente Kriteriumsvalidität dieser Skalen.

Von untergeordneter Relevanz gegenüber den bereits angeführten sind die Ergebnisse, die die übrigen Modellzusammenhänge betreffen. Von den im konzeptionellen Modell angenommenen Beziehungen bestätigt sich lediglich der Pfad zwischen Ästhetik und Kaufintention in beiden PLS-Modellen. Im Fall des ufosV2r-Modells weist der Pfadkoeffizient  $\beta=0,188$  einen t-Wert von  $t=2,041$  auf, der nur geringfügig über dem kritischen Wert von 2 liegt. Im ufosV2f-Modell ergibt sich für den Pfad ein höherer Koeffizient von  $\beta=0,244$  ( $t=2,149$ ). Die **Hypothese H10** findet folglich Bestätigung. Alle übrigen Pfade des Modells erweisen sich nicht als signifikant, so dass die **Hypothesen H11, H12, H14, H15, H17, H18** und **H20** auf Grundlage der PLS-Analysen nicht angenommen werden können (siehe Tabelle 8-22).

Tab. 8-22: Übersicht über die mit Hilfe der PLS-Modelle überprüften Hypothesen

Nr.	Annahme	Bestätigung
H5a	ufosV2r → Kaufintention	+
H5b	ufosV2f → Kaufintention	+
H7a	ufosV2r → Vertrauen	+
H7b	ufosV2f → Vertrauen	+
H8	Usability → Einkaufsvergnügen	kann nicht überprüft werden
H9a	Ästhetik → ufosV2r	+
H9b	Ästhetik → ufosV2f	+
H10	Ästhetik → Kaufintention	+
H11	Vertrauen → Kaufintention	n.s.
H12	Wahrgenommene Produktanzahl → Kaufintention	n.s.
H13	Einkaufsvergnügen → Kaufintention	kann nicht überprüft werden
H14	Wahrgenommener Preis → Kaufintention	n.s.
H15	Service → Kaufintention	n.s.
H16	Ästhetik → Einkaufsvergnügen	kann nicht überprüft werden
H17	Wahrgenommene Produktanzahl → Vertrauen	n.s.
H18	Bekanntheit → Vertrauen	n.s.
H19	Einkaufsvergnügen x Drucker → Kaufintention	kann nicht überprüft werden
H20	Ästhetik x Drucker → Kaufintention	n.s.

**Anmerkung:** Zusammenhang wird durch die berechneten PLS-Modell bestätigt (+), nicht signifikant (n.s.).



## 8.6 MLR zur Vorhersage der Kaufentscheidung

In einem letzten Schritt der Datenanalyse wird den beiden Hypothesen nachgegangen, die bis zu diesem Punkt der Auswertung noch nicht untersucht wurden. Einerseits soll die Frage beantwortet werden, inwiefern die beiden neuen *ufosV2*-Skalen eine Relevanz für tatsächliches Kaufverhalten besitzen (Hypothese H6 zum Zusammenhang zwischen den Usability-Skalen und der Kaufentscheidung). Zum anderen gilt es zu prüfen, inwiefern sich ein signifikanter Einfluss zwischen der Kaufintention und der Kaufentscheidung ergibt (Hypothese H21). Diese Annahmen konnten mit Hilfe der durchgeführten PLS-Analysen nicht überprüft werden, da PLS keine Einbindung kategorialer Daten auf Ebene der abhängigen Variablen erlaubt. Entsprechend erfolgt eine Untersuchung der Hypothesen H6 und H21 mit Hilfe einer multinomialen logistischen Regression (MLR; Abschnitt 6.7).

In der MLR wird eine Vorhersage berechnet, ob die Teilnehmer a) einen Gutschein wählen, b) im ersten Shop kaufen oder c) im zweiten Online-Shop kaufen. Die AV stellt somit die *Kaufentscheidung* mit den Ausprägungen *Kauf Shop 1*, *Kauf Shop 2* und *Gutscheinwahl* dar. Die Berechnungen werden anhand des Datensatzes mit zwei Shopbewertungen pro Fall durchgeführt.<sup>28</sup> Es wird ein sequentielles Vorgehen gewählt, bei dem eine blockweise Aufnahme von Prädiktoren in die Analyse erfolgt (Tabachnick & Fidell, 2006, S. 481 ff.). Dabei werden für jeden Analyseschritt die Veränderungen in den Gütekriterien des Gesamtmodells betrachtet, um Rückschlüsse auf die Einflussstärke des zuletzt hinzugefügten Variablenblockes zu ziehen. Abbildung 8-4 gibt einen Überblick über den Ablauf der Analyse. Im Einzelnen werden die folgenden Schritte durchlaufen:

*Schritt 1:* Ausgegangen wird von einem Modell, bei dem Variablen zur Demographie, zur Vorerfahrung sowie Kontrollvariablen als Prädiktoren dienen. Im Einzelnen handelt es sich um die drei demographischen Variablen *Alter*, *Geschlecht* und *Einkommen*, die *Anzahl der Online-Einkäufe* zur Vorerfahrung sowie die beiden Kontrollvariablen *Art der Internetverbindung* und *Erhebungsdauer*. Zusätzlich werden die beiden Variablen *Produkt-Involvement* und *Produktart* in diesem Schritt mit aufgenommen. Es wird erwartet, dass die beiden genannten Kontrollvariablen keinen Einfluss auf die AV Kaufentscheidung nehmen. Bezüglich des Einflusses der weiteren Variablen bestehen keine Hypothesen.

*Schritt 2a:* In Schritt 2a der sequentiellen MLR werden die Skalenwerte der *ufosV2r*-Skala für beide Shops aufgenommen. Auf Grundlage des Ergebnisses dieses Analyseschrittes soll eine Aussage zu Hypothese H6a getroffen werden, die die prädiktive Validität der reflektiven Usability-Skala betrifft. Diese Hypothese besagt, die Skala würde substanziell zur Erklärung der Kaufentscheidung beitragen.

---

<sup>28</sup> Als Referenzkategorie wird zunächst die *Kauf Shop 2* gewählt, so dass die beiden Blöcke *Kauf Shop 1 vs. Kauf Shop 2* sowie *Kauf Shop 2 vs. Gutschein* berechnet werden. Die Parameter des dritten Blocks *Kauf Shop 1 vs. Gutschein* werden geschätzt, indem der Gutschein als Referenzkategorie gewählt wird.

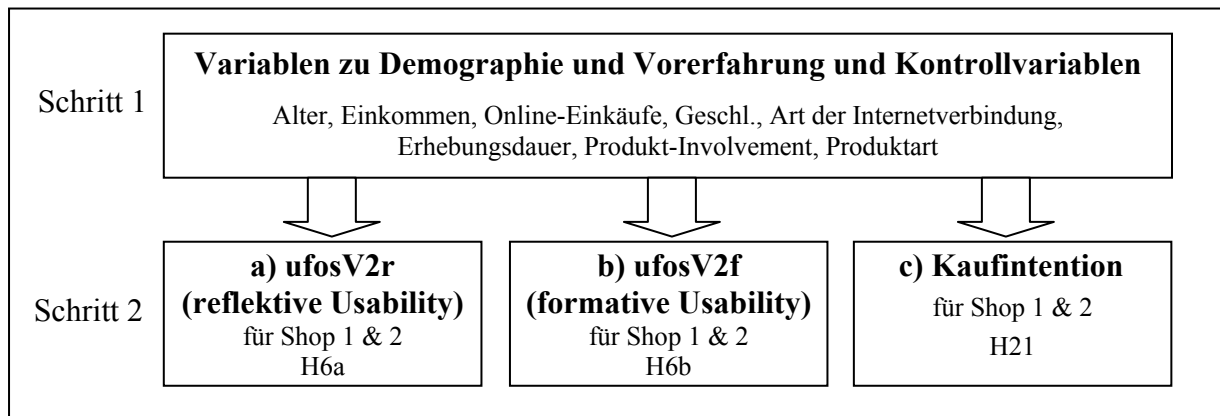


Abb. 8-4: Schritte der sequentiellen multinomialen logistischen Regressionsanalyse unter Angabe der jeweils hinzugefügten Prädiktoren

*Schritt 2b:* Eine analoge Hypothese wurde zur prädiktiven Validität der formativen Usability-Skala *ufosV2f* getroffen (H6b). Entsprechend werden in Schritt 2b anstelle der *ufosV2r*-Skalenwerte die der *ufosV2f*-Skala für beide Shops als Prädiktoren einbezogen. Als Skalenwerte dienen die Latenten Variablen Scores aus der PLS-Analyse zum Zwei-Konstrukt-Modell (Abschnitt 8.4.3).

*Schritt 2c:* Im konzeptuellen Erfolgsmodell wird die Kaufintention als direkter Prädiktor für die Kaufentscheidung angenommen (Hypothese H21). In Schritt 2c der MLR erfolgt eine Untersuchung, inwiefern sich diese Annahme bestätigt. Dazu werden anstelle der Usability-Variablen die Skalenwerte der *Kaufintention* für beide Shops als Prädiktoren hinzugezogen.

### 8.6.1 Prüfung der Voraussetzungen und Ausreißerdiagnostik

Bevor die Berechnung der MLR durchgeführt wird, erfolgt zunächst eine Prüfung der Voraussetzungen sowie eine Ausreißerdiagnose. Zur Prüfung der Multikollinearität zwischen den MRL-Prädiktoren werden die drei in Abschnitt 6.8 dargestellten Diagnoseansätze verfolgt. Zunächst werden die bivariaten Korrelationen zwischen den Prädiktoren analysiert (siehe dazu Tabelle E-29 im Anhang E). Eine vergleichsweise hohe Korrelation besteht zwischen dem Alter und dem Einkommen ( $r=0,58$ ;  $p<0,001$ ). In Tabelle E-30 im Anhang E werden die Toleranz- und VIF-Werte der Prädiktoren für die verschiedenen Schritte der MLR sowie die maximalen Konditionsindizes aufgeführt. Kein Toleranz-Wert liegt unterhalb von 0,4 und kein VIF-Wert oberhalb des kritischen Wertes von 2. Der maximale Konditionsindex liegt bei 13,81 (Schritt 2). Auf Grundlage dieser Analysen kann gefolgert werden, dass kein kritisches Maß an Multikollinearität zwischen den Prädiktoren besteht. Die Anzahl der erwarteten Häufigkeiten für die kategorialen Variablen Branche und Geschlecht ist groß genug, um die Goodness-of-fit-Tests interpretieren zu können (siehe Abschnitt 6.7.3). Die Voraussetzungen zur Berechnung der MLR sind damit gegeben.

Nach der Voraussetzungsprüfung wird untersucht, inwiefern Ausreißer vorhanden sind, die dem im untersuchten Modell unterstellten Zusammenhang nicht entsprechen und das Ergebnis durch ihre Variablenausprägungen verzerren. Für die Ausreißerdiagnose werden für beide

Logit-Modelle die vorhergesagten und beobachteten Variablenausprägungen der AV fallweise miteinander verglichen. Entsprechend erfolgt eine Berechnung von drei binären logistischen Regressionen, in denen die Entscheidungen *Kauf Shop 1 vs. Gutschein*, *Kauf Shop 1 vs. Kauf Shop 2* bzw. *Kauf Shop 2 vs. Gutschein* als Kriterium dienen (siehe Abschnitt 6.7.2). In Tabelle E-31 im Anhang E findet sich eine Auflistung der Probanden, die in mindestens einer der berechneten logistischen Modelle standardisierte Residuenwerte über 1 aufweisen. Aus theoretischer Sicht kann die Entscheidung zur Elimination eines Ausreißers nur dann getroffen werden, wenn es Hinweise darauf gibt, dass der jeweilige Fall nicht der Population angehört (vgl. Tabachnik & Fidell, 2006, S. 73). Entsprechend werden die im Rahmen der Ausreißerdiagnostik identifizierten Fälle mit jenen abgeglichen, die im Zuge der Analyse zum problematischen Antwortverhalten auffällig geworden sind (siehe Abschnitt 8.1.5). Tabelle E-32 im Anhang E listet die Codes der 16 Teilnehmer auf, die im Rahmen beider Analysen identifiziert wurden. Es erfolgt ein Ausschluss dieser Fälle aus der im Folgenden berechneten MLR, so dass sich das N auf 352 reduziert.

### 8.6.2 Sequentielle Berechnung der MLR

In der Tabelle 8-23 werden die Ergebnisse der sequentiellen Berechnung zusammengefasst. Eine ausführliche Darstellung der Ergebnisse jedes einzelnen Schrittes findet sich in den Tabellen E-33 bis E-48 im Anhang E.

*Schritt 1:* Aufgrund eines signifikanten Ergebnisses des Likelihood-Quotiententests zur globalen Anpassungsgüteprüfung ( $\chi^2=71,804$ ;  $df=22$ ;  $p<0,001$ ) kann geschlossen werden, dass im ersten Schritt der MLR mindestens ein Regressionsparameter im Modell von Null abweicht (siehe Tabelle E-33 im Anhang E). Ein *Nagelkerke*  $R^2=0,21$  weist auf eine eher niedrige Gesamtgüte dieses Ausgangsmodells hin. Der Anteil korrekt klassifizierter Fälle liegt bei 51,90%, was im Vergleich zur Zufallswahrscheinlichkeit (PZW) von 35% nicht als hoch anzusehen ist (siehe Tabelle E-34 im Anhang E).<sup>29</sup> Die auf Variablenebene durchgeführten Likelihood-Quotienten-Tests ergeben eine Signifikanz für die Variablen Einkommen ( $\chi^2=14,487$ ;  $df=2$ ;  $p<0,01$ ), Alter ( $\chi^2=15,067$ ;  $df=2$ ;  $p<0,001$ ), Produkt-Involvement ( $\chi^2=12,444$ ;  $df=2$ ;  $p<0,01$ ) und Produktart ( $\chi^2=34,164$ ;  $df=2$ ;  $p<0,001$ ; siehe Tabelle E-35 im Anhang E). Weder das Geschlecht, die Anzahl der Online-Einkäufe, die Erhebungsdauer noch die Art der im Versuch genutzten Internetverbindung erweisen sich als signifikante Prädiktoren für die Kaufentscheidung. Dies spricht dafür, dass kein Einfluss der beiden letztgenannten Kontrollvariablen auf die Versuchsergebnisse besteht.

---

<sup>29</sup> Im Fall der berechneten Analysen ergibt sich ausgehend von der Verteilung der Beobachtungen eine proportionale Zufallswahrscheinlichkeit von  $PZW=0,232+0,422+0,352=0,35$  (siehe Abschnitt 6.7.3).

Tab. 8-23: Kennwerte der sequentiellen MLR – Schritte 1 bis 2c

Prädiktoren	$\chi^2$ -Wert	df	Modell $\chi^2$ -Wert	$\Delta$ Modell $\chi^2$ -Wert	Nagel- kerke $R^2$	$\Delta$ Nagel- kerke $R^2$	Vorher- sage in %
<i>Schritt 1: Demographie etc.</i>							
Alter	15,067***	2					
Einkommen	14,487**	2					
Geschlecht	2,821	2					
Online-Einkäufe	2,387	2					
Erhebungsdauer	1,755	2					
Internetverbindung	1,219	2					
Produkt-Involvement	12,444**	2					
Produktart	34,164***	8					
Alle UVn Modell 1			71,804*** (df=22)		0,214		51,90
<i>Schritt 2a: ufosV2r-Skala</i>							
Alter	17,980***	2					
Einkommen	16,005***	2					
Geschlecht	4,884	2					
Online-Einkäufe	2,044	2					
Erhebungsdauer	1,682	2					
Internetverbindung	0,929	2					
Produkt-Involvement	8,372*	2					
Produktart	30,417***	8					
ufosV2r Shop 1	45,001***	2					
ufosV2r Shop 2	81,243***	2					
Alle UVn Modell 2a im Vergleich zu Schritt 1			196,049*** (df=26)	124,245*** (df=4)	0,494	0,280	67,64
<i>Schritt 2b: ufosV2f-Skala</i>							
Alter	16,628***	2					
Einkommen	14,842**	2					
Geschlecht	3,949	2					
Online-Einkäufe	2,525	2					
Erhebungsdauer	1,412	2					
Internetverbindung	0,774	2					
Produkt-Involvement	8,795*	2					
Produktart	33,314***	8					
ufosV2f Shop 1	46,863***	2					
ufosV2f Shop 2	71,852***	2					
Alle UVn Modell 2b im Vergleich zu Schritt 1			180,299*** (df=26)	108,495*** (df=4)	0,464	0,250	64,72
<i>Schritt 2c: Kaufintention</i>							
Alter	19,711***	2					
Einkommen	13,751**	2					
Geschlecht	4,569	2					
Online-Einkäufe	2,965	2					
Erhebungsdauer	1,728	2					
Internetverbindung	0,749	2					
Produkt-Involvement	8,645**	2					
Produktart	29,697***	8					
Kaufintention Shop 1	131,123***	2					
Kaufintention Shop 2	42,757***	2					
Alle UVn Modell 2c im Vergleich zu Schritt 1			242,430*** (df=26)	170,626*** (df=4)	0,575	0,361	70,55

Anmerkung: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ ,  $\chi^2$ -Wert des Likelihood-Quotienten-Tests auf Variablenebene ( $\chi^2$ -Wert); Freiheitsgrade (df),  $\chi^2$ -Wert des Likelihood-Quotienten-Tests zur globalen Modellanpassung (Modell  $\chi^2$ -Wert), Inkrement des Parameters ( $\Delta$ ), Pseudo-Determinationskoeffizient (Nagelkerke  $R^2$ ), Prozent korrekt vorhergesagter Fälle (Vorhersage in %).

*Schritt 2a:* Das Hinzufügen der Skalenwerte der reflektiven Usability-Skala ufosV2r führt zu einer signifikanten Verbesserung der Vorhersagekraft des Modells. Das Inkrement des  $\chi^2$ -Wertes im globalen Likelihood-Quotiententest ist hoch und signifikant ( $\Delta\chi^2=124,245$ ;  $df=4$ ;  $p<0,001$ ; siehe Tabelle E-37 im Anhang E). Ebenso deutlich fällt das Inkrement zur Varianzaufklärung aus: Das Nagelkerke  $R^2$  des MLR-Modells in Schritt 2a liegt bei einem hohen Wert von 0,494, was einem zusätzlichen Anteil von 28% aufgeklärter Varianz ( $\Delta R^2$ ) entspricht. Auch der Anteil korrekt klassifizierter Fälle ist mit 67,64% deutlich größer als für das Ausgangsmodell (siehe Tabelle E-38 im Anhang E). Beide ufosV2r-Skalenwerte (für Shop 1 und 2) erweisen sich in den auf Variablenebene durchgeführten Likelihood-Quotienten-Tests als signifikant ( $\chi^2=45,001$ ;  $df=2$ ;  $p<0,001$  bzw.  $\chi^2=81,243$ ;  $df=2$ ;  $p<0,001$ ; siehe Tabelle E-39 im Anhang E). **Hypothese H6a**, wonach die Skala ufosV2r eine hohe Vorhersagekraft für die Kaufentscheidung besitzt, ist folglich als bestätigt anzusehen.

*Schritt 2b:* Auch beim Hinzufügen der Latenten Variablen Scores der formativen Usability-Skala ufosV2f für beide Shops erweist sich das Inkrement des  $\chi^2$ -Wertes im globalen Likelihood-Quotiententest als hoch und signifikant ( $\Delta\chi^2=108,495$ ;  $df=4$ ;  $p<0,001$ ; Tabelle E-41 im Anhang E). Das Nagelkerke  $R^2$  steigt auf einen hohen Wert von 0,464 (im Vergleich zum Ausgangsmodell:  $\Delta R^2=0,25$ ). Der Anteil korrekt klassifizierter Fälle ist mit 64,72% deutlich größer als für das Ausgangsmodell (Tabelle E-42 im Anhang E). Die auf Variablenebene durchgeführten Likelihood-Quotienten-Tests weisen die beiden ufosV2f-Skalenwerte (für Shop 1 und 2) als signifikante Prädiktoren aus ( $\chi^2=46,863$ ;  $df=2$ ;  $p<0,001$  bzw.  $\chi^2=71,852$ ;  $df=2$ ;  $p<0,001$ ; Tabelle E-43 im Anhang E). Obwohl die prädiktive Güte für die ufosV2f-Skala ein wenig geringer ausfällt als für die ufosV2r-Skala, bestätigt sich folglich **Hypothese H6b**, dass ufosV2f eine hohe Vorhersagekraft für die Kaufentscheidung besitzt.

*Schritt 2c:* Anstelle der Skalen des ufosV2-Fragebogens werden in Schritt 2c die Skalenwerte der Kaufintention als Prädiktoren hinzugezogen. Erwartungsgemäß verbessert sich die Vorhersagegüte des Modells deutlich. Das Inkrement des  $\chi^2$ -Wertes im globalen Likelihood-Quotiententest ist hoch und signifikant ( $\Delta\chi^2=170,626$ ;  $df=4$ ;  $p<0,001$ ; Tabelle E-45 im Anhang E). Für das Nagelkerke  $R^2$  wird ein hoher Wert von 0,575 erreicht, was einem Inkrement von 36% aufgeklärter Varianz entspricht. Auch der Anteil korrekt klassifizierter Fälle liegt mit 70,55% deutlich höher als beim Ausgangsmodell (Tabelle E-46 im Anhang E). In den Likelihood-Quotienten-Tests auf Variablenebene zeigen sich signifikante  $\chi^2$ -Werte für die Skalenwerte zur Kaufintention für beide Shops ( $\chi^2=131,123$ ;  $df=2$ ;  $p<0,001$  bzw.  $\chi^2=42,757$ ;  $df=2$ ;  $p<0,001$ ; Tabelle E-46 im Anhang E). Somit bestätigt sich **Hypothese H21**, wonach die Kaufintention einen signifikanten Prädiktor für die Kaufentscheidung darstellt.

Als Fazit der MLR kann festgehalten werden, dass sowohl die ufosV2r- als auch die ufosV2f-Skala eine hohe Vorhersagekraft in Bezug auf das Kaufverhalten der Vpn besitzt. Damit bestätigen sich die Hypothesen H6a und H6b. Weiterhin wurde ein starker Einfluss der Kaufintention auf die Kaufentscheidung nachgewiesen und damit Hypothese H21 bestätigt.

## 8.7 Übersicht der Befunde

(1) *Festlegung der Items beider ufosV2-Skalen:* In die Datenerhebung gingen neun reflektive und 58 formative Usability-Items ein. Auf Grundlage einer EFA wurde ein reflektiver Indikator eliminiert (siehe Abschnitt 8.2.1), so dass sich die endgültige ufosV2r-Skala aus acht Items zusammensetzt; sie werden in Tabelle 8-24 wiedergegeben.

Tab. 8-24: Items der endgültigen reflektiven 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.

Basierend auf den Anteilen fehlender Werte und aufgrund inhaltlicher Überschneidungen wurden im Rahmen der Auswertung insgesamt zwölf der 58 formativen Indikatoren eliminiert (siehe Abschnitt 8.4.1). Tabelle 8-35 enthält die 46 Items der endgültigen formativen Skala ufosV2f.

Tab. 8-25.1: Items der endgültigen 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.

Tab. 8-25.2: Items der endgültigen formativen Skala ufosV2f

<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.

(2) *Ergebnisse zur Skalengüte:* Es wurden diverse Hypothesen zur Güte der ufosV2-Skalen empirisch untersucht. Tabelle 8-26 gibt einen Überblick, inwiefern die in Kapitel drei vorgestellten inhaltlichen Hypothesen H1 bis H9 zur Skalengüte bestätigt werden konnten.

Tab. 8-26: Hypothesen zur Güte der ufosV2-Skalen

Nr.	Hypothese	Methode / Abschnitt	Befund	empirische Bestätigung
H1	Interne Konsistenz der Skala ufosV2r	8.2.2	$\alpha=0,936$	✓
		PLS 8.4.3	Composite Reliability =0,949	✓
H2	Faktorielle Validität der Skala ufosV2r	EFA 8.2.1	Faktorenstruktur	eingeschr. ✓
		CFA 8.2.3	Fitmaße	eingeschr. ✓
		PLS 8.5.1	Kreuzladungen der Indikatoren mit LVn	✓
H3	Differenzierungsfähigkeit der Skala ufosV2r zwischen Online-Shops	ANOVA 8.2.4	$F=3,836$ ( $p<0,001$ )	✓
H4	ufosV2f → ufosV2r	PLS 8.4.3	$\beta=0,876$ (sign.); $R^2=0,767$	✓
H5a	ufosV2r → Kaufintention (konkurrenente Kriteriumsvalidität der Skala)	PLS 8.5.2	$\beta=0,342$ (sign.)	✓
H5b	ufosV2f → Kaufintention (konkurrenente Kriteriumsvalidität der Skala)		$\beta=0,238$ (sign.)	✓
H6a	ufosV2f → Kaufentscheidung (prädiktive Validität der Skala)	MLR 8.6.2	$\Delta\chi^2=124,3$ ; $\Delta R^2=0,28$ ; $AkkF=67,64\%$	✓
H6b	ufosV2f → Kaufentscheidung (prädiktive Validität der Skala)		$\Delta\chi^2=108,5$ ; $\Delta R^2=0,25$ ; $AkkF=64,72\%$	✓
H7a	ufosV2r → Vertrauen	PLS 8.5.2	$\beta=0,475$ (sign.)	✓
H7b	ufosV2f → Vertrauen	8.5.2	$\beta=0,534$ (sign.)	✓
H8	Usability → Einkaufsvergnügen	-	-	k.A.
H9a	Ästhetik → ufosV2r	PLS 8.5.2	$\beta=0,627$ (sign.)	✓
H9b	Ästhetik → ufosV2f		$\beta=0,685$ (sign.)	✓

Anmerkung: Hypothese bestätigt (✓), Hypothese unter Einschränkung bestätigt (eingeschr. ✓), keine Aussage möglich (k.A.), nicht signifikant und somit nicht bestätigt (n.s.), Anteil korrekt klassifizierter Fälle (AkkF).

(3) Ergebnisse zu den Zusammenhänge zwischen den übrigen Konstrukten innerhalb des konzeptuellen Erfolgsmodells: Weiterhin wurden einige Hypothesen (H10 bis H21) des konzeptuellen Erfolgsmodells zu Konstruktzusammenhängen geprüft, die die Usability nicht betreffen. Die Ergebnisse zu diesen Annahmen werden in Tabelle 8-27 zusammengefasst.

Tab. 8-27: Weitere Hypothesen zu Zusammenhängen des konzeptuellen Erfolgsmodells

Nr.	Hypothese	Methode / Abschnitt	Befund	empirische Bestätigung
H10	Ästhetik → Kaufintention	PLS 8.5.2	$\beta=0,188$ ; $\beta=0,244$ (beide sign.)	✓
H11	Vertrauen → Kaufintention		$\beta=0,074$ ; $\beta=0,103$	n.s.
H12	Wahrge. Produktanzahl → Kaufintention		$\beta=0,134$ ; $\beta=0,141$	n.s.
H13	Einkaufsvergnügen → Kaufintention	-	-	k.A.
H14	Wahrge. Preis → Kaufintention	PLS 8.5.2	$\beta=0,145$ ; $\beta=0,140$	n.s.
H15	Service → Kaufintention		$\beta=0,104$ ; $\beta=0,111$	n.s.
H16	Ästhetik → Einkaufsvergnügen	-	-	k.A.
H17	Wahrge. Produktanzahl → Vertrauen	PLS 8.5.2	$\beta=0,139$ ; $\beta=0,093$	n.s.
H18	Bekanntheit → Vertrauen		$\beta=0,131$ ; $\beta=0,138$	n.s.
H19	Einkaufsvergnügen x Drucker → Kaufintention	-	-	k.A.
H20	Ästhetik x Drucker → Kaufintention	PLS 8.5.2	$\beta=0,021$ ; $\beta=0,031$	n.s.
H21	Kaufintention → Kaufentscheidung	MLR 8.6.2	$\Delta\chi^2=170,6$ ; $\Delta R^2=0,36$ ; $AkkF=70,55\%$	✓

Anmerkung: Hypothese bestätigt (✓), keine Aussage möglich (k.A.), nicht signifikant und somit nicht bestätigt (n.s.), Anteil korrekt klassifizierter Fälle (AkkF).



## 9. Diskussion

Im Rahmen der Diskussion wird zunächst eine kritische Beurteilung des methodischen Vorgehens vorgenommen. Dazu erfolgt eine Analyse bzgl. eventueller Einschränkungen der Interpretierbarkeit bzw. Übertragbarkeit der Untersuchungsergebnisse (Abschnitt 9.1). Auf dieser Grundlage ist eine angemessene Interpretation der Untersuchungsergebnisse möglich, indem zunächst die einzelnen Hypothesen aufgegriffen werden (Abschnitt 9.2), die sich anschließend in Form allgemeiner Schlussfolgerungen zur Güte der ufosV2-Skalen integrieren lassen (Abschnitt 9.3). In Abschnitt 9.4 erfolgt eine Interpretation der wesentlichen Ergebnisse aus der Perspektive der Erfolgsfaktorenforschung. Die Implikationen der Befunde für Praxis und Forschung sind Gegenstand von Abschnitt 9.5. Die Arbeit schließt mit einem Ausblick auf weiterführende Fragestellungen (Abschnitt 9.6).

### 9.1 Bewertung des methodischen Vorgehens

Die Beurteilung des methodischen Vorgehens erstreckt sich über verschiedene Bereiche. Da eine quasi-experimentelle Form der Datenerhebung gewählt wurde, gilt es, sowohl die interne als die externe Validität der Untersuchung zu bewerten. Zudem wird das Vorgehen der statistischen Auswertung hinsichtlich potenzieller Vor- und Nachteile beleuchtet.

#### 9.1.1 Interne Validität des Quasi-Experiments

Bei der Durchführung eines Experiments ist stets zu hinterfragen, inwiefern eine angemessene interne Validität erreicht wird, also inwiefern die Effekte der UVn auf die AV durch systematische Störeinflüsse verzerrt werden (vgl. Rack & Christophersen, 2006, S. 31). Zur Erfüllung dieses Gütekriteriums wurden in der durchgeführten quasi-experimentellen Studie diverse versuchsplanerische Maßnahmen ergriffen, die positiv hervorzuheben sind.

*Einfluss der Produktpräferenz:* Zum Zeitpunkt der Kaufentscheidung befand sich stets dasselbe Produkt in den Warenkörben beider Shops. Hätte es sich um zwei unterschiedliche Produkte gehandelt, so wäre ein Einfluss der Präferenz für eines der beiden Produkte auf die Kaufentscheidung nicht auszuschließen. Die Produktpräferenz wurde folglich konstant gehalten und damit angemessen kontrolliert.

*Kontrolle des Produkt-Involvements:* Auf Grundlage des hohen Mittelwertes ( $M=4,74$ ) der Skala Produkt-Involvement kann mit Vorsicht geschlossen werden, dass die meisten Teilnehmer dem im Vorfeld gewählten Produkt eine hohe Relevanz zumaßen. Eine Standardabweichung von  $SD=1,26$  weist jedoch darauf hin, dass in der Stichprobe Unterschiede bestanden, als wie wichtig die Teilnehmer die Produkte wahrnahmen. Diesem Umstand konnte im Rahmen der sequentiellen MLR dadurch Rechnung getragen werden, dass die Variable Produkt-Involvement im ersten Schritt als Prädiktor für die Kaufentscheidung einbezogen wurde (siehe Abschnitt 8.6.2). Die Ergebnisse der Analyse lassen darauf schließen, dass die Variable tatsächlich in signifikantem Maße Einfluss auf die AV nahm.

*Einflüsse von Durchführungsdauer und Art der Internetverbindung:* Die Durchführungsdauer des Quasi-Experimentes variierte in erheblichem Maße (siehe Abschnitt 8.1.5). Zudem erfolgte die Versuchsdurchführung im Fall einiger weniger Teilnehmer unter Nutzung einer recht langsamen Verbindung zum Internet (siehe Abschnitt 7.2). Es wäre sowohl für die Variable *Dauer der Erhebung* als auch für die *genutzte Internetverbindung* denkbar, dass sie einen störenden Einfluss auf die Ergebnisse genommen haben könnte (siehe Abschnitt 5.4.3.2). Um diese Störeinflüsse auszuschließen, erfolgte in Schritt 1 der sequentiellen MLR eine Einbeziehung dieser Kontrollvariablen als Prädiktoren für die Kaufentscheidung. Für beide Variablen ergab sich kein signifikanter Zusammenhang zur AV. Es kann folglich davon ausgegangen werden kann, dass sie keinen wesentlichen Einfluss auf die Ergebnisse genommen haben.

*Einflüsse der Itempräsentationsreihenfolge:* In diversen Studien, in denen eine fragebogenbasierte Erhebung von Produktbewertungen erfolgte, zeigten sich Einflüsse der Präsentationsreihenfolge von Items auf die Untersuchungsergebnisse (vgl. Konradt et al., 2006; Oishi et al., 2003; Schwarz et al., 1991; Strack & Werth, 2002; siehe Abschnitt 5.6). Um derartigen Effekten entgegen zu wirken bzw. sie kontrollieren zu können, wurden im Versuchsplan zwei Maßnahmen ergriffen: Zum einen erfolgte die Darbietung der formativen Usability-Items untereinander randomisiert (siehe Abschnitt 5.6.1), zum anderen wurde die Reihenfolge der Itemblöcke der ufosV2f- und ufosV2r-Skalen systematisch variiert (Abschnitt 5.6.2). Eventuelle Effekte dieser Reihenfolge auf die Beantwortung der Items, die auf Ebene der einzelnen Teilnehmer aufgetreten sein können, wurden folglich über den gesamten Datensatz hinweg ausbalanciert.

*Reihenfolge der Shoppräsentation:* Wie in den Abschnitten 5.6.3 und 5.7 zum Versuchsplan zu entnehmen ist, wurde die Präsentationsreihenfolge der Online-Shops ausbalanciert. Durch diese Maßnahme sollte dem Umstand Rechnung getragen werden, dass die Usability-Bewertung des zweiten Shops anders ausfallen könnte, je nachdem wie gebrauchstauglich ein zuvor präsentierter Shop vom Nutzer wahrgenommen wird. Eventuelle Effekte dieser Art dürften sich durch die vorgenommene Ausbalancierung über den gesamten Datensatz hinweg ausgleichen. Aufgrund der in Abschnitt 7.1.2 erwähnten Akquiseprobleme für die Ticket-Shops konnten jedoch nicht alle Bedingungen des Versuchsplans gefüllt werden, so dass die Ausbalancierung nicht als vollständig anzusehen ist. Wie das Ergebnis eines t-Tests jedoch zeigt, hat die Reihenfolge der Shopdarbietung keinen Einfluss auf die Ergebnisse der Studie genommen (siehe Abschnitt 8.1.1).

*Einflüsse der Versuchsumgebung:* Die Hypertextversuchsumgebung lief sehr stabil und erwies sich für die Probanden als leicht zu handhaben. Es ergaben sich keine Anzeichen, dass die Teilnehmerinstruktionen unverständlich ausgefallen sind. Aufgrund der computer-gestützten Präsentationsform der Instruktionen können Versuchsleitereffekte weitgehend ausgeschlossen werden.

*Operationalisierung der Kaufentscheidung:* Eine weitere Kritik ließe sich dahingehend äußern, dass die Kaufentscheidung im Quasi-Experiment erhoben wurde, *nachdem* die Teilnehmer ihre Shop-Bewertungen abgegeben hatten. Streng genommen ist nicht auszuschließen, dass die Kaufentscheidung durch diesen Umstand beeinflusst wurde. Eine kognitive Auseinandersetzung mit Fragebogenitems kann die jeweils dahinter stehenden Bewertungsaspekte beim Befragten salient machen. In Folge dessen wäre es möglich, dass die bewerteten Einzelaspekte zum Zeitpunkt der Kaufentscheidung eine veränderte Relevanz besessen haben als im Normalfall. Diese Gefahr wird jedoch als relativ gering angesehen.

*Erhebungszeitraum:* Wie in Kapitel 7 erwähnt, zog sich die Erhebung der Daten über einen Zeitraum von zehn Monaten hin. Es ist zu vermuten, dass sich das Design einiger der 35 Online-Shops innerhalb dieser Zeit geändert hat, z.B. durch die Einbindung weihnachtlicher Motive im Dezember. Modifikationen in der Shop-Gestaltung könnten Veränderungen in den Ausprägungen der Usability und der weiteren betrachteten kundennahen Konstrukte zur Folge haben. Dieser Umstand erweist sich in Bezug auf die Ergebnisse der Arbeit jedoch als unproblematisch. Für die Validierung des Fragebogens ließe sich sogar argumentieren, dass eine erhöhte Unterschiedlichkeit der Bewertungsgegenstände zu begrüßen ist, da sie in einer Erhöhung der Datenvarianz resultieren kann. Lediglich für die varianzanalytische Überprüfung der Hypothese H3 hätte eine Änderung der Shopinterfaces ein Problem darstellen können. In dieser Hypothese wurde angenommen, die ufosV2r-Skala sei in der Lage, Unterschiede zwischen den Shops zu identifizieren. Das signifikante Ergebnis der ANOVA (siehe Abschnitte 8.2.4) entkräftet Befürchtungen bzgl. negativer Konsequenzen des langen Erhebungszeitraums.

Insgesamt wurden diverse versuchsplanerische Maßnahmen ergriffen, um eine hohe interne Konsistenz des Quasi-Experimentes zu gewährleisten. Es ist zu hinterfragen, inwiefern weitere Störeinflüsse eine Rolle gespielt haben können, die nicht erhoben bzw. kontrolliert wurden. Unter Umständen haben einzelne Probanden bereits im Vorfeld der Versuchsdurchführung von der Möglichkeit zur Wahl des Gutscheins erfahren – etwa aufgrund von Empfehlungen anderer Teilnehmer. Dies könnte sich negativ auf die Motivation der betroffenen Probanden ausgewirkt haben, einen Online-Kauf zu tätigen. Im Rahmen der Datenerhebung bzw. -auswertung ergaben sich darüber hinaus keine Hinweise auf verzerrende Einflüsse durch Störvariablen, die nicht erhoben bzw. kontrolliert wurden.

### **9.1.2 Externe Validität des Quasi-Experiments**

Experimente zeichnen sich einerseits dadurch aus, dass sie gute Möglichkeiten zur Gewährleistung einer hohen internen Validität bieten; andererseits ergeben sich bei dieser Form der Datenerhebung häufig Einschränkungen der externen Validität. Dieses Gütekriterium betrifft die Frage, inwiefern sich die Ergebnisse einer Studie auf andere Situationen und Personen verallgemeinern lassen (vgl. Rack & Christophersen, 2006, S. 31). Hinsichtlich des durchgeführten Versuchs kann hinterfragt werden, inwiefern a) die

hergestellte Einkaufssituation zu künstlich ausfiel, b) die Kaufentscheidung als realistisch gelten kann, c) die Stichprobenziehung als angemessen anzusehen ist bzw. d) ein störender Einfluss aufgrund der getroffenen Shopauswahl befürchtet werden muss.

#### 9.1.2.1 Realitätsbezug der experimentellen Einkaufssituation

Zur Bewertung der externen Validität soll zunächst der Frage nachgegangen werden, ob die experimentelle Einkaufssituation zu große Unterschiede zu einem gewöhnlichen Kauf im Internet aufwies, als dass eine Verallgemeinerung der Ergebnisse gerechtfertigt ist. Die Tatsache, dass 291 der insgesamt 378 Teilnehmer im Versuch einen tatsächlichen Online-Einkauf vornahmen, spricht eindeutig gegen einen mangelnden Realitätsbezug (siehe Abschnitt 8.3.2). Andererseits können gegenüber einer normalen Einkaufssituation diverse Unterschiede angeführt werden.

*Standardisierung der Interaktion:* Zum einen wurde der Ablauf des Shopbesuchs durch die gestellten Aufgaben bis zu einem gewissen Grad standardisiert. Dies erschien nötig, um zu gewährleisten, dass die Shop-Bewertungen auf einem ähnlichen Erfahrungsschatz der Probanden beruhen.<sup>30</sup> Zudem sollten die Teilnehmer über eine angemessene Beurteilungsgrundlage für die diversen abgefragten Bewertungsaspekte verfügen. Das Maß an Standardisierung war jedoch bewusst begrenzt, da den Teilnehmern bei der Umsetzung der Aufgaben ein eher großer Spielraum gegeben wurde. Die Probanden erhielten also ausgiebig Gelegenheit, die Shops auch gemäß den eigenen Interessen zu erkunden.

*Angemessenheit der Aufgaben:* Neben dem Ausmaß an Standardisierung der Interaktion stellt sich auch die Frage nach der Angemessenheit der gestellten Aufgaben. Sicherlich kann die Art der Aufgabenstellung einen erheblichen Einfluss auf die Ergebnisse nehmen. Im Fall von Aufgaben zu komplexeren Benutzungsvorgängen (etwa „Vergleichen Sie alle Produkte, die die Eigenschaften a, b und c aufweisen bzgl. des Kriteriums xy“) würden sich zwangsläufig sehr spezifische und hohe Anforderungen an die Benutzerfreundlichkeit eines Online-Shops ergeben. Es wäre daher denkbar, dass Shops bei komplexeren Aufgaben tendenziell als weniger benutzerfreundlich wahrgenommen werden. Zudem wäre ein Großteil der Probanden nicht in der Lage, die Aufgabe vollständig zu erledigen. Die Wahl von prototypischen Einkaufsabläufen (gezielte Suche, Stöbern, Warenkorbinteraktion, Durchlaufen des Bestellvorgangs), wie sie in der vorliegenden Arbeit getroffen wurde, erscheint angemessen, da ein Shop diese Aufgaben unterstützen *muss*, damit der Kunde seinen Einkauf erfolgreich abwickeln kann.

---

<sup>30</sup> Die Standardisierung kann somit auch als Maßnahme zur Gewährleistung der internen Validität angesehen werden.

*Dauer und Umfang der Interaktion:* Die maximale Dauer der Interaktion mit den Shops wurde den Teilnehmern freigestellt. Durch die gestellten Aufgaben ergab sich eine minimale Dauer, die bei ca. 15 Minuten pro Schnittstelle lag, was gemessen am gesamten Funktionsumfang vieler Shops sicher kurz ist. Der große Spielraum des Interaktionsumfangs für den Probanden stützt jedoch die externe Validität der Untersuchung. Schließlich liegt es auch bei einem normalen Online-Einkauf in der Hand der Kunden, wie lange sie sich im Online-Shop aufhalten.

*Zeitpunkt der Shopbewertung:* Es ließe sich kritisch anmerken, dass die Shop-Bewertungen im Versuchsablauf erfolgten, noch bevor die Probanden ihre Bestellung in einem der beiden Shops abgeschlossen hatten. Zwar wurde diesem Umstand dadurch Rechnung getragen, dass die Teilnehmer im Rahmen der zweiten Aufgabe gebeten wurden, sich über den Bestellvorgang zu informieren und ihn probeweise zu beginnen. Es muss jedoch in Zweifel gezogen werden, ob die Erfahrungen in einem solchen Testlauf mit denen einer tatsächlichen Bestellung vergleichbar sind. Entsprechend hätte die Shop-Bewertung im Einzelfall anders ausfallen können, wenn sie erst nach Abschluss der Bestellung erhoben worden wäre. Ein Anzeichen für einen derartigen Einfluss des Versuchsszenarios auf die Beantwortung der Items kann in dem recht hohen Anteil fehlender Werte von 10,6% für das Item uf47 gesehen werden, das die Dateneingabe im Bestellvorgang betrifft.

Es kann resümiert werden, dass die Anzahl der getätigten Einkäufe eindeutig für die externe Validität der experimentellen Einkaufssituation spricht. Andererseits müssen gegenüber einer normalen Einkaufssituation einige Unterschiede eingeräumt werden, die die Übertragbarkeit der Untersuchungsergebnisse einschränken könnten. Über Art und Ausmaß dieser Einschränkungen ließe sich allenfalls spekulieren. Nach Auffassung des Autors ist die externe Validität der Einkaufssituation jedoch insgesamt als hoch anzusehen.

#### 9.1.2.2 Operationalisierung der Kaufentscheidung

Wie in Abschnitt 2.1.6.1 dargestellt wurde, kann von einer *Kaufentscheidung* dann gesprochen werden, wenn mindestens drei verschiedene Entscheidungsalternativen bestehen und ein Mindestmaß an kognitiver Beteiligung auf Seiten des Käufers aufgebracht wird (Wiswede, 1995). Beide Bedingungen werden durch die Operationalisierung dieser Erfolgsgröße im Rahmen der Untersuchung erfüllt. Zum einen konnten sich die Teilnehmer zwischen einem Kauf in einem von zwei Shops und der Wahl eines Gutscheins entscheiden. Zum anderen ist davon auszugehen, dass es sich bei den Kaufentscheidungen nicht um *Gewohnheits-* bzw. *Spontankäufe* handelte, da sich die Teilnehmer bereits im Vorfeld der Untersuchung auf ein Produkt festlegten. Es wurden folglich *überlegte Käufe* getätigt, die eine kognitive Beteiligung der Probanden erforderten. Gemäß den Kriterien von Wiswede (1995) kann also von einer extern validen Kaufentscheidung gesprochen werden. Unkritisch erscheint, dass die Teilnehmer gebeten wurden, sich bereits im Vorfeld auf ein Produkt festzulegen. Entsprechend handelt es sich um eine geplante Form des Einkaufs.

Vielfach finden sich Berichte darüber, dass Online-Kunden ihren Einkauf abbrechen, nachdem sie den Bestellvorgang bereits begonnen haben. Um diesem Umstand Rechnung zu tragen, erhielten die Teilnehmer die Möglichkeit, eine zunächst getroffene Entscheidung bis zum Abschicken der Bestellung zu revidieren und erneut vorzunehmen. 26 Probanden machten von dieser Möglichkeit Gebrauch. Die operationalisierte Kaufentscheidung kann auch in dieser Hinsicht durchaus als realistisch beurteilt werden.

Gewissermaßen unrealistisch erscheint jedoch, dass die Probanden die bei einer Bestellung im Versuch anfallenden Rechnungsbeträge nicht bzw. nur zum Teil mit ihrem eigenen Geld beglichen, da ihnen 15 Euro als Teilnahmeentlohnung von der Versuchsleitung ausbezahlt wurden. Die anfallenden Kosten fielen recht unterschiedlich hoch aus. In einigen Fällen lag der Betrag unterhalb der 15 Euro, so dass die Teilnehmer den jeweiligen Differenzbetrag behalten konnten. Vielfach ergaben sich auch Rechnungsbeträge, die oberhalb der Grenze von 15 Euro lagen. Nur in diesen Fällen mussten die Probanden für den Restbetrag selber aufkommen. Entschieden sie sich *gegen* eine Bestellung, so erhielten sie einen Gutschein über 15 Euro für eine große Warenhauskette. Diese Option besteht beim normalen Online-Einkauf natürlich nicht, da der Kunde sein eigenes Geld behält, wenn er nichts bestellt.

Zusammenfassend ist festzuhalten, dass die im Rahmen des Quasi-Experiments operationalisierte Kaufentscheidung gewisse Unterschiede gegenüber einer realen aufweist. Inwiefern dies Einschränkungen der externen Validität dieser Erfolgsgröße mit sich bringt, bleibt unklar. Nach Auffassung des Autors ist jedoch nicht von schwerwiegenden Folgen auszugehen. Eine Sichtung thematisch verwandter Studien zeigt im Übrigen, dass in den wenigsten Fällen *überhaupt* tatsächliches Kaufverhalten erfasst und auf abhängiger Variablenseite in die Auswertung einbezogen wurde. Zumeist erfolgt lediglich eine Abfrage der Kaufintention. Die Güte dieser Erfolgsgröße bleibt jedoch unklar, solange kein Nachweis geführt wird, dass sie auch eine Verhaltensrelevanz besitzt. Zudem kann in vielen Untersuchungen ein Common Method Bias nicht ausgeschlossen werden, da sowohl auf Ebene der Prädiktoren als auch auf Ebene der abhängigen Variablen ausschließlich subjektive Fragebogendaten operationalisiert wurden (siehe Abschnitt 9.1.3).

### 9.1.2.3 Repräsentativität der Stichprobe

Zur Bewertung der externen Validität des Quasi-Experimentes gilt es, Effekte der Stichprobenspezifität auszuschließen (vgl. Kaya & Himme, 2006, S. 90). Wie berichtet wurde (Abschnitt 7.1.3), unterschied sich die gezogene Stichprobe hinsichtlich einiger Merkmale gegenüber der Population deutscher Internetnutzer. Neben diesen Unterschieden hat bei der Ziehung der Stichprobe möglicherweise eine Selbstselektion der Teilnehmer stattgefunden. Ein Teil der Probanden hat sich u.U. deswegen zur Teilnahme bereit erklärt, da ein besonderes Interesse am Untersuchungsgegenstand bestand. Für einige Vpn stellte möglicherweise die Entlohnung von 15 Euro einen besonderen Anreiz zur Teilnahme dar. In Abschnitt 2.7.7.2 wurden Befunde zu Einflüssen personenbezogener Merkmale im

Zusammenhang des Online-Shoppings angeführt. Eine Übertragbarkeit der Ergebnisse dieser Studie auf andere Länder ist angesichts kultureller Unterschiede fraglich (vgl. Mahmood et al., 2004; Noiwan & Norcio, 2006), was angesichts des Umstands, dass eine deutschsprachigen Version des ufosV2-Fragebogens entwickelt wurde, allerdings trivial erscheint. Es lassen sich darüber hinaus keine konkreten Einschränkungen der Ergebnisse benennen, die sich aus den Unterschieden zwischen Stichprobe und Population bzw. den angeführten Formen von Selbstselektion ergeben.

Im Vergleich zu thematisch verwandten Untersuchungen erscheinen die Einschränkungen der Stichprobenrepräsentativität für das durchgeführte Quasi-Experiment relativ gering. Am durchgeführten Versuch nahm eine vergleichsweise große Anzahl an Vpn teil. Neben der Stichprobengröße ist auch ihre Zusammensetzung zu würdigen. Im Gegensatz zu vielen anderen Laborexperimenten setzt sich die Stichprobe nicht ausschließlich aus Studenten zusammen, und der Anteil an Probanden in schwer zu akquirierenden Altersgruppen fällt verhältnismäßig hoch aus. In einer Reihe von Untersuchungen wurden Online-Befragungen durchgeführt. Diese Form der Datenerhebung ist jedoch mit einer Reihe von Problemen behaftet (Evans & Mathur, 2005). Unter anderem treten oft starke Selbstselektionseffekte auf, die dazu führen, dass die Stichproben nicht repräsentativ ausfallen (Ray & Tabor, 2003).

#### 9.1.2.4 Abhängigkeit der Ergebnisse von der Shopauswahl

Neben der Repräsentativität der Vpn-Stichprobe gilt es, die der Shop-Auswahl zu betrachten (Monk, 2004). Es stellt sich die Frage, inwiefern die Ergebnisse der Studie davon abhängen, welche Benutzerschnittstellen zur Variation der UV *Online-Shop* herangezogen wurden. Insgesamt liegt die Anzahl der einbezogenen Online-Shops bei 35, die fünf verschiedenen Produktarten zuzuordnen sind. Diese Anzahl ist im Verhältnis zu anderen Studien zur Validierung von Usability-Fragebögen als groß anzusehen, was positiv zu werten ist. Die Wahl der Produktarten bzw. Shops erfolgte anhand verschiedener Kriterien (spezifizierbare Produkte im Preisbereich 15 Euro, große Produktanzahl, hohe Beliebtheit der Produkte; siehe Abschnitt 5.1.1). Es fand also keine zufällige Ziehung der Shops statt. Es wurde jedoch darauf geachtet, dass in der Auswahl der Shops sowohl bekannte als auch unbekannte sowie Shops mit eher guter bzw. schlechter Usability-Ausprägung enthalten waren (siehe Abschnitt 5.1.2). Zudem unterscheiden sich die fünf Produktarten untereinander deutlich hinsichtlich ihrer Produkteigenschaften. Mit Druckerpatronen wurden Verbrauchsgüter berücksichtigt, die sich primär durch Sucheigenschaften auszeichnen (siehe Abschnitt 2.7.7.1). Bücher, CDs und DVDs besitzen demgegenüber einen hohen Entertainment-Charakter. Dies gilt auch für Veranstaltungstickets, die jedoch keine langlebigen Produkte darstellen, sondern vielmehr als eine Form von Dienstleistung anzusehen sind. Angesichts dieser Vielfalt unterschiedlicher Shops, die zur Variation der UV herangezogen wurden, ist es unwahrscheinlich, dass eine andere, zufällig gezogene Shopauswahl zu stark divergierenden Ergebnisse geführt hätte.

### 9.1.3 Vorgehen der statistischen Auswertung

Neben der internen und externen Validität des Quasi-Experiments könnten sich Einschränkungen der Interpretierbarkeit von Untersuchungsergebnissen aufgrund methodischer Probleme bei der statistischen Datenauswertung ergeben. Entsprechend wird das Vorgehen der Daten-Analyse sowie die Güte der Messungen einer kritischen Betrachtung unterzogen.

Das methodische Vorgehen kann in vielerlei Hinsicht als gründlich angesehen werden. So wurde die Aufteilung des Datensatzes durch das Ergebnis eines t-Test zum Einfluss der Shop-Reihenfolge angemessen gerechtfertigt (Abschnitt 8.1.1). Mit der Multiplen Imputation wurde eine valide Technik zur Ersetzung fehlender Werte angewendet (vgl. Tabachnik & Fidell, 2006). Der Ausschluss von Ausreißern in der berechneten MLR (siehe Abschnitt 8.6.1) erfolgte nur dann, wenn auf Grundlage einer Prozedur zur Identifikation problematischer Fälle begründete Zweifel bzgl. der Eignung der ausreißenden Fälle bestanden (siehe Abschnitt 8.1.5).

Zentral für eine Bewertung des methodischen Vorgehens ist die Frage, inwiefern ein Common Method Bias die Ergebnisse der Untersuchung beeinflusst haben könnte (vgl. Spector, 2006; siehe Abschnitt 5.3). Es ließe sich vermuten, dass die Ergebnisse der PLS-Analyse einem Common Method Bias unterliegen, da alle LVn des Modells auf Basis derselben Methode erfasst wurden und eine Einbindung der AV Kaufentscheidung in diese Analyse aufgrund des kategorialen Skalenniveaus nicht möglich war (siehe Abschnitt 8.5). Das Ergebnis des Harman's Single-Factor Tests wies darauf hin, dass im Datensatz kein Common Method Bias besteht (siehe Abschnitt 8.1.4). Wie Podsakoff et al. (2003) jedoch darlegen, besitzt ein solches Ergebnis eine deutlich eingeschränkte Aussagekraft. Den entscheidenden Hinweis, dass ein Common Method Bias in der PLS-Analyse zum konzeptuellen Erfolgsmodell keine große Rolle gespielt haben dürfte, liefert das Ergebnis der sequentiell berechneten MLR (siehe Abschnitt 8.6.2). In Schritt 2c der MLR zeigte sich, dass die Kaufintention eine hohe prädiktive Güte in Bezug auf die Kaufentscheidung und damit ein hohes Maß an Verhaltensrelevanz besitzt (Hypothese H21). Der hohe Zusammenhang zwischen beiden Variablen räumt Zweifel bzgl. der Eignung der Kaufintention als AV in der PLS-Analyse weitgehend aus. Schließlich ist ein Common Method Bias im Fall der Kaufentscheidung nicht zu erwarten, da es sich um ein objektives, verhaltensbasiertes Maß handelt und nicht um ein subjektives, mit Hilfe von Fragebogenitems erfasstes.

Neben diesen positiven Aspekten der statistischen Auswertung ergaben sich auch einige methodische Probleme, die im Folgenden angeführt werden. Sowohl die Ergebnisse der EFA als auch der CFA sind insofern nur eingeschränkt interpretierbar, als dass die Anzahl der Indikatoren einzelner latenter Variablen streng genommen nicht hoch genug war, um Berechnungen dieser Analysen durchzuführen (siehe Abschnitte 6.5.1.1 und 6.5.3.1).



Weiterhin liegt zur Erfassung der Konstrukte wahrgenommener Preis und Bekanntheit jeweils nur ein Item vor (siehe Abschnitt 8.2.5). Zwar können Single-Item-Maße durchaus die gängigen Gütekriterien Reliabilität und Validität erfüllen (Drolet & Morrison, 2001; Gardener et al., 1998). Jedoch sollten im Vorfeld ihres Einsatzes Erkenntnisse bzgl. ihrer Messgüte bestehen, was für die verwendeten Single-Items nicht der Fall war. Entsprechend müssen Zweifel bzgl. ihrer Messgüte eingeräumt werden; die inhaltliche Interpretation von Hypothesen, die den wahrgenommenen Preis und die Bekanntheit betreffen, wird damit erschwert. Weiterhin mussten vier von ursprünglich sechs formativen Items zur Erfassung des Service aufgrund zu hoher Anteile fehlender Werte eliminiert werden, was eine Änderung des Bedeutungsinhalts dieses Konstrukt zur Folge hat (siehe Abschnitt 8.2.5).

Ein methodisches Problem ergab sich bei der Analyse des PLS-Zwei-Konstrukt-Modells (siehe Abschnitt 8.4.2). Zwischen den formativen Indikatoren der ufosV2f-Skala bestand ein sehr hohes Maß an Multikollinearität. Um diesem Problem zu begegnen und eine unverzerrte Schätzung der Modellparameter zu ermöglichen, wurden die Indikatoren mit Hilfe einer HKA zu Unterindizes zusammengefasst. Folglich konnte der Einfluss der einzelnen formativen Usability-Aspekte auf die Erfolgsgröße Kaufintention nicht in Form von Gewichten identifiziert werden – was inhaltlich sicherlich von großem Interesse gewesen wäre. Zum anderen lag der Gesamtanteil aufgeklärter Varianz in der HKA bei 55%. Zwar ist dieser Anteil nicht als schlecht zu werten (Heck, 1998, S. 187), jedoch bleibt offensichtlich ein nicht unerheblicher Anteil der Datenvarianz unaufgeklärt. Der damit verbundene Informationsverlust kann Folgen für die berechneten Analysen nach sich gezogen haben (siehe Abschnitt 9.2.1.5).

Es ließe sich weiterhin kritisieren, dass im Rahmen der sequentiellen MLR keine Einbeziehung von weiteren kundennahen Erfolgsfaktoren stattfand. Eine Analyse der Einflussstärke dieser Konstrukte und ihrer Zusammenhänge untereinander stellt für die vorliegende Arbeit nur einen Nebenschauplatz dar, die dem eigentlichen Ziel der Skalenentwicklung und -validierung untergeordnet ist. Eine entsprechende Untersuchung erfolgte im Rahmen einer PLS-Analyse (Abschnitt 8.5). Durch eine erneute Analyse der Beziehungen mit Hilfe der MLR würde dem Gegenstandsbereich der Erfolgsfaktorenforschung innerhalb der Arbeit eine zu hohe Priorität zukommen. Zudem wäre eine Interpretation der Ergebnisse der MLR deutlich erschwert, da sich bei einer Einbeziehung dieser Konstrukte eine sehr hohe Anzahl an Prädiktoren ergäbe.

Letztlich ist anzumerken, dass dem Datensatz dieser Arbeit Querschnittsdaten zugrunde liegen. Eine kausale Interpretation der beobachteten Konstruktbeziehungen in den berechneten Strukturgleichungsmodellen ist daher nur unter Vorbehalt möglich. So kann sich in einem PLS-Modell zwar ein hoher und signifikanter Pfadkoeffizient zwischen zwei Konstrukten zeigen; die angenommene Richtung der Kausalität ist dadurch jedoch nicht als belegt anzusehen.

## 9.2 Diskussion der Befunde

Nachdem das methodische Vorgehen der experimentellen Untersuchung einer kritischen Betrachtung unterzogen wurde, kann eine Interpretation der Ergebnisse vorgenommen werden. Zunächst werden die Befunde zu den einzelnen Gütekriterien der neu entwickelten ufosV2-Skalen beleuchtet, anschließend erfolgt eine Betrachtung der Ergebnisse zu den weiteren Zusammenhängen des konzeptuellen Erfolgsmodells.

### 9.2.1 Befunde zur Güte der ufosV2-Skalen

#### 9.2.1.1 Interne Konsistenz der reflektiven ufosV2r-Skala

Die Hypothese H1 betrifft die Frage, inwiefern die aus acht Items bestehende reflektive Skala ufosV2r eine ausreichende interne Konsistenz aufweist, als dass eine reliable Erfassung des Konstruktes möglich ist. Zwei Befunde sprechen für diese Annahme: Zum einen zeigte sich ein sehr hohes Cronbachs Alpha von  $\alpha=0,94$  (siehe Abschnitt 6.5.2). Der zweite Befund zur internen Konsistenz entstammt der PLS-Analyse zum Zwei-Konstrukt-Modell, in der eine sehr hohe Composite Reliability von 0,95 erreicht wurde (siehe Abschnitt 6.6.2.1). Die Hypothese H1 wird damit bestätigt. Mit der Gewährleistung einer hohen Messgenauigkeit wird eine wesentliche Voraussetzung für die praktische Anwendung der Skala zum Zwecke der Evaluation von Online-Shops erfüllt. Sollte die Anzahl von acht Items in einem bestimmten Untersuchungskontext als zu hoch angesehen werden, kann in Erwägung gezogen werden, nur drei der reflektiven Items heranzuziehen, für die sich immer noch ein Cronbachs Alpha von  $\alpha=0,90$  ergibt (siehe Abschnitt 8.2.2). Ob die Einschränkung der Messgenauigkeit vertretbar erscheint, ist vor dem jeweiligen Forschungshintergrund zu beurteilen. Der Anwender sollte sich jedoch darüber im Klaren sein, dass die weiteren Skalengütekriterien lediglich für die ufosV2r-Skala mit allen acht Items untersucht wurden.

#### 9.2.1.2 Faktorielle Validität der reflektiven ufosV2r-Skala

In der Hypothese H2 wurde angenommen, die reflektive Skala ufosV2r erweise sich gegenüber Skalen zur Erfassung weiterer kundennaher Erfolgsfaktoren als faktoriell valide. Es wurden drei statistische Verfahren angewandt, auf Grundlage derer Aussagen bzgl. dieser Annahme getroffen werden können.

In einer EFA ergab sich ein gemeinsamer Faktor für die acht Usability-Items sowie die beiden Items zur Ästhetik, wenn zusätzlich die Items zu drei weiteren Skalen einbezogen wurden (vierte berechnete HAA in Tabelle 8-6; siehe Abschnitt 8.2.1). Dieses Ergebnis spricht zunächst gegen eine angemessene faktorielle Validität von ufosV2r. Wurden die Items der drei weiteren Skalen nicht in die Analyse aufgenommen, so zeigte sich die Eindimensionalität des Usability/Ästhetik-Faktors jedoch nicht (fünfte HAA in Tabelle 8-7). Vielmehr ergaben sich gemäß Hypothese H2 zwei Faktoren. Es ist zu berücksichtigen, dass bei einer EFA mehr als zwei Items pro erwarteten Faktor vorhanden sein sollten, um eine Berechnung

durchführen zu können (vgl. Bühner, 2004, S. 157). Diese Voraussetzung war im Fall der berechneten EFA nicht erfüllt. Entsprechend ist der Eindimensionalität des Faktors Usability/Ästhetik, die sich in der vierten HAA ergab, nicht zu viel Gewicht beizumessen. Die Ergebnisse der EFA weisen bereits indirekt auf einen hohen Zusammenhang zwischen der Ästhetik und der Usability hin. Diese Beziehung wurde in einer der weiteren Hypothesen ausdrücklich angenommen (H9) und im Rahmen einer PLS-Analyse empirisch bestätigt (siehe Abschnitt 9.2.1.6).

In einer CFA wurde ein angemessener Modell-Fit erst nach Zulassen von Kovariationen der Messfehler auf latenter Variablenebene erreicht (siehe Abschnitt 8.2.3), was unter Einschränkung für die Annahme einer faktoriellen Validität von ufosV2r im Sinne von Hypothese H2 spricht. Auch im Fall dieses Verfahrens ist jedoch zu berücksichtigen, dass die Anzahl der Indikatoren für drei der vier LVn höher hätte liegen müssen, um eine Berechnung durchführen zu können (vgl. Marsh et al., 1998). Daher sind die Ergebnisse der CFA ebenfalls nur unter einem gewissen Vorbehalt interpretierbar.

Eindeutig für eine faktorielle Validität der Skala ufosV2r sprechen die Ergebnisse der PLS-Analyse zum konzeptuellen Erfolgsmodell. Bei einer Betrachtung der Korrelationen zwischen den einbezogenen Indikatoren mit den LVn des Modells zeigten sich keine unerwarteten Nebenladungen. Zusammenfassend wird die faktorielle Validität der reflektiven Usability-Skala ufosV2r durch die Ergebnisse der EFA und CFA unter leichter Einschränkung und durch die PLS-Analyse uneingeschränkt untermauert. Die Hypothese H2 wird insgesamt als bestätigt angesehen.

### 9.2.1.3 Diskriminationsfähigkeit der reflektiven ufosV2r-Skala

Um die Diskriminationsfähigkeit der reflektiven ufosV2r-Skala zwischen mehreren Online-Shops zu testen (Hypothese H3), erfolgte die Berechnung einer ANOVA (siehe Abschnitt 8.2.4). Aufgrund des signifikanten Ergebnisses kann auch dieses Gütekriterium der Skala ufosV2r als erfüllt gelten. Es ergeben sich für mehrere Online-Shops also offensichtlich unterschiedliche Skalenausprägungen – was für die praktische Anwendung dieser Globalskala sicher als eine Minimalanforderung anzusehen ist. Schließlich erscheint ein Einsatz einer Evaluationsskala nur dann sinnvoll, wenn erwartet werden kann, dass sich mit ihrer Hilfe Unterschiede in der Usability-Ausprägung aufzeigen lassen. Inwiefern unterschiedliche Skalenwerte von ufosV2r auch tatsächlich eine unterschiedlich hoch wahrgenommene Usability widerspiegeln, kann erst auf Grundlage der weiteren Ergebnisse zur Validität der Skala beurteilt werden.

#### 9.2.1.4 Statistischer Zusammenhang der ufosV2-Skalen

Gemäß dem in Abschnitt 2.8.3.4 vorgestellten Vorgehen wurde die formative Skala ufosV2f in einem PLS-Zwei-Konstrukt-Modell in Beziehung zur reflektiven Skala ufosV2r gesetzt (siehe Abschnitt 8.4.3). Es ergab sich ein sehr hoher, signifikanter Pfadkoeffizient mit einem Wert von  $\beta=0,876$  sowie eine sehr hohe aufgeklärte Varianz in der latenten Variable ufosV2r ( $R^2=0,767$ ). Mit Hilfe zweier zusätzlicher Analysen wurde gezeigt, dass sich dieses Ergebnis in Hinblick auf die Methode zur Bestimmung der formativen Unterindizes als stabil erweist. Die Hypothese H4, in der ein hoher Zusammenhang zwischen beiden LVn angenommen wurde, findet somit Bestätigung. Offensichtlich deckt das Set der 46 formativen ufosV2f-Items das Usability-Globalurteil der ufosV2r-Skala sehr gut ab. Es kann folglich von einer inhaltlich validen Abbildung des Konstruktes durch die ufosV2f-Skala ausgegangen werden.

Aufgrund der hohen Korrelationen zwischen den formativen Indikatoren ließe sich annehmen, dass sich bei einem Ausschluss einzelner Items aus der ufosV2f-Skala ebenfalls ein hoher Zusammenhang zum Globalurteil der ufosV2r-Skala ergeben würde. Es stellt sich jedoch die Frage, ob es sinnvoll wäre, weitere Items aus der Skala zu eliminieren. Das Anwendungsziel der ufosV2f-Skala ist darin zu sehen, auf Grundlage der Ausprägung von Einzelitems Rückschlüsse auf die Usability der jeweils dahinter stehenden Gestaltungsmerkmale zu ziehen. Ein Ausschluss von Items würde allerdings zu einer Einschränkung der Evaluationsbreite der ufosV2f-Skala führen, was dem Anwendungsziel entgegenstünde. In Abschnitt 2.4.3.6 wurde auf den Trade-Off zwischen dem Aufwand, der mit der Beantwortung weiterer Items verbunden ist, und dem inhaltlichen Informationszugewinn hingewiesen. Der durch 46 formative Items erreichte Trade-Off wird als sinnvoll erachtet, da einerseits keine lange Dauer der Beantwortung zu erwarten ist und andererseits die wesentlichen Einzelaspekte der Usability von Online-Shops abgefragt werden.

#### 9.2.1.5 Kriteriumsvalidität der ufosV2-Skalen

Die konkurrente Kriteriumsvalidität der neuen ufosV2-Skalen ist Gegenstand der Hypothesen H5a und H5b. Darin wurde ein hoher Zusammenhang zwischen der ufosV2r- bzw. der ufosV2f-Skala und der Erfolgsgröße Kaufintention angenommen. In beiden berechneten PLS-Modellen erwiesen sich die Pfade zwischen der exogenen LV Usability und der endogenen LV Kaufintention als signifikant. Dabei fiel der Pfadkoeffizient im Fall der ufosV2r-Skala höher aus als für die ufosV2f-Skala. Die unterschiedliche Höhe dieser Parameterausprägungen wird am Ende dieses Abschnitts diskutiert.

Eine Einbindung der Variable Kaufentscheidung in ein PLS-Modell ist aufgrund des kategorialen Skalenniveaus dieser AV nicht möglich. Entsprechend wurde eine sequentielle MLR berechnet, um die Hypothesen H6a und H6b statistisch analysieren zu können (siehe Abschnitt 8.6.2). In diesen Hypothesen wurde ein substanzieller Einfluss der reflektiven Skala ufosV2r (H6a) bzw. der formativen Skala ufosV2f (H6b) auf die Erfolgsgröße

Kaufentscheidung angenommen. Es wurde ein konservatives Vorgehen gewählt: Im ersten Schritt der MLR dienten diverse Variablen zur Demographie, Vorerfahrung und Kontrolle als Prädiktoren; die Einbeziehung der ufosV2r- bzw. ufosV2f-Skalenwerte für beide Shops erfolgte erst im zweiten Analyseschritt (2a bzw. 2b). Durch das Hinzufügen sowohl der reflektiven ufosV2r- (Schritt 2a) als auch der formativen ufosV2f-Variablen (Schritt 2b) konnte eine signifikante Verbesserung der Vorhersagekraft des Modells erreicht werden. Auf Grundlage dieser Analysen bestätigen sich somit die Hypothesen H6a bzw. H6b, wonach die Skalen ufosV2r bzw. ufosV2f im Sinne einer prädiktiven Validität eine hohe Vorhersagekraft für die Kaufentscheidung besitzen.

Die Ausprägungen von drei Parametern lassen jedoch darauf schließen, dass die reflektive ufosV2-Skala eine leicht bessere Vorhersage ermöglicht als die formative: (a) fiel das Inkrement des  $\chi^2$ -Wertes im globalen Likelihood-Quotiententest höher aus, (b) war der Zuwachs aufgeklärter Varianz in der AV größer (*Nagelkerke*  $R^2=0,28$  für ufosV2r; *Nagelkerke*  $R^2=0,25$  für ufosV2f), und (c) erwies sich der Anteil korrekt klassifizierter Fälle mit 67,64% für die reflektive ufosV2-Skala als leicht höher als für die formative Skala mit 64,72%. Bemerkenswert ist, dass die Güte der Vorhersage durch die ufosV2r-Skala nicht wesentlich schlechter ausfällt als für die Skala zur Kaufintention (70,55%).

Wie dargestellt wurde, zeigte sich sowohl in der PLS-Analyse als auch in der sequentiellen MLR ein leicht geringerer Zusammenhang zum jeweiligen Kriterium für die formative ufosV2f-Skala als für die reflektive Skala ufosV2r. Dies könnte den Schluss nahe legen, dass ufosV2f eine geringere Kriteriumsvalidität besitzt als ufosV2r. Es gilt jedoch, das in Abschnitt 9.1.3 angesprochene Problem zu berücksichtigen, dass eine Bildung von Unterindizes der formativen ufosV2f-Items mittels einer HKA erfolgen musste, um die Multikollinearität zwischen den Indikatoren zu behandeln. Möglicherweise liegt die leicht geringere Kriteriumsvalidität der ufosV2f-Skala in dem Informationsverlust begründet, der aus der Zusammenfassung der Indikatoren zu Unterindizes resultiert. Es ist durchaus bedauerlich, dass kein verlässliches statistisches Analyseverfahren zur Verfügung steht, durch das eine Gewichtung der einzelnen Indikatoren bei bestehender Multikollinearität möglich wäre.

Insgesamt bestätigen die durchgeführten Analysen für beide Skalen des ufosV2-Fragebogens sowohl die konkurrente Kriteriumsvalidität in Bezug auf die Kaufintention als auch die prädiktive Validität in Bezug auf die Kaufentscheidung. Diese Ergebnisse besitzen wichtige Implikationen für die Relevanzeinschätzung der Maße. Zum einen erweisen sich die Ausprägungen der Skalen dahingehend als relevant, ob ein Online-Kunde eine Intention ausbildet, einen Einkauf im jeweiligen Shop zu tätigen. Zum anderen – und dies ist sicherlich höher zu bewerten – kann beiden Maßen eine Relevanz für das Kaufverhalten von Online-Kunden beigemessen werden. Für die praktische Anwendung der ufosV2-Skalen bedeutet

dies, dass sich verbesserte Skalenausprägungen mit hoher Wahrscheinlichkeit in einer erhöhten Anzahl getätigter Käufe widerspiegeln werden.

Es ließe sich kritisieren, dass die Ergebnisse der MLR als ein Nachweis der prädiktiven Validität gewertet werden, obwohl das Zeitintervall zwischen der Erfassung der Prädiktoren und der AV Kaufentscheidung im Versuch nicht sehr groß ausfiel (vgl. Bühner, 2004, S. 31). Einer solchen Kritik folgend stellen die MLR-Befunde vielmehr eine weitere Untermauerung der Übereinstimmungsvalidität (konkurrente Kriteriumsvalidität) dar. Es erscheint aus Sicht des Verfassers jedoch durchaus schlüssig, angesichts des gewählten Versuchsszenarios von einer *Prädiktion* der AV Kaufentscheidung und entsprechend einer Form der Vorhersagevalidität zu reden. Inhaltlich würde der Begriff *Übereinstimmung* nicht der Analyseform eines Choice-Modells (MLR) gerecht werden, bei dem eine Vorhersage eines kategorialen, zeitlich versetzten Wahlverhaltens auf Grundlage von Fragebogendaten erfolgte.

#### 9.2.1.6 Nomologische Validität der ufosV2-Skalen

Die Hypothesen H7, H8 und H9 beziehen sich auf die Zusammenhänge der ufosV2-Skalen zu den Konstrukten innerhalb des konzeptuellen Erfolgsmodells. Die Überprüfung dieser Annahmen erfolgte mit Hilfe des PLS-Strukturgleichungsverfahrens (siehe Abschnitt 8.5.2).

In den Hypothesen H7a und H7b wurde ein Einfluss der Skalen ufosV2r und ufosV2f auf das Vertrauen in den Anbieter auf Seiten der Online-Kunden angenommen. In beiden PLS-Analysen zum konzeptuellen Erfolgsmodell bestätigt sich dieser Zusammenhang in Form sehr hoher, signifikanter Pfadkoeffizienten. Dieses Ergebnis steht im Einklang mit diversen Studien, die die Vertrauenswirksamkeit der Usability im Online-Shopping bereits empirisch untermauert haben (siehe Abschnitt 2.7.1). Aufgrund der Tatsache, dass der Analyse lediglich Querschnittsdaten zugrunde liegen, könnten Zweifel hinsichtlich der angenommenen Kausalrichtung zwischen den Konstrukten geäußert werden. Allerdings erscheint es wenig schlüssig, dass sich das Vertrauen in den Anbieter auf die wahrgenommene Usability auswirkt.

Die Annahme eines positiven Einflusses der Ästhetik auf die ufosV2-Skalen wurde in den Hypothesen H9a und H9b getroffen. H9a wird durch einen signifikanten Pfadkoeffizienten zwischen der Ästhetik und der Usability-LV ufosV2r bestätigt. Der Pfad Ästhetik → ufosV2f zu H9b findet ebenfalls empirische Bestätigung, wobei der  $\beta$ -Koeffizient höher ausfiel als für die reflektive Skala. Die sehr hohen Parameterausprägungen weisen auf einen sehr engen Zusammenhang zwischen den Konstrukten hin. Dieser deutete sich bereits in der berechneten EFA an (siehe Abschnitt 9.2.1.2). In den meisten der in Abschnitt 2.7.3 genannten Studien wird unterstellt, die Ästhetik beeinflusse die wahrgenommene Usability. Entsprechend wurde diese Richtung der Kausalität zwischen beiden LVn auch in der vorliegenden Arbeit im Rahmen der PLS-Analysen angenommen. Aufgrund des Vorliegens von Querschnittsdaten,

kann die Signifikanz des Zusammenhangs jedoch nicht als ein Beleg dieser Richtungsannahme gewertet werden.

Die Hypothese H8 zur Beziehung zwischen der Usability und dem Einkaufsvergnügen konnte nicht untersucht werden, da die Skala zum Einkaufsvergnügen keine ausreichende faktorielle Validität gegenüber den anderen reflektiven Skalen aufwies. Zusammenfassend finden abgesehen von H8 alle Hypothesen des konzeptuellen Erfolgsmodells, die die LV Usability betreffen, empirische Bestätigung. Diese Ergebnisse sprechen für eine hohe nomologische Validität der beiden ufosV2-Maße: Beziehungen der Usability zu anderen kundennahen Konstrukten, die auf Grundlage zurückliegender Studien als empirisch weitgehend gesichert gelten können, finden auch bei Verwendung der ufosV2-Skalen zur Erfassung der Usability Bestätigung. Folglich erscheinen die Skalen geeignet, die Zusammenhänge zwischen der Usability und weiteren Konstrukten in empirischen Studien näher zu untersuchen.

### **9.2.2 Befunde zu den weiteren Hypothesen im konzeptuellen Erfolgsmodell**

Neben den Hypothesen des konzeptuellen Erfolgsmodells, die sich auf die LV Usability beziehen, wurden in den inhaltlichen Hypothesen H10 bis H20 Zusammenhänge zwischen den übrigen LVn des Modells angenommen. Eine Bestätigung dieser Annahmen in den PLS-Analysen würde den nomologischen Rahmen zur Validierung der ufosV2-Maße gewissermaßen stärken (siehe Abschnitt 8.5.2). Zur Einschätzung der ufosV2-Skalengüte, also der wesentlichen Ziele der Studie, sind sie allerdings nur von indirekter Relevanz. Daher erfolgt keine umfangreiche Diskussion der Einzelbefunde.

Die Hypothese H10 betrifft den Einfluss der Ästhetik auf die Kaufintention. Auf Grundlage signifikanter Pfadkoeffizienten in den PLS-Modellen wurde diese Annahme bestätigt. Für die folgenden sieben weiteren Hypothesen ergaben sich hingegen keine signifikanten Ergebnisse: H11 zum Einfluss des Vertrauens auf die Kaufintention, H12 zum Einfluss der wahrgenommenen Produktanzahl auf die Kaufintention, H14 zum Einfluss des wahrgenommenen Preises auf die Kaufintention, H15 zum Einfluss des Service auf die Kaufintention, H17 zum Einfluss der wahrgenommenen Produktanzahl auf das Vertrauen, H18 zum Einfluss der Bekanntheit auf das Vertrauen sowie H20 zum moderierenden Einfluss der Produktart auf den Zusammenhang zwischen der Ästhetik und der Kaufintention. Aufgrund der mangelnden faktoriellen Validität der Skala zum Einkaufsvergnügen konnten drei weitere Hypothesen nicht untersucht werden. Betroffen sind die Hypothesen H13 zum Einfluss des Einkaufsvergnügens auf die Kaufintention, H16 zum Einfluss der Ästhetik auf das Einkaufsvergnügen sowie H19 zum moderierenden Einfluss der Produktart auf den Zusammenhang zwischen dem Einkaufsvergnügen und der Kaufintention.

Es stellt sich die Frage, welche Erklärungen sich für die Insignifikanz der Hypothesen H11, H12, H14, H15, H17, H18 und H20 finden lassen. Im Fall der Hypothesen H14 (Einfluss wahrgenommener Preis auf Kaufintention) und H18 (Einfluss Bekanntheit auf Vertrauen)

könnte vermutet werden, dass die Messgüte der Single-Item-Skalen zum wahrgenommenen Preis und zur Bekanntheit nicht ausreichend hoch ausfiel (siehe Abschnitt 9.1.3). Unter Umständen ließe sich der insignifikante Einfluss des Service auf die Kaufintention (H15) darauf zurückführen, dass vier der ursprünglich sechs service-bezogenen Items aufgrund zu hoher Anteile fehlender Werte eliminiert werden mussten (siehe Abschnitt 9.1.3).

Erstaunlich ist, dass der Einfluss des Vertrauens auf die Kaufintention im Rahmen der PLS-Analysen nicht bestätigt wurde (H11). Dieser Zusammenhang wurde in der Vergangenheit durch diverse andere Studien untermauert (siehe Abschnitt 2.6.2). Wie der Korrelationsmatrix für die LVn der PLS-Analysen entnommen werden kann, besteht eine mittlere bis hohe Korrelation von  $r=0,47$  zwischen Vertrauen und Kaufintention. Eine mögliche Erklärung, dass sich die Pfadkoeffizienten im Modell trotzdem nicht als signifikant erwiesen, könnte in der Beziehung zwischen Vertrauen und Usability vermutet werden. Die Usability erwies sich in den Analysen als ein signifikanter Prädiktor für das Vertrauen (siehe Abschnitt 9.2.1.6). Es ließe sich vermuten, dass Suppressoreffekte hierbei eine Rolle spielen (MacKinnon, Krull & Lockwood, 2000). Eine genauere Analyse dieser Beziehungen liegt jedoch eindeutig außerhalb des Fokus der vorliegenden Arbeit. Ähnliche Effekte könnten auch für die weiteren Zusammenhänge angenommen werden, die in den PLS-Modellen keine Bestätigung fanden (H12, H14).

Von größerem Interesse als die Hypothesen H10 bis H20 ist die Hypothese H21 zum Zusammenhang zwischen den beiden Erfolgsgrößen Kaufintention und -entscheidung, deren Untersuchung nicht im Rahmen der PLS-Modelle sondern mittels der MLR-Analyse erfolgte. Die Skalenwerte der Kaufintention wurden in Schritt 2c der sequentiellen MLR als Prädiktoren für die Kaufentscheidung herangezogen (siehe Abschnitt 8.6.2). Es ergab sich ein signifikantes Inkrement im globalen Likelihood-Quotiententest, ein großer Anteil zusätzlich aufgeklärter Varianz (*Nagelkerke*  $\Delta R^2=0,36$ ) und ein beachtlicher Gesamtanteil korrekt klassifizierter Fälle von 70,55%. Die Hypothese H21 kann folglich angenommen werden. Offensichtlich besitzt das Maß der Kaufintention eine hohe Relevanz für das Kaufverhalten der Teilnehmer am Quasi-Experiment. Dieses Ergebnis bestätigt nicht nur eine Kernannahme des konzeptuellen Erfolgsmodells, sondern unterstreicht auch die Güte beider Erfolgsgrößen (siehe Abschnitt 9.1.3).



### 9.3 Zusammenfassende Bewertung der Ergebnisse zu ufosV2

Im Folgenden soll ein Resümee zur Entwicklung des ufosV2-Fragebogens gezogen werden. Insgesamt können die Ziele der Untersuchung als erreicht angesehen werden. Ausgehend vom ufosV1-Fragebogen wurden sowohl eine formative als auch eine reflektive Skala zur Erfassung der Usability von Online-Shops aus Nutzersicht entwickelt. Die reflektive Skala ufosV2r besteht aus acht Items und ermöglicht die Erfassung eines Globalurteils des Konstruktes. Die formative Skala ufosV2f setzt sich aus 46 Items zusammen und greift die wesentlichen Facetten der Usability von Online-Shops auf Ebene der Einzelitems auf.

Alle untersuchten Gütekriterien der ufosV2-Skalen finden empirische Bestätigung. Die reflektive Skala ufosV2r erweist im Sinne der internen Konsistenz als hoch reliabel. Sie ist zudem in der Lage, zwischen verschiedenen Online-Shops zu differenzieren. Zwischen der formativen ufosV2f- und der reflektiven ufosV2r-Skala besteht trotz des sehr unterschiedlichen Abstraktionsniveaus der Items ein äußerst hoher Zusammenhang, was als ein Hinweis auf die Inhaltsvalidität der Skalen gewertet werden kann. Die Kriteriumsvalidität beider Skalen wird durch die Untersuchungsergebnisse in zweierlei Hinsicht untermauert. Zum einen bestätigt sich in einem komplexen PLS-Modell, dass für beide ufosV2-Maße starke Beziehungen zur Kaufintention bestehen. Diese Ergebnisse weisen auf eine hohe konkurrente Kriteriumsvalidität hin. Mit Hilfe eines Choice-Modells (MLR) konnte zudem gezeigt werden, dass beide ufosV2-Skalen eine hohe Relevanz für das Online-Kaufverhalten besitzen, da sie eine gute Vorhersage der Kaufentscheidung ermöglichen. Entsprechend kann auch die prädiktive Validität als hoch eingestuft werden. Weiterhin bestätigen sich die angenommenen Zusammenhänge beider Usability-Skalen zu den Erfolgsgrößen Vertrauen und Ästhetik, was die nomologische Validität des ufosV2-Instruments stützt. Die faktorielle Validität der reflektiven Skala ufosV2r kann als gegeben angesehen werden. In einer EFA sowie einer CFA erfährt dieses Kriterium zwar nur eingeschränkte Bestätigung, was jedoch mit hoher Wahrscheinlichkeit auf Voraussetzungsverletzungen zur Berechnung dieser Analysen zurückzuführen ist.

Eine gewisse Schwächung der Ergebnisse ließe sich darin sehen, dass im konzeptuellen Erfolgsmodell zwar alle Hypothesen bestätigt wurden, die die Usability betreffen, diverse weitere Zusammenhangsannahmen zu anderen Erfolgsfaktoren jedoch keine empirische Unterstützung fanden. Als Erklärung können zum einen Einschränkungen der Messgüte bei der Erfassung der Konstrukte wahrgenommener Preis, Bekanntheit sowie Service und zum anderen Suppressoreffekte vermutet werden.

Die vorliegende Untersuchung zeichnet sich gegenüber anderen Studien zum Bereich Online-Shopping qualitativ hinsichtlich mehrerer Aspekte aus. Eine Sichtung relevanter Beiträge zeigt, dass in den meisten Fällen Online-Befragungen durchgeführt werden, um verschiedene Qualitätsdimensionen von E-Commerce-Schnittstellen aus Sicht von Kunden abzufragen und anschließend statistisch miteinander in Beziehung zu setzen. Diese Art der Erhebung ist

jedoch mit einer Reihe potenzieller Probleme behaftet, in Folge derer die Datenqualität und damit die Aussagekraft der Ergebnisse leiden kann (vgl. Evans & Mathur, 2005; Ray & Tabor, 2003). Neben Online-Befragungsstudien finden sich diverse Untersuchungen, in denen Experimente durchgeführt wurden. Auch gegenüber den meisten dieser Studien weist das durchgeführte Laborexperiment Vorteile auf. Zum einen ist die Anzahl der untersuchten Teilnehmer als vergleichsweise hoch anzusehen. Zudem werden in den meisten Untersuchungen ausschließlich Studenten befragt, so dass starke Zweifel bzgl. der Übertragbarkeit auf die Population der Internetnutzer bestehen müssen. Im Vergleich zu anderen Studien bestehen in der vorliegenden Untersuchung dagegen nur geringe Unterschiede zwischen Stichprobe und Population.

Der wesentliche Vorteil des durchgeführten Quasi-Experiments liegt aus methodischer Sicht darin, dass tatsächliches Kaufverhalten operationalisiert und als AV in statistische Analysen einbezogen wurde. In Folge dessen fiel die Einkaufssituation im Vergleich zu vielen anderen Experimenten realistisch aus, auch wenn gewisse Unterschiede zu normalen Einkaufssituationen eingeräumt werden müssen. Zum anderen wurde durch die Erhebung der Kaufentscheidung eine wichtige Maßnahme zur Reduktion des Common Method Bias getroffen, da Prädiktoren und abhängige Variablen mit Hilfe unterschiedlicher Methoden erfasst wurden. Letztlich kann vermutet werden, dass die Ergebnisse (insbesondere verglichen mit anderen Validierungsstudien) nur in geringem Maße von der Auswahl der Benutzerschnittstellen abhängen, da eine hohe Anzahl von Online-Shops in die Untersuchung einbezogen wurde.

Es stellt sich die Frage, inwiefern sich Einschränkungen der genannten Versuchsergebnisse aufgrund des methodischen Vorgehens ergeben. Zur Gewährleistung einer hohen internen Validität der Studie wurde eine Reihe versuchsplanerischer Maßnahmen ergriffen. Hinweise auf Einschränkungen der internen Validität in Form unberücksichtigter Störeffekte ergaben sich jedoch nicht. Für eine hohe externe Validität der im Quasi-Experiment hergestellten Einkaufssituation spricht die Tatsache, dass ein sehr großer Teil der Probanden im Rahmen des Versuchs einen Einkauf tätigte. In welcher Hinsicht die Übertragbarkeit der Ergebnisse durch die gewissen Unterschiede zwischen der im Versuch operationalisierten und einer im Realkontext gefällten Kaufentscheidung eingeschränkt wird, bleibt unklar. Die Teilnehmerstichprobe kann zwar nicht als uneingeschränkt repräsentativ für die Population deutscher Internetnutzer angesehen werden, im Abgleich mit Befunden zu möglichen Einflüssen durch personenbezogene Merkmale ergeben sich jedoch keine klaren Konsequenzen bzgl. einer geminderten Aussagekraft der Ergebnisse (siehe Abschnitt 2.7.7.2). Befunde von Studien zu kulturellen Unterschieden legen Zweifel nahe, inwiefern sich die Ergebnisse auf andere Länder übertragen lassen, was allerdings unkritisch erscheint, da ein deutschsprachiger Fragebogen entwickelt wurde. Der verfolgte Validierungsansatz könnte letztlich dahingehend kritisiert werden, dass dem konzeptuellen Erfolgsmodell implizit die Annahme zugrunde liegt, Kaufentscheidungen von Online-Kunden würden auf Grundlage rationaler Erwägungen getroffen. Einige Autoren hegen diesbezüglich jedoch Zweifel (z.B. Shugan, 2006).

## 9.4 Erfolgswirksamkeit der Usability im Online-Shopping

Nachdem die Untersuchungsergebnisse in Hinblick auf die Güte der entwickelten Skalen bewertet wurden, lassen sich die Befunde auch in Bezug auf die Erfolgswirksamkeit der Usability im Rahmen des Online-Shoppings interpretieren. Die durchgeführten PLS- und MLR-Analysen bestätigen nicht nur die Validitäten der neu entwickelten Maße zur Erfassung der Gebrauchstauglichkeit, sondern untermauern darüber hinaus die gewichtige Rolle der Usability für die Handelsform Business-to-Consumer. Wie die hohen positiven Pfadkoeffizienten zwischen der Usability und der Kaufintention in den PLS-Modellen zum konzeptuellen Erfolgsmodell zeigen, wird die Kaufintention von Online-Kunden maßgeblich durch ihre Wahrnehmung der Usability-Ausprägung des jeweiligen Shops geprägt. Auch das tatsächliche Kaufverhalten steht offensichtlich in einem starken Zusammenhang dazu, inwieweit das Shopinterface als gebrauchstauglich bewertet wird, da die Usability-Skalen gemäß den Ergebnissen der MLR-Analyse eine hohe Vorhersagekraft auf die Kaufentscheidung besitzen.

Der Einfluss der Usability auf den Erfolg von Online-Anbietern wurde zuvor bereits in diversen anderen Studien empirisch bestätigt (siehe Abschnitt 2.5.2). Viele dieser Untersuchungen aus dem Bereich der Erfolgsfaktorenforschung lassen sich jedoch in methodischer Hinsicht kritisieren; in Abschnitt 2.1.5.3 wurden verschiedene Kritikpunkte dargestellt. Im Lichte dieser Kritik gewinnt die Aussagekraft der vorliegenden Arbeit weiteres Gewicht, da sich das methodische Vorgehen im Vergleich zu vielen anderen Untersuchungen durch die bereits geschilderten Vorzüge auszeichnet. Auch besitzt die Studie ein hohes Maß an Praxisrelevanz. So wird nicht nur aufgezeigt, dass die Usability gegenüber anderen kunden-nahen Erfolgsfaktoren als besonders wichtig anzusehen ist, sondern zugleich ein Instrument entwickelt, das dabei helfen kann, der Wichtigkeit dieser Variable in der Praxis Rechnung zu tragen.

## 9.5 Implikationen für Forschung und Praxis

Aus den Ergebnissen der Arbeit lassen sich einige Implikationen ableiten, die im Folgenden betrachtet werden sollen. Diese betreffen zum einen die Möglichkeiten und Vorteile einer Anwendung beider neu entwickelter ufosV2-Skalen. Zum zweiten erwachsen Konsequenzen für die fragebogenbasierte Erfassung der Usability im Allgemeinen. Schließlich können zum dritten einige Erwägungen zur theoretischen Fundierung des Konstruktes Usability getroffen werden.

### 9.5.1 Verwendung der ufosV2-Skalen

Durch die Erfüllung der untersuchten Gütekriterien der ufosV2-Skalen sind die Voraussetzungen für ihre Anwendung gegeben. Beide Skalen zeichnen sich durch unterschiedliche Eigenschaften aus, die ihre Eignung in verschiedenen Anwendungskontexten bestimmen. Da die reflektive Skala ufosV2r lediglich acht Items umfasst, ist ihr Einsatz mit wenig Aufwand verbunden. Die Kürze der Skala erweist sich als vorteilhaft, da vielfach nur eine begrenzte Anzahl an Items verwendet werden kann. Dies trifft beispielsweise auf Online-Befragungen zu, da auf Seiten von Internetnutzern zumeist keine hohe Bereitschaft zur Beantwortung vieler Items besteht.

Dabei ist die Aussage zur Ausprägung der Usability, die sich auf Grundlage von ufosV2r treffen lässt, summativer Natur, da ein Globalurteil von Seiten des Nutzers abgefragt wird. Rückschlüsse auf konkrete Usability-Probleme oder mangelhaft gestaltete Bereiche eines Online-Shops ermöglicht die Skala nicht. Häufig liegt das Ziel der Anwendung von Usability-Skalen jedoch nicht in der Identifikation konkreter Gestaltungsmängel. Der Einsatz von ufosV2r bietet sich daher beispielsweise an, wenn eine Qualitätsbeurteilung konkurrierender Shop-Prototypen, eine Evaluation im zeitlichen Verlauf oder ein Abgleich mit vorher definierten Benchmarks erfolgen soll (vgl. Floyd, Reisin & Schmidt, 1989; Kim, Shaw & Schneider, 2003). Auch im Rahmen wissenschaftlicher Untersuchungen, etwa in Studien im Bereich der Erfolgsfaktorenforschung, liegt das Interesse oft nicht in Aussagen zu spezifischen Aspekten der Usability, sondern eher in einer reliablen und validen Erfassung einer Gesamtausprägung des Konstruktes. Die reflektive ufosV2r-Skala erscheint daher sowohl für eine Anwendung im Rahmen praktischer Evaluationsmaßnahmen (summativ) als auch zu Forschungszwecken geeignet.

Im Gegensatz zur reflektiven Skala zeichnet sich die formative Skala ufosV2f dadurch aus, dass sie differenzierte Aussagen zur Usability einzelner Merkmale eines Online-Shops ermöglicht. Entsprechend bietet sich ihre Anwendung beispielsweise dann an, wenn es im Sinne eines Screenings gilt, Gestaltungsbereiche aufzuzeigen, in denen Designmängel bestehen (vgl. Barnes & Vidgen, 2003). Durch die Erstellung von Itemprofilen lassen sich sowohl verschiedene Versionen desselben als auch unterschiedliche Shops auf spezifischer Ebene miteinander vergleichen. Aufbauend auf den Ergebnissen der ufosV2f-Skala können

als mangelhaft identifizierte Gestaltungsbereiche mit Hilfe weiterer Usability-Evaluationsmethoden wie Nutzer-Tests oder der heuristischen Evaluation genauer adressiert werden (vgl. Dzida et al., 2000).

Die Anwendung von Usability-Fragebögen ist im Allgemeinen mit einer Reihe von Vor- und Nachteilen verbunden (siehe Abschnitt 2.4.5). Ein Nachteil dieser Verfahrensgruppe ist darin zu sehen, dass ihrer Anwendung im Rahmen formativer Evaluationsmaßnahmen Grenzen gesetzt sind. Entsprechend ist weder mit Hilfe der reflektiven noch der formativen Skala des ufosV2-Fragebogens die Ableitung konkreter Gestaltungsempfehlungen möglich (Hamborg & Gediga, 2006). Im Vergleich zu vielen anderen Usability-Fragebögen zeichnen sich die beiden ufosV2-Skalen jedoch dadurch aus, dass eine klare Spezifikation des jeweiligen Messmodells vorgenommen wurde. Die damit verbundenen Vorteile werden im folgenden Abschnitt zusammengefasst.

### 9.5.2 Formative und reflektive Erfassung der Usability in Fragebögen

Wie im Rahmen dieser Arbeit aufgezeigt wurde, ist die Unterscheidung zwischen formativen und reflektiven Messmodellen bei der Entwicklung von Usability-Fragebögen bislang ignoriert worden. In Folge dessen liegt eine Vielzahl an Skalen vor, die sowohl formative als auch reflektive Usability-Items beinhalten, jedoch auf einer reflektiven Skalenentwicklungsprozedur gemäß der klassischen Testtheorie basieren (vgl. Hamborg & Gediga, 2006). Im Sinne der Dichotomie formativ vs. reflektiv kann in diesen Fällen somit von einer Fehlspezifikation der zugrunde gelegten Messmodelle gesprochen werden (vgl. Jarvis et al., 2003).

In Abschnitt 2.8.5 wurde dargestellt, dass Fehlspezifikationen von Messmodellen zu deutlichen Einschränkungen der Skalensvalidität führen können (vgl. MacKenzie et al., 2005). Entsprechend ließe sich auch die Inhaltsvalidität vieler vorhandener Usability-Fragebögen anzweifeln. Eine mangelnde Ausprägung dieses Gütekriteriums kann einige ernstzunehmende Folgen nach sich ziehen. Es kommen Zweifel an der Gültigkeit der Ergebnisse von Forschungsarbeiten auf, in denen fehlspezifizierte Usability-Skalen Verwendung finden. Ferner ist zu erwarten, dass auf Grundlage derartiger Maße in der praktischen softwareergonomischen Tätigkeit falsche Schlussfolgerungen gezogen werden. So wäre beispielsweise der Extremfall denkbar, dass beim Vergleich zweier Prototypen eine Entscheidung zugunsten eines weniger benutzerfreundlichen Interfaces getroffen wird, da dies durch ein fehlspezifiziertes und entsprechend invalides Usability-Maß angezeigt wird. Über das tatsächliche Ausmaß der Validitätseinschränkungen fehlspezifizierter Usability-Fragebögen kann nur spekuliert werden. Es ist davon auszugehen, dass zwischen formativen Usability-Aspekten oft hohe Korrelationen bestehen; die Ergebnisse der vorliegenden Arbeit bestätigen diese Annahme. Mit anderen Worten fällt die Usability-Ausprägung von Benutzerschnittstellen häufig *insgesamt* graduell besser oder schlechter aus. Entsprechend könnte vermutet werden, dass die Nicht-Berücksichtigung einzelner Usability-Facetten nicht

generell schwerwiegende Folgen nach sich ziehen muss. Diese Annahme wird in anderen Studien durch die positiven Befunde zur Validität vorhandener Usability-Fragebögen gestützt, wie auch im Fall des ufosV1-Fragebogens (siehe Abschnitt 2.4.1.9; Konradt et al., 2003).

Ein Problem, das mit der Unterscheidung zwischen formativen und reflektiven Messmodellen verbunden ist, liegt darin, dass es sich vielfach als schwierig erweist, im Nachhinein zu bestimmen, ob ein Item eher der einen oder der anderen Messmodellart zuzuordnen ist (vgl. Eggert & Fassott, 2003; Jarvis et al., 2003). Die Erfahrungen bei der nachträglichen Einschätzung für die Items des ufosV1-Fragebogens belegen dies (siehe Abschnitt 4.2.2.2). Somit erscheint es umso wichtiger, zum Zeitpunkt der Itemformulierung die Unterschiede beider Formen von Messmodellen zu berücksichtigen, um einer mangelnden Konstruktvalidität der Skala vorzubeugen.

### 9.5.3 Theoretisches Fundament der Usability

Aus den Ergebnissen dieser Arbeit bzw. der Unterscheidung zwischen formativen und reflektiven Messmodellen ergeben sich weiterhin Implikationen für die theoretische Untermauerung des Konstrukts Usability. „Zu einem Fragebogen gehört ein zugrunde liegendes Modell“, formulieren Hassenzahl, Burmester und Koller (2003, S. 189). Spontan ist dieser Aussage natürlich zustimmen – denn auf welcher Grundlage kann sonst festlegt werden, *was* eine Skala messen soll, wenn nicht basierend auf einer Theorie? Als problematisch erweist sich, dass für das Konstrukt Gebrauchstauglichkeit bis dato keine schlüssige Theorie vorliegt (vgl. Diaper & Sanger, 2006; Gediga & Hamborg, 2002). Gelegentlich wird auf den Ansatz der mentalen Modelle als theoretisches Fundament der Usability verwiesen. Dessen großer Interpretationsspielraum lässt allerdings Zweifel bzgl. seiner tatsächlichen Eignung aufkommen. Vielfach ziehen die Entwickler von Usability-Fragebögen die in der DIN EN ISO 9241 festgeschriebenen Prinzipien der Dialoggestaltung als quasi-theoretische Grundlage des Skalenaufbaus heran. Zweifelsohne besitzen die Gestaltungsprinzipien der DIN-Norm eine hohe Augenscheinvalidität, da eine Benutzerschnittstelle, die nicht aufgabenangemessen, selbsterklärend, individualisierbar etc. gestaltet ist, nicht gebrauchstauglich sein kann. Wie bereits erwähnt wurde, gehen die DIN-Kriterien auf eine empirische Studie von Dzida et al. (1978) zurück. An dem methodischen Vorgehen von Dzida und Kollegen kann jedoch Kritik geübt werden. So wurde in der besagten Studie eine faktorenanalytische Identifikation von Usability-Faktoren vorgenommen, obwohl viele der in die Analyse einbezogenen Usability-Aspekte aus messtheoretischer Sicht als formative Indikatoren des Konstrukts anzusehen sind. Entsprechend ist der inhaltliche Aussagewert die extrahierten Dimensionen stark zu bezweifeln. Die empirische Grundlage der Prinzipien zur Dialoggestaltung wird damit in Frage gestellt.

Die Entwickler vieler Usability-Fragebögen machen aus der Not, dass keine klare Theorie der Usability besteht, sprichwörtlich eine Tugend, so auch im Fall des ufosV1-Fragebogens. Sie „erfinden“ kurzerhand ihr eigenes Usability-Modell, indem sie dem Beispiel von Dzida et al.

(1978) folgend zunächst eine große Anzahl an formativen und reflektiven Items sammeln, eine Datenerhebung mit einigen (oft jedoch wenigen) Benutzerschnittstellen durchführen und eine Faktorenanalyse zur Identifikation vermeintlich bestehender Usability-Dimensionen anwenden. Durch die Elimination von Items werden die Eindeutigkeit der Faktorenstruktur sowie die internen Konsistenz-Werte der extrahierten Skalen optimiert. Anschließend erfolgt eine inhaltliche Interpretation und entsprechende Benennung der Skalen. Dieses Vorgehen führt zu einer Identifikation von Faktoren bzw. Konstrukten, deren Validität aus messtheoretischer Sicht anzuzweifeln ist, da Methoden der klassischen Skalenentwicklung angewandt wurden, obwohl die einbezogenen Indikatoren nicht (ausschließlich) reflektiver Natur sind.

Ein daraus resultierendes Problem kann in einer Art „Konstrukt-Inflation“ gesehen werden: es werden immer wieder Studien veröffentlicht, die ihr eigenes Usability-bezogenes Konstrukt vorstellen. In der vorliegenden Arbeit wurde anhand der Konstrukte Perceived Ease of Use, Perceived Usefulness und Usability aufgezeigt, dass eine klare inhaltliche (geschweige denn theoretische) Abgrenzung vieler inhaltlich verwandter Konstrukte im Bereich Mensch-Maschine-Interaktion nicht gezogen werden kann (siehe Abschnitt 2.2.2.3). Ein hohes Maß an Intransparenz in der relevanten Literatur ist die Folge. Anhand der Benennungen von Konstrukten geht nur selten klar hervor, was die jeweiligen Autoren inhaltlich unter diesen Variablen verstehen. Daher ist man beim Lesen neuer Publikationen in der Regel genötigt, sich die genauen Operationalisierungen der Variablen vor Augen zu führen, um überhaupt eine inhaltliche Einordnung der Untersuchung an sich und ihrer Ergebnisse vornehmen zu können. Die Vergleichbarkeit der Befunde verschiedener Studien wird dadurch stark eingeschränkt. Der Nutzen vieler neu vorgestellter Konstrukte steht daher in Frage. Bei vielen Evaluationsskalen, insbesondere aus dem Bereich des Marketings, fällt zudem auf, dass keine klare Trennung zwischen verschiedenen Konstrukten vorgenommen wird (z.B. Wolfinger, & Gilly, 2003; Yoo & Donthu, 2001).

Trotz der angeführten Kritik einer „Konstrukt-Inflation“ werden auch in der vorliegenden Arbeit gewissermaßen zwei „neue“ Formen der Usability vorgestellt, nämlich die formative und die reflektive. Zwar handelt es sich bei der formativen und der reflektiven Usability im messtheoretischen Sinne um zwei verschiedene latente Variablen. In Bezug auf das dahinter stehende Konstrukt Usability unterscheiden sie sich jedoch nur hinsichtlich des inhaltlichen Abstraktionsniveaus der Items: Während eine reflektive Usability-Skala Items zum Globaleindruck des Nutzers abfragt, bezieht sich eine formative Skala auf spezifische Aspekte der Usability. Sie stellen also auf theoretischer Ebene keine neuen Formen der Usability dar. Die Unterscheidung liegt somit gewissermaßen im Betrachtungswinkel.

## 9.6 Weiterführende Fragestellungen

Ausgehend von der Interpretation der Befunde lassen sich einige Fragestellungen ableiten, die an die vorliegende Arbeit anknüpfen. Es könnten Erweiterungen des ufosV2-Instruments in Erwägung gezogen werden. Zudem sollte empirisch untersucht werden, inwiefern sich die im Rahmen der Arbeit geäußerte Vermutung bestätigt, dass fehlspezifizierte Usability-Fragebögen gegenüber korrekt spezifizierten Instrumenten eine geringere Validität aufweisen.

Bei der Untersuchung des Zusammenhangs beider ufosV2-Skalen trat das methodische Problem auf, dass sich die Bestimmung individueller Gewichte der einzelnen formativen ufosV2f-Indikatoren aufgrund eines hohen Maßes an Multikollinearität zwischen diesen manifesten Variablen als statistisch unmöglich erwies. Zwar konnte über den Umweg der Bildung von Unterindizes die wesentliche Fragestellung beantwortet werden, wie stark die Beziehung zwischen beiden ufos-Skalen ausfällt. Jedoch wäre darüber hinaus interessant, die Einflüsse der einzelnen formativen Usability-Indikatoren zu bestimmen, um Rückschlüsse auf ihre Relevanz in verschiedenen Nutzungskontexten ziehen zu können. Für die praktische Anwendung des Fragebogens würde eine Identifikation der Einflussstärken die Perspektive eröffnen, den Verbesserungsbedarf verschiedener Schwachstellen priorisieren zu können. Eine Möglichkeit der Relevanzeinschätzung der einzelnen Items läge darin, zusätzlich zur Bewertung der Ausprägung („stimme gar nicht zu“ bis „stimme voll“) für jedes Item eine Abfrage der Wichtigkeit vorzunehmen, wie z.B. im IsoMetrics<sup>L</sup> von Gediga et al. (1999; siehe Abschnitt 2.4.3.5). Entsprechend ließe sich untersuchen, inwiefern die Identifikation besonders gravierender Gestaltungsmängel durch die Einbindung einer solchen zusätzlichen Abfrage in die ufosV2f-Skala erleichtert würde.

Untersuchenswert erscheint weiterhin, inwiefern sich ein reflektives bzw. formatives Maß zur Erfassung der nutzerseitig erlebten Usability gegenüber fehlspezifizierten Skalen, bei denen nicht zwischen beiden Messmodellannahmen unterschieden wird, als eine validere Form der Erhebung erweist. Dies könnte beispielsweise durch eine Gegenüberstellung der prognostischen Skalengüte der verschiedenen konkurrierenden Skalen in Bezug auf ein geeignetes Kriterium erfolgen. Zeigt sich eine höhere Varianzaufklärung im Kriterium im Falle der reflektiven und formativen Skala als für fehlspezifizierte Skalen, so würde dies die Relevanz einer klaren Unterscheidung zwischen beiden Formen von Messmodellen bei der Entwicklung neuer Usability-Skalen unterstreichen.

Zu untersuchen wäre darüber hinaus, inwiefern eine Ableitung von Aussagen zur Usability einzelner Gestaltungsmerkmale einer Benutzerschnittstelle auf Basis von Ergebnissen eines formativen Fragebogenmaßes besser möglich ist als im Fall von reflektiven bzw. fehlspezifizierten Maßen.



## Zusammenfassung der Untersuchung

Die Erfassung latenter Konstrukte kann auf Grundlage zwei verschiedener Formen von Messmodellen erfolgen. Sie unterscheiden sich bzgl. der angenommenen Kausalitätsrichtung zwischen der latenten Variable und den Indikatoren (Bollen, 1989). Einer traditionellen Skalenentwicklungsprozedur liegt zumeist implizit eine reflektive Messmodellannahme zugrunde. Dabei stellen die Indikatoren verschiedene Folgen der latenten Variable dar (Churchill, 1979). Die Höhe der Korrelationen zwischen den reflektiven Indikatoren lässt auf die Skalenreliabilität schließen. In formativen Messmodellen dagegen wird die latente Variable durch die Indikatoren verursacht. Hohe Korrelationen zwischen formativen Indikatoren können auftreten, müssen es aber nicht. Folglich kann weder die Faktorenanalyse noch Cronbachs Alpha zur Beurteilung der Messgüte formativer Maße herangezogen werden. In der Vergangenheit wurden viele Skalen basierend auf einer reflektiven Messmodellannahme konstruiert, obwohl die Konstrukte eigentlich formative Messmodelle erfordern. Derartige Fehlspezifikationen können zu Einschränkungen der Skalengültigkeit führen.

Bei der Entwicklung von Usability-Fragebögen fand die Unterscheidung der beiden Messmodellarten bislang keine Berücksichtigung. Ziel der Arbeit war es, eine Fehlspezifikation im Messmodell des Usability Fragebogens für Online-Shops (ufos) von Konradt et al. (2003) auszuräumen. Es wurden sowohl eine formative Skala (ufosV2f, 46 Items) als auch eine reflektive Skala (ufosV2r, 8 Items) zur Erfassung der Usability von Online-Shops aus Kundensicht entwickelt. Um die Gütekriterien dieser Skalen zu prüfen, wurde ein Laborexperiment mit  $N=378$  Teilnehmern durchgeführt, die jeweils zwei von 35 Online-Shops kennen lernten. Neben der Usability und weiteren Erfolgsfaktoren des Bereichs Internet-Handel erfolgte die Erhebung einer Skala zur Kaufintention. Zudem wurde das Kaufverhalten in Form einer Kaufentscheidung erfasst, die an die Entlohnung zur Versuchsteilnahme gekoppelt war.

Mehr als zwei Drittel der Probanden nahmen eine Bestellung vor, die restlichen wählten einen Gutschein. Für die reflektive Skala ufosV2r ergab sich eine hohe interne Konsistenz von  $\alpha=0,94$ . Die Skala erwies sich darüber hinaus als faktoriell valide und ermöglichte eine Differenzierung zwischen mehreren Online-Shops. Gemäß einer Validierungsprozedur von Chin (1998) wurde die formative Usability-Skala in einem PLS-Zwei-Konstrukt-Strukturgleichungsmodell zur reflektiven in Beziehung gesetzt. Den Ergebnissen zufolge bildet die formative Skala ufosV2f ein valides Set an Indikatoren zur Erfassung des Konstruktes. In weiteren PLS-Modellen ergaben sich für beide Usability-Skalen positive Effekte auf die Kaufintention, was eine hohe konkurrente Kriteriumsvalidität anzeigt. Es bestätigten sich weiterhin Zusammenhänge zwischen den Usability-Skalen und den Konstrukten Vertrauen bzw. Ästhetik. In einer sequentiellen multinomialen Regression nahmen beide ufos-Skalen

zudem starken Einfluss auf die Kaufentscheidung der Teilnehmer; die Befunde dieser Analyse stützen somit die Annahme einer hohen prädiktiven Validität.

Insgesamt erfüllen sowohl die formative und als auch die reflektive ufos-Skala alle im Vorfeld geforderten Gütekriterien. Das gewählte Versuchsdesign gewährleistet eine hohe externe Validität der Ergebnisse. Das Vorliegen eines Common Method Bias ist weitgehend auszuschließen.

Die Unterscheidung zwischen formativen und reflektiven Messmodellen bringt wichtige Implikationen für die Forschung zur Mensch-Maschine-Interaktion mit sich. Zum einen geraten viele vorhandene Usability-Fragebögen hinsichtlich ihrer Inhaltsvalidität auf den Prüfstand, da ihnen fehlspezifizierte Messmodelle zugrunde liegen. Zudem erscheint eine theoretische Fundierung der Usability unangemessen, die auf einer Anwendung der Faktorenanalyse zur Aufschlüsselung des Konstruktes beruht. Unter anderem gehen auch die Prinzipien der Dialoggestaltung in Teil 10 der DIN EN ISO 9241 (1996) auf eine empirische Studie zurück, bei der eine faktorenanalytische Einteilung verschiedener Usability-Aspekte erfolgte (Dzida et al., 1978).

## Summary

### **Usability in Electronic Commerce: Development of a Questionnaire (UFOS-V2) under Consideration of Formative and Reflective Measurement Models**

For assessing latent constructs, two kinds of measurement models might be applied which differ in their underlying assumption of causal relationship between the latent variable and its manifest indicators (Bollen, 1989). Traditional scale development usually draws on reflective measurement models where the observed indicators are assumed to be caused by the latent variable (Churchill, 1979). As a consequence, high correlations between the indicators suggest a reflective scale to be reliable. In contrast, in case of a formative construct it is assumed that the indicators cause the latent variable. High correlations between formative indicators might occur, but are not generally expected. Hence, factor analysis and Cronbachs Alpha are inappropriate methods of evaluating a formative measure. Studies show that in case of many scales, the measurement model has been specified as being reflective unintentionally although the construct affords a formative model. As a potential consequence of this kind of misspecification, reduced scale validity might result.

Until now, the distinction between formative and reflective measurement has been completely ignored by developers of usability questionnaires. Aim of the study was to correct a misspecification of the Usability Questionnaire for Online Stores (UFOS) by Konradt et al. (2003). Both a formative (UFOS-V2f, 46 items) and a reflective usability scale (UFOS-V2r, 8 items) for the assessment of subjective online store usability from the customers' point of view were developed. In order to determine reliability and validity, 378 subjects participated in a laboratory experiment, each exploring two of a total of 35 online stores. Besides to usability and other success factors in online-shopping, the intention to buy was assessed as a dependent variable for both stores. In addition, actual purchase behaviour was observed by combining the subject reward with the decision to buy.

More than two thirds of the participants actually made a purchase, whereas the rest chose a voucher. A Cronbachs Alpha of  $\alpha=0.94$  indicates high internal consistency of the reflective measure UFOS-V2r. The scale proved to be factorial valid and enabled a differentiation between online stores. Following a validation procedure by Chin (1998), the formative usability scale was used as a predictor for the reflective usability measure in a two-construct PLS structural equation model. Results indicate that the formative scale UFOS-V2f forms a valid set of items for the assessment of subjective online store usability. In additional PLS-models, both UFOS scales showed positive effects on the intention to buy, suggesting high concurrent criterion validity. Relationships of UFOS scales with both trust and aesthetics were supported. Furthermore, a sequential multinomial regression analysis was performed on the decision to buy as outcome. Again, both UFOS-V2 measures provided good prediction of the criterion, confirming high predictive validity.

In summary, the formative and the reflective UFOS measures fulfil all required scale criteria. The experimental design of the study assured high external validity of the results. The probability of common method bias within the data is rather low.

The differentiation between formative and reflective measurement has important implications for human-computer-interaction research. First, the content validity of many existing usability questionnaires is challenged because their measurement models are specified wrongly. Moreover, a theoretical foundation of usability seems inappropriate that is based upon a factor analytical breakdown of the construct. Amongst others, the dialogue design criteria in part 10 of the DIN ISO EN 9241 (1996) go back to an empirical study in which factor analysis was applied for the purpose of classifying different usability aspects (Dzida et al., 1978).

## Literaturverzeichnis

- Adkisson, H. (2002). *Identifying De-Facto Standards for E-Commerce Web Sites*. Master-Arbeit an der University of Washington. Online im Internet: URL: [http://www.hpadkisson.com/papers/hpa\\_thesis\\_final2.pdf](http://www.hpadkisson.com/papers/hpa_thesis_final2.pdf) (Stand 27.11.06)
- AGOF (2006). *Berichtsband - Teil 1 zur Internet Facts 2006-I*. Arbeitsgemeinschaft Online-Forschung e.V.. Online im Internet: URL: <http://agof.de/studie.353.html> (Stand 27.11.06)
- Ahuja, J.S. & Webster, J. (2001). Perceived Disorientation: An Examination of a New Measure to Assess Web Design Effectiveness. *Interacting with Computers*, 14, 15-29.
- 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. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Psychological Bulletin*, 84(5), 888-918.
- Ajzen, I. & Driver, B.L. (1992). Application of the Theory of Planned Behavior to Leisure Choice. *Journal of Leisure Research*, 24, 207-224.
- 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(3), 38-53.
- Albers, S. & Hildebrandt, L. (2006). Methodische Probleme bei der Erfolgsfaktorenforschung – Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. *Zeitschrift für betriebswirtschaftliche Forschung*, 58, 2-33.
- Albers, S., Panten, G. & Schäfers, B. (Hrsg.) (2002). *Die eCommerce-Gewinner – Wie Unternehmen im Web profitabel wurden – 10 Erfolgsgeschichten aus erster Hand*. Frankfurt am Main: FAZ-Institut.
- Anandalingam, G., Day, R.W. & Raghavan, S. (2005). The Landscape of Electronic Market Design. *Management Science*, 51(3), 316-327.
- Anderson, E.W. & Sullivan, M.W. (1993). The Antecedents and Consequences of Consumer Satisfaction for Firms. *Marketing Science*, 12(2), 125-143.
- Anderson, E.W., Fornell, C. & Lehmann, D.R. (1994). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing*, 58(3), 53-66.

- Anderson, J.C. & Gerbing, D.W. (1991). Predicting the Performance of Measures in a Confirmatory Factor Analysis with Pretest Assessment of their Substantive Validities. *Journal of Applied Psychology*, 76, 732-740.
- Anderson, R.E. & Srinivasan, S.S. (2003). E-Satisfaction and E-Loyalty: A Contingency Framework. *Psychology & Marketing*, 20, 123-138.
- Andrews, D., Nonnecke, B. & Preece, J. (2003). Electronic Survey Methodology: A Case Study in Reaching Hard-to-Involve Internet Users. *International Journal of Human-Computer Interaction*, 16, 185-210.
- Andrews, J.C. & Durvasula, S. (1991). Suggestions for Manipulating and Measuring Involvement in Advertising Message Content. *Advances in Consumer Research*, 18, 194-201.
- Apliki (2006). *Test Online-Apotheken: Kunden vermissen Arzt und Apotheker*. Pressemitteilung vom 28. September. Online im Internet: URL: <http://www.openpr.de/news/101932.html> (Stand 27.11.06)
- Arbuckle, J.L. & Wothke, W. (1999). *AMOS 4.0 User's Guide*. Chicago, IL: Small Waters.
- Ariely, D. & Simonson, I. (2003). Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions. *Journal of Consumer Psychology*, 13, 113-123.
- Arnold, M.J. & Reynolds, K.E. (2003). Hedonic Shopping Motivations. *Journal of Retailing*, 79, 77-95.
- Assael, H. (2005). A Demographic and Psychographic Profile of Heavy Internet Users and Users by Type of Internet Usage. *Journal of Advertising Research*, 45(1), 93-123.
- Athanassopoulos, A.D. (2000). Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior. *Journal of Business Research*, 47(3), 191-207.
- Auger, P. (2005). The Impact of Interactivity and Design Sophistication on the Performance of Commercial Websites for Small Businesses. *Journal of Small Business Management*, 43(2), 119-137.
- Babin, B.J., Darden, W.R. & Griffin, M. (1994). Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. *Journal of Consumer Research*, 20, 644-656.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (10. Aufl.). Berlin: Springer.
- Bagozzi, R.P. (1983). Issues in the Application of Covariance Structure Analysis: A Further Comment. *The Journal of Consumer Research*, 9(4), 449-450.
- Bailey, J. & Pearson, S.W. (1983). Development of a Tool for Measuring and Analysing Computer User Satisfaction. *Management Science*, 29(5), 530-545.

- Bakos, J.Y. (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science*, 43(12), 1676-1692.
- Balabanis, G. & Reynolds, N.L. (2001). Consumer Attitudes Towards Multi-Channel Retailers' Web-Sites: The Role of Involvement, Brand Attitude, Internet Knowledge and Visit Duration. *Journal of Business Strategies*, 18(2), 105-131.
- Balazs, B. (2002). *Entwicklung eines Fragebogeninstrumentes zur nutzerbasierten Beurteilung der Usability von Online-Shops*. Unveröffentlichte Diplomarbeit, Humboldt-Universität, Berlin.
- Bapna, R., Goes, P. & Gupta, A. (2003). Analysis and Design of Business-to-Consumer Online Auctions. *Management Science*, 49(1), 85-101.
- Barclay, D., Thompson, R. & Higgins, C. (1995). The Partial Least Square (PLS) Approach to Causal Modeling, Personal Computer Adoption and Use as an Illustration. *Technology Studies*, 2(2), 285-324.
- Barnes, S.J. & Vidgen, R. (2001). An Evaluation of Cyber-Bookshops: The WebQual Method. *International Journal of Electronic Commerce*, 6(1), 11-31.
- Barnes, S.J. & Vidgen, R.T. (2002). An Integrative Approach to the Assessment of E-Commerce Quality. *Journal of Electronic Commerce Research*, 3, 114-127.
- Barnes, S.J. & Vidgen, R.T. (2003). Measuring Web Site Quality Improvements: A Case Study of the Forum on Strategic Management Knowledge Exchange. *Industrial Management & Data Systems*, 10(5/6), 297-309.
- Bauer, H.H. & Sauer, N.E. (2004). Die Erfolgsfaktorenforschung als schwarzes Loch? *Die Betriebswirtschaft (DBW)*, 64(4), 621-623.
- Baumol, W.E. & Ide, E.A. (1956). Variety in Retailing. *Management Science*, 3(1), 93-101.
- Becker, S.A. (2002). An Exploratory Study on Web Usability and the Internationalization of US E-Businesses. *Journal of Electronic Commerce Research*, 3, 265-278.
- Bei, L. & Chiao, Y. (2001). An Integrated Model for the Effects of Perceived Product, Perceived Service Quality, and Perceived Price Fairness on Consumers Satisfaction and Loyalty. *Journal of Customer Satisfaction, Dissatisfaction, and Complaining Behavior*, 14, 125-40.
- Beimel, J., Schindler, R. & Wandke, H. (1994). Do Human Factors Experts Accept the ISO 9241 Part 10-Dialog Principle-Standard? *Behaviour & Information Technology*, 13(4), 299-308.
- Belk, R.W. (1975). Situational Variables and Consumer Behavior. *Journal of Consumer Research*, 1, 157-164.

- Bellman, S. & Rossiter, R. (2004). The Website Schema. *Journal of Interactive Advertising*, 4(2). Online im Internet: URL: <http://www.jiad.org/vol4/no2/bellman/index.htm> (Stand 27.11.06)
- Bellman, S., Lohse, G.L. & Johnson, E.J. (1999). Predictors of Online Buying Behaviour. *Communications of the ACM*, 42(12), 32-38.
- Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
- Benyon, D. & Imaz, M. (1999). Metaphors and Models: Conceptual Foundations of Representations in Interactive Systems Development. *Human-Computer Interaction*, 14, 159-189.
- Berkowitz, L. (1978). Whatever Happened to the Frustration-Aggression Hypothesis? *American Behavioral Scientist*, 21(5), 691-708.
- Bernard, M. (2001). Developing Schemas for the Location of Common Web Objects. *Usability News 3.1*. Online im Internet: URL: [http://psychology.wichita.edu/surl/usabilitynews/3W/web\\_object.htm](http://psychology.wichita.edu/surl/usabilitynews/3W/web_object.htm) (Stand 27.11.06)
- Bernard, M. (2002). Examining User Expectations for the Location of Common E-Commerce. Web Objects. *Usability News 4.1*. Online im Internet: URL: [http://psychology.wichita.edu/surl/usabilitynews/41/web\\_object-ecom.htm](http://psychology.wichita.edu/surl/usabilitynews/41/web_object-ecom.htm) (Stand 27.11.06)
- Bernard, M. & Sheshadri, A. (2004). Preliminary Examination of Global Expectations of Users' Mental Models for E-Commerce Web Layouts. *Usability News 6.2*. Online im Internet: URL: [http://psychology.wichita.edu/surl/usabilitynews/62/web\\_object\\_international.htm](http://psychology.wichita.edu/surl/usabilitynews/62/web_object_international.htm) (Stand 27.11.06)
- Bevan, N. & Macleod, M. (1994). Usability Measurement in Context. *Behaviour & Information Technology*, 13(1/2), 132-145.
- Bhatt, G. (2004). Bringing Virtual Reality for Commercial Web Sites. *International Journal of Human Computer Studies*, 60(1), 1-15.
- Bhattacharjee, A. (2001). An Empirical Analysis of the Antecedents of Electronic Commerce Service Continuance. *Decision Support Systems*, 32, 201-214.
- Bhattacharjee, A. (2002). Individual Trust in Online Firms: Scale Development and Initial Test. *Journal of Management Information Systems*, 19, 211-241.
- Bias, R.G. (1994). The Puralistic Usability Walkthrough: Coordinated Empathies. In J. Nielsen & R.L. Mack (Eds.), *Usability Inspection Methods* (63-76). New York: Wiley.
- Bias, R.G. & Mayhew, D.J. (Eds.) (2005). *Cost Justifying Usability. An Update for the Internet Age*. San Francisco, CA: Morgan Kaufmann.



- Blackwell, R.D., Miniard, P.W. & Engel, F. (2001). *Consumer Behaviour* (4<sup>th</sup> ed.). Fort Worth, Tex: Harcourt College Publ.
- 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.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. 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.
- Brooke, J. (1996). SUS – a Quick and Dirty Usability Scale. In P.W. Jordan, B. Thomas, B.A. Weerdemeester & I.L. McClelland (Eds.), *Usability Evaluation in Industry* (189-194). London: Taylor & Francis.
- Brosius, F. (2004). *SPSS 12*. Bonn: Mitp.
- Brown, M. & Muchira, R. (2004). Investigating the Relationship between Internet Privacy Concerns and Online Purchase Behavior. *Journal of Electronic Commerce Research*, 5, 62-70.
- Bruner, G.C., II. & Kumar, A. (2000). Web Commercials and Advertising Hierarchy-of-Effects. *Journal of Advertising Research*, 40(1/2), 35-42.
- Bryant, A. & Colledge, B. (2002). Trust in Electronic Commerce Business Relationships. *Journal of Electronic Commerce Research*, 3(2), 32-39.
- Brynjolfsson, E. & Smith, M.D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 46, 563-585.
- Brynjolfsson, E., Smith, M.D. & Hu, Y. (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science*, 49, 1580-1596.
- Bucher, H.-J. & Jäckel, M. (2002). E-Business-Plattformen im Usability-Test - Methoden und Befunde. In H.-J. Bucher & M. Jäckel (Hrsg.). *Die Kommunikationsqualität von E-Business-Plattformen. Empirische Untersuchungen zu Usability und Vertrauen von Online-Angeboten. Schriftenreihe zum E-Business* (21-50). Trier: Competence Center E-Business.

- Bucher, H.-J., Büffel, S. & Wollscheid, S. (2002). Wie E-Business-Plattformen genutzt werden - qualitative Befunde zur Usability von Online-Angeboten. In H.-J. Bucher & M. Jäckel (Hrsg.). *Die Kommunikationsqualität von E-Business-Plattformen. Empirische Untersuchungen zu Usability und Vertrauen von Online-Angeboten. Schriftenreihe zum E-Business* (89-141). Trier: Competence Center E-Business.
- Büffel, S. (2002). Usability und Vertrauen bei der Nutzung von Internetangeboten. In H.-J. Bucher & M. Jäckel (Hrsg.). *Die Kommunikationsqualität von E-Business-Plattformen. Empirische Untersuchungen zu Usability und Vertrauen von Online-Angeboten. Schriftenreihe zum E-Business* (145-174). Trier: Competence Center E-Business.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Burke, R.R. (2002). Technology and the Customer Interface: What Consumers Want in the Physical and Virtual Store. *Journal of the Academy of Marketing Science*, 30, 411-432.
- Büttner, O.B., Schulz, S. & Silberer, G. (2006). Vertrauen, Risiko und Usability bei der Nutzung von Internet-Apotheken. In H. Bauer, M.M. Neumann & A. Schüle (Hrsg.), *Konsumentenvertrauen. Konzepte und Anwendungen für ein nachhaltiges Kundenbindungsmanagement* (355-366). München: Verlag Vahlen.
- Byrne, B.M. (2001). *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum.
- Calisir, F. & Calisir, F. (2004). The Relation of Interface Usability Characteristics, Perceived Usefulness, and Perceived Ease of Use to End-User Satisfaction with Enterprise Resource Planning (ERP) Systems. *Computers in Human Behavior*, 20, 505-515.
- Carroll, J.M. (1997). Human-Computer Interaction: Psychology as a Science of Design. *International Journal of Human-Computer Studies*, 46, 501-522.
- Cassel, C., Hackl, P. & Westlund, A.H. (1999). Robustness of Partial Least-Squares Method for Estimating Latent Variable Quality Structures. *Journal of Applied Statistics*, 26, 435-446.
- Cassel, J. (2003). Genderizing HCI. In J.A. Jacko & A.Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (401-412). Mahwah, N.J.: Lawrence Erlbaum.
- Caulton, D.A. (2001). Relaxing the Homogeneity Assumption in Usability Testing. *Behaviour & Information Technology*, 20(1), 1-7.
- Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J. & Shneiderman, B. (2004). Determining Causes and Severity of End-User Frustration. *International Journal of Human-Computer Interaction*, 17(3), 333-356.

- Chan, T. & Ahern, T.C. (1999). Targeting Motivation: Adapting Flow Theory to Instructional Design. *Journal of Educational Computing Research*, 21, 151-193.
- Chang, J. (2003). Online Shopping: Advantages over the Offline Alternative. *Journal of Internet Banking and Commerce*, 8(2). Online im Internet: URL: <http://www.arraydev.com/commerce/JIBC/0311-07.htm> (Stand 27.11.06)
- Chaparro, B. (2002). Top Ten Mistakes of Shopping Cart Design. *Usability News* 4.2. Online im Internet: URL: <http://psychology.wichita.edu/surl/usabilitynews/42/shoppingcart.htm> (Stand 27.11.06)
- Charitou, A., Georgiou, I. & Soteriou, A. (2003). E-Valuating the Role of E-Service Quality on the Pricing of Internet Stocks. *EFMA 2004 Basel Meetings Paper*. Online im Internet: URL: <http://ssrn.com/abstract=498062> (Stand 27.11.06)
- Chignell, M.H. (1990). A Taxonomy of User Interface Terminology. *ACM SIGCHI Bulletin*, 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 E. Soloway, D. Frye & S.B. Sheppard (Eds.), *CHI'88 Proceedings. Human Factors in Computing Systems* (213-218). New York: ACM Press.
- Chin, W.W. (1998). The Partial Least Squares Approach for Structural Equation Modelling. In G.A. Marcoulides (Ed.), *Modern Methods for Business Research* (295-336). Mahaw, N.J.: Lawrence Erlbaum.
- Cho, Y., Im, I., Fjermestad, J. & Hiltz, S.R. (2003). The Impact of Product Category on Customer Dissatisfaction in Cyberspace. *Business Process Management Journal*, 9(5), 635-651.
- Choi, B., Lee, C., Lee, H. & Subramani, M. (2004). Effects of Web Retail Service Quality and Product Categories on Consumer Behavior: A Research Model and Empirical Exploration. In *Proceedings of the 37th Hawaii International Conference on System Science IEEE Computer Society*. Online im Internet: URL: <http://csdl2.computer.org/comp/proceedings/hicss/2004/2056/07/205670177c.pdf> (Stand 27.11.06)
- Christophersen, T. (2002). *Usability im Kontext des Online-Shoppings: Einflüsse auf die Kaufentscheidung*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität, Kiel.
- 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: DUV.

- Christophersen, T. & Konradt, U. (2004). Wettbewerbsvorteile im Online-Shopping durch kundenorientierte Gestaltung der Benutzungsoberfläche. In S. Albers, V. Haßmann, F. Somm & T. Tomczak (Hrsg.), *Verkauf: Kundenmanagement, Vertriebssteuerung, E-Commerce. Loseblattwerk und Online-Dienst*. Gabler: Wiesbaden. Online im Internet: URL: <http://www.verkauf-aktuell.de> (Stand 27.11.06)
- Chun, S.-H. & Kim, J.-C. (2005). Pricing Strategies in B2C Electronic Commerce: Analytical and Empirical Approaches. *Decision Support Systems*, 40(2), 375-388.
- Churchill, G.A. Jr. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16, 64-73.
- Clawson, C. (1971). How Useful are 90-Day Purchase Probabilities? *Journal of Marketing Research*, 35(4), 43-47.
- Clement, M., Peters, K. & Preiß, F.J. (1999). Electronic Commerce. In S. Albers (Hrsg.), *eCommerce: Einstieg, Strategie und Umsetzung im Unternehmen* (49-64). Frankfurt am Main: FAZ-Institut für Management- Markt- und Medieninformationen.
- Clore, G.C. (1994). Why Emotions are Felt. In P. Ekman & R.J. Davidson (Eds.), *The Nature of Emotions: Fundamental Questions* (103-111). New York: Oxford University Press.
- Cockton, G., Lavery, D. & Woolrych, A. (2003). Inspection-based Evaluations. In J.A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (1118-1138). Mahwah, N.J.: Lawrence Erlbaum.
- Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3<sup>rd</sup> ed.). Mahwah: Lawrence Erlbaum.
- Cole, M., O'Keefe, R.M. & Siala, H. (2000). From the User Interface to the Consumer Interface. *Information Systems Frontiers*, 1, 349-361.
- Constantinides, E. (2004). Influencing the Online Consumer's Behavior: the Web Experience. *Internet Research: Electronic Networking Applications and Policy*, 14, 111-126.
- 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.
- Couper, M.P., Traugott, M.W. & Lamias, M.J. (2001). Web Survey Design and Administration. *Public Opinion Quarterly*, 65, 230-253.
- Coyle, J.R. & Gould, S.J. (2002). How Consumers Generate Clickstreams through Web Sites: An Empirical Investigation of Hypertext, Schema and Mapping Theoretical Explanations. *Journal of Interactive Advertising*, 2(2). Online im Internet: URL: <http://www.jiad.org/vol2/no2/coyle/index.htm> (Stand 27.11.06)

- Crisp, C.B., Jarvenpaa, S.L. & Todd, P.A. (1997). *Individual Differences and Internet Shopping Attitudes and Intentions*. Working Paper. Graduate School of Business, University of Texas. Online im Internet: URL: [http://ccwf.cc.utexas.edu/~crisp/Indiv\\_Shop.htm](http://ccwf.cc.utexas.edu/~crisp/Indiv_Shop.htm) (Stand 26.3.01)
- Cronbach, L.J. (1951). Croefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 12, 1-16.
- Cronbach, L.J. (1971). Test Validation. In R.L. Thorndike (Ed.), *Educational Measurement* (443-507). Washington, DC: American Council on Education.
- Cronin Jr., J.J., Brady, M.K. & Hult, G.T.M. (2000). Assessing the Effects of Quality, Value, and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments. *Journal of Retailing*, 76(2), 193-218.
- Curran, P.J., West, S.G. & Finch, J.F. (1996). The Robustness of Test Statistics to Non-normality and Specification Error in Confirmatory Factor Analysis. *Psychological Methods*, 1(1), 16-29.
- Danaher, P.J., Wilson, I.W. & Davis, R.A. (2003). A Comparison of Online and Offline Consumer Brand Loyalty. *Marketing Science*, 22(4), 461-476.
- Danielson, D.R. (2003). Transitional Volatility in Web Navigation. *IT & Society*, 1(3), 131-158.
- 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.
- Dawson, S., Bloch, P.-H. & Ridgway, N.-M. (1990). Shopping Motives, Emotional States, and Retail Outcomes. *Journal of Retailing*, 66, 408-427.
- Dayton, C.M. (1970). *The Design of Educational Experiments*. New York: McGraw-Hill.
- 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.
- DeLone, W.H. & McLean, E.R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information 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.

- DeRouvray, C. & Couper, M.P. (2002). Designing a Strategy for Reducing “No Opinion” Responses in Web-Based Surveys. *Social Science Computer Review*, 20(1), 3-9.
- Detlor, B., Sproule, S. & Gupta, C. (2003). Pre-Purchase Online Information Seeking: Search versus Browse. *Journal of Electronic Commerce Research*, 4(2), 72-84.
- Devaraj, S., Fan, M. & Kohli, R. (2002). Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics. *Information Systems Research*, 13, 316-333.
- Devaraj, S., Fan, M. & Kohli, R. (2003). Performance Impacts of Information Technology: Is Actual Usage the Missing Link? *Management Science*, 49(3), 273-289.
- Diamantopoulos, A. & Siguaw, J.A. (2002). *Formative vs. Reflective Indicators in Measure Development: Does the Choice of Indicators Matter?* Cornell School of Hotel Administration, Working Paper.
- Diamantopoulos, A. & Winklhofer, H.M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38, 269-277.
- Diaper, D. & Sanger, C. (2006). Tasks for and Tasks in Human-Computer Interaction. *Interacting with Computers, Special Theme Papers from Special Editorial Board Members*, 18(1), 117-138.
- Diller, S., Lin, L. & Tashjian, V. (2003). Evolving Role of Security, Privacy, And Trust In A Digitized World. In Julie A. Jacko & Andrew Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (1213-1225). Mahwah, N.J.: Lawrence Erlbaum.
- DIN 66284-8 (1988). *Bildschirmarbeitsplätze - Grundsätze ergonomischer Dialoggestaltung*. Berlin: Beuth.
- DIN EN ISO 9241-10 (1996). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 10: Grundsätze der Dialoggestaltung*. 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.
- DIN EN ISO 9241-12 (2000). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 12: Informationsdarstellung*. Berlin: Beuth.
- DIN EN ISO 9241-13 (2000). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 13: Benutzerführung*. Berlin: Beuth.

- DIN EN ISO 9241-14 (2000). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 14: Dialogführung mittels Menüs*. Berlin: Beuth.
- DIN EN ISO 9241-15 (1999). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 15: Dialogführung mittels Kommandosprache*. Berlin: Beuth.
- DIN EN ISO 9241-16 (2000). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 16: Dialogführung mittels direkter Manipulation*. Berlin: Beuth.
- DIN EN ISO 9241-17 (2000). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 18: Dialogführung mittels Bildschirmformularen*. Berlin: Beuth.
- DIN EN ISO 9241-2 (1992). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 2: Anforderungen an die Arbeitsaufgabe – Leitsätze*. Berlin: Beuth.
- Dinev, T. & Hart, P. (2005). Internet Privacy Concerns and Social Awareness as Determinants of Intention to Transact. *International Journal of Electronic Commerce*, 10(2), 7-29.
- Dion, K., Berscheid, E. & Walster, E. (1972). What is Beautiful is Good. *Journal of Personality and Social Psychology*, 24(3), 285-290.
- Döhl, M. (2001). Walkthrough vs. Videokonfrontation - Vergleich zweier Methoden zur formativen Software-Evaluation. In H. Oberquelle, R. Oppermann & R. Krause (Hrsg.), *Mensch & Computer 2001*, Berichte des German Chapter of the ACM, 55 (441-442). Stuttgart: B.G. Teubner.
- Doll, W.J. & Torkzadeh, G. (1988). The Measurement of Ender-User Computing Satisfaction. *MIS Quaterly*, 12(2), 259-274.
- Doll, W.J., Deng, X., Raghunathan, T.S., Torkzadeh, G. & Xia, W. (2004). The Meaning and Measurement of User Satisfaction: A Multi-group Invariance Analysis of the End-User Computing Satisfaction Instrument. *Journal of Management Information Systems*, 2(1), 227-262.
- Doney, P.M. & Cannon, J.P. (1997). An Examination of the Nature of Trust in the Buyer-Seller Relationship. *Journal of Marketing*, 51(2), 35-51.
- Donovan, R.J. & Rossiter, J.R. (1982). Store Atmosphere: An Environmental Psychology Approach. *Journal of Retailing*, 58, 34-57.
- Donthu, N. & Garcia, A. (1999). The Internet Shopper. *Journal of Advertising Research*, 3, 52-58.
- Drolet, A. & Morrison, D. (2001). Do We really Need Multiple-Item Measures in Service Research? *Journal of Service Research*, 3(3), 196-204.

- Dumas, J.S. (2003). User-Based Evaluations. In J.A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (1093-1117). Mahwah, N.J.: Lawrence Erlbaum.
- Dumas, J.S. & Redish, J.C. (1993). *A Practical Guide to Usability Testing*. Norwood, NJ: Ablex Pub. Corp.
- Dutke, S. (1994). *Mentale Modelle: Konstrukte des Wissens und Verstehens. Kognitionspsychologische Grundlagen für die Software-Ergonomie*. Göttingen: Hogrefe.
- Dzida, W. (1994). Qualitätssicherung durch software-ergonomische Normen. In E. Eberleh, H. Oberquelle & R. Oppermann (Hrsg.), *Einführung in die Software-Ergonomie* (373-406). Berlin: de Gruyter.
- Dzida, W. (1995). Standards for User-Interfaces. *Computer Standards & Interfaces*, 17, 89-97.
- 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)*. Göttingen: Hogrefe.
- Dzida, W., Herda, S. & Itzfeldt, W.D. (1978). User Perceived Quality of Interactive Systems. *IEEE Transactions on Software Engineering*, 4(4), 270-276.
- Dzida, W., Hofmann, B., Freitag R., Redtenbacher, W., Baggen, R., Geis, T., Beimel, J., Zurheiden, C., Hampe-Neteler, W., Hartwig, R. & Peters, H. (2000). Gebrauchstauglichkeit von Software – ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11. *Arbeitsschutz – Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Forschung, F 1693*. Dortmund.
- Dzida, W., Wiethoff, M. & Arnold, A.G. (1993). *ERGOGuide - The Quality Assurance Guide to Ergonomic Software*. Delft (NL), Sankt Augustin: Delft University of Technology, Gesellschaft für Mathematik und Datenverarbeitung mbh Bonn (GMD).
- Eagly, A.H., Ashmore, R.D., Makhijani, M.G. & Longo, L.C. (1991). What is Beautiful is Good, but ...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype. *Psychological Bulletin*, 110(1), 109-128.
- East, R. (1993). Investment Decisions and the Theory of Planned Behavior. *Journal of Economic Psychology*, 14, 337-375.
- 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.



- Edwards, J.R. & Bagozzi, R.P. (2000). On the Nature and Direction of Relationships between Constructs and Measures. *Psychological Methods*, 5, 155-174.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, 1-26.
- 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.
- Egger, F.N. (2003). *From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce*. PhD Thesis, Eindhoven University of Technology (The Netherlands).
- 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).
- Einwiller, S. (2003). When Reputation Engenders Trust: An Empirical Investigation in Business-to-Consumer Electronic Commerce. *Electronic Markets*, 13, 196-209.
- Eroglu, S.A., Machleit, K.A. & Davis, L.M. (2003). Empirical Testing of a Model of Online Store Atmospherics and Shopper Responses. *Psychology & Marketing*, 20, 139- 150.
- Evans, J.R. & Mathur, A. (2005). The Value of Online Surveys. *Internet Research: Electronic Networking Applications and Policy*, 15(2), 195-219.
- Everard, A. & Galletta, D.F. (2003). The Effect of Presentation Flaws in On-line Stores' Web Sites in Perceived Quality and Consumer Trust of the On-line Store and Intention to Purchase. In Janice DeGross (Ed.), *Proceedings of the Ninth Americas Conference on Information Systems* (2118-2122). Tampa, FL.
- Eysenck, H.J. & Eysenck, M.W (1987). *Persönlichkeit und Individualität. Ein naturwissenschaftliches Paradigma*. München-Weinheim: PVU.
- Faulkner, L. (2003). Beyond the Fiver-User Assumption: Benefits of Increased Sample Sizes in Usability Testing. *Behavior Research Methods, Instruments & Computers*, 35(3), 379-383.
- Featherman, M.S. & Pavlou, P.A. (2003). Predicting E-Services Adoption: A Perceived Risk Facets Perspective. *International Journal of Human-Computer Studies*, 59, 451-474.
- Ferber, R. & Piskie, R. (1965). Subjective Probabilities and Buying Intentions. *Review of Economics and Statistics*, 4(7), 322-325.
- Fickel, N. (2001). *Sequenzialregression*. Berlin: Vwf.

- Finn, A. & Kayande, U. (1997). Reliability Assessment and Optimization of Marketing Measurement. *Journal of Marketing Research*, 34, 262-275.
- Fiore, A.M., Kim, J. & Lee, H.-H. (2005). Effect of Image Interactivity Technology on Consumer Responses toward the Online Retailer. *Journal of Interactive Marketing*, 19(3), 38-53.
- Fishbein, M. & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Fletcher, D. & Flechter, M.E. (2003). Persistent Cart Design: A Review of Implementations. In C. Stephanidis & J. Jacko (Eds.), *Proceedings of the 10th International Conference HCI International 2003 on Human-Computer Interaction. Volume 2: Theory and Practice* (1188-1192). Mahwah, N.J.: Lawrence Erlbaum.
- Fornell, C. (1983). Issues in the Application of Covariance Structure Analysis: A Comment. *Journal of Consumer Research*, 9(4), 443-448.
- Fornell, C. & Bookstein, F.L. (1982). Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research*, 19, 440-452.
- Fornell, C. & Cha, J. (1994). Partial Least Squares. In R.P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (52-78). Oxford: Blackwell Publishers.
- Fornell, C. & Larcker, D.F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18, 39-50.
- Forsythe, S., Liu, C., Shannon, D. & Gardner, L.C. (2006). Development of a Scale to Measure the Perceived Benefits and Risks of Online Shopping. *Journal of Interactive Marketing*, 20(2), 55-75.
- Foucault, B.E. & Scheufele, D.A. (2002). Web vs Campus Store? Why Students Buy Textbooks Online. *Journal of Consumer Marketing*, 19, 409-423.
- Frijada, N.H. (1994). Varieties of affect: Emotions and Episodes, Moods, and Sentiments. In P. Ekman & R.J. Davidson (Eds.), *The Nature of Emotion* (59-67). New York: Oxford University Press.
- Fritz, W. (1990). Marketing – ein Schlüsselfaktor des Unternehmenserfolgs? *Marketing ZFP*, 2, 91-110.
- Fritz, W. (1995). *Marketing-Management und Unternehmenserfolg* (2. Aufl.). Stuttgart: Schäffer-Poeschel.
- Fritz, W. (2004). Die Erfolgsfaktorenforschung – ein Misserfolg? *Die Betriebswirtschaft (DBW)*, 64(5), 623-625.

- Frohlich, D.M. (2004). Beauty as a Design Prize. *Human-Computer Interaction*, 19(4), 359-366.
- Fu, L. & Salvendy, G. (2002). The Contribution of Apparent and Inherent Usability to a User's Satisfaction in a Searching and Browsing Task on the Web. *Ergonomics*, 45, 415-424.
- Fu, L., Salvendy, G. & Turley, L. (2002). Effectiveness of User Testing and Heuristic Evaluation as a Function of Performance Classification. *Behaviour & Information Technology*, 21(2), 137-143.
- Furnell, S. & Karweni, T. (1999). Security Implications of Electronic Commerce: a Survey of Consumers and Business. *Internet Research: Electronic Networking Applications and Policy*, 9, 372-382.
- Fürntratt, E. (1969). Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in der Faktorenanalyse psychologischer Daten. *Diagnostika*, 15, 62-75.
- Gammack, J. & Hodkinson, C. (2003). Virtual Reality, Involvement and the Consumer Interface. *Journal of End User Computing*, 15(4), 78-96.
- Garbarino, E. & Strahilevitz, M. (2004). Gender Differences in the Perceived Risk of Buying Online and the Effects of Receiving Site Recommendation. *Journal of Business Research*, 57, 768-75.
- Gardener, D.G., Cummings, L.L., Dunham, R.B. & Pierce, J.L. (1998). Single-Item versus Multiple-Item Measurement Scales: an Empirical Comparison. *Educational and Psychological Measurement*, 58, 898-915.
- Gediga, G. & Hamborg, K.-C. (1997). Heuristische Evaluation und IsoMetrics: Ein Vergleich. In R. Liskowsky, B.M. Velichkovsky & W. Wünschmann (Hrsg.), *Software-Ergonomie '97, Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (145-155). Stuttgart: Teubner.
- Gediga, G. & Hamborg, K.-C. (1999). IsoMetrics: An Usability Inventory Supporting Summative and Formative Evaluation of Software Systems. In H.-J. Bullinger & J. Ziegler (Hrsg.), *Human-Computer Interaction, Ergonomics and User Interfaces, Proceedings of HCI International '99*, Vol. 1 (1018-1022). Mahwah, N.J.: Lawrence Erlbaum.
- Gediga, G. & Hamborg, K.C. (2002). Ergonomische Evaluation von Software: Methoden und Modelle im Software-Entwicklungsprozess. *Zeitschrift für Psychologie*, 210(1), 40-57.
- Gediga, G., Hamborg, K.C. & Düntsch, I. (1999). The IsoMetrics Usability Inventory: an Operationalization of ISO 9241-10 Supporting Summative and Formative Evaluation of Software Systems. *Behaviour & Information Technology*, 18, 151-164.

- Gefen, D. (2002). Customer Loyalty in E-Commerce. *Journal of the Association for Information Systems*, 3, 27-51.
- Gefen, D. & Straub, D. (2000). The Relative Importance of Perceived Ease of Use in IS Adoption: A study of E-Commerce Adoption. *Journal of the Association for Information Systems*, 1(8), 1-30.
- Gefen, D. & Straub, D. (2003). Managing User Trust in B2C e-Services. *e-Service Journal*, 2(2), 7-24.
- Gefen, D., Karahanna, E. & Straub, D. (2003). Inexperience and Experience With Online Shops: The Importance of TAM and Trust. *Transactions on Engineering Management*, 50(3), 307-321.
- Gefen, D., Straub, D.W. & Boudreau, M.-C. (2000). Structural Equation Models and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, 4(7), 1-79.
- Gehrke, D. & Turban, E. (1999). Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness. In *Proceedings of the 32nd Hawaii International Conference on System Sciences IEEE*. Maui, HI, USA.
- George, J.F. (2004). The Theory of Planned Behavior and Internet Purchasing. *Internet Research: Electronic Networking Applications and Policy*, 14, 198-212.
- Gerbing, D.W. & Anderson, J.C. (1984). On the Meaning of within-Factor Correlated Measurement Errors. *The Journal of Consumer Research*, 11(1), 572-580.
- Gerbing, D.W. & Anderson, J.C. (1993). Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models* (40-65). Newbury Park: Sage.
- Ghani, J.A., Supnick, R. & Rooney, P. (1991). The Experience of Flow in Computer-Mediated and in Face-to-Face Groups. In J.I. DeGross, I. Benbasat, G. DeSanctis & C.M. Beath (Eds.), *Proceedings of the 12th International Conference on Information Systems* (229-237). New York.
- Gladstone, K., Rundle, C. & Alexander, T. (2002). Accessibility and Usability of Ecommerce Systems. In K. Miesenberger, J. Klaus & W. Zagler (Eds.), *Computer Helping People with Special Needs: 8th International Conference, ICCHP 2002*, July, Linz, Austria.
- Goonetilleke, R.S., Duffy, C. & Jacques, D. (2001). Customer Needs in Web Based Interaction: A Macro View of Usability. In M.J. Smith, G. Salvendy, D. Harris & R.J. Koubek (Eds.), *Proceedings of the HCI International 2001. Usability Evaluation and Interface design*. New Orleans, Louisiana, USA.

- Gordon, M.E., McKeage, K. & Fox, M.A. (1998). Relationship Marketing Effectiveness: The Role of Involvement. *Psychology & Marketing*, 15(5), 443-459.
- 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: DUV.
- 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.
- Graeff, T.R. (2003). Exploring Consumers' Answers to Survey Questions: Are Uninformed Responses Truly Uninformed? *Psychology & Marketing*, 20, 643-667.
- Graff, M. (2005). Individual Differences in Hypertext Browsing Strategies. *Behaviour & Information Technology*, 24(2), 93-99.
- Graham, J.W., Cumsille, P.E. & Elek-Fisk, E. (2002). Methods for Handling Missing Data. In J.A. Schinka & W.F. Velicer (Eds.), *Research Methods in Psychology*. Volume 2 of Handbook of Psychology (87-114). New York: John Wiley & Sons.
- Greving, B. (2006). Messen und Skalieren von Sachverhalten. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (73-88). Wiesbaden: DUV.
- Grewal, R. & Cote, J.A. & Baumgartner, H. (2004). Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. *Marketing Science*, 23(4), 519-529.
- Grice, J.W. & Harris, R.J. (1998). A Comparison of Regression and Loading Weights for the Computation of Factor Scores. *Multivariate Behavioral Research*, 33(2), 221-247.
- Grünig, R., Heckner, F. & Zeus, A. (1996). Methoden zur Identifikation strategischer Erfolgsfaktoren. *Die Unternehmung*, 50(1), 3-12.
- Hacker, W. & Richter, P. (1980). *Psychische Fehlbeanspruchung: Psychische Ermüdung, Monotonie, Sättigung und Streß*. Berlin: Deutscher Verlag der Wissenschaften.
- Haenecke, H. (2002). Methodenorientierte Systematisierung der Kritik an der Erfolgsfaktorenforschung. *Zeitschrift für Betriebswirtschaft*, 2, 165-183.
- Hagel III., J. & Rayport, J.F. (1997). The Coming Battle for Customer Information. *Harvard Business Review*, 75(1), 53-65.
- Hall, R.H. & Hanna, P. (2004). The Impact of Web Page Text-Background Colour Combinations on Readability, Retention, Aesthetics and Behavioural Intention. *Behaviour & Information Technology*, 23(3), 183-195.

- Hamborg, K.-C. (2002). Gestaltungsunterstützende Evaluation von Software: Zur Effektivität und Effizienz des IsoMetrics (L) Verfahrens. In M. Herczeg, W. Prinz & H. Oberquelle (Hrsg.), *Mensch und Computer 2002: Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten* (303-312). Stuttgart: B.G. Teubner.
- 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)*. Göttingen: Hogrefe.
- Hamborg, K.-C. & Greif, S. (1999). Heterarchische Aufgabenanalyse. In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (147-177). Zürich: VDF.
- 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.
- Hansen, G. (1987). Multikollinearität und Prognosefehler. *Jahrbücher für Nationalökonomie und Statistik*, 203, 357-370.
- 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> (Stand 27.11.06)
- Harris, L.C. & Goode, M.M.H. (2004). The Four Levels of Loyalty and the Pivotal Role of Trust: a Study of Online Service Dynamics. *Journal of Retailing*, 80, 139-158.
- Hassanein, K. & Head, M. (2005). The Impact of Infusing Social Presence in the Web Interface: An Investigation across Product Types. *International Journal of Electronic Commerce*, 10(2), 31-55.
- Hassenzahl, M. (2004a). Beautiful Objects as an Extension of the Self: A Reply. *Human-Computer Interaction*, 19(4), 377-386.
- Hassenzahl, M. (2004b). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19(4), 319-349.
- 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.
- Hassenzahl, M., Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Hrsg.), *Mensch & Computer 2003. Interaktion in Bewegung* (187-196). Stuttgart: B.G. Teubner.

- Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000). Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. In *Proceedings of the Chi '00 Conference on Human Factors in Computing Systems* (201-208). Den Haag: ACM Press.
- Hauptverband des Deutschen Einzelhandels (2005). *E-Commerce-Umsatz 2006: HDE erwartet 16,3 Milliarden Euro*. Pressemitteilung vom 7. Dezember. Online im Internet: URL: <http://www.hde.de/servlet/PB/menu/1056964/index.html> (Stand 03.05.06)
- Hawes, J.M., Mast, K.E. & Swan, J.E. (1989). Trust Earning Perceptions of Sellers and Buyers. *Journal of Personal Selling & Sales Management*, 9(1), 1-8.
- Hawkins, D.I. & Coney, K.A. (1981). Uniformed Response Error in Survey Research. *Journal of Marketing Research*, 18(3), 370-374.
- Heck, R.H. (1998). Factor Analysis: Exploratory and Confirmatory Approaches. In G.A. Marcoulides (Ed.), *Modern Methods for Business Research* (177-215). Mahaw, N.J.: Lawrence Erlbaum.
- Helander, M.G. & Khalid, H.M. (2000). Modeling the Customer in Electronic Commerce. *Applied Ergonomics*, 31, 609-619.
- Henderson, R. & Divett, M.J. (2003). Perceived Usefulness, Ease of Use and Electronic Supermarket Use. *International Journal of Human Computer Studies*, 59, 383-395.
- Herrmann, A., Huber, F. & Kressmann, F. (2006). Varianz- und kovarianzbasierte Strukturgleichungsmodelle – Ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung. *Zeitschrift für betriebswirtschaftliche Forschung*, 58, 34-66.
- Hermanns, A. & Sauter, M. (2001). Electronic Commerce – Grundlagen, Potentiale, Marktteilnehmer und Transaktionen. In A. Hermanns & M. Sauter (Hrsg.), *Management-Handbuch Electronic Commerce: Grundlagen, Strategien, Praxisbeispiele* (15-32). München: Vahlen.
- Hertzum, M. & Jacobsen, N.E. (2003). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1), 183-204.
- Hix, D. & Hartson, H.R. (1993). *Developing User Interfaces*. New York: John Wiley & Sons.
- Hoch, S.J., Bradlow, E.T. & Wansink, B. (1999). The Variety of an Assortment. *Marketing Science*, 18(4), 527-546.
- Hodkinson, C., Kiel, G. & McColl-Kennedy, J.R. (2000). Consumer Web Search Behaviour: Diagrammatic Illustration of Wayfinding on the Web. *International Journal of Human-Computer Studies*, 52, 805-830.

- Hoffman, D.L. & Novak, T.P. (1996). Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing*, 60(3), 50-69.
- Hoffman, D.L., Novak, T.P. & Peralta, M. (1998). Building Consumer Trust Online. *Communications of the ACM*, 42, 80-85.
- Holbrook, M.B. & Hirschman, E.C. (1982). The Experimental Aspects of Consumption: Consumer Fantasies, Feelings and Fun. *Journal of Consumer Research*, 9, 132-140.
- Homburg, C. & Baumgartner, H. (1995). Beurteilung von Kausalmodellen – Bestandsaufnahme und Anwendungsempfehlungen. *Marketing ZFP*, 17(3), 162-176.
- 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. *Marketing ZFP*, 18, 5-24.
- Homburg, C. & Krohmer, H. (2004). Die Fliegenpatsche als Instrument des wissenschaftlichen Dialogs. *Die Betriebswirtschaft (DBW)*, 64(5), 626-631.
- Homburg, C. & Pflesser, C. (2000). Konfirmatorische Faktorenanalyse. In A. Herrmann & C. Homburg (Hrsg.), *Marktforschung* (2. Aufl., 415-436). Wiesbaden: Gabler.
- Homburg, C., Hoyer, W.D. & Fassnacht, M. (2002). Service Orientation of a Retailer's Business Strategy: Dimensions, Antecedents, and Performance Outcomes. *Journal of Marketing*, 66(4), 86-101.
- Hong, W., Thong, J.L. & Tam, K.Y. (2004). Does Animation Attract Online Users' Attention? The Effects of Flash on Information Search Performance and Perceptions. *Information Systems Research*, 15(1), 60-86.
- Horan, P.M., DiStefano, C. & Motl, R.W. (2003). Wording Effects in Self-Esteem Scales: Methodological Artifact or Response Style? *Structural Equation Modeling*, 10(3), 435-455.
- Hornbæk, K. (2006). Current Practice in Measuring Usability: Challenges to Usability Studies and Research. *International Journal of Human-Computer Studies*, 64(2), 79-102.
- Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression* (2<sup>nd</sup> ed.). New York: Wiley.
- Hradil, S. (2005). *Soziale Ungleichheit in Deutschland* (8. Aufl.). Wiesbaden: Vs.
- Hu, L. & Bentler, P.M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Structural Equation Modeling*, 3(4), 424-453.



- Hudlicka, E. (2003). To Feel or not to Feel: The Role of Affect in Human-Computer Interaction. *International Journal of Human-Computer Studies*, 59, 1-32.
- Iyer, G. & Pazgal, A. (2003). Internet Shopping Agents: Virtual Co-Location and Competition. *Marketing Science*, 22(1), 85-106.
- Jacko, J.A., Sears, A. & Borella, M.S. (2000). The Effect of Network Delay and Media on User Perceptions of Web Resources. *Behaviour & Information Technology*, 19, 447-439.
- Jahng, J.J., Jain, H. & Ramamurthy, K. (2002). Personality Traits and Effectiveness of Presentation of Product Information in E-Business Systems. *European Journal of Information Systems*, 11, 181-195.
- Jarvenpaa, S.L. & Tiller, E.H. (2001). Customer Trust in Virtual Environments: a Managerial Perspective. *Boston University Law Review*, 3, 664-685.
- Jarvenpaa, S.L. & Todd, P.A. (1997). Consumer Reactions to Electronic Shopping on the World Wide Web. *International Journal of Electronic Commerce*, 1(2), 59-88.
- Jarvenpaa, S.L. & Tractinsky, N. (1999). Consumer Trust in an Internet Store: A Cross-Cultural Validation. *Journal of Computer Mediated Communication*, 5(2). Online im Internet: URL: <http://jcmc.indiana.edu/vol5/issue2/jarvenpaa.html> (Stand 27.11.06)
- Jarvenpaa, S.L., Tractinsky, N. & Vitale, M. (2000). Consumer Trust in an Internet Store. *Information Technology and Management, Special Issue on Electronic Commerce*, 1, 45-71.
- Jarvis, C.B., MacKenzie, S.B. & Podsakoff, P.M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecifications in Marketing and Consumer Research. *Journal of Consumer Research*, 30, 199-218.
- Jayawardhena, C., Wright, L.T. & Masterson, R. (2003). An Investigation of Online Consumer Purchasing. *Qualitative Market Research*, 6(1), 58-65.
- Jee, J. & Lee, W.-N. (2002). Antecedents and Consequences of Perceived Interactivity: An Exploratory Study. *Journal of Interactive Advertising*, 3(1). Online im Internet: URL: <http://www.jiad.org/vol3/no1/jee/index.htm> (Stand 27.11.06)
- Jeffries, R., Miller, J.R., Wharton, C. & Uyeda, K.M. (1991). User Interface Evaluation in the Real World: a Comparison of Four Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (119-124). New York: ACM.
- John, B.E. & Marks, S.J. (1997). Tracking the Effectiveness of Usability Evaluation Methods. *Behaviour & Information Technology*, 16(4/5), 188-202.
- Jordan, P.W. (1997). Human Factors for Pleasure in Product Use. *Applied Ergonomics*, 29, 25-33.

- Jöreskog, K. & Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Scientific Software International Inc.
- Kahn, B. & Lehmann, D.R. (1991). Modeling Choice among Assortments. *Journal of Retailing*, 67(3), 274-299.
- Kaiser, H.F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34, 111-117.
- Karat, C.-M., Karat, J., Vergo, J., Pinhanez, C., Riecken, D. & Cofino, T. (2002). That's Entertainment! Designing Streaming, Multimedia Web Experiences. *International Journal of Human-Computer Interaction*, 14, 369-384.
- Karvonen, K. (2000). The Beauty of Simplicity. In *Proceedings of the ACM Conference on Universal Usability (CUU 2000)*, November 16-17, 2000, Washington, DC.: ACM Press.
- Kau, A.K., Tang, Y.E. & Ghose, S. (2003). Typology of Online Shoppers. *Journal of Consumer Marketing*, 20, 139-156.
- Keeney, R.L. (1999). The Value of Internet Commerce to the Customer. *Management Science*, 45, 533-542.
- Khalifa, M. & Limayem, M. (2003). Virtual Extension: Drivers of Internet Shopping. *Communications of the ACM*, 46, 233-239.
- Kim, J. & Lee, J. (2002). Critical Design Factors for Successful E-Commerce Systems. *Behaviour & Information Technology*, 21, 185-199.
- Kim, J. & Moon, J.Y. (1998). Designing towards Emotional Usability in Customer Interfaces – Trustworthiness of Cyber-Banking System Interfaces. *Interacting with Computers*, 10, 1-29.
- 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, 899-940.
- Kim, J., Lee, J., Han, K. & Lee, M. (2002). Business as Buildings: Metrics for the Architectural Quality of Internet Businesses. *Information Systems Research*, 13(3), 239-254.
- 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., Williams, R. & Lee, Y. (2003). Attitudes Toward Online Shopping and Retail Website Quality: A Comparison of US and Korean Consumers. *Journal of International Consumer Marketing*, 16(1), 89-111.

- 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. (1987). The Computer User Satisfaction Inventory. *IEE Colloquium on Evaluation Techniques for Interactive System Design II*, London.
- Kirakowski, J. (1997). *The Use of Questionnaire Methods for Usability Assessment – Background Notes on the SUMI Questionnaire*. Online im Internet: URL: <http://sumi.ucc.ie/sumipapp.html> (Stand 27.11.06)
- Kirakowski, J. & Cierlik, B. (1998). Measuring the Usability of Web Sites. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Vol. 1* (424-428). Santa Monica, CA: HFES.
- Kirakowski, J. & Claridge, N. (2001). A Professional Tool for Evaluating Web Sites. *HM-HCI Conference*. Lille, France.
- Kirakowski, J. & Corbett, M. (1990). *Effective Methodology for the Study of HCI*. Amsterdam: North Holland.
- Kirakowski, J. & Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), 210-212.
- Kirakowski, J., Claridge, N. & Whitehand, R. (1998). Human Centered Measures of Success in Web Site Design. In *Proceedings of the 4th Conference on Human Factors & the Web*. Basking Ridge, NJ.
- Klein, L.R. (1998). Evaluation the Potential of Interactive Media through a New Lens: Search versus Experience Products. *Journal of Business Research*, 41, 195-203.
- Kleinginna, P.R., Jr. & Kleinginna, A.M. (1981). A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. *Motivation & Emotion*, 5(4), 345-379.
- Kline, R.B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- Kohavi, R. & Round, M. (2004). *Emetrics Summit 2004: Front Line Internet Analytics at Amazon.com*. Online im Internet: URL: <http://ai.stanford.edu/~ronnyk/emetricsAmazon.pdf> (Stand 27.11.06)
- Kohli, R., Devaraj, S. & Mahmood, A. (2004). Understanding Determinants of Online Consumer Satisfaction: A Decision Process Perspective. *Journal of Management Information Systems*, 21, 115-135.
- Konradt, U., Christophersen, T. & Schäffer-Külz, 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.
- Kotha, S., Rajgopal, S. & Venkatachalam, M. (2004). The Role of Online Buying Experience as a Competitive Advantage: Evidence from Third-Party Ratings for E-Commerce Firms. *Journal of Business*, 77(2), 109-133.
- Koufaris, M. (2002). Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. *Information Systems Research*, 13, 205-223.
- Koufaris, M. & Hampton-Sosa, W. (2004). The Development of Initial Trust in an Online Company by new Customers. *Information & Management*, 41, 377-397.
- Koufaris, M., Kambil, A. & LaBarbera, P.A. (2002). Consumer Behavior in Web-Based Commerce: An Empirical Study. *International Journal of Electronic Commerce*, 6, 115-139.
- Krafft, M., Götz, O., Liehr-Gobbers, K. (2005). Die Validierung von Strukturgleichungsmodellen mit Hilfe des Partial-Least-Squares (PLS)-Ansatzes. In F. Bliemel, A. Eggert, G. Fassott & J. Henseler (Hrsg.), *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele* (71-86). Schäffer-Poeschel: Stuttgart.
- Krone, A., Hamborg, K.C. & Gediga, G. (2002). Untersuchungen zu emotionalen Reaktionen bei Nutzungsfehlern in der Mensch-Computer Interaktion. *Zeitschrift für Arbeits- und Organisationspsychologie*. Online im Internet: URL: <http://www.people.freenet.de/gediga/fehler.pdf> (Stand 27.11.06)
- Kulviwat, S., Guo, C. & Engchanil, N. (2004). Determinants of Online Information Search: a Critical Review and Assessment. *Internet Research: Electronic Networking Applications and Policy*, 14, 245-253.
- Kumar, R.L., Smith, M.A. & Bannerjee, S. (2004). User Interface Features Influencing Overall Ease of Use and Personalization. *Information & Management*, 41, 289-302
- 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.
- Kwon, O.B, Kim, C.R. & Lee, E.J. (2002). Impact of Website Information Design Factors on Consumer Ratings of Web-Based Auction Sites. *Behaviour & Information Technology*, 21, 387-402.
- Lance, C.E., Butts, M.M. & Michels, L.C. (2006). The Sources of Four Commonly Reported Cutoff Criteria – What Did They Really Say? *Organizational Research Methods*, 9(2), 202-220.

- Lavery, D., Cockton, G. & Atkinson, M.P. (1997). Comparison of Evaluation Methods Using Structured Usability Problem Reports. *Behaviour & Information Technology*, 16, 246-266.
- Lavie, T. & Tractinsky, N. (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *International Journal of Human-Computer Studies*, 60, 269-298.
- Law, K. & Wong, C.-S. (1999). Multidimensional Constructs in Structural Equation Analysis: An Illustration Using the Job Perception and Job Satisfaction Constructs. *Journal of Management*, 25, 143-160.
- Lazar, J., Jones, A., Hackley, M. & Shneiderman, B. (2006). Severity and Impact of Computer User Frustration: A Comparison of Student and Workplace Users. *Interacting with Computers*, 18(2), 139-326.
- Lee, B.-K. & Lee, W.-N. (2004). The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment. *Psychology & Marketing*, 21, 159-183.
- Lee, J., Kim, J. & Moon, J.Y. (2000). What Makes Internet Users Visit Cyber Stores Again? Key Design Factors for Customer Loyalty. In T. Turner & G. Szwillus (Eds.), *Proceedings of the CHI 2000 Human Factors in Computing Systems*, April, The Hague, Amsterdam (305-312). New York: ACM Press.
- Lee, K.S. & Tan, S.J. (2003). E-retailing versus Physical Retailing: A Theoretical Model and Empirical Test of Consumer Choice. *Journal of Business Research*, 56(11), 877-885.
- Lee, M.K.O. & Turban, E. (2001). A Trust Model for Consumer Internet Shopping. *International Journal of Electronic Commerce*, 6(1), 75-91.
- Lee, S.-J., Lee, W.-N., Kim, H. & Stout, P.A. (2004). A Comparison of Objective Characteristics and User Perception of Web Sites. *Journal of Interactive Advertising*, 4(2). Online im Internet: URL: <http://www.jiad.org/vol4/no2/lee/index.htm> (Stand 27.11.06)
- Levin, A.M., Levin, I.P. & Weller, J.A. (2005). A Multi-Attribute Analysis of Preferences for Online and Offline Shopping: Differences across Products, Consumers, and Shopping Stages. *Journal of Electronic Commerce Research*, 6(4), 281-290.
- Levy, B.I. (1984). Research into the Psychological Meaning of Color. *American Journal of Art Therapy*, 23, 58-62.
- Lewis, C. (1982). Using the "Thinking Aloud" Method in Cognitive Interface Design. *IBM Research Reports RC 9265 (40713)*, IBM Thomas J. Watson Research Center. New York: Yorktown Heights.

- Lewis, C. & Wharton, C. (1997). Cognitive Walkthroughs. In M. Helander, T.K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2<sup>nd</sup> ed., 717-732). North-Holland: Elsevier.
- Lewis, J.R. (1991). Psychometric Evaluation of an After-Scenario Questionnaire for Computer Usability Studies: the ASQ. *SIGCHI Bull*, 23(1), 78-81.
- Lewis, J.R. (1992). Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In *Proceedings of the Human Factors Society 36<sup>th</sup> Annual Meeting* (1259-1263). Santa Monica, CA: Human Factors Society.
- 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: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7, 57-78.
- Lewis, J.R. (2002). Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human-Computer Interaction*, 14, 463-488.
- Lewis, M., Singh, V. & Fay, S. (2006). An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions. *Marketing Science*, 25(1), 51-64.
- Liang, T.P. & Lai, H.-J. (2002). Effect of Store Design on Consumer Purchases: an Empirical Study of On-line Bookstores. *Information & Management*, 39, 431-444.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Peltz PVU.
- Lightner, N.J. (2003). What Users Want in E-commerce Design: Effects of Age, Education and Income. *Ergonomics*, 46, 153-168.
- Lightner, N.J., Yenisey, M.M., Ozok, A.A. & Salvendy, G. (2002). Shopping Behaviour and Preferences in E-Commerce of Turkish and American University Students: Implications from Cross-Cultural Design. *Behaviour & Information Technology*, 21, 371-385.
- 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.
- Lindgaard, G. & Dudek, C. (2003). What is this Evasive Beast we Call User Satisfaction? *Interacting with Computers*, 15, 429-452.
- Liu, C. & Arnett, K.P. (2000). Exploring the Factors Associated with Web Site Success in the Context of Electronic Commerce. *Information & Management*, 38, 23-33.

- Liu, C., Marchewka, J.T., Lu, J. & Yu, C.-S. (2005). Beyond Concern – a Privacy-Trust-Behavioral Intention Model of Electronic Commerce. *Information & Management*, 42(2), 289-304.
- Lohmöller, J.-B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica.
- Lohse, G.L. & Spiller, P. (1998). Electronic Shopping: The Effect of Customer Interfaces on Traffic and Sales. *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> (Stand 14.09.04)
- 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* (3. Aufl.). Stuttgart: Enke.
- Lunneborg, C.E. (2000). *Data Analysis by Resampling: Concepts and Application*. Pacific Grove.
- Lwin, M.O. & Williams, J.D. (2006). Promises, Promises: How Consumers Respond to Warranties in Internet Retailing. *Journal of Consumer Affairs*, 40(2), 236-260.
- Lynch, P.D., Kent, R.J. & Srinivasan, S.S. (2001). The Global Internet Shopper: Evidence from Shopping Tasks in Twelve Countries. *Journal of Advertising Research*, 41(3), 15-23.
- MacCallum, R.C., Widaman, K.F., Zhang, S. & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods*, 4(11), 84-99.
- MacKenzie, S.B., Podsakoff, P.M. & Jarvis, C.B. (2005). The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology*, 90, 710-730.
- MacKinnon, D.P., Krull, J.L. & Lockwood, C.M. (2000). Equivalence of the Mediation, Confounding and Suppression Effect. *Prevention Science*, 1(4), 173-181.
- Magenat, P. (1999). *Shopping On The Internet: Usability of 9 Swiss E-Commerce Sites*. Online im Internet: URL: [http://www-3.ibm.com/ibm/easy/eou\\_ext.nsf/publish/878](http://www-3.ibm.com/ibm/easy/eou_ext.nsf/publish/878) (Stand 04.09.03)

- Maguire, M. & Pearce, L.-J. (2001). An Investigation into Attitudes to, and Experience of Internet Shopping. In C. Stephanidis (Ed.), *Universal Access in HCI: Towards an Information Society for All. Proceedings of the 1st International Conference on Universal Access in Human-Computer Interaction*, New Orleans, USA, August (1088-1092). Mahwah, N.J.: Lawrence Erlbaum.
- Mahmood, M.A., Bagchi, K. & Ford, T.C. (2004). On-line Shopping Behaviour: Cross-Country Empirical Research. *International Journal of Electronic Commerce*, 9(1), 9-30.
- March, J.G. & Sutton, R.I. (1997). Organization Performance as a Dependent Variable. *Organization Science*, 6, 698-706.
- Marsh, H.W., Hau, K.T., Balla, J.R. & Grayson, D. (1998). Is more ever too much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, 33(2), 181-220.
- Mathwick, C., Malhotra, N. & Rigdon, E. (2001). Experiential Value: Conceptualization, Measurement and Application in the Catalog and Internet Shopping Environment. *Journal of Retailing*, 77, 39-56.
- Mayhew, D.J. (2003). Requirements Specifications within The Usability Engineering Life Cycle. In J.A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (913-921). Mahwah, N.J.: Lawrence Erlbaum.
- McGaughey, R.E. & Mason, K.H. (1998). The Internet as a Marketing Tool. *Journal of Marketing Theory and Practice. Journal of Marketing Theory & Practice*, 6(3), 1-11.
- McKinney, V., Yoon, K. & Zahedi, F. (2002). The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. *Information Systems Research*, 13, 296-315.
- McKnight, D.H., Choudhury, V. & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, 13, 334-359.
- 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.
- Mellewigt, T. & Matiaske, W. (2000). Zur Messung des Unternehmens- und Kooperationserfolgs. *Die Betriebswirtschaft*, 60, 125-128.



- Menon, S., Kahn, B.E. (2000). *Cross-Category Effects of Induced Arousal and Pleasure on the Internet Shopping Experience*. Workingpaper, Department of Marketing, The Wharton School, University of Pennsylvania.
- Metzger, W. (1953). *Gesetze des Sehens* (2. Aufl.). Frankfurt: Kramer.
- Meuter, M.L., Ostrom, A.L., Bitner, M.J. & Roundtree, R. (2003). The Influence of Technology Anxiety on Consumer Use and Experiences with Self-Service Technologies. *Journal of Business Research*, 56(11), 899-907.
- Mittal, B. (1989). A Theoretical Analysis of Two Recent Measures of Involvement. *Advances in Consumer Research*, 16, 697-702.
- Moe, W.W. (2003). Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology*, 13, 29-40.
- Moe, W.W. & Fader, P.S. (2004). Dynamic Conversion Behaviour at E-Commerce Sites. *Management Science*, 50(3), 326-335.
- Molich, R., Ede, M.R., Kassgaards, K. & Karyukin, B. (2004). Comparative Usability Evaluation. *Behaviour & Information Technology*, 23(1), 65-74
- Moll, T. (1987). Über Methoden zur Analyse und Evaluation interaktiver Computersysteme. In K.-P. Fähnrich (Hrsg.), *Software-Ergonomie* (179-190). München: Oldenbourg.
- Monroe, K.B. (1973). Buyers' Subjective Perceptions of Price. *Journal of Marketing Research*, 10(1), 70-80.
- Montgomery, A.L., Li, S., Srinivasan, K. & Liechty, L.J. (2002). Predicting Online Purchase Conversion Using Web Path Analysis. *Marketing Science*, 23(4), 579-595.
- Moorman, C., Deshpandé, R. & Zaltman, G. (1993). Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*, 57(1), 81-101.
- 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.
- Motl, R.W. & DiStefano, C. (2002). Longitudinal Invariance of Self-Esteem and Method Effects Associated with Negatively Worded Items. *Structural Equation Modeling*, 9(4), 562-578.
- Muller, M.J., Matheson, L., Page, C. & Gallup, R. (1998). Participatory Heuristic Evaluation. *Interactions*, 3(4), 13-18.
- Murphy, K.R., Anhalt, R.L. & Jako, R. (1993). Nature and Consequences of Halo Error: a Critical Analysis. *Journal of Applied Psychology*, 78(2), 218-225.

- Muthitacharoen, A., Palvia, P.C., Brooks, L.D., Krishnan, B.C., Otondo, R.F. & Retzlaff-Robert, D. (2006). Reexamining Technology Acceptance in Online Task Behaviours. *Electronic Markets*, 16(1), 4-15.
- Nagy, M.S. (2002). Using a Single-Item Approach to Measure Facet Job Satisfaction. *Journal of Occupational and Organizational Psychology*, 75, 77-86.
- Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 78, 311-329.
- Newman, M.S. (1966). Return on Investment: an Analysis of the Concept. *Management Services*, 3(4), 15-23.
- Nicolai, A. & Kieser, A. (2002). Trotz eklatanter Erfolglosigkeit: Die Erfolgsfaktorenforschung weiter auf Erfolgskurs. *Die Betriebswirtschaft (DBW)*, 64(5), 579-596.
- Nicolai, A. & Kieser, A. (2004). Von Konsensgenerierungsmaschinen, Nebelkerzen und »the Operation called ›Verstehen‹«. *Die Betriebswirtschaft (DBW)*, 64(5), 631-635.
- Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press.
- Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen & R.L. Mack (Eds.), *Usability Inspection Methods* (25-62). New York: Wiley.
- Nielsen, J. (2005a). Durability of Usability Guidelines. *Jakob Nielsens Alertbox vom 17. Januar 2005*. Online im Internet: URL: <http://www.useit.com/alertbox/20050117.html> (Stand 27.11.06)
- Nielsen, J. (2005b). Top Ten Web Design Mistakes in 2005. *Jakob Nielsens Alertbox vom 3. Oktober 2005*. Online im Internet: URL: <http://www.useit.com/alertbox/designmistakes.html> (Stand 27.11.06)
- Nielsen, J. (2006). Quantitative Studies: How Many Users to Test? *Jakob Nielsens Alertbox vom 26. Juli 2006*. Online im Internet: URL: [http://www.useit.com/alertbox/quantitative\\_testing.html](http://www.useit.com/alertbox/quantitative_testing.html) (Stand 27.11.06)
- Nielsen, J. & Landauer, T.K. (1993). A Mathematical Model of the Finding of Usability Problems. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (Eds.), *Proceedings ACM InterCHI '93 Conference, Amsterdam* (206-213). New York: ACM Press.
- Nielsen, J. & Mack, R.L. (2004). *Usability Inspection Methods*. New York: Wiley.
- 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.
- Nielsen, J., Molich, R., Snyder, C. & Farrell, S. (2001). *E-Commerce User Experience*. Fremont: Nielsen Norman Group.

- Nisbett, R.E. & Wilson, T.D. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, 35, 250-256.
- Noiwan, J. & Norcio, A.F. (2006). Cultural Differences on Attention and Perceived Usability: Investigating Color Combinations of Animated Graphics. *International Journal of Human-Computer Studies*, 64(2), 103-122.
- Norman, D.A. (2004a). Beauty, Goodness, and Usability. *Special Section of Human-Computer Interaction*, 19(4), 311-318.
- Norman, D.A. (2004b). *Emotional Design - Why We Love (or Hate) Everyday Things*. New York, NY: Basic Books.
- Norman, K.L., Friedman, Z., Norman, K. & Stevenson, R. (2001). Navigational Issues in the Design of Online Self-Administered Questionnaires. *Behaviour & Information Technology*, 20, 37-45.
- Notani, A.S. (1997). Perceptions of Affordability: Their Role in Predicting Purchase Intent and Purchase. *Journal of Economic Psychology*, 18, 525-546.
- Novak, T.P., Hoffman, D.L. & Yung, Y.-F. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science*, 19(1), 22-42.
- Nunnally, J.C. (1978). *Psychometric Theory* (2<sup>nd</sup> ed.). New York: McGraw-Hill.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory* (3<sup>rd</sup> ed.). New York: McGraw-Hill.
- Oishi, S., Schimmack, U. & Colcombe, S.J. (2003). The Contextual and Systematic Nature of Life Satisfaction Judgments. *Journal of Experimental Social Psychology*, 39(3), 232-247.
- Ollermann, F. (2004). Verhaltensbasierte Validierung von Usability-Fragebögen. In R. Keil-Slawik, H. Selke & G. Szwillus (Hrsg.), *Mensch & Computer 2004: Allgegenwärtige Interaktion* (55-64). München: Oldenbourg-Verlag.
- Olsen, J. (1981). What is an Esthetic Response? In E.C. Hirschmann & M.B. Holbrook (Eds.), *Symbolic Consumer Behavior* (71-74). Ann Arbor, NY: Association of Consumer Research.
- Öörni, A. (2003). Consumer Search in Electronic Markets: an Experimental Analysis of Travel Services. *European Journal of Information Systems*, 12(1), 30-40.
- Oppermann, R. & Reiterer, H. (1997). Software Evaluation Using the 9241 Evaluator. *Behaviour & Information Technology*, 16(4/5), 232-245.
- Oppermann, R., Murchner, B., Reiterer, H. & Koch, M. (1992). *Software-ergonomische Evaluation. Der Leitfaden EVADIS II*. Berlin: Walter de Gruyter.

- Otter, M. & Johnson, H. (2000). Lost in Hyperspace: Metrics and Mental Models. *Interacting with Computers*, 13, 1-40.
- Overbeeke, K. & Wensveen, S. (2004). Beauty in Use. *Human-Computer Interaction*, 19(4), 367-369.
- Ozok, A.A. & Salvendy, G. (2001). How Consistent is Your Web Design? *Behaviour & Information Technology*, 20, 433-447.
- Palmer, J.W. (2002). Web Site Usability, Design, and Performance Metrics. *Information Systems Research*, 13(2), 151-167.
- Panten, G. (2005). *Internet-Geschäftsmodell Virtuelle Community - Analyse zentraler Erfolgsfaktoren unter Verwendung des Partial-Least-Squares (PLS)-Ansatzes*. Wiesbaden: DUV.
- Panten, G. & Thies, S. (2006). Analyse kausaler Wirkungszusammenhänge mit Hilfe von Partial Least Squares (PLS). In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (311-329). Wiesbaden: DUV.
- Parasuraman, A., Zeithaml, V.A. & Malhotra, A. (2005). E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality. *Journal of Service Research*, 7(3), 213-233.
- Park, C. & Jun, J.-K. (2003). A Cross-Cultural Comparison of Internet Buying Behaviour. Effects of Internet Usage, Perceived Risks and Innovativeness. *International Marketing Review*, 20(5), 534-553.
- Park, C.-W. & Moon, B.-J. (2003). The Relationship between Product Involvement and Product Knowledge: Moderating Roles of Product Type and Product Knowledge Type. *Psychology & Marketing*, 20, 977-997.
- Park, S., Choi, D. & Kim, J. (2004). Critical Factors for the Aesthetic Fidelity of Web Pages: Empirical Studies with Professional Web Designers and Users. *Interacting with Computers*, 16, 351-376.
- Partala, T. & Surakka, V. (2004). The Effects of Affective Interventions in Human-Computer Interaction. *Interacting with Computers*, 16, 295-309.
- 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), 101-134.
- 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.
- Peters, T.J. & Waterman, R.H. (2000). *Auf der Suche nach Spitzenleistungen: Was man von den bestgeführten US-Unternehmen lernen kann* (8. Aufl.). Landsberg am Lech: MVG.

- Petre, M., Minocha, S. & Roberts, D. (2006). Usability beyond the Website: an Empirically-Grounded E-Commerce Evaluation Instrument for the Total Customer Experience. *Behaviour & Information Technology*, 25(2), 189-203.
- Pickering, J. & Isherwood, B. (1974). Purchase Probabilities and Consumer Buying Behavior. *Journal of Marketing Research Society*, 16, 203-226.
- Pilke, E.M. (2004). Flow Experiences in Information Technology Use. *International Journal of Human-Computer Studies*, 61, 347-357.
- 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
- Porteous, M.A. & Kirakowski, J. (1992). *SUMI - Subjective Usability Measurement Inventory*. Department of Applied Psychology, University College Cork.
- Porteous, M.A., Kirakowski, J. & Corbett, M. (1993). *SUMI User Handbook*. Human Factors Research Group, University College Cork, Ireland.
- Porter, M.E. (1980). *Competitive Strategy*. New York: Free Press.
- Preece, J. (1994). *Human-Computer Interaction*. Harlow: Addison-Wesley.
- Proctor, R.W. & Van Zandt, T. (1994). *Human Factors in Simple and Complex Systems*. Needham Heights, MA: Allyn & Bacon.
- Prümper, J. (1993). Software-Evaluation based upon ISO 9241 Part 10. In T. Grechenig & M. Tscheligi (Eds.), *Human Computer Interaction* (255-265). Berlin: Springer
- Prümper, J. (1997). Der Benutzerfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In R. Liskowsky, B.M. Velichkovsky & W. Wünschmann (Hrsg.), *Software-Ergonomie '97. Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (253-262). Stuttgart: Teubner.
- 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.
- Rack, O. & Christophersen, T. (2006). Experimente. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (19-36). Wiesbaden: DUV.
- Ranaweera, C., McDougall, G. & Bansal, H. (2005). A Model of Online Customer Behavior: Moderating Effects of Customer Characteristics. *Marketing Theory*, 5(1), 51-74.
- Rauterberg, M. (1992). Läßt sich die Gebrauchstauglichkeit interaktiver Software messen? Und wenn ja, wie? *Ergonomie und Informatik*, 16, 3-18.

- Ray, N.M. & Tabor, S.W. (2003). Cyber Surveys Come of Age. *Marketing Research*, 15(2), 32-37.
- Rehman, A. (2000). *Holiday 2000 E-Commerce*. Creative Good. Online im Internet: URL: <http://scholar.google.de/scholar?hl=de&lr=&q=cache:Urx5VQgSeQ8J:csg.chcsolutions.com/csg/research/usability/cg-holiday2000.pdf+Holiday+2000+E-Commerce+rehman> (Stand 27.11.06)
- Reibstein, D.J. (2002). What Attracts Customers to Online Stores, and What Keeps Them Coming Back? *Journal of the Academy of Marketing Science*, 30, 465-473.
- Reimer, K. (2006). Bootstrapping und andere Resampling-Methoden. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (401-418). Wiesbaden: DUV.
- Resnick, M.L. & Montania, R. (2003). Perceptions of Customer Service, Information Privacy, and Product Quality From Semiotic Design Features in an Online Web Store. *International Journal of Human-Computer Interaction*, 16, 211-234.
- 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.
- Robinson, J., Shaver, P. & Wrightsman, L. (1991). Criteria for Scale Selection and Evaluation. In J. Robinson, P. Shaver & L. Wrightsman (Eds.), *Measures of Personality and Social Psychology Attitudes* (395-415). San Diego.
- Rodgers, S. & Harris, M.A. (2003). Gender and E-Commerce: An Exploratory Study. *Journal of Advertising Research*, 43, 322-330.
- Rohn, A. (2006). Subjektive versus objektive Erfolgsmaße. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (99-114). Wiesbaden: DUV.
- 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: DUV.
- 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.
- Rowley, J. (2000). Product Search in e-Shopping: A Review and Research Propositions. *Journal of Consumer Marketing*, 17(1), 20-35.

- Roy, M.C., Dewit, O. & Aubert, B.A. (2001). The Impact of Interface Usability on Trust in Web Retailers. *Internet Research: Electronic Networking Applications and Policy*, 11, 388-398.
- Rubin, D.B. (1976). Inference and Missing Data. *Biometrika*, 63, 581-592.
- Rubin, J. (1994). *Handbook of Usability Testing*. New York: John Wiley & Sons, Inc.
- Satzinger, J.W. & Olfman, L. (1998). User Interface Consistency across End-User Applications: The Effects on Mental Models. *Journal of Management Information Systems*, 14, 167-193.
- Schafer, J.L. (1999). *NORM: Multiple Imputation of Incomplete Multivariate Data under a Normal Model, Version 2*. Software for Windows 95/98/NT. Online im Internet: URL: <http://www.stat.psu.edu/~jls/misoftwa.html> (03.06.06)
- Schafer, J.L. & Graham, J.W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods*, 7, 147-177.
- Schafer, J.L. & Olsen, M.K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: a Data Analyst's Perspective. *Multivariate Behavioral Research*, 33, 545-571.
- Schaupp, L.C. & Bélanger, F. (2005). A Conjoint Analysis of Online Consumer Satisfaction. *Journal of Electronic Commerce Research*, 6(2), 95-111.
- Schenkman, B.N. & Jönsson, F.U. (2000). Aesthetics and Preferences of Web Pages. *Behaviour & Information Technology*, 19, 367-377.
- Schiffman, L.G., Sherman, E. & Long, M.M. (2003). Toward a Better Understanding of the Interplay of Personal Values and the Internet. *Psychology & Marketing*, 20, 169-186.
- Schimmel, K. & Morris, R. (2003). Gender Differences and E-commerce Behavior and Perceptions. *Journal of Internet Banking and Commerce*, 8(2). Online im Internet: URL: <http://www.arraydev.com/commerce/JIBC/0306-01.htm> (Stand 27.11.06)
- Schindler, R.M., Morrin, M. & Bechwati, N.N. (2005). Shipping Charges and Shipping-Charge Skepticism: Implications for Direct Marketers' Pricing Formats. *Journal of Interactive Marketing*, 19(1), 41-53.
- Schlosser, A.E. (2003). Computers as Situational Cues: Implications for Consumers' Product Cognitions and Attitudes. *Journal of Consumer Psychology*, 13(1/2), 103-113.
- Schlosser, A.E., White, T.B. & Lloyd, S.M. (2006). Converting Web Site Visitors into Buyers: How Web Site Investment Increases Consumer Trusting Beliefs and Online Purchase Intentions. *Journal of Marketing*, 70(2), 133-148.
- Schmitz, S.W. & Latzer, M. (2002). Competition in B2C e-Commerce: Analytical Issues and Empirical Evidence. *Electronic Markets*, 12(3), 163-174.

- 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: DUV.
- Scholderer, J. & Balderjahn, I. (2006). Was unterscheidet harte und weiche Strukturgleichungsmodelle nun wirklich? Ein Klärungsversuch zur LISREL-PLS-Frage. *Marketing ZFP*, 28(1), 57-70.
- Schriesheim, C.A. & Eisenbach, R.J. (1995). An Exploratory and Confirmatory Factor-Analytic Investigation of Item Wording Effects on the Obtained Factor Structures of Survey Questionnaire Measures. *Journal of Management*, 21(6), 1177-1193.
- Schubert, P. (2002). Extended Web Assessment Method (EWAM): Evaluation of Electronic Commerce Applications from the Customer's Viewpoint. *International Journal of Electronic Commerce*, 7(2), 51-80.
- Schulz, S., Mau, G. & Silberer, G. (in Druck). Benutzerfreundlichkeit von Reisekatalogen - Die Anwendung von Usability-Fragebögen auf Printkataloge. *transfer - Werbeforschung & Praxis*, 51(4).
- Schurr, P.H. & Ozanne, J.L. (1985). Influences on Exchange Processes: Buyers' Preconceptions of a Seller's Trustworthiness and Bargaining Toughness. *Journal of Consumer Research*, 11(4), 939-953.
- 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(1), 3-23.
- Sellin, N. & Keeves, J.P. (1994). Path Analysis with Latent Variables. In T. Husen & T.N. Postlethwaite (Eds.), *International Encyclopedia of Education* (2<sup>nd</sup> ed., 4352-4359). London: Elsevier Publishers.
- Shankar, V., Smith, A.K. & Rangaswamy, A. (2003). Customer Satisfaction and Loyalty in Online and Offline Environments. *International Journal of Research in Marketing*, 20(2), 153-175.
- Sherman, E. & Smith, R.B. (1986). Mood State of Shoppers and Store Image: Promising Interactions and Possible Behavioral Effects. In M. Wallendorf & P. Anderson (Eds.), *Advances in Consumer Research*, 13. Provo, UT: Association for Consumer Research.
- Shim, S., Eastlick, M.A., Lotz, S.L. & Washington, P. (2001). An Online Pre-purchase Intentions Model: The Role of Intention to Search. *Journal of Retailing*, 77, 397-416.
- Shneiderman, B. (1987). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading, MA: Addison-Wesley.



- Shneiderman, B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (2<sup>nd</sup> ed.). Reading, MA: Addison-Wesley.
- Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (3<sup>rd</sup> ed.). Reading, MA: Addison Wesley.
- Shugan, S.M (2006). Are Consumers Rational? Experimental Evidence? *Marketing Science*, 25(1), 1-7.
- Sing, C.K. (2004). The Measurement, Analysis, and Application of the Perceived Usability of Electronic Stores. *Singapore Management Review*, 26(2), 49-64.
- Slama, M.E. & Tashchian, A. (1985). Selected Socioeconomic and Demographic Characteristics Associated with Purchasing Involvement. *Journal of Marketing*, 49(1), 72-82.
- Smith, M.D., Bailey, J. & Brynjolfsson, E. (1999). Understanding Digital Markets: Review and Assessment. In E. Brynjolfsson & B. Kahin (Eds.), *Understanding the Digital Economy: Data, Tools, and Research* (99-136), Cambridge: MIT Press.
- Smyth, J.D., Dillman, D.A., Christian, L.M. & Stern, M.J. (2006). Effects of Using Visual Design Principles to Group Response Options in Web Surveys. *International Journal of Internet Science*, 1(1), 5-15.
- Sotgiu, F. & Ancarani, F. (2005). The Drivers of E-tailers' Price Levels. *International Review of Retail, Distribution & Consumer Research*, 15(1), 75-89.
- Späth, N. (2002). *Usability von eCommerce-Anwendungen: Eine Untersuchung zu nutzerseitigen Bedürfnissen und Erwartungen an Online-Shops*. Unveröffentlichte Diplomarbeit. Hamburg.
- Spector, P.E. (2006). Method Variance in Organizational Research: Truth or Urban Legend? *Organizational Research Methods*, 9(2), 221-232.
- Spinas, P. (1987). *Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen*. Unveröffentlichte Dissertation, Institut für Arbeitspsychologie, Zürich.
- Srivihok, A. (1999). *An Assessment Tool for Electronic Commerce: End User Evaluation of Web Commerce Sites*. Vortrag auf der Conference on the Measurement of Electronic Commerce, Singapore.
- Stafford, T.F., Turan, A. & Raisinghani, M.S. (2004). International and Cross-Cultural Influences on Online Shopping Behavior. *Journal of Global Information Technology Management*, 7(2), 70-88.
- Statistisches Bundesamt - Pressestelle (2006). *Informationstechnologie in Unternehmen und Haushalten 2005*. Wiesbaden: Statistisches Bundesamt. Online im Internet: URL: [http://www.destatis.de/presse/deutsch/pk/2006/ikt\\_2005i.pdf](http://www.destatis.de/presse/deutsch/pk/2006/ikt_2005i.pdf) (Stand 27.11.06)

- Stefani, A. & Xenos, M. (2001). A Model for Assessing the Quality of E-Commerce Systems. In *Proceedings of the PC-HCI 2001 Conference on Human Computer Interaction* (105-109). New Orleans, Louisiana, USA.
- Steinle, C., Kirschbaum, J. & Kirschbaum, V. (1996). *Erfolgreich überlegen - Erfolgsfaktoren und ihre Gestaltung in der Praxis*. Frankfurt am Main: FAZ (Verlagsbereich Wirtschaftsbücher).
- Stibel, J.M. (2005). Mental Models and Online Consumer Behaviour. *Behaviour & Information Technology*, 24(2), 147-150.
- Strack, F. & Werth, L. (2002). Wie zufrieden sind Sie eigentlich? Psychologische Determinanten von Zufriedenheitsmessungen. In K. Ferger, N. Gissel & J. Schwier (Hrsg.), *Sportspiele, Trainieren, Vermitteln, Erleben. Schriftenreihe der Deutschen Vereinigung für Sportwissenschaft* (13-24). Hamburg: Czwalina.
- Sudman, S., Bradburn, N.M. & Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tabachnik, B. & Fidell, L.S. (2006). *Using Multivariate Statistics* (5<sup>th</sup> ed.). Boston: Pearson.
- Tauber, E.M. (1975). Predictive Validity in Consumer Research. *Journal of Advertising Research*, 15(5), 59-64.
- Temme, D. (2006). Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen. *Marketing ZFP*, 28(3), 183-196.
- Teo, T.S.H. & Lim, V.K.G. (2000). Gender Differences in Internet Usage and Task Preferences. *Behaviour & Information Technology*, 19, 283-295.
- Thayer, R.E. (1989). *The Biopsychology of Mood and Arousal*. New York: Oxford University Press.
- Thorndike, E.L. (1920). A Constant Error on Psychological Rating. *Journal of Applied Psychology*, 4, 25-29.
- Tilson, R. & Dong, J. (1998). Factors and Principles Affecting the Usability of Four E-commerce Sites. In *Proceedings of the Fourth Conference on Human Factors and the Web*. Online im Internet: URL: <http://www.research.microsoft.com/users/marycz/hfweb98/tilson> (Stand 27.11.06)
- Timmers, P. (1998). Business Models for Electronic Markets. *Electronic Markets*, 8(2), 3-8.
- TNS Infratest (2005). *(N)Onliner Atlas 2005. Eine Typographie des digitalen Grabens durch Deutschland*. München. Online im Internet: Verfügbar unter URL: <http://www.nonliner-atlas.de/archiv.asp> (Stand 27.11.06)

- Tognazzini, B. (2004). *Top Ten Reasons to Not Shop On Line*. Ask Tog Kolumne Januar 2004. Online im Internet: URL: <http://www.asktog.com/columns/062top10ReasonsToNotShop.html> (Stand 27.11.06)
- Toms, E.G. (2000). Understanding and Facilitating the Browsing of Electronic Text. *International Journal of Human-Computer Studies*, 52(3), 423-452.
- Tractinsky, N. (1997). Aesthetic and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In S. Pemberton (Ed.), *CHI 97 Proceedings*, Atlanta (115-123). New York: ACM Press.
- Tractinsky, N. (2004). A Few Notes on the Study of Beauty in HCI. *Human-Computer Interaction*, 19, 351-357.
- Tractinsky, N. & Meyer, J. (2001). Task Structure and the Apparent Duration of Hierarchical Search. *International Journal of Human-Computer Studies*, 55, 845-860.
- Tractinsky, N. & Rao, V.S. (2001). Incorporating Social Dimensions in Web-Store Design. *Human Systems Management*, 20, 105-121.
- Tractinsky, N., Katz, A.S. & Ikar, D. (2000). What is Beautiful is Usable. *Interacting with Computers*, 13(2), 127-145.
- Trifts, V. & Häubl, G. (2003). Information Availability and Consumer Preference: Can Online Retailers Benefit From Providing Access to Competitor Price Information? *Journal of Consumer Psychology*, 13, 149-160.
- Ulich, E. (1981). Subjektive Tätigkeitsanalyse als Voraussetzung autonomieorientierter Arbeitsgestaltung. In F. Frei & E. Ulich (Hrsg.), *Beiträge zur psychologischen Arbeitsanalyse* (327-347). Bern: Huber.
- Van der Heijden, H. (2003). Factors Influencing the Usage of Websites: the Case of a Generic Portal in the Netherlands. *Information & Management*, 40, 541-549.
- Van der Heijden, H. & Verhagen, T. (2002). Measuring and Assessing Online Store Image: A Study of Two Online Bookshops in the Benelux. In R.H. Sprague (Ed.), *Proceedings of the 35<sup>th</sup> Hawaii International Conference on System Sciences 2002*. Big Island, Hawaii, CD-ROM, Los Alamitos CA: IEEE Computer Society Press.
- Van der Heijden, H. & Verhagen, T. (2004). Online Store Image: Conceptual Foundations and Empirical Measurements. *Information & Management*, 41, 609-617
- Van der Heijden, H., Verhagen, T. & Creemers, M. (2001). Predicting Online Purchase Behavior: Replications and Tests of Competing Models. In *Proceedings of the 34<sup>th</sup> Annual Hawaii International Conference on System Sciences (HICSS-34)*: IEEE Computer Society.

- Van der Heijden, H., Verhagen, T. & Creemers, M. (2003). Understanding Online Purchase Intentions: Contributions from Technology and Trust Perspectives. *European journal of information systems*, 12(1), 41-48.
- Van der Veer, G.C. & del Carmen Puerta Melguizo, M. (2003). Mental Models. In J.A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (52-80). Mahwah, N.J.: Lawrence Erlbaum.
- Van Dyk, T. & Renaud, K. (2002). Task Analysis for E-Commerce and the Web. In D. Diaper & N. Stanton (Eds.), *The Handbook of Task Analysis for Human-Computer Interaction*, Chapter 12. Fairfax, Virginia: Techbooks.
- Van Herpen, E. & Pieters, R. (2002). The Variety of an Assortment: An Extension to the Attribute-Based Approach. *Marketing Science*, 21(3), 331-341.
- Van Vliet, P.J.A. & Pota, D. (2000). Understanding Online Retail: a Classification of Online Retailers. *Journal of Computer Information Systems*, 41(2), 23-29.
- Vanderwerf, P. & Mahon, J.F. (1997). Meta-Analysis of the Impact of Research Methods on Findings of First-Mover Advantages. *Management Science*, 43, 1510-1519.
- Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*, 11(4), 342-365.
- Venkatesh, V. & Davis, F.D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46, 186-204.
- Venkatesh, V. & Morris, M.G. (2000). Why Don't Men ever Stop to Ask for Directions? Gender, Social Influence, and their Role in Technology Acceptance and Usage Behaviour. *MIS Quarterly*, 24(1), 115-139.
- Verzyer Jr., R.W. (1993). Aesthetic Response and the Influence of Design Principles on Product Preferences. *Advances in Consumer Research*, 20, 224-228.
- Virzi, R.A. (1997). Usability Inspection Methods. In M. Helander, T.K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2<sup>nd</sup> ed., 705-715). Amsterdam: Elsevier.
- Wade, M.R. & Nevo, S. (2005). Development and Validation of a Perceptual Instrument to Measure E-Commerce Performance. *International Journal of Electronic Commerce*, 10(2), 123-146.
- Wang, Y.-S., Tang, T.-I. & Tang, J.T.E. (2001). An Instrument for Measuring Customer Satisfaction Toward Web Sites That Market Digital Products and Services. *Journal of Electronic Commerce Research*, 2(3), 89-102.

- Wanous, J., Reichers, A.E. & Hudy, M. (1997). Overall Job Satisfaction: How Good Are Single-Item Measures? *Journal of Applied Psychology*, 82, 247-252.
- Warden, C., Wu, W.Y. & Tsai, D.C. (2006). Online Shopping Interface Components: Relative Importance as Peripheral and Central Cues. *CyberPsychology & Behavior*, 9(3), 285-296.
- Warshaw, P.R. (1980). Predicting Purchase and other Behavior from General and Contextually Specific Intentions. *Journal of Marketing Research*, 17(1), 26-33.
- Weathers, D. & Makienko, I. (2006). Assessing the Relationships between E-tail Success and Product and Web Site Factors. *Journal of Interactive Marketing*, 20(2), 41-54
- Weening, M.W.H. & Maarleveld, M. (2002). The Impact of Time Constraint on Information Search Strategies in Complex Choice Tasks. *Journal of Economic Psychology*, 23(6), 689-692.
- Weiber, R. & Adler, J. (1995). Informationsökonomisch begründete Typologisierung von Kaufprozessen. *Zeitschrift für betriebswirtschaftliche Forschung*, 1, 43-65.
- Westbrook, R.A. & Black, W.C. (1985). A Motivation Based Shopper Typology. *Journal of Retailing*, 61(1), 78-103.
- Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen & R.L. Mack (Eds.), *Usability Inspection Methods* (105-140). New York: Wiley.
- Whiteside, J., Bennet, J. & Holtzblatt, K. (1988). Usability Engineering: Our Experience and Evolution. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (791-817). Amsterdam: Elsevier.
- Widaman, K. (1993). Common Factor Analysis versus Principal Component Analysis: Differential Bias in Representing Model Parameters? *Multivariate Behavioral Research*, 28(3), 263-311.
- Wigand, R.T. (1996). Electronic Commerce: Definition, Theory, and Context. *The Information Society*, 13(1), 1-16.
- Williams, L.J., Edwards, J.R. & Vandenberg, R.J. (2003). Recent Advances in Causal Modeling Methods for Organizational and Management Research. *Journal of Management*, 29(6), 903-936
- Willumeit, H., Gediga, G. & Hamborg, K.-C. (1995). Validation of the IsoMetrics Usability Inventory. *Forschungsberichte Nr. 105, Fachbereich Psychologie*. Universität Osnabrück

- Winand, U. & Pohl, W. (2000). Die Vertrauensproblematik in elektronischen Netzwerken. In J. Link (Hrsg.), *Wettbewerbsvorteile durch Online Marketing. Die strategischen Perspektiven elektronischer Märkte* (2. Aufl., 261-277). Berlin: Springer.
- Wirtz, J. & Bateson, J.E.G. (1995). An Experimental Investigation of Halo Effects in Satisfaction Measures of Service Attributes. *International Journal of Service Industry Management*, 6(3), 84-103.
- Wiswede, G. (1995). *Einführung in die Wirtschaftspsychologie*. München: Reinhardt.
- 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.
- Wolfenbarger, M. & Gilly, M.C. (2001). Shopping Online for Freedom, Control and Fun. *California Management Review*, 43(2), 34-55.
- Wolfenbarger, M. & Gilly, M.C. (2003). eTailQ: Dimensionalizing, Measuring and Predicting etail Quality. *Journal of Retailing*, 79, 183-198.
- Xenos, M. (1999). Lessons Learned on Measuring Perceived Software Quality. In *Proceedings of the FESMA99 International Conference, Federation of European Software Measurement Associations* (349-356). Amsterdam.
- Yang, Z. & Peterson, R.T. (2004). Customer Perceived Value, Satisfaction, and Loyalty: The Role of Switching Costs. *Psychology & Marketing*, 21(10), 799-822
- Yang, Z., Cai, Z., Zhou, Z. & Zhou, N. (2005). Development and Validation of an Instrument to Measure User Perceived Service Quality of Information Presenting Web Portals. *Information & Management*, 42, 575-589.
- 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 R. Keil-Slawik, H. Selke & G. Szwillus (Hrsg.), *Mensch & Computer 2004: Allgegenwärtige Interaktion* (43-53). München: Oldenbourg.
- Yoo, B. & Donthu, N. (2001). Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, 2 (1), 31-45.
- 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. (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing*, 52(3), 2-22.
- Zhang, P. & von Dran, G.M. (2001). Expectations and Rankings of Website Quality Features: Results of Two Studies on User Perceptions. In *Proceedings of the Thirty-Fourth Annual Hawaii International Conference on System Sciences (HICSS34)*. Maui, Hawaii.
- Zhang, P., von Dran, G.M., Blake, P. & Pipithsuksunt, V. (2001). Important Design Features in Different Web Site Domains. *E-Service Journal*, 1(1), 77-91.





## **Anhang A**

### **Ergänzende Darstellungen zum Hintergrund**

---

Abb. A-1.1: ufosV1-Fragebogen (Seite 1) .....	283
Abb. A-1.2: ufosV1-Fragebogen (Seite 2) .....	284
Abb. A-1.3: ufosV1-Fragebogen (Seite 3) .....	285
Abb. A-1.4: ufosV1-Fragebogen (Seite 4) .....	286
Abb. A-1.5: ufosV1-Fragebogen (Seite 5) .....	287
Abb. A-1.6: ufosV1-Fragebogen (Seite 6) .....	288
Abb. A-1.7: ufosV1-Fragebogen (Seite 7) .....	289
Abb. A-1.8: ufosV1-Fragebogen (Seite 8) .....	290
Tab. A-1.1: Item-Pool des ufosV1-Fragebogens (Items u1 bis u54).....	291
Tab. A-1.2: Item-Pool des ufosV1-Fragebogens (Items u55 bis u80).....	292



Usability Fragebogen  
für  
Online-Shops (Version 1.1)

Dipl.-Psych. Timo Christophersen

Christian-Albrechts-Universität zu Kiel  
Institut für Psychologie  
Lehrstuhl für Arbeits-, Organisations-  
und Marktpsychologie; Graduiertenkolleg  
Westring 425  
D-24098 Kiel  
Tel.: 0431- 8801535  
E-mail: [christophersen@bwl.uni-kiel.de](mailto:christophersen@bwl.uni-kiel.de)

Dipl.-Psych. Björn Balazs

Humboldt-Universität zu Berlin  
Institut für Psychologie  
Lehrstuhl für Kognitive Ergonomie und  
Ingenieurpsychologie  
Wolfgang-Köhler-Haus  
Rudower Chaussee 18  
D-12489 Berlin  
Tel.: 0179 - 4541949  
E-mail: [b@lazzs.de](mailto:b@lazzs.de)

Abb. A-1.1: ufosV1-Fragebogen (Seite 1)

## Kurzübersicht zu Durchführung und Auswertung

Mehrere Nutzer (min.n=15) sollten den Fragebogen nach umfassender Exploration (ca.20 min) des zu bewertenden Shops beantworten. Bei einer Anwendung im Rahmen von Nutzertests wird empfohlen, den Versuchsteilnehmern drei Aufgaben zu stellen, um eine angemessene Form der Exploration zu gewährleisten. Dabei sollten die Teilnehmer a) gezielt nach einem vorgegebenen Produkt suchen, b) ungezielt, dem eigenen Interesse folgend durch den Shop stöbern und c) den Warenkorb bedienen (z.B. durch Entfernen von Produkten) bzw. den Bestellvorgang probeweise beginnen. Denkbar ist weiterhin, jeden x-ten Shopbesucher um die Beantwortung des Fragebogens zu bitten.

Die Kodierung der Ergebnisse erfolgt auf einer Skala von 1 (stimme voll zu) bis 5 (stimme gar nicht zu). Ein globales Maß der Einschätzung der Benutzerfreundlichkeit ergibt sich durch die Bildung eines gemittelten Summenscores (TOTAL) über alle Items. Außerdem können zur detaillierteren Analyse gemittelte Summenscores der sieben Subskalen des Fragebogens berechnet werden:

	Faktor	Fragen
1.	Allgemeiner Usabilityeindruck	01-14
2.	Informationen zu den Rahmenbedingungen	15-21
3.	Produktsuche	22-28
4.	Warenkorbhandling/Bestellung	29-35
5.	Produktüberblick	36-39
6.	Selbstbeschreibungsfähigkeit	40-43
7.	Produktbeschreibungen	44-47

Die Interpretation der Ergebnisse sollte vergleichend zu Ergebnissen anderer Shops (bzw. alternativer Prototypen) erfolgen.

Für nähere Informationen stehen die Autoren gerne zur Verfügung. Auch können die Diplomarbeiten, in deren Kontext der Fragebogen entwickelt wurde, unter [www.ergopool.de](http://www.ergopool.de) heruntergeladen werden. Die Entwicklung des Fragebogens wird außerdem dargestellt bei:

Konradt, U., Wandke, H., Balazs, B. & Christophersen, T. (2003). Usability in online shops: scale construction, validation and the influence on the buyer's intention and decision. Behaviour and Information Technology, 22, 165-174.

### Rechtlicher Hinweis:

Das Fragebogenverfahren einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung der Autoren außerhalb der strengen Grenzen des Urheberrechtsgesetzes unzulässig und strafbar. Dies gilt insbesondere für die Anwendung des Verfahrens zu kommerziellen Zwecken, Übersetzungen und Veränderungen. Die Nutzung im privaten bzw. wissenschaftlichen Rahmen wird von den Autoren dagegen begrüßt, erfordert jedoch ebenso eine Absprache mit den Autoren. Alle Daten, die mit Hilfe des Fragebogens erhoben werden, müssen den Autoren zur Weiterentwicklung des Instruments in Kopie zur Verfügung gestellt werden. Eine unveränderte Vervielfältigung und Weitergabe dieses Dokuments ist ausdrücklich und ohne weitere Zustimmung der Autoren erlaubt.

Abb. A-1.2: ufosV1-Fragebogen (Seite 2)

Sehr geehrte Untersuchungsteilnehmerin, sehr geehrter Untersuchungsteilnehmer,

bitte beurteilen Sie, inwieweit Sie den folgenden 47 Aussagen über den gerade besuchten Shop zustimmen. Nutzen Sie dazu die 5 Auswahlfelder von »Stimme voll zu« bis »Stimme gar nicht zu«. Wenn es Ihnen überhaupt nicht möglich sein sollte, eine Aussage zu beurteilen, kreuzen Sie bitte das Feld »Kann ich nicht beurteilen« an.

1. Dieser Shop macht das Einkaufen im Netz leicht.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

2. Es ist mir zu kompliziert, diesen Shop zu benutzen.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

3. Man kann sich schnell einen Überblick über den Shop verschaffen.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

4. Der Umgang mit dem Shop ist leicht zu erlernen.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

5. Dieser Shop regt mich an, ihn weiter zu erkunden.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

6. Der Einkauf kann zügig erledigt werden.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen


7. Ich kann jederzeit Informationen zur Benutzung des Shops erhalten.

Stimme voll zu

Stimme gar nicht zu

Kann ich nicht beurteilen

Abb. A-1.3: ufosV1-Fragebogen (Seite 3)


Usability Fragebogen für Online-Shops

---

**8. Dieser Shop bietet nützliche Hilfe zur Benutzung an.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**9. Es ist leicht, sich innerhalb des Shops zu bewegen.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**10. Der Shop ist in sich stimmig aufgebaut.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**11. Dieser Shop ist gut strukturiert.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**12. Ich kann die wichtigen Informationen immer gleich erkennen.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**13. Ich kann gut auf die für mich wichtigen Informationen zugreifen.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**14. Die Gestaltung unterstützt mich beim Benutzen des Shops.**


Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**15. Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

Seite 04
© 2002-2004 B. Balazs & T. Christophersen

Abb. A-1.4: ufosV1-Fragebogen (Seite 4)

 Usability Fragebogen für Online-Shops

16. Ich erhalte nützliche Informationen über die Lieferbedingungen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

17. Ich erhalte nützliche Informationen zur Rückgabe von Produkten.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

18. Ich erhalte nützliche Informationen über die verschiedenen Zahlungsmöglichkeiten.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

19. Ich erhalte nützliche Informationen über die Lieferzeiten der einzelnen Produkte.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

20. Ich erhalte nützliche Informationen über die allgemeinen Geschäftsbedingungen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

21. Ich erhalte nützliche Informationen über den Umgang mit meinen persönlichen Daten.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

22. Die Benennungen sind so gewählt, dass ich gut nach Produkten suchen kann.


Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

23. Ich kann leicht herausfinden, ob der Shop die Produkte führt, die ich suche.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

Seite 05 © 2002-2004 B. Balazs & T. Christophersen

Abb. A-1.5: ufosV1-Fragebogen (Seite 5)


Usability Fragebogen für Online-Shops

---

24. Es ist leicht, nach Produkten zu suchen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

25. Ich kann gut auf die Suchfunktionen des Shops zugreifen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

26. Die Suchfunktionen sind einfach zu bedienen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

27. Die Suchfunktion ermöglicht mir, schnell bestimmte Produkte zu finden.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

28. Die Produktsuche liefert mir sinnvolle Ergebnisse.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

29. Der Warenkorb ist übersichtlich.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

30. Ich weiß stets, welche Produkte sich gerade im Warenkorb befinden.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

31. Es ist einfach, Produkte in den Warenkorb zu legen.


Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

---

Seite 06
© 2002-2004 B. Balazs & T. Christophersen

Abb. A-1.6: ufosV1-Fragebogen (Seite 6)



 Usability Fragebogen für Online-Shops

32. Ich kann den Inhalt meines Warenkorbes einfach verändern.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

33. Der gesamte Bestellvorgang ist leicht zu überschauen.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

34. Die Eingabeaufforderungen sind hilfreich.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

35. Ich erkenne sofort, ob meine Eingabe erfolgreich war.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

36. Die Produktlisten im Shop sind übersichtlich.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

37. Die Produkte sind in sinnvoll gewählte Kategorien eingeteilt.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

38. Die Produkte sind sinnvoll angeordnet.


Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

39. Ich habe stets den Überblick, was die Produkte kosten.

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

Seite 07 © 2002-2004 B. Balazs & T. Christophersen

Abb. A-1.7: ufosV1-Fragebogen (Seite 7)


Usability Fragebogen für Online-Shops

---

**40. Es ist mir klar, wie ein Einkauf in diesem Shop abläuft.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**41. Ich benötige keine Erläuterungen, um in diesem Shop einkaufen zu können.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**42. Mir ist immer klar, welche Schritte ich als nächstes unternehmen kann.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**43. Wenn ich auf etwas klicke, verbirgt sich dahinter das, was ich erwarte.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**44. Ich kann angeben, welche Eigenschaften ein von mir gesuchtes Produkt erfüllen soll.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**45. Ich kann gut überprüfen, ob ein Produkt meinen Vorstellungen entspricht.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**46. Ich bin mit den Beschreibungen der Produkte zufrieden.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

**47. Ich finde die verwendeten Bilder und Graphiken hilfreich.**

Stimme voll zu	Stimme gar nicht zu	Kann ich nicht beurteilen
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>

Vielen Dank für Ihre Mühe!

---

Seite 08
© 2002-2004 B. Balazs & T. Christophersen

Abb. A-1.8: ufosV1-Fragebogen (Seite 8)

Tab. A-1.1: Item-Pool des ufosV1-Fragebogens (Items u1 bis u54)

Item-code	Item
u1	Dieser Shop macht das Einkaufen im Netz leicht.
u2	Es ist mir zu kompliziert, diesen Shop zu benutzen.
u3	Man kann sich schnell einen Überblick über den Shop verschaffen.
u4	Es ist schwer, den Überblick zu behalten.
u5	Es dauert lange, bis man gelernt hat, sich in dem Shop zurechtzufinden.
u6	Der Umgang mit dem Shop ist leicht zu erlernen.
u7	Dieser Shop regt mich an, ihn weiter zu erkunden.
u8	Der Einkauf kann zügig erledigt werden.
u9	Ich kann in der Reihenfolge vorgehen, die ich am sinnvollsten finde.
u10	Es ist mir klar, wie ein Einkauf in diesem Shop abläuft.
u11	Ich benötige keine Erläuterungen, um in diesem Shop einkaufen zu können.
u12	Ich muss immer wieder nach Informationen zur Benutzung des Shops suchen.
u13	Ich kann jederzeit Informationen zur Benutzung des Shops erhalten.
u14	Der Shop bietet nützliche Hilfe zur Benutzung an.
u15	In diesem Shop unterlaufen einem leicht Fehler.
u16	Es ist leicht, Fehler zu korrigieren..
u17	Fehlermeldungen sagen mir deutlich, was ich machen soll.
u18	Ich muss mir zu viel merken.
u19	Der Shop hilft mir, meinen Benutzernamen und mein Passwort nicht zu vergessen.
u20	Es ist leicht, sich innerhalb des Shops zu bewegen.
u21	Ich finde nur schwer meinen Weg durch den Shop.
u22	Ich kann einfach zwischen den verschiedenen Seiten hin- und herspringen.
u23	Ich kann jederzeit auf zentrale Seiten im Shop zurückkehren (z.B. die Startseite).
u24	Ich kann jede Aktion jederzeit wieder rückgängig machen.
u25	Mir ist immer klar, welche Schritte ich als nächstes unternehmen kann.
u26	Ich weiß oftmals nicht, wo ich mich innerhalb des Shops gerade befinde.
u27	Der Shop ist in sich stimmig aufgebaut.
u28	Gleiche Informationen werden stets auf die gleiche Art dargestellt.
u29	Dieser Shop ist gut strukturiert.
u30	Es sind zu viele Schritte nötig, um zu einem Produkt zu gelangen.
u31	Die einzelnen Seiten sind übersichtlich.
u32	Ich kann mich auf den Seiten des Shops gut zurechtfinden.
u33	Ich kann die wichtigen Informationen immer gleich erkennen.
u34	Ich kann gut auf die für mich wichtigen Informationen zugreifen.
u35	Es werden häufig zu viele Informationen dargestellt.
u36	Ich kann bestimmen, welche Informationen angezeigt werden und welche nicht.
u37	Die Texte sind nur schwer lesbar.
u38	Die Gestaltung unterstützt mich beim Benutzen des Shops.
u39	Ich finde das Meiste dort, wo ich es erwarte.
u40	Der Shop bietet mir alle Möglichkeiten, die ich von Online-Shops erwarte.
u41	Wenn ich auf etwas klicke, verbirgt sich dahinter das, was ich erwarte.
u42	Ich kann den Shop so benutzen, wie ich es erwarte.
u43	Ich finde die im Shop verwendete Sprache leicht verständlich.
u44	Die Benennungen sind so gewählt, dass ich gut nach Produkten suchen kann.
u45	Ich kann leicht herausfinden, ob der Shop die Produkte führt, die ich suche.
u46	Es ist leicht, nach Produkten zu suchen.
u47	Ich kann gut auf die Suchfunktionen des Shops zugreifen.
u48	Die Suchfunktionen sind einfach zu bedienen.
u49	Die Suchfunktion ermöglicht mir, schnell bestimmte Produkte zu finden.
u50	Ich kann angeben, welche Eigenschaften ein von mir gesuchtes Produkt erfüllen soll.
u51	Die Produktsuche liefert mir sinnvolle Ergebnisse.
u52	Die Suche liefert auch dann nützliche Ergebnisse, wenn mein Suchbegriff ungenau ist.
u53	Nach einer erfolglosen Suche erhalte ich nützliche Tipps zum weiteren Vorgehen.
u54	Die Produktlisten im Shop sind übersichtlich.

Tab. A-1.2: Item-Pool des ufosV1-Fragebogens (Items u55 bis u80)

u55	Die Produkte sind in sinnvoll gewählte Kategorien eingeteilt.
u56	Die Produkte sind sinnvoll angeordnet.
u57	Ich kann verschiedene Produkte gut miteinander vergleichen.
u58	Ich kann gut überprüfen, ob ein Produkt meinen Vorstellungen entspricht.
u59	Ich bin mit den Beschreibungen der Produkte zufrieden.
u60	Ich finde die verwendeten Bilder und Graphiken hilfreich.
u61	Der Shop empfiehlt mir interessante Produkte, nach denen ich ursprünglich nicht gesucht habe.
u62	Ich habe stets den Überblick, was die Produkte kosten.
u63	Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.
u64	Ich erhalte nützliche Informationen über die Lieferbedingungen.
u65	Ich erhalte nützliche Informationen zur Rückgabe von Produkten.
u66	Ich erhalte nützliche Informationen über die verschiedenen Zahlungsmöglichkeiten.
u67	Ich erhalte nützliche Informationen über die Lieferzeiten der einzelnen Produkte.
u68	Ich erhalte nützliche Informationen über die allgemeinen Geschäftsbedingungen.
u69	Ich erhalte nützliche Informationen über den Umgang mit meinen persönlichen Daten.
u70	Der Warenkorb ist übersichtlich.
u71	Ich weiß stets, welche Produkte sich gerade im Warenkorb befinden.
u72	Es ist einfach, Produkte in den Warenkorb zu legen.
u73	Ich kann den Inhalt meines Warenkorbes einfach verändern.
u74	Der gesamte Bestellvorgang ist leicht zu überschauen.
u75	Ich finde es schwer, mich anzumelden.
u76	Der Shop macht mir die Eingabe von Daten einfach (z.B. Anschrift, Kreditkartennummer).
u77	Die Eingabeaufforderungen sind hilfreich.
u78	Ich weiß bei einigen Eingaben nicht, warum ich sie machen muss.
u79	Ich erkenne sofort, ob meine Eingabe erfolgreich war.
u80	Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.

## **Anhang B**

### **Ergänzende Darstellungen zur ufosV2-Entwicklung**

Tab. B-1.1: Modifikationen und Eliminationen der ufosV1-Items (u1 bis u17).....	295
Tab. B-1.2: Modifikationen und Eliminationen der ufosV1-Items (u18 bis u31).....	296
Tab. B-1.3: Modifikationen und Eliminationen der ufosV1-Items (u32 bis u44).....	297
Tab. B-1.4: Modifikationen und Eliminationen der ufosV1-Items (u45 bis u58).....	298
Tab. B-1.5: Modifikationen und Eliminationen der ufosV1-Items (u59 bis u70).....	299
Tab. B-1.6: Modifikationen und Eliminationen der ufosV1-Items (u71 bis u80).....	300
Tab. B-2: Quellen für neue formative Usability-Items der ufosV2f-Skala .....	301

Tab. B-1.1: Modifikationen und Eliminationen der ufosV1-Items (Items u1 bis u17)

Item-code	Item	$p_m$	% Missing	Messmodellannahme
u1	<i>Dieser Shop macht das Einkaufen im Netz leicht.</i> unveränderte Weiterverwendung → ur1	0,41	0	reflektiv
u2	<i>Es ist mir zu kompliziert, diesen Shop zu benutzen.</i> unveränderte Weiterverwendung → ur2	0,36	0	reflektiv
u3	<i>Man kann sich schnell einen Überblick über den Shop verschaffen.</i> unveränderte Weiterverwendung → ur3	0,43	0	reflektiv
u4	<i>Es ist schwer, den Überblick zu behalten.</i> Elimination: Problem einer unklaren Messmodellannahme; hohe Korrelation zu u3 von $r=0,78$	0,59	0,8	unklar
u5	<i>Es dauert lange, bis man gelernt hat, sich in dem Shop zurechtzufinden.</i> Elimination: Problem einer unklaren Messmodellannahme; hohe Korrelation zu u6 von $r=0,66$	0,64	2,5	unklar
u6	<i>Der Umgang mit dem Shop ist leicht zu erlernen.</i> unveränderte Weiterverwendung → ur6	0,36	1,7	reflektiv
u7	<i>Dieser Shop regt mich an, ihn weiter zu erkunden.</i> unveränderte Weiterverwendung → ur7	0,54	1,7	reflektiv
u8	<i>Der Einkauf kann zügig erledigt werden.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → ur9: <i>In diesem Shop kann ein Einkauf zügig erledigt werden.</i>	0,40	3,3	reflektiv
u9	<i>Ich kann in der Reihenfolge vorgehen, die ich am sinnvollsten finde.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf1: <i>Beim Einkauf in diesem Shop kann ich in der Reihenfolge vorgehen, die ich selbst am sinnvollsten finde.</i>	0,39	1,7	formativ
u10	<i>Es ist mir klar, wie ein Einkauf in diesem Shop abläuft.</i> Elimination: Problem einer unklaren Messmodellannahme	0,28	0	unklar
u11	<i>Ich benötige keine Erläuterungen, um in diesem Shop einkaufen zu können.</i> Elimination: Problem einer unklaren Messmodellannahme	0,39	1,7	unklar
u12	<i>Ich muss immer wieder nach Informationen zur Benutzung des Shops suchen.</i> Elimination: Problem einer unklaren Messmodellannahme	0,65	2,5	unklar
u13	<i>Ich kann jederzeit Informationen zur Benutzung des Shops erhalten.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf2: <i>Es sind nützliche Informationen zur Benutzung des Shops abrufbar.</i>	0,41	5,8	formativ
u14	<i>Der Shop bietet nützliche Hilfe zur Benutzung an.</i> Umformulierung des Items, um Aspekt der Assistenz mehr zu betonen → uf3: <i>Bei auftretenden Problemen bietet der Shop von sich aus nützliche Hilfe an.</i>	0,49	7,5	formativ
u15	<i>In diesem Shop unterlaufen einem leicht Fehler.</i> Elimination: schlechtes Item, da Nutzern keine Fehler unterlaufen – es gibt nur fehlerhafte Software; Problem einer unklaren Messmodellannahme	0,55	3,3	unklar
u16	<i>Es ist leicht, Fehler zu korrigieren.</i> → uf4: <i>Es ist in diesem Shop leicht, versehentlich durchgeführte Aktionen zu korrigieren.</i>	0,36	1,7	formativ
u17	<i>Fehlermeldungen sagen mir deutlich, was ich machen soll.</i> Elimination: zu hoher Anteil fehlender Werte; inhaltlich wird dieser Aspekt durch u16 bzw. uf4 abgedeckt	0,55	30,0	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).

Tab. B-1.2: Modifikationen und Eliminationen der ufosV1-Items (Items u18 bis u31)

Item-code	Item	$p_m$	% Missing	Messmodellannahme
u18	<i>Ich muss mir zu viel merken.</i> konkrete Bezugnahme auf den Bewertungsgegenstand und Umformulierung, um die Itemschwierigkeit zu optimieren → uf5: <i>Ich muss mir viel merken, wenn ich in dem Shop einkaufe.</i>	0,74	0	formativ
u19	<i>Der Shop hilft mir, meinen Benutzernamen und mein Passwort nicht zu vergessen.</i> Elimination: zu hoher Anteil fehlender Werte; zu spezifischer Aspekt, der im Zweifel auch dichotom (ja/nein) durch einen Experten beantwortet werden kann	0,54	45,0	formativ
u20	<i>Es ist leicht, sich innerhalb des Shops zu bewegen.</i> unveränderte Weiterverwendung → uf6	0,35	2,5	formativ
u21	<i>Ich finde nur schwer meinen Weg durch den Shop.</i> Elimination: inhaltliche Redundanz zu u20 bzw. uf6 bestätigt durch eine Korrelation von $r=0,8$	0,68	0	formativ
u22	<i>Ich kann einfach zwischen den verschiedenen Seiten hin- und herspringen.</i> Umformulierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) und konkrete Bezugnahme auf den Bewertungsgegenstand → uf7: <i>Es ist einfach, zwischen den verschiedenen Seiten des Shops hin- und herzuspringen.</i>	0,39	5,0	formativ
u23	<i>Ich kann jederzeit auf zentrale Seiten im Shop zurückkehren (z.B. die Startseite).</i> Umformulierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf8: <i>Zentrale Seiten des Shops (z.B. die Startseite) sind jederzeit schnell zu erreichen.</i>	0,25	2,5	formativ
u24	<i>Ich kann jede Aktion jederzeit wieder rückgängig machen.</i> Umformulierung, da <i>jederzeit</i> überflüssig ist → uf9: <i>Ich kann jede Aktion wieder rückgängig machen.</i>	0,30	1,7	formativ
u25	<i>Mir ist immer klar, welche Schritte ich als nächstes unternehmen kann.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf10: <i>Mir ist beim Einkauf in dem Shop stets klar, welche Handlungen ich als nächstes durchführen kann.</i>	0,44	0,8	formativ
u26	<i>Ich weiß oftmals nicht, wo ich mich innerhalb des Shops gerade befinde.</i> Umformulierung, um Itemschwierigkeit zu reduzieren → uf11: <i>Ich weiß manchmal nicht, wo ich mich innerhalb des Shops befinde.</i>	0,65	0,8	formativ
u27	<i>Der Shop ist in sich stimmig aufgebaut.</i> Elimination: inhaltliche Redundanz zu u29 bzw. uf12 bestätigt durch eine Korrelation von $r=0,79$	0,39	3,3	formativ
u28	<i>Gleiche Informationen werden stets auf die gleiche Art dargestellt.</i> Elimination: zu geringe Itemschwierigkeit; Umformulierung schwer vorzunehmen; inhaltlich wird dieser Konsistenzaspekt auch durch andere Items abgedeckt	0,27	7,5	formativ
u29	<i>Dieser Shop ist gut strukturiert.</i> unveränderte Weiterverwendung → uf12	0,44	1,7	formativ
u30	<i>Es sind zu viele Schritte nötig, um zu einem Produkt zu gelangen.</i> unveränderte Weiterverwendung → uf13	0,60	0,8	formativ
u31	<i>Die einzelnen Seiten sind übersichtlich.</i> Elimination: inhaltliche Redundanz zu u32 bzw. uf14 bestätigt durch eine Korrelation von $r=0,69$	0,39	2,5	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).



Tab. B-1.3: Modifikationen und Eliminationen der ufosV1-Items (Items u32 bis u44)

Item-code	Item	$p_m$	% Missing	Messmodellannahme
u32	<i>Ich kann mich auf den Seiten des Shops gut zurechtfinden.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf14: <i>Ich finde mich auf den einzelnen Seiten des Shops gut zurecht.</i>	0,36	0	formativ
u33	<i>Ich kann die wichtigen Informationen immer gleich erkennen.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf15: <i>Ich finde in diesem Shop schnell die Informationen, die für meinen Einkauf wichtig sind.</i>	0,49	0	formativ
u34	<i>Ich kann gut auf die für mich wichtigen Informationen zugreifen.</i> Elimination: inhaltliche Redundanz zu u33 bzw. uf15 bestätigt durch eine Korrelation von $r=0,72$	0,43	2,5	formativ
u35	<i>Es werden häufig zu viele Informationen dargestellt.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf16: <i>Es werden in diesem Online-Shop häufig zu viele Informationen dargestellt.</i>	0,59	0,8	formativ
u36	<i>Ich kann bestimmen, welche Informationen angezeigt werden und welche nicht.</i> konkrete Bezugnahme auf den Bewertungsgegenstand; Itemschwierigkeit zwar zu hoch, aber durch Umformulierung schwer zu entschärfen → uf17: <i>Ich kann selbst bestimmen, welche Informationen vom Shop angezeigt werden und welche nicht.</i>	0,67	8,3	formativ
u37	<i>Die Texte sind nur schwer lesbar.</i> Umformierung, um Itemschwierigkeit zu reduzieren → uf18: <i>Die Texte sind manchmal nur schwer lesbar.</i>	0,73	0	formativ
u38	<i>Die Gestaltung unterstützt mich beim Benutzen des Shops.</i> Item erscheint nicht spezifisch genug, <i>Gestaltung</i> ist eher schwammig, daher Fokussierung auf das Design → uf19: <i>Das optische Design des Shops stört mich bei meinem Einkauf.</i>	0,51	1,7	unklar
u39	<i>Ich finde das Meiste dort, wo ich es erwarte.</i> Elimination: Inhaltlicher Aspekt der Selbstbeschreibungsfähigkeit wird durch Item u42 bzw. ur9 abgedeckt	0,43	3,3	formativ
u40	<i>Der Shop bietet mir alle Möglichkeiten, die ich von Online-Shops erwarte.</i> Umformulierung → ur8: <i>Der Shop bietet alle Möglichkeiten, die ich mir von einem Online-Shop wünsche.</i>	0,48	2,5	reflektiv
u41	<i>Wenn ich auf etwas klicke, verbirgt sich dahinter das, was ich erwarte.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf20: <i>Wenn ich in diesem Shop auf etwas klicke, passiert genau das, was ich erwarte.</i>	0,37	0,8	formativ
u42	<i>Ich kann den Shop so benutzen, wie ich es erwarte.</i> unveränderte Weiterverwendung → ur9	0,44	0	reflektiv
u43	<i>Ich finde die im Shop verwendete Sprache leicht verständlich.</i> Umformierung, um Itemschwierigkeit zu erhöhen; Item bietet sich für negative Polung an → uf21: <i>Manchmal finde ich die im Shop verwendeten Texte inhaltlich unverständlich.</i>	0,29	0	formativ
u44	<i>Die Benennungen sind so gewählt, dass ich gut nach Produkten suchen kann.</i> Umformierung, um die Güte der Benennungen im gesamten Shop abzudecken → uf22: <i>Die Benennungen sind so gewählt, dass ich den Shop gut benutzen kann.</i>	0,38	4,2	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).

Tab. B-1.4: Modifikationen und Eliminationen der ufosV1-Items (Items u45 bis u58)

Item-code	Item	$p_m$	% Missing	Messmodellannahme
u45	<i>Ich kann leicht herausfinden, ob der Shop die Produkte führt, die ich suche.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf23: <i>Es ist schnell erkennbar, welche Produkte ich in dem Online-Shop kaufen kann und welche nicht.</i>	0,37	0,8	formativ
u46	<i>Es ist leicht, nach Produkten zu suchen.</i> Es erscheint inhaltlich trivial, dass es leicht ist, nach Produkten zu suchen; entscheidender ist, ob gesuchte Produkte auch gefunden werden können → uf24: <i>Es ist leicht, gesuchte Produkte zu finden.</i>	0,37	0,8	formativ
u47	<i>Ich kann gut auf die Suchfunktionen des Shops zugreifen.</i> Umformierung, um Itemschwierigkeit zu erhöhen → uf25: <i>Ich kann sehr gut auf die Suchfunktionen des Shops zugreifen.</i>	0,29	1,7	formativ
u48	<i>Die Suchfunktionen sind einfach zu bedienen.</i> Elimination: inhaltliche Redundanz zu u49 bzw. uf26 bestätigt durch eine Korrelation von $r=0,67$	0,25	0,8	formativ
u49	<i>Die Suchfunktion ermöglicht mir, schnell bestimmte Produkte zu finden.</i> Umformulierungen, um konsistent mit Begriffen umzugehen (Suchfunktion vs. Suchfunktionen) → uf26: <i>Die Suchfunktionen ermöglichen mir, schnell bestimmte Produkte zu finden.</i>	0,34	0,8	formativ
u50	<i>Ich kann angeben, welche Eigenschaften ein von mir gesuchtes Produkt erfüllen soll.</i> Elimination: kann auch durch Experten beantwortet werden (Beantwortung ja/nein)	0,54	7,5	formativ
u51	<i>Die Produktsuche liefert mir sinnvolle Ergebnisse.</i> Umformulierungen, um konsistent mit Begriffen umzugehen (Produktsuche vs. Suchfunktionen) → uf27: <i>Die Suchfunktionen liefern mir sinnvolle Ergebnisse.</i>	0,40	0	formativ
u52	<i>Die Suche liefert auch dann nützliche Ergebnisse, wenn mein Suchbegriff ungenau ist.</i> Elimination: zu hoher Anteil fehlender Werte, kann auch schnell durch Experten beantwortet werden	0,57	11,7	formativ
u53	<i>Nach einer erfolglosen Suche erhalte ich nützliche Tipps zum weiteren Vorgehen.</i> Elimination: zu hoher Anteil fehlender Werte	0,69	29,2	formativ
u54	<i>Die Produktlisten im Shop sind übersichtlich.</i> unveränderte Weiterverwendung → uf29	0,45	2,5	formativ
u55	<i>Die Produkte sind in sinnvoll gewählte Kategorien eingeteilt.</i> Umformulierung, da <i>gewählte</i> überflüssig ist → uf29: <i>Die Produkte sind in sinnvolle Kategorien eingeteilt.</i>	0,40	4,2	formativ
u56	<i>Die Produkte sind sinnvoll angeordnet.</i> Elimination: inhaltliche Redundanz zu u55 bzw. uf29 bestätigt durch eine Korrelation von $r=0,8$	0,42	0	formativ
u57	<i>Ich kann verschiedene Produkte gut miteinander vergleichen.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich kann...</i> ) → uf30: <i>Ein Vergleich verschiedener Produkte ist leicht.</i>	0,57	12,5	formativ
u58	<i>Ich kann gut überprüfen, ob ein Produkt meinen Vorstellungen entspricht.</i> Elimination: inhaltliche Redundanz zu u59 bzw. uf31 bestätigt durch eine Korrelation von $r=0,66$	0,48	5,8	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).

Tab. B-1.5: Modifikationen und Eliminationen der ufosV1-Items (Items u59 bis u70)

Item-code	Item	$p_m$	% Missing	Messmodell-annahme
u59	<i>Ich bin mit den Beschreibungen der Produkte zufrieden.</i> unveränderte Weiterverwendung → uf31	0,46	0,8	formativ
u60	<i>Ich finde die verwendeten Bilder und Graphiken hilfreich.</i> unveränderte Weiterverwendung → uf32	0,41	2,5	formativ
u61	<i>Der Shop empfiehlt mir interessante Produkte, nach denen ich ursprünglich nicht gesucht habe.</i> unveränderte Weiterverwendung → uf33	0,53	7,5	formativ
u62	<i>Ich habe stets den Überblick, was die Produkte kosten.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich habe stets den Überblick...</i> ) und Erhöhung der Itemschwierigkeit → uf34: <i>Die Preise der einzelnen Produkte sind sofort ersichtlich.</i>	0,17	0	formativ
u63	<i>Ich habe stets den Überblick, welche zusätzlichen Kosten für Porto und Verpackung entstehen.</i> unveränderte Weiterverwendung → uf35	0,52	0,8	formativ
u64	<i>Ich erhalte nützliche Informationen über die Lieferbedingungen.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ) → uf36: <i>Der Shop bietet mir nützliche Informationen zu den Lieferbedingungen.</i>	0,37	0,8	formativ
u65	<i>Ich erhalte nützliche Informationen zur Rückgabe von Produkten.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ) → uf37: <i>Die Informationen zur Rückgabe von Produkten sind verständlich.</i>	0,47	7,5	formativ
u66	<i>Ich erhalte nützliche Informationen über die verschiedenen Zahlungsmöglichkeiten.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ) → uf38: <i>Ich werde gut über die verschiedenen Zahlungsmöglichkeiten informiert.</i>	0,32	4,2	formativ
u67	<i>Ich erhalte nützliche Informationen über die Lieferzeiten der einzelnen Produkte.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ), Item bietet sich für negative Polung an → uf39: <i>Man muss lange nach Angaben zu den Lieferzeiten der Produkte suchen.</i>	0,48	5,0	formativ
u68	<i>Ich erhalte nützliche Informationen über die allgemeinen Geschäftsbedingungen.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ) → uf40: <i>Es sind nützliche Informationen zu den allgemeinen Geschäftsbedingungen verfügbar.</i>	0,39	5,0	formativ
u69	<i>Ich erhalte nützliche Informationen über den Umgang mit meinen persönlichen Daten.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich erhalte nützliche...</i> ) → uf41: <i>Ich werde ausreichend darüber informiert, wie mit meinen persönlichen Daten umgegangen wird.</i>	0,47	8,3	formativ
u70	<i>Der Warenkorb ist übersichtlich.</i> unveränderte Weiterverwendung → uf42	0,21	1,7	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).

Tab. B1.6: Modifikationen und Eliminationen der ufosV1-Items (Items u71 bis u80)

Item-code	Item	$p_m$	% Missing	Messmodell-annahme
u71	<i>Ich weiß stets, welche Produkte sich gerade im Warenkorb befinden.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf43: <i>Ich weiß in diesem Shop stets, welche Produkte sich gerade im Warenkorb befinden.</i>	0,27	0	formativ
u72	<i>Es ist einfach, Produkte in den Warenkorb zu legen.</i> konkrete Bezugnahme auf den Bewertungsgegenstand → uf44: <i>Es ist in diesem Shop ganz einfach, Produkte in den Warenkorb zu legen.</i>	0,18	0,8	formativ
u73	<i>Ich kann den Inhalt meines Warenkorbes einfach verändern.</i> Umformierung, um Ähnlichkeit zwischen Items zu reduzieren (häufige Verwendung des Beginns <i>Ich habe stets den Überblick...</i> ) und Erhöhung der Itemschwierigkeit → uf45: <i>Der Inhalt des Warenkorbes kann ganz einfach verändert werden.</i>	0,25	0	formativ
u74	<i>Der gesamte Bestellvorgang ist leicht zu überschauen.</i> Umformulierung → uf46: <i>Der Bestellvorgang kann problemlos durchlaufen werden.</i>	0,35	0	formativ
u75	<i>Ich finde es schwer, mich anzumelden.</i> Elimination: zu hoher Anteil fehlender Werte; zu spezifischer Aspekt und eher weniger relevant	0,63	17,5	formativ
u76	<i>Der Shop macht mir die Eingabe von Daten einfach (z.B. Anschrift, Kreditkartennummer).</i> Anteil fehlender Werte ist hoch; Item aber hoch relevant, daher Umformulierung → uf47: <i>Die Eingabe meiner persönlichen Daten (z.B. Anschrift, Kreditkartennummer) während des Bestellvorgangs ist unkompliziert.</i>	0,27	19,2	formativ
u77	<i>Die Eingabeaufforderungen sind hilfreich.</i> Elimination: einerseits eher spezifischer Bewertungsaspekt, andererseits zu schwammig formuliert; wird inhaltlich auch durch u76 bzw. uf47 abgedeckt	0,35	9,2	formativ
u78	<i>Ich weiß bei einigen Eingaben nicht, warum ich sie machen muss.</i> Umformulierung → uf48: <i>Während des Bestellvorgangs weiß ich bei einigen persönlichen Angaben nicht, wofür sie benötigt werden.</i>	0,52	6,7	formativ
u79	<i>Ich erkenne sofort, ob meine Eingabe erfolgreich war.</i> Elimination: Inhaltlicher Aspekt wird durch das Item u76 bzw. uf47 abgedeckt	0,34	7,5	formativ
u80	<i>Es dauert zu lange, bis der Shop auf meine Eingaben reagiert.</i> unveränderte Weiterverwendung → uf49	0,62	0	formativ

Anmerkung: Itemschwierigkeit ( $p_m$ ), Anteil fehlender Werte (% Missing) aus Studie von Konradt et al. (2003).

Tab. B-2: Quellen für neue formative Usability-Items der ufosV2f-Skala

<i>Neuer Item Code</i>	<i>ufos-Item Originalitem (Quelle)</i>
uf50	<i>Der Ablauf des Bestellvorgangs gefällt mir.</i> <i>Show the steps of the process</i> (Guideline Nr. 172; Nielsen, Molich & Snyder, 2001)
uf51	<i>Die Auflistungen der Produkte lassen sich bequem nach meinen Bedürfnissen ordnen.</i> <i>Allow customers to sort products by the factors they care about most</i> (Guideline Nr. 97; Nielsen et al. 2001)
uf52	<i>Ich erhalte nützliche Informationen zum Unternehmen, das den Shop betreibt.</i> <i>Show detailed company info</i> (Guideline Nr. 48; Nielsen et al. 2001)
uf53	<i>Ich bin mit den Informationen zur Sicherheit des Datentransfers zufrieden.</i> <i>Include a prominent link to privacy and security policies</i> (Guideline Nr. 183; Nielsen et al. 2001)
uf54c	<i>Die Werbung auf den Seiten des Shops lenkt mich vom Einkaufen ab.</i> (Bucher & Jäckel, 2002)
uf55	<i>Die Startseite dieses Shops stellt einen guten Ausgangspunkt für meinen Einkauf dar.</i> (Bucher & Jäckel, 2002)
uf56c	<i>Bei der Benutzung des Shops treten störende technische Probleme auf.</i> <i>Pursue technical problems vigilantly - and fix them</i> (Guideline Nr. 58; Nielsen et al. 2001)
uf57c	<i>Ich vermisse wichtige Informationen, die ich für meine Kaufentscheidung benötige.</i> Selbst entwickeltes Items als Zusammenfassung für alle Items zur Verfügbarkeit von Informationen
uf58c	<i>Die Ladezeiten der Seiten sind bei diesem Online-Shop zu lang.</i> <i>The Web Site does not take long to load</i> (Charitou, Georgiou & Soteriou (2003) <i>It does not take a long time to load the front page of the site.</i> (Kim, Lee, Han & Lee, 2002)



## **Anhang C**

### **Ergänzende Darstellungen zur Untersuchungsplanung**

---

Tab. C-1: Originalitems zur reflektiven Erfassung des Konstruktes Vertrauen.....	305
Tab. C-2: Originalitems zur reflektiven Erfassung des Konstruktes Einkaufsvergnügen...	305
Tab. C-3: Originalitems zur reflektiven Erfassung des Konstruktes Ästhetik .....	305
Tab. C-4: Originalitems zur Erfassung des Konstruktes wahrgenommener Preis .....	306
Tab. C-5: Originalitems zur reflektiven Erfassung des Konstruktes Service.....	306
Tab. C-6: Originalitems zur formativen Erfassung des Konstruktes Kaufintention .....	306
Tab. C-7: Originalitems zur reflektiven Erfassung des Produkt Involvement (Skala „Importance“ von McQuarrie & Munson, 1992).....	306
Tab. C-8: Ausschnitt des Versuchsplans – 84 Versuchsbedingungen der Produktart Bücher .....	307



Tab. C-1: Originalitems zur reflektiven Erfassung des Konstruktes Vertrauen

Neuer Item Code	Verwendetes Item zum Vertrauen Originalitem (Quelle)	% Missing	$p_m$	$r_{it}$
t1	Dieser Online-Shop ist vertrauenswürdig. (Aktuelle Untersuchung)			
	This store is trustworthy. (Jarvenpaa et al., 2000)	-	-	-
	Dieser Online-Shop ist vertrauenswürdig. (Konradt et al., 2003)	5,8	0,51	0,61
t2	Ich bin mir sicher, dass der Online-Shop seine Versprechen und Verpflichtungen einhalten wird. (Aktuelle Untersuchung)			
	This store wants to be known as one who keeps promises and commitments. (Jarvenpaa et al., 2000)	-	-	-
	Es ist dem Online-Shop wichtig, dass er Versprechen und Verpflichtungen einhält. (Konradt et al., 2003)	26,7	0,50	0,53
t3	Ich vertraue darauf, dass dieser Shop 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. (Konradt et al., 2003)	5	0,53	0,59
t4	Ich habe das Gefühl, dieser Online-Shop ist ehrlich. (Aktuelle Untersuchung)			
	I would characterize LegalAdvice.com as honest. (McKnight, 2002)	-	-	-

Anmerkung: Anteil fehlender Werte (% Missing), Itemschwierigkeit ( $p_m$ ), korrigierte Trennschärfe ( $r_{it}$ ).

Tab. C-2: Originalitems zur reflektiven Erfassung des Konstruktes Einkaufsvergnügen

Neuer Item Code	Verwendetes Item zum Einkaufsvergnügen Originalitem (Quelle)
se1	Es macht wirklich Spaß, bei diesem Online-Shop einzukaufen. (Aktuelle Untersuchung)
	It's really fun to shop at this website. (Wolfenbarger & Gilly, 2003)
se2c	Die Benutzung dieses Online-Shops frustriert mich. (Aktuelle Untersuchung)
	During my last visit to Booksamillion.com I felt frustrated. (Ghani, Supnick & Rooney, 1991; Koufaris, 2002)

Tab. C-3: Originalitems zur reflektiven Erfassung des Konstruktes Ästhetik

Neuer Item Code	Verwendetes Item zur Ästhetik Originalitem (Quelle)
a1	Ich finde es optisch ansprechend, wie dieser Online-Shop seine Produkte anbietet. (Aktuelle Untersuchung)
	XYZ's Internet site is aesthetically appealing. (Mathwick, Malhotra & Rigdon, 2001)
a2	Der Internetauftritt dieses Online-Shops ist ästhetisch gestaltet. (Aktuelle Untersuchung)
	The Way XYZ displays its products is attractive. (Mathwick, Malhotra & Rigdon, 2001)
	The site's design is beautiful. (Lowengart & Tractinsky (2001)

Tab. C-4: Originalitems zur Erfassung des Konstruktes wahrgenommener Preis

Neuer Item Code	Verwendetes Item zum wahrgenommenen Preis Originalitem (Quelle)
subpc	Die Preise in diesem Shop sind mir zu hoch. (Aktuelle Untersuchung)
	Overall, I am happy with XYZ's prices. (Mathwick, Malhotra & Rigdon, 2001)
	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)

Tab. C-5: Originalitems zur formativen Erfassung des Konstruktes Service

Neuer Item Code	Verwendetes Item zum Service Originalitem (Quelle)
sv3	Ich finde, der Online-Shop hat eine gute Regelung zur Rückgabe von Produkten. (Aktuelle Untersuchung)
	Returning items is relatively straightforward. (Wolfenbarger & Gilly, 2003)
	The return policy at this site is reasonable. (Wolfenbarger & Gilly, 2003)
sv4c	Die Lieferzeiten der Produkte empfinde ich bei diesem Online-Shop als zu lang. (Aktuelle Untersuchung)
	You get your merchandise quickly when you order. (Wolfenbarger & Gilly, 2003)
sv5c	Die Kosten, die der Online-Shop für Porto und Verpackung berechnet, sind mir zu hoch. (Aktuelle Untersuchung)
	The website has reasonable shipping and handling costs. (Wolfenbarger & Gilly, 2003)

Tab. C-6: Originalitems zur reflektiven Erfassung des Konstruktes Kaufintention

Neuer Item Code	Verwendetes Item zur Kaufintention Originalitem (Quelle)	% Missing	$p_m$	$r_{it}$
ki1	Ich werde diesen Online-Shop nach dem Versuch bald wieder aufsuchen. (Aktuelle Untersuchung)			
	How likely is it that you would return to this store's web site? (Jarvenpaa et al., 2000)	-	-	-
	Wie wahrscheinlich ist es, dass Sie diesen Online-Shop wieder aufsuchen werden? (Konradt et al., 2003)	0,8	0,68	0,74
ki2	Ich denke, dass ich innerhalb der nächsten drei Monate in diesem Online-Shop bestellen werde. (Aktuelle Untersuchung)			
	How likely is that you would consider purchasing from this store in the next 3 months? (Jarvenpaa et al., 2000)	-	-	-
	Wie wahrscheinlich ist es, dass sie innerhalb der nächsten drei Monate etwas in diesem Shop bestellen? (Konradt et al., 2003)	0	0,79	0,74
ki3	Es ist recht wahrscheinlich, dass ich im Rahmen des laufenden Versuchs in diesem Shop bestelle. (Aktuelle Untersuchung)			
	For this purchase, how likely is it that you buy from this store? (Jarvenpaa et al., 2000)	-	-	-
	Wie wahrscheinlich ist es, dass Sie im Rahmen des laufenden Versuches etwas in diesem Shop bestellen? (Konradt et al., 2003)	2,5	0,70	0,49

Anmerkung: Anteil fehlender Werte (% Missing), Itemschwierigkeit ( $p_m$ ), korrigierte Trennschärfe ( $r_{it}$ ).

Tab. C-7: Originalitems zur reflektiven Erfassung des Produkt Involvement (Skala „Importance“ von McQuarrie &amp; Munson, 1992)

We would like to know how interested you are in books. Please use the series of descriptive words listed below to indicate your level of interest in books:	
Important .....	Unimportant
Irrelevant .....	Relevant
Means a lot to me .....	Means nothing to me
Of no concern to me .....	Of concern to me
Matters to me .....	Doesn't matter to me

Tab. C-8.1: Ausschnitt des Versuchsplans – 84 Versuchsbedingungen der Produktart Bücher

<i>Teilnehmer- code</i>	<i>Erster präsentierter Online-Shop</i>	<i>Zweiter präsentierter Online-Shop</i>	<i>Itempräsentations- reihenfolge</i>
b1_2_1	Karstadt.de	JPC.de	reflektiv-formativ
b1_2_2	Karstadt.de	JPC.de	formativ-reflektiv
b1_3_1	Karstadt.de	Libri.de	reflektiv-formativ
b1_3_2	Karstadt.de	Libri.de	formativ-reflektiv
b1_4_1	Karstadt.de	Booxtra.de	reflektiv-formativ
b1_4_2	Karstadt.de	Booxtra.de	formativ-reflektiv
b1_5_1	Karstadt.de	BOL.de	reflektiv-formativ
b1_5_2	Karstadt.de	BOL.de	formativ-reflektiv
b1_6_1	Karstadt.de	Buchkatalog.de	reflektiv-formativ
b1_6_2	Karstadt.de	Buchkatalog.de	formativ-reflektiv
b1_7_1	Karstadt.de	Buch24.de	reflektiv-formativ
b1_7_2	Karstadt.de	Buch24.de	formativ-reflektiv
b2_1_1	JPC.de	Karstadt.de	reflektiv-formativ
b2_1_2	JPC.de	Karstadt.de	formativ-reflektiv
b2_3_1	JPC.de	Libri.de	reflektiv-formativ
b2_3_2	JPC.de	Libri.de	formativ-reflektiv
b2_4_1	JPC.de	Booxtra.de	reflektiv-formativ
b2_4_2	JPC.de	Booxtra.de	formativ-reflektiv
b2_5_1	JPC.de	BOL.de	reflektiv-formativ
b2_5_2	JPC.de	BOL.de	formativ-reflektiv
b2_6_1	JPC.de	Buchkatalog.de	reflektiv-formativ
b2_6_2	JPC.de	Buchkatalog.de	formativ-reflektiv
b2_7_1	JPC.de	Buch24.de	reflektiv-formativ
b2_7_2	JPC.de	Buch24.de	formativ-reflektiv
b3_1_1	Libri.de	Karstadt.de	reflektiv-formativ
b3_1_2	Libri.de	Karstadt.de	formativ-reflektiv
b3_2_1	Libri.de	JPC.de	reflektiv-formativ
b3_2_2	Libri.de	JPC.de	formativ-reflektiv
b3_4_1	Libri.de	Booxtra.de	reflektiv-formativ
b3_4_2	Libri.de	Booxtra.de	formativ-reflektiv
b3_5_1	Libri.de	BOL.de	reflektiv-formativ
b3_5_2	Libri.de	BOL.de	formativ-reflektiv
b3_6_1	Libri.de	Buchkatalog.de	reflektiv-formativ
b3_6_2	Libri.de	Buchkatalog.de	formativ-reflektiv
b3_7_1	Libri.de	Buch24.de	reflektiv-formativ
b3_7_2	Libri.de	Buch24.de	formativ-reflektiv
b4_1_1	Booxtra.de	Karstadt.de	reflektiv-formativ
b4_1_2	Booxtra.de	Karstadt.de	formativ-reflektiv
b4_2_1	Booxtra.de	JPC.de	reflektiv-formativ
b4_2_2	Booxtra.de	JPC.de	formativ-reflektiv
b4_3_1	Booxtra.de	Libri.de	reflektiv-formativ
b4_3_2	Booxtra.de	Libri.de	formativ-reflektiv
b4_5_1	Booxtra.de	BOL.de	reflektiv-formativ
b4_5_2	Booxtra.de	BOL.de	formativ-reflektiv
b4_6_1	Booxtra.de	Buchkatalog.de	reflektiv-formativ
b4_6_2	Booxtra.de	Buchkatalog.de	formativ-reflektiv
b4_7_1	Booxtra.de	Buch24.de	reflektiv-formativ
b4_7_2	Booxtra.de	Buch24.de	formativ-reflektiv
b5_1_1	BOL.de	Karstadt.de	reflektiv-formativ
b5_1_2	BOL.de	Karstadt.de	formativ-reflektiv
b5_2_1	BOL.de	JPC.de	reflektiv-formativ
b5_2_2	BOL.de	JPC.de	formativ-reflektiv
b5_3_1	BOL.de	Libri.de	reflektiv-formativ
b5_3_2	BOL.de	Libri.de	formativ-reflektiv

Tab. C-8.2: Ausschnitt des Versuchsplans – 84 Versuchsbedingungen der Produktart Bücher

<i>Teilnehmer- code</i>	<i>Erster präsentierter Online-Shop</i>	<i>Zweiter präsentierter Online-Shop</i>	<i>Itempräsentations- reihenfolge</i>
b5_4_1	BOL.de	Booxtra.de	reflektiv-formativ
b5_4_2	BOL.de	Booxtra.de	formativ-reflektiv
b5_6_1	BOL.de	Buchkatalog.de	reflektiv-formativ
b5_6_2	BOL.de	Buchkatalog.de	formativ-reflektiv
b5_7_1	BOL.de	Buch24.de	reflektiv-formativ
b5_7_2	BOL.de	Buch24.de	formativ-reflektiv
b6_1_1	Buchkatalog.de	Karstadt.de	reflektiv-formativ
b6_1_2	Buchkatalog.de	Karstadt.de	formativ-reflektiv
b6_2_1	Buchkatalog.de	JPC.de	reflektiv-formativ
b6_2_2	Buchkatalog.de	JPC.de	formativ-reflektiv
b6_3_1	Buchkatalog.de	Libri.de	reflektiv-formativ
b6_3_2	Buchkatalog.de	Libri.de	formativ-reflektiv
b6_4_1	Buchkatalog.de	Booxtra.de	reflektiv-formativ
b6_4_2	Buchkatalog.de	Booxtra.de	formativ-reflektiv
b6_5_1	Buchkatalog.de	BOL.de	reflektiv-formativ
b6_5_2	Buchkatalog.de	BOL.de	formativ-reflektiv
b6_7_1	Buchkatalog.de	Buch24.de	reflektiv-formativ
b6_7_2	Buchkatalog.de	Buch24.de	formativ-reflektiv
b7_1_1	Buch24.de	Karstadt.de	reflektiv-formativ
b7_1_2	Buch24.de	Karstadt.de	formativ-reflektiv
b7_2_1	Buch24.de	JPC.de	reflektiv-formativ
b7_2_2	Buch24.de	JPC.de	formativ-reflektiv
b7_3_1	Buch24.de	Libri.de	reflektiv-formativ
b7_3_2	Buch24.de	Libri.de	formativ-reflektiv
b7_4_1	Buch24.de	Booxtra.de	reflektiv-formativ
b7_4_2	Buch24.de	Booxtra.de	formativ-reflektiv
b7_5_1	Buch24.de	BOL.de	reflektiv-formativ
b7_5_2	Buch24.de	BOL.de	formativ-reflektiv
b7_6_1	Buch24.de	Buchkatalog.de	reflektiv-formativ
b7_6_2	Buch24.de	Buchkatalog.de	formativ-reflektiv

## **Anhang D**

### **Ergänzende Darstellungen zur Durchführung**

### Verwendete Werbemittel

Email zur Teilnehmerakquise.....	312
Abb. D-1: Flyer zur Untersuchung.....	313
Pressemitteilung der Universität Kiel .....	314
Abb. D-2.1: Screenshot der Werbepage - Startseite.....	315
Abb. D-2.2: Screenshot der Werbepage - Kontaktformular.....	315
Abb. D-2.3: Screenshot der Werbepage - Übersicht über den Teilnehmerbedarf .....	316
Abb. D-2.4: Screenshot der Werbepage - Bilder zur Untersuchung.....	316
Abb. D-2.5: Screenshot der Werbepage - Impressum.....	317
Tab. D-1: Versuchsbedingungen zur Produktart Veranstaltungsticket, für die keine Probanden erhoben wurden.....	317

### Nähere Beschreibung der Stichprobe im Vergleich zur Population

Tab. D-2: Geschlechterverteilung in Altersgruppen in der Population - Angaben in Prozent .....	318
Tab. D-3: Kreuztabelle Produktart x Alter x Geschlecht.....	318
Tab. D-4: Einkommen in der Population.....	319
Tab. D-5: Einkommen in der Stichprobe.....	319
Tab. D-6: Bildungsstand in der Population .....	319
Tab. D-7: Bildungsstand in der Stichprobe .....	319
Tab. D-8: Beschäftigung in der Population .....	319
Tab. D-9: Beschäftigung in der Stichprobe .....	319
Tab. D-10: Interneterfahrung in der Population .....	319
Tab. D-11: Interneterfahrung der Teilnehmer in Jahren.....	319
Tab. D-12: PC-Nutzungserfahrung der Teilnehmer in Jahren .....	320
Tab. D-13: PC-Nutzungshäufigkeit der Teilnehmer in Tagen pro Monat.....	320
Tab. D-14: PC-Nutzungshäufigkeit der Teilnehmer in Stunden/Tag.....	320
Tab. D-15: Internet-Nutzungshäufigkeit der Teilnehmer in Stunden/Tag .....	320
Tab. D-16: Internet-Nutzungshäufigkeit der Teilnehmer in Tagen/Monat .....	320
Tab. D-17: Online-Einkäufe der Teilnehmer in der Vergangenheit.....	320

---

Tab. D-18: Online-Einkäufe für die jeweilige Produktart .....	320
---	-----

### **Screenshots zentraler Seiten der Hypertext-Versuchsumgebung**

Abb. D-3.1: Hypertext-Versuchsumgebung Seite 1 - Begrüßung.....	321
Abb. D-3.2: Hypertext-Versuchsumgebung Seite 2 - Festlegung auf Produkt .....	321
Abb. D-3.3: Hypertext-Versuchsumgebung Seite 3 - Überblick über den Versuchsablauf.	322
Abb. D-3.4: Hypertext-Versuchsumgebung Seite 6 - erster Shop Aufgabe 1 (Teil 1) .....	322
Abb. D-3.5: Hypertext-Versuchsumgebung Seite 7 - erster Shop Aufgabe 1 (Teil 2) .....	323
Abb. D-3.6: Hypertext-Versuchsumgebung Seite 12 - Bewertung erster Shop (Teil 1).....	323
Abb. D-3.7: Hypertext-Versuchsumgebung Seite 24 - Kaufentscheidung .....	324
Abb. D-3.8: Hypertext-Versuchsumgebung Seite 25 - Bestellung .....	324
Abb. D-3.9: Hypertext-Versuchsumgebung Seite 27 - Wahl des Gutscheins.....	325

### **Informationen zum Versuchsablauf**

Einverständniserklärung zur Teilnahme an der Untersuchung .....	325
Teilnehmerinstruktionen der Hauptuntersuchung.....	326

## **Verwendete Werbemittel**

### **Email zur Teilnehmerakquise**

Betreff: Studie zum Einkaufen im Internet - 15 Euro bei Teilnahme

Hallo,

im Rahmen einer Doktorarbeit an der Kieler Uni werden zurzeit Teilnehmer für eine Untersuchung zum Einkaufen im Internet gesucht. Wer mitmacht, bekommt 15 Euro bei einem Online-Kauf erstattet.

Die Untersuchung dauert ca. eine Stunde. Zur Versuchsdurchführung müssten Sie in einen der Computerräume der Kieler Uni kommen. Sie werden dort zwei Online-Shops kennen lernen. Am Ende des Versuchs sollen Sie einen der Shops auswählen und ein von Ihnen frei gewähltes Produkt bestellen. Als Belohnung für die Teilnahme erstatten wir anteilig 15 Euro des Bestellwertes. Es ist notwendig, dass Sie sich im Vorfeld festlegen, welches Produkt Sie bestellen möchten. Möglich ist die Bestellung eines beliebigen Produktes aus den Kategorien: Musik-Tonträger (CDs, LPs etc.), Bücher, DVDs, Druckerpatronen und Veranstaltungs- & Konzerttickets.

Leider suchen wir in einigen Altersgruppen bereits keine weiteren Teilnehmer mehr. Konkrete Informationen und ein Anmeldeformular finden Sie im Internet auf der Seite [www.uni-kiel.de/studie-internet-einkauf/](http://www.uni-kiel.de/studie-internet-einkauf/). Sie können sich auch per Email ([christophersen@bwl.uni-kiel.de](mailto:christophersen@bwl.uni-kiel.de)) oder Telefon (0431 - 8801592) an uns wenden.

Hoffentlich bis bald!

Beste Grüße,

Timo Christophersen



Christian-Albrechts-Universität zu Kiel



**15,-  
Euro**

Februar 2006

## Studie zum Einkaufen im Internet

### ★ Kauf von Druckerzubehör ★

Hast Du Interesse, in einem ganz normalen Online-Shop für Druckerzubehör eine Tintenpatrone oder Tonerkartusche Deiner Wahl zu bestellen und als Gegenleistung **15 Euro** in bar ausgezahlt zu bekommen?

Im Rahmen der ca. einstündigen Untersuchung wirst Du zwei verschiedene Online-Shops zum Kauf von Druckerzubehör kennen lernen und anhand eines Fragebogens bewerten. Am Ende der Untersuchung sollst Du einen der beiden Shops auswählen und dort eine Patrone oder Kartusche Deiner Wahl bestellen.

Wir suchen Teilnehmer in allen Altersgruppen!

**Als Belohnung für die Teilnahme an der Untersuchung erhältst Du von uns 15 Euro, die Du dazu nutzen kannst, die Kosten für die Bestellung zu begleichen.**

Es ist notwendig, dass Du Dich bereits im Vorfeld des Untersuchungstermins festlegst und uns mitteilst, welche Patrone oder Kartusche Du bestellen möchtest. Die Durchführung der Untersuchung erfolgt in einem der Computerräume an der Uni Kiel. Weitere Informationen sowie ein elektronisches Anmeldeformular findest Du unter der folgenden Internetadresse:

**[www.uni-kiel.de/studie-internet-einkauf/](http://www.uni-kiel.de/studie-internet-einkauf/)**

Dipl.-Psych. Timo Christophersen  
Email: christophersen@bwl.uni-kiel.de; Tel: 0431 / 880 1592

---

Die Studie verfolgt rein wissenschaftliche Ziele; es besteht keinerlei Kontakt zu den Online-Shops. Das Projekt wird im Rahmen einer Doktorarbeit und mehrerer Diplomarbeiten bearbeitet und von Seiten der Deutschen Forschungsgemeinschaft (DFG) finanziell unterstützt. Doktorvater ist Prof. Dr. Konradt, Lehrstuhlsthlinhaber für Arbeits, Organisations & Marktpsychologie.

Abb. D-1: Flyer zur Untersuchung

## **Pressemitteilung der Universität Kiel**

Einkaufen im Netz

Uni Kiel sucht Teilnehmer für Studie

Für eine Studie zum Interneteinkauf sucht die Kieler Universität noch zirka 300 Teilnehmer im Alter über 18 Jahren, die gelegentlich das Internet nutzen. Dabei sollen sie vor Ort in einem der Kieler Universitäts-Computerräume ein Produkt ihrer Wahl (CD, DVD, Buch, Veranstaltungsticket, Druckerpatrone) online bestellen, auf das sie sich vorab festgelegt haben. Diplom-Psychologe Timo Christophersen untersucht für seine Doktorarbeit, welche Faktoren für Kunden beim Einkaufen im Internet wichtig sind, zum Beispiel Angebotspalette, Preis oder Bedienungsfreundlichkeit des Online-Shops. Jeder Teilnehmer erhält für die zirka einstündige Sitzung und als Ausgleich für die Bestellung 15 Euro. Weitere Informationen und den Anmeldebogen gibt es im Internet unter [www.uni-kiel.de/grad-kolleg-bwl/online-shopping/werbepage](http://www.uni-kiel.de/grad-kolleg-bwl/online-shopping/werbepage) oder beim wissenschaftlichen Mitarbeiter Oliver Harm, Tel.: 0431/94521.

Ziel der Untersuchungen ist es, einen Fragebogen zu entwickeln, mit dem Online-Shops bewertet werden können. Die Ergebnisse der Studie sollen im Februar nächsten Jahres vorliegen. Der Studienleiter Timo Christophersen promoviert im Graduiertenkolleg "Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business", in dem mehrere Institute der Christian-Albrechts-Universität zu Kiel interdisziplinär forschen. In diesem Rahmen kann er das Thema, das Marketing und Psychologie vereint, mit Betreuern aus beiden Bereichen bearbeiten. Informationen zum Graduiertenkolleg gibt es unter: [www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/de/index.html](http://www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/de/index.html).

Kontakt:

Christian-Albrechts-Universität zu Kiel

Graduiertenkolleg "Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business"

Timo Christophersen

e-mail: [christophersen@bwl.uni-kiel.de](mailto:christophersen@bwl.uni-kiel.de)

The screenshot shows a web browser window with the URL <http://www.uni-kiel.de/grad-kolleg-bwl/online-shopping/werbepage/index.php>. The page title is "Untersuchung zum Einkaufen im Internet an der Uni Kiel". The navigation menu includes "Startseite", "Kontakt", and "Impressum".

**Haben Sie Interesse, bei einer Bestellung in einem Online-Shop 15 Euro des Bestellwertes erstattet zu bekommen?**

Dann lesen Sie weiter:

Im Rahmen der ca. einstündigen Untersuchung werden Sie zwei verschiedene Online-Shops kennen lernen. Am Ende der Untersuchung sollen Sie einen der beiden Shops auswählen und dort ein von Ihnen frei gewähltes Produkt bestellen. Als Belohnung für Ihre Teilnahme erstatten wir Ihnen 15 Euro des Bestellwertes.

Es ist notwendig, dass Sie sich bereits im Vorfeld des Untersuchungstermins auf ein ganz bestimmtes Produkt festlegen, welches Sie während Ihrer Teilnahme bestellen möchten. Bitte teilen Sie uns Ihr ausgewähltes Produkt möglichst präzise beschrieben mit (z.B. Künstler, Veranstaltungsort und -zeit, Druckerbezeichnung und Patrone schwarz oder Farbe, Titel und Autor). Es kann ein spezifisches Produkt aus den folgenden Kategorien frei ausgewählt werden (die Links in den Klammern sollen als eventuelle Inspirationsquellen dienen):

xxx (xxx)  
xxx (xxx)

Da bereits einige Untersuchungstermine stattgefunden haben, suchen wir in bestimmten Altersgruppen keine Teilnehmer mehr. Bitte sehen Sie hier nach, für welche Produktkategorien in Ihrer Altersgruppe noch Teilnehmer benötigt werden: Übersicht "Bedarf" an Untersuchungsteilnehmern.

Das Mindestalter zur Teilnahme ist 18 Jahre. Aufgrund einer hohen Anzahl notwendiger Teilnehmer wird die Untersuchung vermutlich frühestens im Februar 2006 abgeschlossen sein. Bei Interesse an einer Teilnahme füllen Sie bitte das **Kontaktformular** aus. Alle Angaben werden streng vertraulich behandelt; es erfolgt keine Weitergabe an Dritte.

Die Untersuchung wird an der Christian-Albrechts-Universität zu Kiel im Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), EDV-Raum des Psychologischen Instituts, 4. Stock, Raum 401 b, Olshausenstr. 62 durchgeführt.

So finden Sie den Raum 401 b ganz leicht: Wenn Sie im IPN-Gebäude im 4. Stock aus einem der beiden Fahrstühle treten, wenden Sie sich bitte nach rechts und gehen durch die Glastür. Dann gehen Sie den Gang bis ganz zum Ende und durch eine weitere Glastür. Die linke Tür gegenüber den Schließfächern ist der EDV-Raum 401 b.

Hier finden Sie eine **Wegbeschreibung** zum IPN.

**Sie können uns auch direkt kontaktieren:**

**Dirk-Psych. Timo Christophersen** **Oliver Harm**

Abb. D-2.1: Screenshot der Werbepage - Startseite

The screenshot shows the contact form page with the URL <http://www.uni-kiel.de/grad-kolleg-bwl/online-shopping/werbepage/kontakt.php>. The page title is "Kontaktformular / Anmeldung". The navigation menu includes "Startseite", "Kontakt", and "Impressum".

Bitte füllen Sie die mit \* gekennzeichneten Felder in jedem Fall aus.

**Aktueller 'Bedarf' an Teilnehmern**

Vorname: \*   
Name: \*   
eMail: \*   
Telefon 1:   
Telefon 2 (z.B. Handy):   
Bevorzugte Termine:   
Alter:   
Beruf:   
Geschlecht: männlich   
Ausgewähltes Produkt (genau beschrieben):   
Wodurch sind Sie auf uns aufmerksam geworden?

"Bitte beachten Sie, dass die Teilnahme an der Studie in einem der Computerräume der Universität Kiel und nicht etwa online stattfindet."

[↑ top](#)

Abb. D-2.2: Screenshot der Werbepage - Kontaktformular

Übersicht 'Bedarf' an Untersuchungsteilnehmern

	Altersklassen									
	18-29		30-39		40-49		50-59		60 und älter	
	w	m	w	m	w	m	w	m	w	m
1.) xxx										
2.) xxx										

Admin

<span style="color: red;">■</span>	Es sind bereits genug Untersuchungsteilnehmer vorhanden
<span style="color: green;">■</span>	Es werden noch Untersuchungsteilnehmer gesucht
w	weibliche Teilnehmer
m	männliche Teilnehmer

[zurück zum Kontaktformular](#)

Startseite | Kontakt | Impressum | Letzte Aktualisierung: 16. 12. 2005.

Abb. D-2.3: Screenshot der Werbepage - Übersicht über den Teilnehmerbedarf

Untersuchung zum Einkaufen im Internet an der Uni Kiel

Einige Bilder

© Ole Hagemann (oha)

© Ole Hagemann (oha)

© Ole Hagemann (oha)

© Ole Hagemann (oha)

Startseite | Kontakt | Impressum |

Abb. D-2.4: Screenshot der Werbepage - Bilder zur Untersuchung



Abb. D-2.5: Screenshot der Werbepage - Impressum

Tab. D-1: Versuchsbedingungen zur Produktart Veranstaltungsticket, für die keine Probanden erhoben wurden

Vpn-Code	Shop 1	Shop 2
t1 4 2	Eventim.de	Kartenhaus.de
t1 5 2	Eventim.de	Karten-Online.de
t1 6 1	Eventim.de	TicketOnline.de
t1 6 2	Eventim.de	TicketOnline.de
t1 7 1	Eventim.de	TicketCenter.de
t1 7 2	Eventim.de	TicketCenter.de
t2 4 1	Getgo.de	Kartenhaus.de
t2 4 2	Getgo.de	Kartenhaus.de
t2 5 1	Getgo.de	Karten-Online.de
t2 5 2	Getgo.de	Karten-Online.de
t2 6 2	Getgo.de	TicketOnline.de
t2 7 1	Getgo.de	TicketCenter.de
t2 7 2	Getgo.de	TicketCenter.de
t3 5 2	Kartenkaufen.de	Karten-Online.de
t3 6 1	Kartenkaufen.de	TicketOnline.de
t3 6 2	Kartenkaufen.de	TicketOnline.de
t3 7 2	Kartenkaufen.de	TicketCenter.de
t4 1 2	Kartenhaus.de	Eventim.de
t4 2 2	Kartenhaus.de	Getgo.de
t4 7 2	Kartenhaus.de	TicketCenter.de
t5 1 2	Karten-Online.de	Eventim.de
t5 2 1	Karten-Online.de	Getgo.de

Vpn-Code	Shop 1	Shop 2
t5 2 2	Karten-Online.de	Getgo.de
t5 3 1	Karten-Online.de	Kartenkaufen.de
t5 3 2	Karten-Online.de	Kartenkaufen.de
t5 4 1	Karten-Online.de	Kartenhaus.de
t5 4 2	Karten-Online.de	Kartenhaus.de
t5 6 2	Karten-Online.de	TicketOnline.de
t5 7 2	Karten-Online.de	TicketCenter.de
t6 1 2	TicketOnline.de	Eventim.de
t6 2 2	TicketOnline.de	Getgo.de
t6 3 1	TicketOnline.de	Kartenkaufen.de
t6 3 2	TicketOnline.de	Kartenkaufen.de
t6 5 1	TicketOnline.de	Karten-Online.de
t6 5 2	TicketOnline.de	Karten-Online.de
t7 1 2	TicketCenter.de	Eventim.de
t7 2 1	TicketCenter.de	Getgo.de
t7 2 2	TicketCenter.de	Getgo.de
t7 3 1	TicketCenter.de	Kartenkaufen.de
t7 3 2	TicketCenter.de	Kartenkaufen.de
t7 4 2	TicketCenter.de	Kartenhaus.de
t7 5 2	TicketCenter.de	Karten-Online.de
t7 6 2	TicketCenter.de	TicketOnline.de

## Nähere Beschreibung der Stichprobe im Vergleich zur Population

Tab. D-2: Geschlechterverteilung in Altersgruppen in der Population – Angaben in Prozent

	Altersgruppen						Gesamt
	14-19	20-29	30-39	40-49	50-59	60+	
männlich	51,3	53,7	53,8	57,2	57,5	64,5	55,7
weiblich	48,7	46,3	46,2	42,8	42,5	35,5	44,3

Anmerkung: Die Angaben zur Population entstammen der Studie Internet Facts 2006-III (AGOF, 2006).

Tab. D-3: Kreuztabelle Produktart x Alter x Geschlecht

Produktart	Alter		Geschlecht		Gesamt
			♀	♂	
Bücher	18-29	n	14	17	31
		%	45,2	54,8	36,9
	30-39	n	11	14	25
		%	44	56	29,8
	40-49	n	10	6	16
		%	62,5	37,5	19,0
	50-59	n	5	2	7
%		71,4	28,6	8,3	
60+	n	3	2	5	
%	60	40	6		
Gesamt	n	43	41	84	
%	51,2	48,8	100		
DVDs	-18	n	1	0	1
		%	100	0	1,2
	18-29	n	31	23	54
		%	57,4	42,6	63,5
	30-39	n	6	12	18
		%	33,3	66,7	21,2
	40-49	n	2	5	7
%		28,6	71,4	8,2	
50-59	n	1	2	3	
	%	33,3	66,7	3,5	
60+	n	0	2	2	
%	0	100	2,4		
Gesamt	n	41	44	85	
%	48,2	51,8	100		
CDs	-18	n	0	3	3
		%	0	100	3,6
	18-29	n	12	11	23
		%	52,2	47,8	27,4
	30-39	n	9	12	21
		%	42,9	57,1	25
	40-49	n	9	11	20
%		45	55	23,8	
50-59	n	5	6	11	
	%	45,5	54,5	13,1	
60+	n	3	3	6	
%	50	50	7,2		
Gesamt	n	38	46	84	
%	45,2	54,8	100		
Druckerpatronen	18-29	n	28	22	50
		%	56	44	59,5
	30-39	n	3	11	14
		%	21,4	78,6	16,7
	40-49	n	3	7	10
		%	30	70	11,9
	50-59	n	1	3	4
%		25	75	4,8	
60+	n	2	4	6	
%	33,3	66,7	7,2		
Gesamt	n	37	47	84	
%	44	56	100		
Tickets	18-29	n	13	16	29
		%	44,8	55,2	70,7
	30-39	n	3	4	7
		%	42,9	57,1	17,1
	40-49	n	1	0	1
		%	100	0	2,4
	50-59	n	2	0	2
%		100	0	4,9	
60-69	n	1	1	2	
%	50	50	4,9		
Gesamt	n	20	21	41	
%	48,8	51,2	100		
Gesamt	-18	n	1	3	4
		%	25	75	1,1
	18-29	n	98	89	187
		%	52,4	47,6	49,5
	30-39	n	32	53	85
		%	37,6	62,4	22,5
	40-49	n	25	29	54
%		46,3	53,7	14,3	
50-59	n	14	13	27	
	%	51,9	48,1	7,1	
60+	n	9	12	21	
%	42,9	57,1	5,5		
Gesamt	n	179	199	378	
%	47,4	52,6	100		

Tab. D-4: Einkommen in der Population

<i>Haushaltsnettoeinkommen</i>	<i>% in Population</i>
<i>bis 1000 Euro</i>	2,8
<i>1000 bis 1500 Euro</i>	7,3
<i>1500 bis 2000 Euro</i>	13,5
<i>2000 bis 2500 Euro</i>	19,3
<i>2500 bis 3000 Euro</i>	17,8
<i>mehr als 3000 Euro</i>	16,8
<i>keine Angabe</i>	22,5

Anmerkung: Die Angaben zur Population entstammen dem (N)ONLINER Atlas (TNS Infratest, 2005).

Tab. D-6: Bildungsstand in der Population

<i>Höchster Schulabschluss</i>	<i>% in Population</i>
<i>Schüler</i>	9,3
<i>Volksschule</i>	29,6
<i>weiterbildende Schule ohne Abitur</i>	37,2
<i>Abitur, Fachhochschulreife</i>	10,7
<i>abgeschlossenes Studium</i>	13,3
<i>Gesamt</i>	100

Anmerkung: Angaben zur Population entstammen dem (N)ONLINER Atlas (TNS Infratest, 2005).

Tab. D-8: Beschäftigung in der Population

<i>Berufliche Tätigkeit</i>	<i>% in Population</i>
<i>Student</i>	3,8
<i>Beamter höherer/ gehobener Dienst</i>	2,8
<i>Leitender Angestellter</i>	6,2
<i>Schüler, Azubi, Zivi., Wehrdienst</i>	14,2
<i>Beamter einfacher/ mittlerer Dienst</i>	2,5
<i>Selbstständig/Freiberufler</i>	8,6
<i>Angestellter</i>	28,9
<i>Arbeiter/Handwerker</i>	11,7
<i>z. Zt. Arbeitslos</i>	5,1
<i>Hausfrau, -mann</i>	6,9
<i>Rentner, Pensionär</i>	9,1

Anmerkung: Angaben zur Population entstammen dem (N)ONLINER Atlas (TNS Infratest, 2005).

Tab. D-10: Interneterfahrung in der Population

<i>Interneterfahrung in Jahren</i>	<i>% in Population</i>
<i>weniger als 0,5 Jahre</i>	4,9
<i>0,5-1 Jahre</i>	6,4
<i>1-2 Jahre</i>	12,9
<i>2-3 Jahre</i>	15,6
<i>mehr als 3 Jahre</i>	60,3

Anmerkung: Angaben zur Population entstammen der Studie Internet Facts 2006-III (AGOF, 2006).

Tab. D-5: Einkommen in der Stichprobe

<i>Nettoeinkommen des Teilnehmers</i>	<i>absolute Häufigkeit</i>	<i>%</i>
<i>bis 300 Euro</i>	59	15,6
<i>300 bis 750 Euro</i>	128	33,9
<i>750 bis 1500 Euro</i>	101	26,7
<i>1500 bis 2000 Euro</i>	38	10,1
<i>2000 bis 2500 Euro</i>	22	5,8
<i>2500 und mehr Euro</i>	23	6,1
<i>gültige Fälle</i>	371	98,1
<i>Missing</i>	7	1,9
<i>Gesamt</i>	378	100

Tab. D-7: Bildungsstand in der Stichprobe

<i>Bildungsstand</i>	<i>absolute Häufigkeit</i>	<i>%</i>
<i>Abitur</i>	303	80,2
<i>Mittlere Reife</i>	44	11,6
<i>Hauptschule</i>	7	1,9
<i>keiner</i>	4	1,1
<i>Sonstiges</i>	20	5,3
<i>Gesamt</i>	378	100

Tab. D-9: Beschäftigung in der Stichprobe

<i>Berufliche Tätigkeit</i>	<i>absolute Häufigkeit</i>	<i>%</i>
<i>Angestellter/Arbeiter</i>	101	26,7
<i>Selbstständig</i>	22	5,8
<i>Beamter</i>	20	5,3
<i>Student</i>	179	47,4
<i>Schüler</i>	4	1,1
<i>keine</i>	21	5,6
<i>Sonstiges</i>	31	8,2
<i>Gesamt</i>	378	100

Tab. D-11: Interneterfahrung der Teilnehmer in Jahren

<i>WWW-Nutzung in Jahren</i>	<i>absolute Häufigkeit</i>	<i>%</i>
<i>0,5-2 Jahre</i>	23	6,1
<i>3-5 Jahre</i>	137	36,2
<i>6-10 Jahre</i>	205	54,2
<i>11-15 Jahre</i>	11	2,9
<i>16-20 Jahre</i>	1	0,3
<i>gültige Fälle</i>	377	99,7
<i>Missing</i>	1	0,3
<i>Gesamt</i>	378	100

Tab. D-12: PC-Nutzungserfahrung der Teilnehmer in Jahren

PC-Nutzung	absolute Häufigkeit	%
1-2 Jahre	4	1,1
3-5 Jahre	41	10,6
6-10 Jahre	171	45,2
11-15 Jahre	115	30,4
16-20 Jahre	32	8,5
21-25 Jahre	12	3,2
26-40 Jahre	3	0,8
Gesamt	378	100

Tab. D-13: PC-Nutzungshäufigkeit der Teilnehmer in Tagen pro Monat

PC-Nutzung	absolute Häufigkeit	%
1-5 Tage im Monat	15	4,0
6-10 Tage im Monat	11	2,9
11-20 Tage im Monat	62	16,4
21-29 Tage im Monat	113	29,9
jeden Tag	176	46,6
gültige Fälle	377	99,7
Missing	1	0,3
Gesamt	378	100

Tab. D-14: PC-Nutzungshäufigkeit der Teilnehmer in Stunden/Tag

PC-Nutzung	absolute Häufigkeit	%
gar nicht	2	0,5
1 Stunde pro Tag	90	23,8
2-3 Stunden pro Tag	147	38,9
4-5 Stunden pro Tag	53	14,0
6-9 Stunden pro Tag	68	18,0
10-12 Stunden p. Tag	17	4,5
gültige Fälle	377	99,7
Missing	1	0,3
Gesamt	378	100

Tab. D-15: Internet-Nutzungshäufigkeit der Teilnehmer in Stunden/Tag

PC-Nutzung	absolute Häufigkeit	%
gar nicht	3	0,8
0,5-1 Stunden pro Tag	222	58,7
2-3 Stunden pro Tag	106	28,0
4-5 Stunden pro Tag	27	7,1
6-9 Stunden pro Tag	14	3,7
10-13 Stunden p. Tag	2	0,5
gültige Fälle	374	99,2
Missing	4	0,8
Gesamt	378	100

Tab. D-16: Internet-Nutzungshäufigkeit der Teilnehmer in Tagen/Monat

Internet-Nutzung	absolute Häufigkeit	%
kein Tag im Monat	1	0,3
1-5 Tage im Monat	29	7,7
6-10 Tage im Monat	18	4,7
11-20 Tage im Monat	80	21,2
21-29 Tage im Monat	106	28,0
jeden Tag	143	37,8
gültige Fälle	377	99,7
Missing	1	0,3
Gesamt	378	100

Tab. D-17: Online-Einkäufe der Teilnehmer in der Vergangenheit

Anzahl Online-Käufe	absolute Häufigkeit	%
gar nicht	39	10,3
1-5	72	19,1
6-10	43	11,4
11-20	53	14,0
21-50	108	28,6
51-100	36	9,5
101-250	20	5,3
251-500	6	1,6
Missing	1	0,3
Gesamt	378	100

Tab. D-18: Online-Einkäufe für die jeweilige Produktart

Anzahl Einkäufe in der Produktkategorie	absolute Häufigkeit	%
gar nicht	191	50,5
1-5	101	26,7
6-10	32	8,5
11-20	29	7,7
21-50	20	5,3
51-100	3	0,8
101-250	2	0,5
Gesamt	378	100



## Screenshots zentraler Seiten der Hypertext-Versuchsumgebung

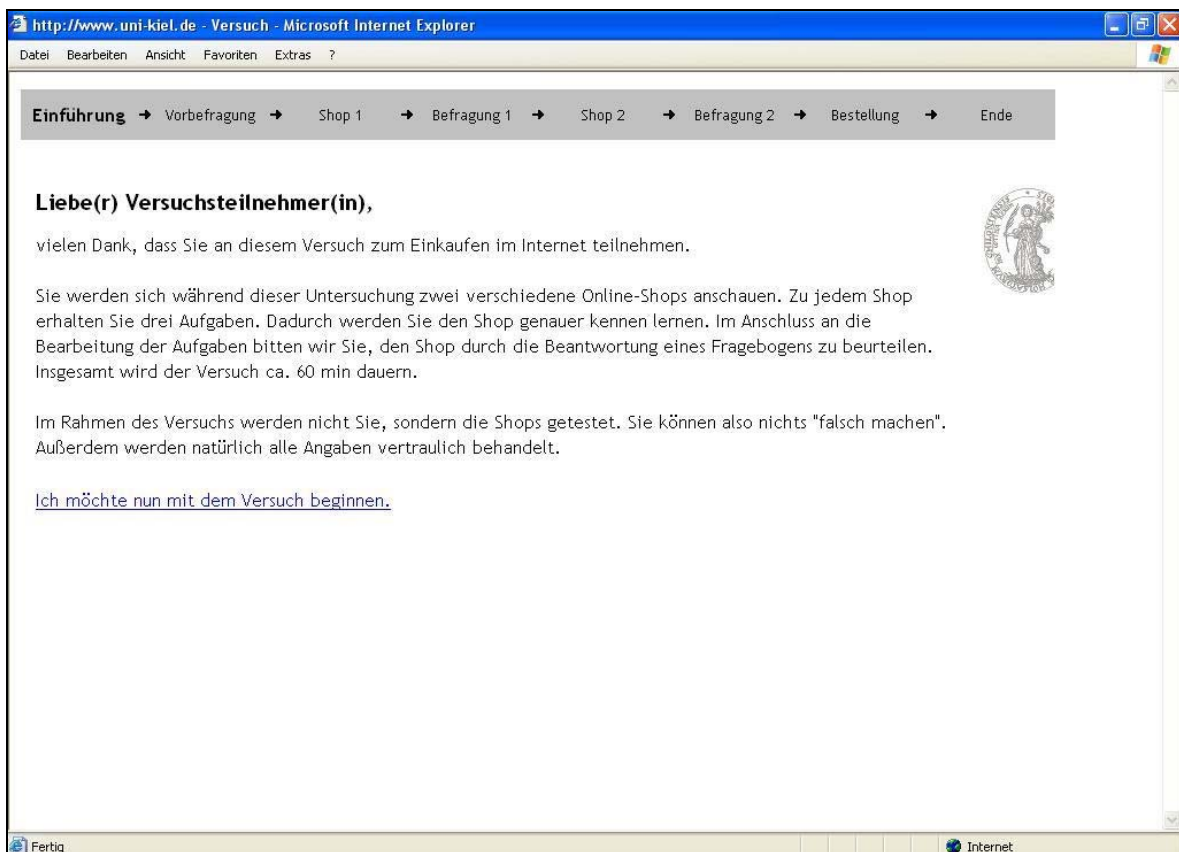


Abb. D-3.1: Hypertext-Versuchsumgebung Seite 1 – Begrüßung

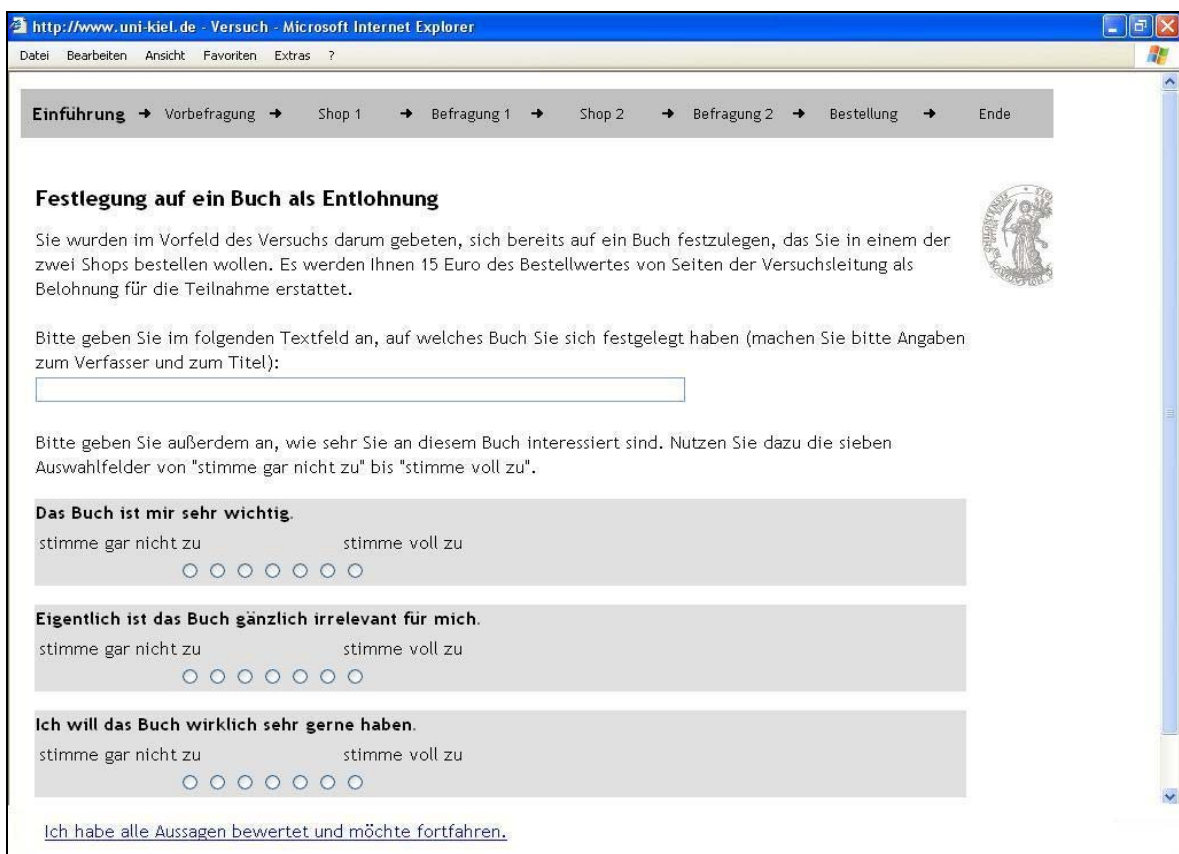


Abb. D-3.2: Hypertext-Versuchsumgebung Seite 2 – Festlegung auf Produkt

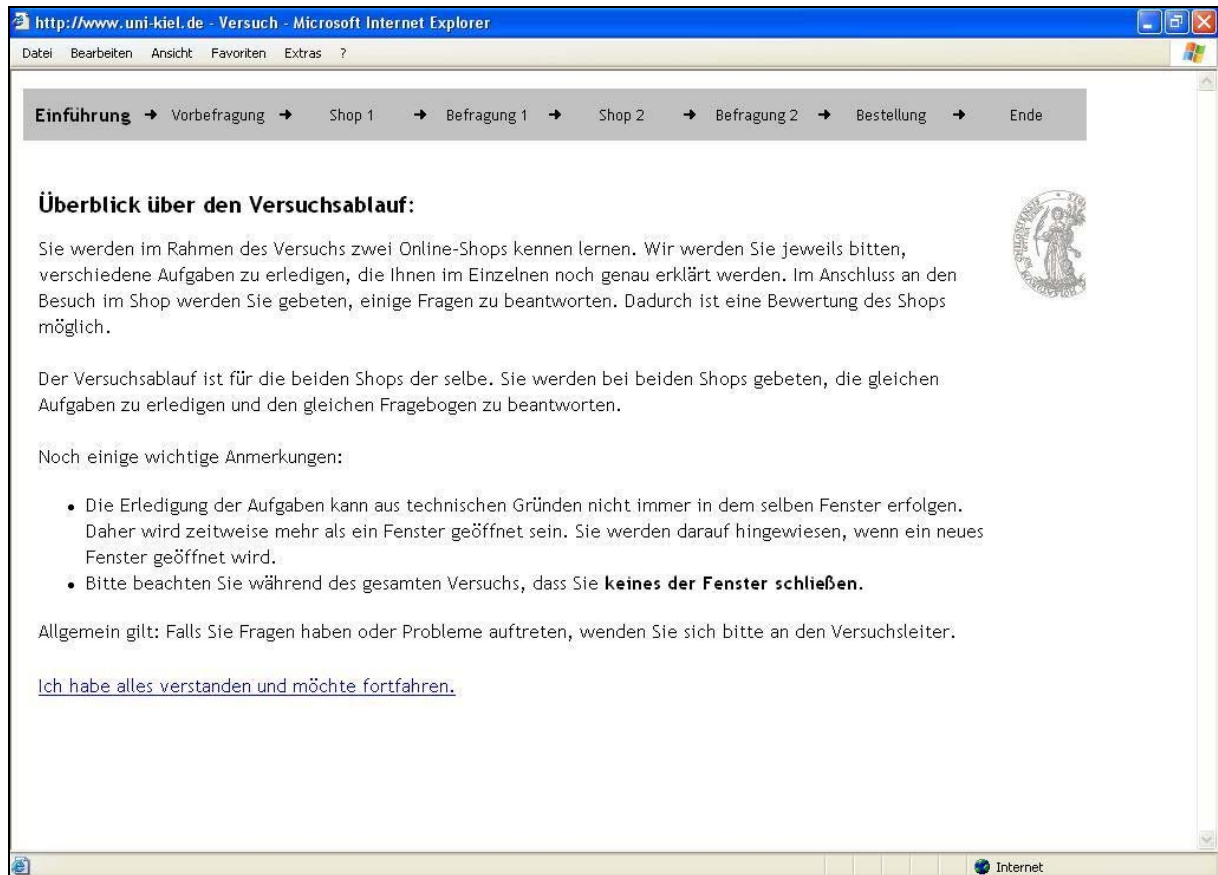


Abb. D-3.3: Hypertext-Versuchsumgebung Seite 3 – Überblick über den Versuchsablauf



Abb. D-3.4: Hypertext-Versuchsumgebung Seite 6 – erster Shop Aufgabe 1 (Teil 1)

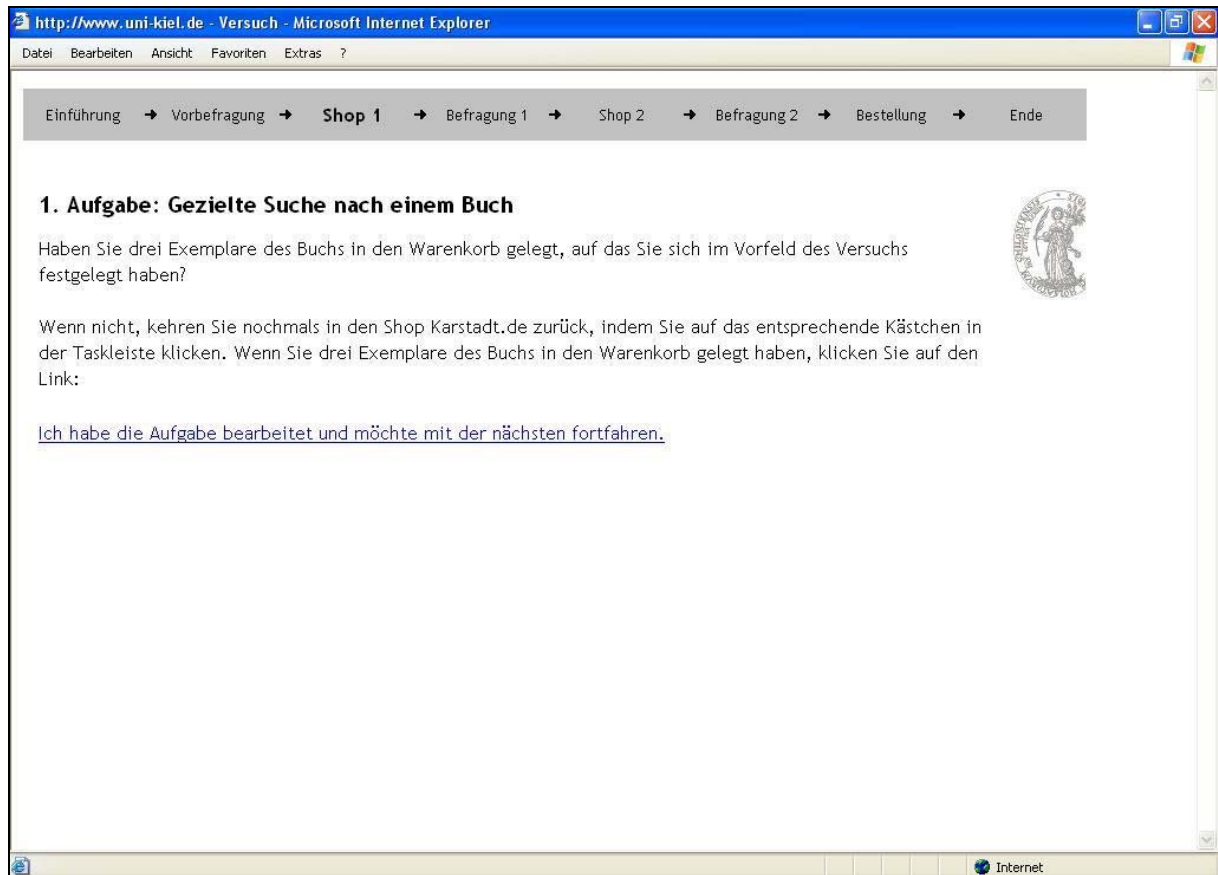


Abb. D-3.5: Hypertext-Versuchsumgebung Seite 7 – erster Shop Aufgabe 1 (Teil 2)

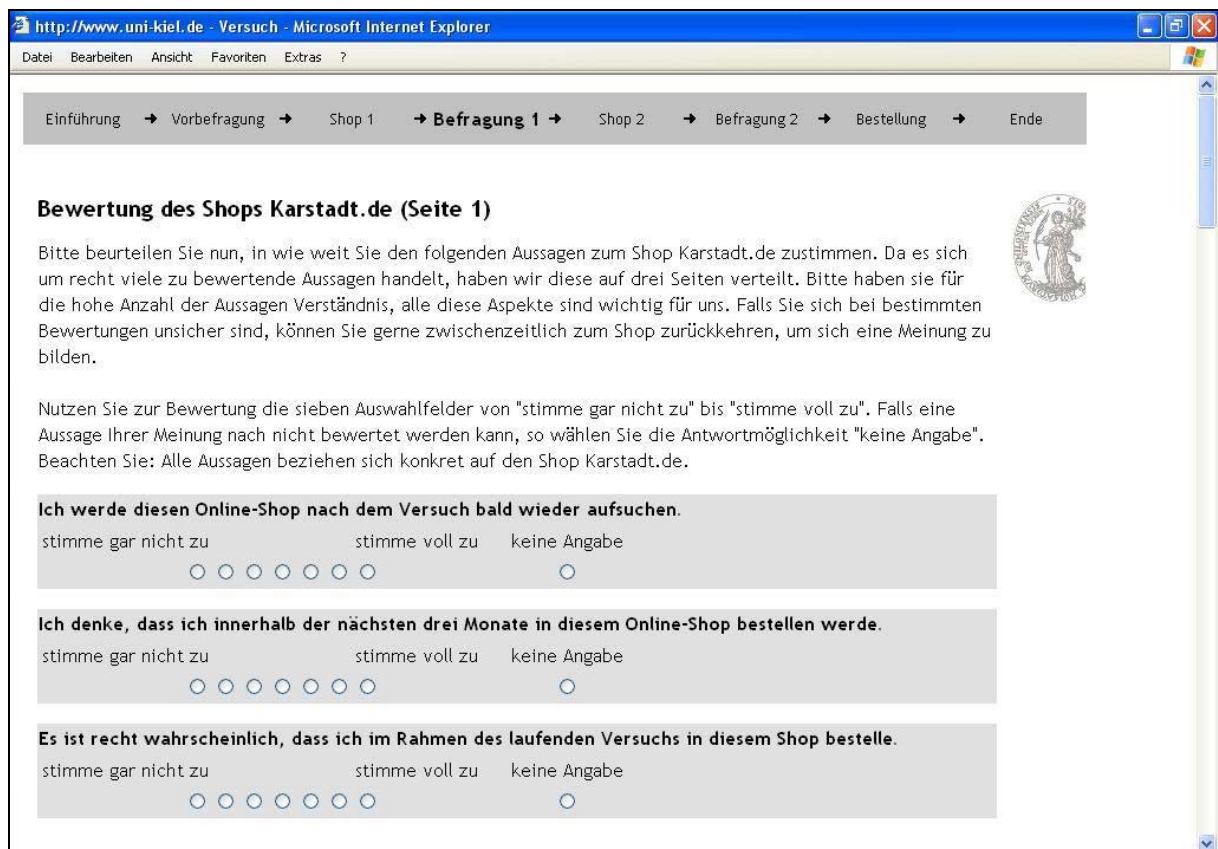


Abb. D-3.6 Hypertext-Versuchsumgebung Seite 12 – Bewertung erster Shop (Teil 1)



Abb. D-3.7: Hypertext-Versuchsumgebung Seite 24 – Kaufentscheidung

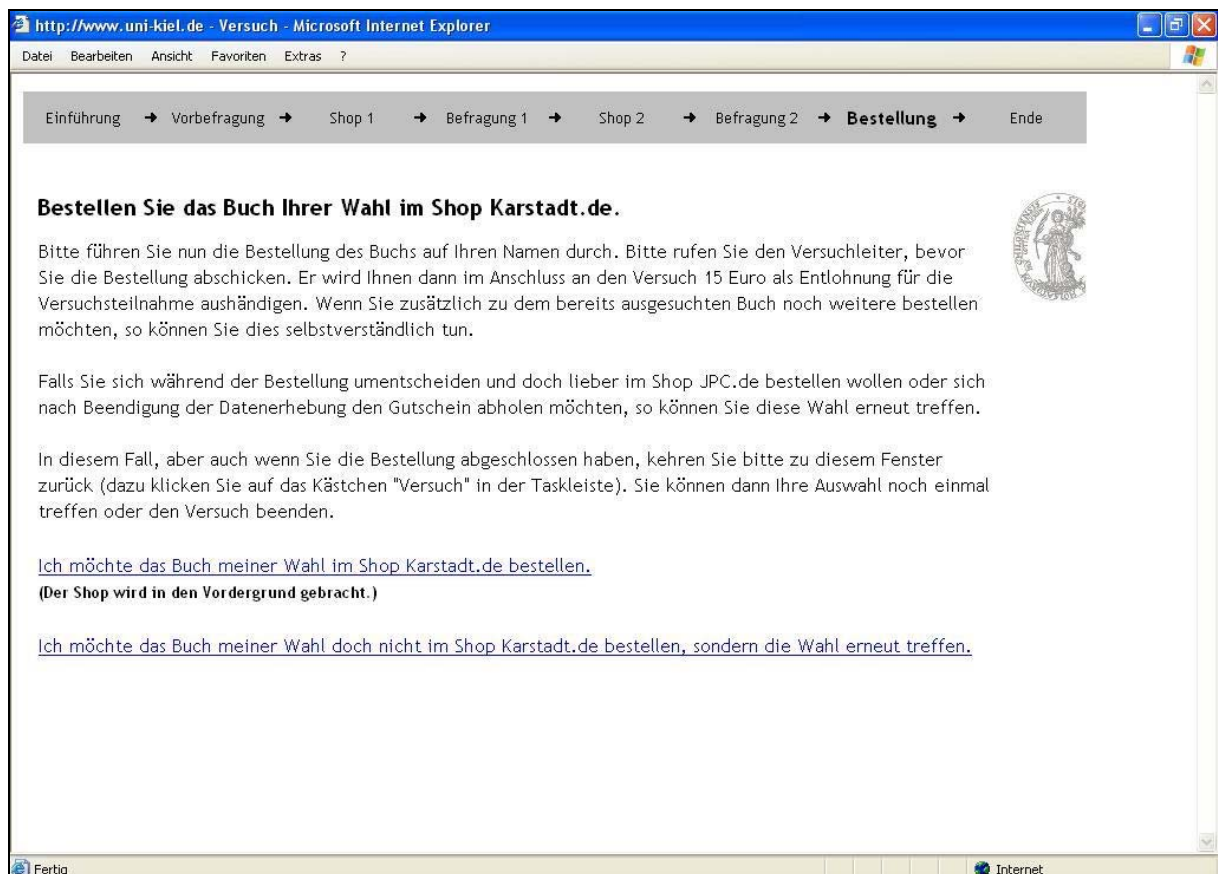


Abb. D-3.8: Hypertext-Versuchsumgebung Seite 25 – Bestellung



Abb. D-3.9: Hypertext-Versuchsumgebung Seite 27 – Wahl des Gutscheins

## Informationen zum Versuchsablauf

### Einverständniserklärung zur Teilnahme an der Untersuchung

Titel der Studie: Bewertung von Online-Shops aus Sicht des Kunden

---

(Vor- und Nachname der Versuchsteilnehmerin/ des Versuchsteilnehmers)

Ich bin über die Dauer und den allgemeinen Verlauf der Studie informiert worden. Ich werde im Anschluss an diese Studie weitere Fragen stellen können und auf Wunsch weitere Informationen zu der Studie erhalten. Ich weiß, dass meine Studienteilnahme freiwillig ist und dass ich jederzeit ohne Angabe von Gründen meine Teilnahme beenden kann.

Ich bin damit einverstanden, dass im Rahmen der Studie meine anonymisierten Daten (d.h. ohne Namensnennung) über Alter, Geschlecht und Ausbildung sowie meine weiteren Angaben, die während der Studie erfasst werden, zur Auswertung der Ergebnisse verwendet werden. Alle im Rahmen der Studie erhobenen Daten werden strikt vertraulich gemäß dem Datenschutz behandelt. Einer wissenschaftlichen Auswertung der anonymisierten Daten und einer möglichen Veröffentlichung der Ergebnisse stimme ich zu. Ich kann jederzeit verlangen, dass meine Daten gelöscht werden. Ich gebe hiermit meine freiwillige Zustimmung zur Teilnahme an dieser Studie.

---

Datum, Unterschrift der Versuchsteilnehmerin/ des Versuchsteilnehmers

## Teilnehmerinstruktionen der Hauptuntersuchung

### Seite 1 (Phase Einführung)

#### Liebe(r) Versuchsteilnehmer(in),

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

Sie werden sich während dieser Untersuchung zwei verschiedene Online-Shops anschauen. Zu jedem Shop erhalten Sie drei Aufgaben. Dadurch werden Sie den Shop genauer kennen lernen. Im Anschluss an die Bearbeitung der Aufgaben bitten wir Sie, den Shop durch die Beantwortung eines Fragebogens zu beurteilen. Insgesamt wird der Versuch ca. 60 min dauern.

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

[Link zu Seite 2:] [Ich möchte nun mit dem Versuch beginnen.](#)

### Seite 2 (Phase Einführung)

#### Festlegung auf eine CD als Entlohnung

Sie wurden im Vorfeld des Versuchs darum gebeten, sich bereits auf eine CD festzulegen, die Sie in einem der zwei Shops bestellen wollen. Es werden Ihnen 15 Euro des Bestellwertes von Seiten der Versuchsleitung als Belohnung für die Teilnahme erstattet.

Bitte geben Sie im folgenden Textfeld an, auf welche CD Sie sich festgelegt haben (machen Sie bitte Angaben zum Interpreten und Titel): [freies Eingabefeld]

Bitte geben Sie außerdem an, wie sehr Sie an dieser CD interessiert sind. Nutzen Sie dazu die sieben Auswahlfelder von "stimme gar nicht zu" bis "stimme voll zu".

*Die CD ist mir sehr wichtig.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Eigentlich ist die CD gänzlich irrelevant für mich.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Ich will die CD wirklich sehr gerne haben.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

[Link zu Seite 3:] [Ich habe alle Aussagen bewertet und möchte fortfahren.](#)

### Seite 3 (Phase Einführung)

#### Überblick über den Versuchsablauf:

Sie werden im Rahmen des Versuchs zwei Online-Shops kennen lernen. Wir werden Sie jeweils bitten, verschiedene Aufgaben zu erledigen, die Ihnen im Einzelnen noch genau erklärt werden. Im Anschluss an den Besuch im Shop werden Sie gebeten, einige Fragen zu beantworten. Dadurch ist eine Bewertung des Shops möglich.

Der Versuchsablauf ist für die beiden Shops derselbe. Sie werden bei beiden Shops 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*.

Allgemein gilt: Falls Sie Fragen haben oder Probleme auftreten, wenden Sie sich bitte an den Versuchsleiter.

[Link zu Seite 4:] Ich habe alles verstanden und 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,  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

*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 bisher im Internet eingekauft?* ungefähr [freies Eingabefeld] mal

*Haben Sie schon einmal vor diesem Versuch Musik-CDs im Internet eingekauft?*  Ja, ich habe in der Vergangenheit ungefähr [freies Eingabefeld] mal Musik-CDs im Internet eingekauft,  Nein

Und noch eine Frage zum Versuch:

*Mit was für einer Verbindung gehen Sie während dieses Versuchs ins Internet?  33,6K-Modem oder langsamer,  56K-Modem,  ISDN,  schneller als ISDN (auch anzukreuzen bei Versuchsdurchführung an der Uni)*

[Link zu Seite 6:] Ich habe alle Fragen beantwortet und möchte mit dem Versuch fortfahren.

## Seite 6 (Phase Shop 1)

### 1. Aufgabe: Gezielte Suche nach einer CD

Sie werden nun den Online-Shop X kennen lernen. Ihre erste Aufgabe ist:

*Suchen Sie in diesem Shop genau die CD heraus, die Sie sich im Vorfeld des Versuchs ausgesucht haben (test). Legen Sie drei Exemplare dieser CD in den Warenkorb (also dreimal genau dasselbe Produkt!), bestellen Sie aber bitte noch nichts.*

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 Seite 7:] Ich möchte jetzt mit der Suche nach der CD beginnen. (Der Shop wird in einem neuen Fenster geöffnet.)

## Seite 7 (Phase Shop 1)

### 1. Aufgabe: Gezielte Suche nach einer CD

Haben Sie drei Exemplare der CD in den Warenkorb gelegt, auf die Sie sich im Vorfeld des Versuchs festgelegt haben?

Wenn nicht, kehren Sie nochmals in den Shop X zurück, indem Sie auf das entsprechende Kästchen in der Taskleiste klicken. Wenn Sie drei Exemplare der CD in den Warenkorb gelegt haben, klicken Sie auf den Link:

[Link zu Seite 8:] Ich habe die Aufgabe bearbeitet und möchte mit der nächsten fortfahren.

## Seite 8 (Phase Shop 1)

### 2. Aufgabe: Stöbern Sie durch den Shop

Stöbern Sie nun bitte weiter durch den Shop X. Bitte bearbeiten Sie dabei die folgenden Teilaufgaben:

- *Legen Sie drei unterschiedliche Produkte in den Warenkorb, die Sie interessieren.*
- *Überlegen Sie: Welche Informationen sind für Ihre Entscheidung wichtig, in diesem Shop zu bestellen?*
- *Informieren Sie sich darüber, wie der Bestellvorgang abläuft, indem Sie ihn probeweise beginnen. Bitte schließen Sie die Bestellung jedoch nicht ab.*

Sie haben zehn Minuten für die Bearbeitung der Aufgabe Zeit. Bitte kehren Sie selbständig nach Ablauf der zehn Minuten zu diesem Fenster zurück; beachten Sie dazu die Uhrzeit, die der Computer anzeigt (in der Taskleiste in der rechten unteren Bildschirmecke).

[Link zu Seite 9:] Ich möchte mit dem Stöbern beginnen und die verschiedenen Teilaufgaben erledigen. (Der Shop wird in den Vordergrund gebracht.)



## Seite 9 (Phase Shop 1)

**2. Aufgabe: Stöbern Sie durch den Shop**

Haben Sie drei verschiedene Produkte in den Warenkorb gelegt, die Sie interessieren? Haben Sie sich darüber hinaus über die Bestellbedingungen und den Ablauf der Bestellung informiert? Haben Sie alle Informationen eingesehen, die Ihnen für eine Bestellung wichtig sind?

Wenn nicht, kehren Sie nochmals in den Shop X zurück. Wenn Sie die oben genannten Teilaufgaben erledigt haben, klicken Sie auf den Link:

[Link zu Seite 10:] Ich habe die Teilaufgaben erledigt und möchte mit der nächsten Aufgabe fortfahren.

## Seite 10 (Phase Shop 1)

**3. Aufgabe: Kennenlernen des Warenkorbs**

Im Warenkorb des Shops X sollten sich im Moment mehrere Produkte befinden.

*Verändern Sie den Warenkorb nun so, dass sich darin nur noch ein Exemplar der CD befindet, auf die Sie sich im Vorfeld der Untersuchung festgelegt haben (test). Falls Sie bestellen möchten, tun sie dies bitte noch nicht.*

Sobald Sie diese Aufgabe erledigt haben, kehren Sie bitte zu diesem Fenster zurück. Dann können Sie mit der Bewertung des Shops X beginnen.

[Link zu Seite 11:] Ich möchte jetzt mit der Veränderung des Warenkorbes beginnen. (Der Shop wird in den Vordergrund gebracht.)

## Seite 11 (Phase Shop 1)

**3. Aufgabe: Kennenlernen des Warenkorbes**

Befindet sich nur ein Exemplar der CD im Warenkorb, nach der Sie im Rahmen der ersten Aufgabe gezielt gesucht haben?

Wenn nicht, kehren Sie nochmals in den Shop X zurück, indem Sie auf das entsprechende Kästchen in der Taskleiste klicken. Wenn Sie den Warenkorb entsprechend verändert haben, klicken Sie auf den Link:

[Link zu Seite 12:] Ich habe die Aufgabe erledigt und möchte den Versuch fortsetzen.

## Seite 12 (Phase Bewertung 1)

**Bewertung des Shops X (Seite 1)**

Bitte beurteilen Sie nun, in wie weit Sie den folgenden Aussagen zum Shop X zustimmen. Da es sich um recht viele zu bewertende Aussagen handelt, haben wir diese auf drei 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 zum Shop 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 den Shop X.

[Link zu Seite 13:] Ich möchte meine Bewertung fortsetzen.

Seite 13 (Phase Bewertung 1)

### **Bewertung des Shops X (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 den Shop X.

[Link zu Seite 14:] Ich möchte meine Bewertung fortsetzen.

Seite 14 (Phase Bewertung 1)

### **Bewertung des Shops X (Seite 3)**

Bitte bewerten Sie nun die verbleibenden Aussagen. 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 den Shop X.

[Link zu Seite 15:] Ich habe die Bewertung abgeschlossen und möchte mit dem Versuch fortfahren.

Seite 15 (Phase Shop 2)

### **1. Aufgabe: Gezielte Suche nach einer CD**

Sie werden nun den Online-Shop Y kennen lernen. Wie zuvor lautet Ihre erste Aufgabe:

*Suchen Sie in diesem Shop genau die CD heraus, die Sie sich im Vorfeld des Versuchs ausgesucht haben (Name des Produktes). Legen Sie drei Exemplare dieser CD in den Warenkorb (also dreimal genau dasselbe Produkt!), bestellen Sie aber bitte noch nichts.*

Wenn Sie diese Aufgabe bearbeitet haben, kehren Sie bitte wieder zum aktuellen Fenster zurück, indem Sie auf das Kästchen "Versuch" in der Taskleiste klicken. Sie können dann mit der nächsten Aufgabe fortfahren.

[Link zu Seite 16:] Ich möchte jetzt mit der Suche nach der CD beginnen. (Der Shop wird in einem neuen Fenster geöffnet.)

Seite 16 (Phase Shop2)

### **1. Aufgabe: Gezielte Suche nach einer CD**

Haben Sie drei Exemplare der CD in den Warenkorb gelegt, auf die Sie sich im Vorfeld des Versuchs festgelegt haben?

Wenn nicht, kehren Sie nochmals in den Shop Y zurück. Wenn Sie drei Exemplare der CD in den Warenkorb gelegt haben, klicken Sie auf den Link:

[Link zu Seite 17:] Ich habe die Aufgabe bearbeitet und möchte mit der nächsten fortfahren.

Seite 17 (Phase Shop 2)

### **2. Aufgabe: Stöbern Sie durch den Shop**

Stöbern Sie nun bitte weiter durch den Shop Y. Bitte bearbeiten Sie dabei wieder die gleichen Teilaufgaben wie bereits beim ersten Shop:

- *Legen Sie drei unterschiedliche Produkte in den Warenkorb, die Sie interessieren.*
- *Überlegen Sie: Welche Informationen sind für Ihre Entscheidung wichtig, in diesem Shop zu bestellen?*
- *Informieren Sie sich darüber, wie der Bestellvorgang abläuft, indem Sie ihn probeweise beginnen. Bitte schließen Sie die Bestellung jedoch nicht ab.*

Sie haben wiederum zehn Minuten für die Bearbeitung der Aufgabe Zeit. Bitte kehren Sie selbständig nach Ablauf der zehn Minuten zu diesem Fenster zurück; beachten Sie dazu die Uhrzeit, die der Computer anzeigt (in der Taskleiste in der rechten unteren Bildschirmecke).

[Link zu Seite 18:] Ich möchte mit dem Stöbern beginnen und die verschiedenen Teilaufgaben erledigen. (Der Shop wird in den Vordergrund gebracht.)

Seite 18 (Phase Shop 2)

### **2. Aufgabe: Stöbern Sie durch den Shop**

Haben Sie drei verschiedene Produkte in den Warenkorb gelegt, die Sie interessieren? Haben Sie sich darüber hinaus über die Bestellbedingungen und den Ablauf der Bestellung informiert? Haben Sie alle Informationen eingesehen, die Ihnen für eine Bestellung wichtig sind?

Wenn nicht, kehren Sie nochmals in den Shop Y zurück. Wenn Sie die oben genannten Teilaufgaben erledigt haben, klicken Sie auf den Link:

[Link zu Seite 19:] Ich habe die Teilaufgaben erledigt und möchte mit der nächsten Aufgabe fortfahren.

Seite 19 (Phase Shop 2)

### **3. Aufgabe: Kennenlernen des Warenkorbes**

Im Warenkorb des Shops Y sollten sich im Moment mehrere Produkte befinden.

*Verändern Sie den Warenkorb nun so, dass sich darin nur noch ein Exemplar der CD befindet, auf die Sie sich im Vorfeld der Untersuchung festgelegt haben (Name des Produktes). Falls Sie bestellen möchten, tun sie dies bitte noch nicht.*

Sobald Sie diese Aufgabe erledigt haben, kehren Sie bitte zu diesem Fenster zurück. Dann können Sie mit der Bewertung des Shops Y beginnen.

[Link zu Seite 20:] Ich möchte jetzt mit der Veränderung des Warenkorbes beginnen. (Der Shop wird in den Vordergrund gebracht.)

Seite 20 (Phase Shop 2)

### **3. Aufgabe: Kennenlernen des Warenkorbes**

Befindet sich nur ein Exemplar der CD im Warenkorb, nach der Sie im Rahmen der ersten Aufgabe gezielt gesucht haben?

Wenn nicht, kehren Sie nochmals in den Shop Y zurück, indem Sie auf das entsprechende Kästchen in der Taskleiste klicken. Wenn Sie den Warenkorb entsprechend verändert haben, klicken Sie auf den Link:

[[Link zu Seite 21:](#)] Ich habe die Aufgabe erledigt und möchte den Versuch fortsetzen.

Seite 21 (Phase Bewertung 2)

### **Bewertung des Shops Y (Seite 1)**

Bitte beurteilen Sie nun, in wie weit Sie den folgenden Aussagen zum Shop Y zustimmen. Da es sich um recht viele zu bewertende Aussagen handelt, haben wir diese auf drei 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 zum Shop 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 den Shop Y.

[[Link zu Seite 22:](#)] Ich möchte meine Bewertung fortsetzen.

Seite 22 (Phase Bewertung 2)

### **Bewertung des Shops Y (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 den Shop Y.

[[Link zu Seite 23:](#)] Ich möchte meine Bewertung fortsetzen.

Seite 23 (Phase Bewertung 2)

### **Bewertung des Shops Y (Seite 3)**

Bitte bewerten Sie nun die verbleibenden Aussagen. 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 den Shop X.

[[Link zu Seite 24:](#)] Ich habe die Bewertung abgeschlossen und möchte mit dem Versuch fortfahren.

Seite 24 (Phase Bestellung)

### **Entscheiden Sie sich, wie Sie die CD erhalten möchten.**

In den Warenkörben der beiden Shops sollte sich nun jeweils ein Exemplar befinden, die Sie sich im Vorfeld des Versuchs ausgesucht haben (*Name des Produktes*). Sollte dies nicht der Fall sein, so sprechen Sie bitte den Versuchsleiter an.

Sie können die CD nun in einem der beiden Shops auf Ihren Namen bestellen. Der Versuchsleiter händigt Ihnen dann im Anschluss an die Untersuchung 15 Euro in bar aus. Liegt der Bestellwert über diesem Betrag, so müssen Sie für den Rest selber aufkommen. Die Abwicklung der Bestellung muss in jedem Fall von Ihnen selbst erledigt werden.

Falls Sie sich nicht entschließen können, die CD in einem der Shops zu bestellen, so besteht auch die Möglichkeit, dass Sie einen Gutschein über 15 Euro für ein in Kiel ansässiges Kaufhaus erhalten. Der Gutschein kann Ihnen jedoch nicht direkt im Anschluss an den Versuch ausgehändigt werden, sondern muss nach Beendigung der Datenerhebung (ca.

Anfang April) von Ihnen persönlich im Büro der Versuchsleitung (Westring 425 in Kiel) abgeholt werden.

Entscheiden Sie sich:

[Link zu Seite 25:] Ich möchte die CD im Shop X bestellen.

[Link zu Seite 26:] Ich möchte die CD im Shop Y bestellen.

[Link zu Seite 27:] Ich möchte mir nach Beendigung der Datenerhebung einen Gutschein abholen.

### Seite 25 (Phase Bestellung)

#### **Bestellen Sie die CD Ihrer Wahl im Shop X.**

Bitte führen Sie nun die Bestellung der CD auf Ihren Namen durch. Bitte rufen Sie den Versuchleiter, bevor Sie die Bestellung abschicken. Er wird Ihnen dann im Anschluss an den Versuch 15 Euro als Entlohnung für die Versuchsteilnahme aushändigen. Wenn Sie zusätzlich zu der bereits ausgesuchten CD noch weitere bestellen möchten, so können Sie dies selbstverständlich tun.

Falls Sie sich während der Bestellung umentscheiden und doch lieber im Shop Y bestellen wollen oder sich nach Beendigung der Datenerhebung den Gutschein abholen möchten, so können Sie diese Wahl erneut treffen.

In diesem Fall, aber auch wenn Sie die Bestellung abgeschlossen haben, kehren Sie bitte zu diesem Fenster zurück (dazu klicken Sie auf das Kästchen "Versuch" in der Taskleiste). Sie können dann Ihre Auswahl noch einmal treffen oder den Versuch beenden.

[Link zu Seite 28:] Ich möchte die CD meiner Wahl im Shop X bestellen. (Der Shop wird in den Vordergrund gebracht.)

[Link zu Seite 24:] Ich möchte die CD meiner Wahl doch nicht im Shop X bestellen, sondern die Wahl erneut treffen.

### Seite 26 (Phase Bestellung)

#### **Bestellen Sie die CD Ihrer Wahl im Shop Y.**

Bitte führen Sie nun die Bestellung der CD auf Ihren Namen durch. Bitte rufen Sie den Versuchleiter, bevor Sie die Bestellung abschicken. Er wird Ihnen dann im Anschluss an den Versuch 15 Euro als Entlohnung für die Versuchsteilnahme aushändigen. Wenn Sie zusätzlich zu der bereits ausgesuchten CD noch weitere bestellen möchten, so können Sie dies selbstverständlich tun.

Falls Sie sich während der Bestellung umentscheiden und doch lieber im Shop X bestellen wollen oder sich nach Beendigung der Datenerhebung den Gutschein abholen möchten, so können Sie diese Wahl erneut treffen.

In diesem Fall, aber auch wenn Sie die Bestellung abgeschlossen haben, kehren Sie bitte zu diesem Fenster zurück (dazu klicken Sie auf das Kästchen "Versuch" in der Taskleiste). Sie können dann Ihre Auswahl noch einmal treffen oder den Versuch beenden.

Ich möchte die CD meiner Wahl im Shop Y bestellen. (Der Shop wird in den Vordergrund gebracht.)

[Link zu Seite 24:] Ich möchte die CD meiner Wahl doch nicht im Shop Y bestellen, sondern die Wahl erneut treffen.

Seite 27 (Phase Bestellung)

### **Wahl eines Gutscheines als Entlohnung zur Versuchsteilnahme**

Bitte lassen Sie sich vom Versuchsleiter einen Zettel mit der Wegbeschreibung zum Büro der Versuchsleitung aushändigen, damit Sie sich den Gutschein nach Abschluss der Datenerhebung abholen können. Auf diesem Zettel finden Sie auch eine Telefonnummer, unter der Sie bitte zur Vereinbarung des Abholtermins anrufen.

Um die Anonymität zu wahren, geben Sie bitte ein Codewort an. Unter Angabe dieses Codeworts wird Ihnen dann später der Gutschein ausgehändigt. Das vierstellige Codewort setzt sich wie folgt zusammen: die zwei Anfangsbuchstaben des Vornamens Ihrer Mutter und dann die letzten beiden Ziffern Ihres Geburtsjahres (z.B. UT74 für die Kombination Uta und 1974).

Codewort: [freies Eingabefeld]

[Link zu Seite 31:] Ich habe alles ausgefüllt und möchte fortfahren.

Seite 28 (Phase Bestellung)

### **Bestellen Sie die CD Ihrer Wahl im Shop X.**

Möchten Sie die CD doch lieber bei Y bestellen oder sich später den Gutschein abholen? In diesem Fall können Sie noch einmal einen Schritt zurückgehen und diese Auswahl erneut treffen.

[Link zu Seite 24:] Ich möchte die CD doch nicht bei X bestellen und meine Auswahl erneut treffen.

Oder haben Sie die Bestellung der CD Ihrer Wahl bei X abgeschlossen?

[Link zu Seite 30:] Ich habe die Bestellung abgeschlossen und möchte fortfahren.

Seite 29 (Phase Bestellung)

### **Bestellen Sie die CD Ihrer Wahl im Shop Y.**

Möchten Sie die CD doch lieber bei X bestellen oder sich später den Gutschein abholen? In diesem Fall können Sie noch einmal einen Schritt zurückgehen und diese Auswahl erneut treffen.

[Link zu Seite 24:] Ich möchte die CD doch nicht bei Y bestellen und meine Auswahl erneut treffen.

Oder haben Sie die Bestellung der CD Ihrer Wahl bei Y abgeschlossen?

[Link zu Seite 30:] Ich habe die Bestellung abgeschlossen und möchte fortfahren.

Seite 30 (Phase Bestellung)

### **Verkaufspreis der von Ihnen gewählten CD**

Bitte geben Sie nun noch an, wie viel die von Ihnen im Vorfeld der Untersuchung gewählte CD in den beiden Shops (inklusive der Versandkosten) kostet.

[freies Eingabefeld] Euro im Shop X.

Klicken Sie [Link zum Shop X] hier, falls Sie noch einmal im Shop nachsehen müssen. (Der Shop X wird in den Vordergrund gebracht.)

[freies Eingabefeld] Euro im Shop Y.

Klicken Sie [Link zum Shop Y] hier, falls Sie noch einmal im Shop nachsehen müssen.

(Der Shop Y wird in den Vordergrund gebracht.)

Haben Sie zusätzlich zu der bereits im Vorfeld gewählten CD noch weitere Musik-CDs bestellt?

Ja, ich habe weitere Musik-CDs gekauft,

*Wenn ja, wie hoch liegt der Bestellwert (inklusive der Versandkosten etc.) insgesamt?*

bei [freies Eingabefeld] Euro.

Nein, ich habe keine weiteren Musik-CDs gekauft.

[Link zu Seite 32:] Ich möchte fortfahren.

### Seite 31 (Phase Bestellung)

#### **Verkaufspreis der von Ihnen gewählten CD**

Bitte geben Sie nun noch an, wie viel die von Ihnen im Vorfeld der Untersuchung gewählte CD in den beiden Shops (inklusive der Versandkosten) kostet.

[freies Eingabefeld] Euro im Shop X.

Klicken Sie [Link zum Shop X] *hier*, falls Sie noch einmal im Shop nachsehen müssen. (Der Shop X wird in den Vordergrund gebracht.)

[freies Eingabefeld] Euro im Shop Y.

Klicken Sie [Link zum Shop Y] *hier*, falls Sie noch einmal im Shop nachsehen müssen. (Der Shop Y wird in den Vordergrund gebracht.)

[Link zu Seite 32:] Ich möchte fortfahren.

### Seite 32 (Phase Bestellung)

#### **Bekanntheit der Shops**

Bitte machen Sie nun noch Angaben zum Bekanntheitsgrad der beiden Shops. Dazu sollen Sie angeben, in wie weit Sie den folgenden Aussagen zustimmen.

Angaben zu Shop X.

*Der Shop X, den ich im Rahmen des Versuchs besucht habe, ist allgemein sehr bekannt unter Internetnutzern.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Ich selber kannte den Shop X bereits vor diesem Versuch sehr gut.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Ich habe vor der Teilnahme an diesem Versuch im Shop X (ungefähr) [freies Eingabefeld] mal bestellt* (bitte die Zahl 0 angeben, falls Sie noch nicht im besagten Shop bestellt haben).

*Waren Sie im Shop X bereits vor dem Versuch als Kunde registriert?*  Ja,  Nein

Angaben zu Shop Y.

*Der Shop Y, den ich im Rahmen des Versuchs besucht habe, ist allgemein sehr bekannt unter Internetnutzern.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Ich selber kannte den Shop Y bereits vor diesem Versuch sehr gut.* (siebenstufige Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“)

*Ich habe vor der Teilnahme an diesem Versuch im Shop Y (ungefähr) [freies Eingabefeld] mal bestellt* (bitte die Zahl 0 angeben, falls Sie noch nicht im besagten Shop bestellt haben).

*Waren Sie im Shop Y bereits vor dem Versuch als Kunde registriert?*  Ja,  Nein

[Link zu Seite 33:] Ich habe die Angaben gemacht und möchte den Versuch nun beenden.

### Seite 33 (Phase Ende)

Insofern Sie bereit sind, an einer kurzen Nachbefragung (ca. sechs Fragen) per Email nach Beendigung der Hauptdatenerhebung teilzunehmen, so tragen Sie bitte hier Ihre Emailadresse ein (die Angabe ist also freiwillig). Wir versichern Ihnen, dass Ihre Adresse nicht an Dritte weitergegeben wird; es besteht also keine Spam-Gefahr.

Meine Emailadresse lautet [freies Eingabefeld].

[Link zu Seite 34:] Ich möchte den Versuch nun beenden.

### Seite 34 (Phase Ende)

Vielen Dank für Ihre Teilnahme am Versuch!

Falls Sie noch Fragen zum Versuch haben sollten, wenden Sie sich bitte an den Versuchsleiter.



## **Anhang E**

### **Ergänzende Ergebnisdarstellungen**

Tab. E-1:	Mittelwerte für die AV Kaufintention bei Erhebung an erster bzw. zweiter Stelle .....	340
Tab. E-2:	t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Kaufintention.....	340
Tab. E-3:	Ausschluss von Fällen aufgrund eines zu hohen Anteils fehlender Werte .....	340
Tab. E-4:	Anzahl der Fälle mit Ersetzung fehlender Werte durch Multiple Imputation..	340
Tab. E-5:	Indikatoren, die in die Multiple Imputation eingehen.....	341
Tab. E-6:	Harman's Single Factor-Test – Eigenwerte und Gesamtvarianz der Faktorenlösung .....	342
Tab. E-7:	Prüfung der Normalverteilungsannahme für alle erhobenen Indikatoren .....	343
Tab. E-8.1:	Harman's Single Factor-Test - Komponentenmatrix .....	344
Tab. E-8.2:	Harman's Single Factor-Test - Komponentenmatrix (Fortsetzung) .....	345
Tab. E-9.1:	Auflistung aller Probanden mit potenziell problematischen Antwortverhalten.....	346
Tab. E-9.2:	Auflistung aller Probanden mit potenziell problematischen Antwortverhalten (Fortsetzung) .....	347
Tab. E-10:	Mustermatrix der ersten berechneten Hauptachsenanalyse .....	347
Tab. E-11:	Mustermatrix der zweiten berechneten Hauptachsenanalyse.....	348
Tab. E-12:	Mustermatrix der dritten berechneten Hauptachsenanalyse.....	348
Tab. E-13:	Cronbachs-Alpha der reflektiven Usability-Skala ufosV2r für die einzelnen Benutzerschnittstellen .....	349
Tab. E-14:	Rangreihe der Shop-Mittelwerte auf der reflektiven Usability-Skala ufosV2r.....	350
Tab. E-15:	Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zum Vertrauen.....	351
Tab. E-16:	Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur wahrgenommenen Produktanzahl .....	352
Tab. E-17:	Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur Ästhetik .....	353
Tab. E-18:	Rangreihe der Shop-Mittelwerte über das Item zum wahrgenommenen Preis	354
Tab. E-19:	Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur Kaufintention	355
Tab. E-20:	Rangreihe Anzahl der im Versuch getätigten Käufe.....	356
Tab. E-21.1:	Pearson-Korrelationen zwischen den formativen Usability Indikatoren .....	357
Tab. E-21.2:	Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung) .....	358
Tab. E-21.3:	Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung) .....	359
Tab. E-21.4:	Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung) .....	360
Tab. E-22:	Toleranz- und VIF-Werte der formativen ufosV2f-Indikatoren .....	361
Tab. E-23:	Rotierte Komponentenmatrix zur Erstellung der Multikollinearitätsindizes zur formativen Usability-Skala ufosV2f.....	362

Tab. E-24:	Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability (ufosV2r) mit Quadratwurzel des AVE-Maßes.....	363
Tab. E-25:	Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability (ufosV2f) mit Quadratwurzel des AVE-Maßes.....	364
Tab. E-26:	Kreuzladungen der Indikatoren im PLS-Modells bei Verwendung der reflektiven Usability-Skala ufosV2r.....	365
Tab. E-27:	Kreuzladungen der Indikatoren im PLS-Modells bei Verwendung der formativen Usability-Skala ufosV2f.....	366
Tab. E-28:	Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum konzeptuellen Erfolgsmodell .....	367
Tab. E-29:	Korrelationen der MLR-Prädiktoren.....	368
Tab. E-30:	Toleranz und Varance-Inflation-Factors (VIF) der MLR-Prädiktoren .....	369
Tab. E-31:	Teilnehmer mit standardisierten Residuen $z_{Redsid} > 1$ in der sequentiellen MLR .....	370
Tab. E-32:	Ausreißer in der MLR .....	371
Tab. E-33:	Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 1 .....	372
Tab. E-34:	Klassifikationsmatrix – sequentielle MRL Schritt 1 .....	372
Tab. E-35:	Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 1.....	372
Tab. E-36:	Parameterschätzer – sequentielle MRL Schritt 1 .....	373
Tab. E-37:	Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2a.....	374
Tab. E-38:	Klassifikationsmatrix – sequentielle MRL Schritt 2a .....	374
Tab. E-39:	Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2a.....	374
Tab. E-40:	Parameterschätzer – sequentielle MRL Schritt 2a .....	375
Tab. E-41:	Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2b .....	376
Tab. E-42:	Klassifikationsmatrix – sequentielle MRL Schritt 2b.....	376
Tab. E-43:	Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2b.....	376
Tab. E-44:	Parameterschätzer – sequentielle MRL Schritt 2b .....	377
Tab. E-45:	Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2c.....	378
Tab. E-46:	Klassifikationsmatrix – sequentielle MRL Schritt 2c .....	378
Tab. E-47:	Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2c.....	378
Tab. E-48:	Parameterschätzer – sequentielle MRL Schritt 2c .....	379

Tab. E-1: Mittelwerte für die AV Kaufintention bei Erhebung an erster bzw. zweiter Stelle

<i>Variable</i>	<i>Bewerteter Shop an erster oder zweiter Stelle</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE des Mittelwertes</i>
Kaufintention	Shop an erster Stelle	368	1,989	1,629	0,085
	Shop an zweiter Stelle	368	1,943	1,904	0,099

Anmerkung: Anzahl Fälle (N), Mittelwert (M), SD (Standardabweichung), Standardfehler (SE).

Tab. E-2: t-Test für unabhängige Stichproben auf Mittelwertgleichheit der AV Kaufintention

	<i>Levene-Test der Varianzgleichheit</i>		<i>t-Test für die Mittelwertgleichheit</i>						
			<i>t</i>	<i>df</i>	<i>Sign. (2-seitig)</i>	<i>Mittlere Differenz</i>	<i>SE der Differenz</i>	<i>95% Konfidenzintervall der Differenz</i>	
	<i>F</i>	<i>Sign.</i>						<i>Untere</i>	<i>Obere</i>
Varianzen sind gleich	23,464	0,000	0,347	734,00	0,728	0,045	0,131	-0,211	0,302
Varianzen sind nicht gleich			0,347	716,90	0,728	0,045	0,131	-0,211	0,302

Anmerkung: F-Wert (F), t-Wert (t), Freiheitsgrade (df), Standardfehler (SE).

Tab. E-3: Ausschluss von Fällen aufgrund eines zu hohen Anteils fehlender Werte

<i>Teilnehmercode</i>	<i>Produkt</i>	<i>Bewertete Shops</i>	<i>Alter</i>	<i>Geschlecht</i>	<i>Anteil fehlender Daten</i>
b1_2_2	Bücher	Karstadt.de / JPC.de	48	w	25,5%
b5_2_2	Bücher	BOL.de / JPC.de	53	m	20,8%
b6_1_1	Bücher	Buchkatalog.de / Karstadt.de	71	m	21,7%
d4_3_2	DVDs	Bücher.de / inCinea.de	66	m	25,5%
m4_1_1	CDs	Abella.de / WOM.de	43	m	23,1%
m4_2_2	CDs	Abella.de / CyberCD.de	54	m	22,6%
m7_4_2	CDs	abc-CD.de / Abella.de	64	w	20,8%
p4_3_1	Patronen	Alphaco.de / Express-Toner	25	w	23,1%
p5_6_2	Patronen	ACI / Planet-Shop.de	72	w	27,4%
p7_2_2	Patronen	Alternate.de / PrintPower.net	47	m	25,5%

Anmerkung: weiblich (w), männlich (m).

Tab. E-4: Anzahl der Fälle mit Ersetzung fehlender Werte durch Multiple Imputation

<i>Reflektive Skala</i>	<i>Anzahl der Fälle</i>
Usability (ufosV2r)	4
Wahrgenommene Produktanzahl	14
Kaufintention	1
Vertrauen	18
Einkaufsvergnügen	7
Ästhetik	4

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>
zeit	Bearbeitungsdauer	2,45	uf41	Usability formativ Item 41	8,42
shop1	Shop 1 vs. anderer (Dummy)	-	uf42	Usability formativ Item 42	0,54
shop2	Shop 2 vs. anderer (Dummy)	-	uf43	Usability formativ Item 43	1,36
shop3	Shop 3 vs. anderer (Dummy)	-	uf44	Usability formativ Item 44	0,41
...	...	-	uf45	Usability formativ Item 45	0,27
shop35	Shop 35 vs. anderer (Dummy)	-	uf46	Usability formativ Item 46	4,76
for_refl	reflektiv - formativ (Dummy)	-	uf47	Usability formativ Item 47	9,65
pisum	Produkt Involvement Skalenwert	0	uf48	Usability formativ Item 48	8,56
alter	Alter	0	uf49	Usability formativ Item 49	2,45
sex	Geschlecht (Dummy)	0	uf50	Usability formativ Item 50	3,40
pc	PC-Nutzung (in Tagen)	0,27	uf54	Usability formativ Item 54	2,45
www	WWW-Nutzung (in Tagen)	0,27	uf55	Usability formativ Item 55	1,22
shop_er	Anzahl Online-Einkäufe	0,27	uf56	Usability formativ Item 56	5,57
reco	Weiterempfehlungstention	3,40	uf57	Usability formativ Item 57	3,13
uf1	Usability formativ Item 1	4,76	uf58	Usability formativ Item 58	3,67
uf4	Usability formativ Item 4	3,94	sv1	Service Item 1	5,71
uf5	Usability formativ Item 5	1,63	sv5	Service Item 5	8,02
uf6	Usability formativ Item 6	0,14	subp	Wahrgenommener Preis	4,21
uf7	Usability formativ Item 7	1,77	gut	Gutschein (Dummy)	0
uf8	Usability formativ Item 8	2,85	kauf_bs	Kauf im Shop (Dummy)	0
uf9	Usability formativ Item 9	3,53	kauf_as	Kauf anderer S. (Dummy)	0
uf10	Usability formativ Item 10	1,77	buch	Buch (Dummy)	-
uf11	Usability formativ Item 11	0,95	dvd	DVD (Dummy)	-
uf12	Usability formativ Item 12	1,22	cd	CD (Dummy)	-
uf13	Usability formativ Item 13	1,49	pat	Druckerpatrone (Dummy)	-
uf14	Usability formativ Item 14	0,27	tick	Ticket (Dummy)	-
uf15	Usability formativ Item 15	2,45	shop_be	Bekanntheitsgrad d. Shops	0
uf16	Usability formativ Item 16	1,77	ki1	Kaufintention Item 1	0,27
uf18	Usability formativ Item 18	1,22	ki2	Kaufintention Item 2	0,27
uf19	Usability formativ Item 19	1,22	ki4	Kaufintention Item 4	0,27
uf20	Usability formativ Item 20	1,09	t1	Vertrauen Item 1	1,63
uf21	Usability formativ Item 21	5,03	t2	Vertrauen Item 2	1,63
uf22	Usability formativ Item 22	1,77	t3	Vertrauen Item 3	1,63
uf23	Usability formativ Item 23	3,40	t4	Vertrauen Item 4	1,63
uf24	Usability formativ Item 24	0,95	ur1	Usability reflektiv Item 1	0
uf25	Usability formativ Item 25	1,22	ur2	Usability reflektiv Item 2	0
uf26	Usability formativ Item 26	0,82	ur3	Usability reflektiv Item 3	0
uf27	Usability formativ Item 27	1,90	ur4	Usability reflektiv Item 4	0
uf28	Usability formativ Item 28	1,77	ur5	Usability reflektiv Item 5	0
uf29	Usability formativ Item 29	1,77	ur6	Usability reflektiv Item 6	0
uf31	Usability formativ Item 31	1,09	ur7	Usability reflektiv Item 7	0
uf32	Usability formativ Item 32	4,48	ur8	Usability reflektiv Item 8	0
uf33	Usability formativ Item 33	5,43	ur9	Usability reflektiv Item 9	0,27
uf34	Usability formativ Item 34	1,09	a1	Ästhetik Item 1	0,41
uf35	Usability formativ Item 35	2,45	a2	Ästhetik Item 2	0,41
uf36	Usability formativ Item 36	2,31	wa1	W. Produktanzahl Item 1	1,49
uf38	Usability formativ Item 38	3,67	wa2	W. Produktanzahl Item 2	1,49
uf39	Usability formativ Item 39	5,43	se1	Einkaufsvergnügen Item 1	0,68
uf40	Usability formativ Item 40	7,74	se2	Einkaufsvergnügen Item 2	0,68

Anmerkung: Anteil fehlender Werte in Prozent (% Missing).

Tab. E-6: Harman's Single Factor-Test – Eigenwerte und Gesamtvarianz der Faktorenlösung

<i>Komponente</i>	<i>Anfänglicher Eigenwert <math>\lambda</math></i>	<i>Erklärte Gesamtvarianz</i>	
		<i>% der Varianz</i>	<i>Kumulierte %</i>
1	23,26	29,07	29,07
2	3,98	4,97	34,05
3	3,42	4,27	38,32
4	2,66	3,33	41,65
5	2,45	3,07	44,71
6	2,19	2,73	47,45
7	2,11	2,64	50,09
8	1,81	2,26	52,35
9	1,67	2,09	54,44
10	1,64	2,05	56,49
11	1,50	1,87	58,36
12	1,43	1,78	60,14
13	1,38	1,72	61,87
14	1,32	1,65	63,52
15	1,23	1,54	65,06
16	1,20	1,50	66,55
17	1,18	1,47	68,03
18	1,04	1,31	69,33
19	1,01	1,27	70,60

Tab. E-7: Prüfung der Normalverteilungsannahme für alle erhobenen Indikatoren

Indikator	Sch	z-Wert Sch	Ex	z-Wert Ex	Indikator	Sch	z-Wert Sch	Ex	z-Wert Ex
ur1	-0,421	-4,666	-0,968	-5,361	uf19c	-0,438	-4,860	-1,223	-6,795
ur2	-0,785	-8,698	-0,717	-3,972	uf20	-0,813	-9,023	-0,304	-1,691
ur3	-0,388	-4,303	-1,059	-5,865	uf21c	-0,889	-9,865	-0,096	-0,531
ur4	-0,892	-9,882	-0,054	-0,298	uf22	-0,609	-6,755	-0,502	-2,789
ur5	0,355	3,930	-1,267	-7,015	uf23	-0,644	-7,148	-0,662	-3,677
ur6	-0,817	-9,052	-0,482	-2,671	uf24	-0,878	-9,747	-0,386	-2,147
ur7	-0,383	-4,246	-1,11	-6,147	uf25	-1,079	-11,973	0,070	0,389
ur8	0,071	0,781	-1,262	-6,991	uf26	-0,939	-10,424	-0,388	-2,158
ur9	-0,255	-2,825	-1,176	-6,511	uf27	-0,802	-8,905	-0,593	-3,292
t1	-0,459	-5,083	-0,728	-4,034	uf28	-0,370	-4,101	-1,084	-6,024
t2	-0,696	-7,708	-0,175	-0,97	uf29	-0,683	-7,580	-0,641	-3,560
t3	-0,485	-5,377	-0,486	-2,693	uf30	0,165	1,713	-1,020	-5,302
t4	-0,635	-7,03	-0,232	-1,283	uf31	-0,285	-3,168	-1,244	-6,915
wa1	-0,118	-1,309	-1,326	-7,345	uf32	-0,274	-3,046	-0,959	-5,329
wa2	-0,33	-3,65	-1,183	-6,554	uf33	0,248	2,751	-1,263	-7,018
wa3	0,551	6,102	0,178	0,986	uf34	-2,104	-23,350	4,531	25,175
pi1	-0,927	-10,263	0,173	0,961	uf35	-0,500	-5,549	-1,121	-6,227
pi2	-1,687	-18,689	2,455	13,595	uf36	-0,715	-7,940	-0,434	-2,412
pi3	-1,234	-13,67	0,997	5,521	uf37	-0,525	-5,385	-0,964	-4,956
se1	0,196	2,168	-1,062	-5,881	uf38	-0,921	-10,224	-0,175	-0,973
se2	-0,667	-7,391	-0,885	-4,903	uf39c	-0,626	-6,948	-0,960	-5,333
a1	-0,028	-0,312	-1,262	-6,989	uf40	-0,810	-8,987	-0,191	-1,061
a2	0,103	1,14	-1,15	-6,368	uf41	-0,209	-2,320	-1,033	-5,740
ki1	0,447	4,947	-1,113	-6,165	uf42	-1,546	-17,158	1,734	9,635
ki2	0,805	8,912	-0,545	-3,018	uf43	-0,530	-5,884	-1,218	-6,770
ki3	0,029	0,316	-1,450	-7,863	uf44	-2,286	-25,372	5,377	29,877
ki4	0,644	7,131	-0,798	-4,422	uf45	-1,399	-15,529	0,723	4,016
subj	-0,226	-2,513	-1,144	-6,359	uf46	-1,162	-12,894	0,789	4,384
uf1	-0,695	-7,709	-0,428	-2,381	uf47	-1,241	-13,776	1,111	6,175
uf2	-0,238	-2,458	-1,075	-5,565	uf48	-0,641	-7,111	-0,689	-3,827
uf3	0,124	1,088	-1,093	-4,799	uf49c	-1,666	-18,488	2,136	11,867
uf4	-1,009	-11,200	0,084	0,467	uf50	-0,594	-6,587	-0,687	-3,816
uf5c	-1,236	-13,719	1,113	6,185	uf51	0,103	1,066	-1,203	-6,227
uf6	-0,724	-8,030	-0,439	-2,441	uf52	0,284	2,896	-1,196	-6,114
uf7	-0,692	-7,683	-0,553	-3,071	uf53	-0,424	-4,345	-0,998	-5,123
uf8	-1,221	-13,548	0,541	3,008	uf54c	-1,158	-12,854	0,117	0,651
uf9	-1,325	-14,704	0,918	5,099	uf55	-0,400	-4,434	-1,123	-6,237
uf10	-0,677	-7,511	-0,534	-2,969	uf56c	-1,766	-19,597	1,959	10,886
uf11c	-0,783	-8,685	-0,540	-2,999	uf57c	-0,349	-3,869	-1,159	-6,442
uf12	-0,330	-3,663	-1,183	-6,571	uf58c	-1,572	-17,442	1,580	8,780
uf13c	-0,974	-10,813	-0,004	-0,020	sv1	-0,882	-9,769	-0,489	-2,706
uf14	-0,606	-6,729	-0,621	-3,449	sv2	-0,774	-7,247	0,136	0,637
uf15	-0,438	-4,857	-0,906	-5,032	sv3	-0,522	-5,089	-0,600	-2,928
uf16c	-0,810	-8,995	-0,444	-2,469	sv4	0,824	8,356	-0,208	-1,055
uf17	0,307	3,163	-1,127	-5,818	sv5c	-0,360	-3,987	-1,107	-6,131
uf18c	-1,156	-12,832	0,346	1,925	sv6	-0,351	-3,318	-0,935	-4,426

Anmerkung: Schiefe (Sch), Exzess (Ex). Der Kolmogorov-Smirnov-Test auf Normalverteiltheit wird für alle Indikatoren signifikant ( $p < 0,001$ ).

Tab. E-8.1: Harman's Single Factor-Test - Komponentenmatrix

Item	Komponente																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
ur7	0,8																		
ur1	0,8																		
se2c	0,8																		
se1	0,8																		
ur9	0,8																		
uf12	0,8																		
uf14	0,8																		
ur4	0,7																		
a1	0,7																		
ur8	0,7																		
ur6	0,7																		
uf15	0,7																		
ur3	0,7																		
ur6	0,7																		
ur5	0,7	-0,3																	
ur2c	0,7																		
uf28	0,7																		
uf55	0,7																		
uf22	0,7																		
uf24	0,7																		
ki1	0,7	-0,3				-0,3													
a2	0,6																		
uf20	0,6																		
uf10	0,6																		
uf50	0,6														-0,3				
reco	0,6	-0,3																	
ki2	0,6	-0,4																	
uf7	0,6	0,3																	
ki4	0,6	-0,3																	
uf29	0,6																	-0,5	
t1	0,6		0,4		-0,3														
uf26	0,6							-0,4											
lov	0,6																		
uf25	0,6							-0,3											
ki3	0,6			0,3															
uf27	0,6				-0,3														
uf19	0,6			-0,3				0,3											
uf13	0,6	0,3																	
uf32	0,6																		
uf11	0,6																		
t4	0,5		0,5		-0,4														
uf1	0,5																		
uf46	0,5	0,3																	
t3	0,5		0,5		-0,4														
uf31	0,5								0,4										
uf23	0,5																		
uf5	0,5					0,3							-0,3						
uf8	0,4												0,3						
uf57	0,4												-0,4	0,4					





Tab. E-9.1: Auflistung aller Probanden mit potenziell problematischen Antwortverhalten

<i>Teilnehmer- code</i>	<i>Alter</i>	<i>Geschlecht</i>	<i>Bearbeitungs- dauer</i>	<i>SD Shop1</i>	<i>SD Shop 2</i>	<i>Usability- reflektiv</i>	<i>Einkaufs- Vergnügen</i>
b1 7 1	27	m	0:52	3	4,91	2,38	6
b2 1 1	39	m	1:00	1,49	3,65	1,25	6
b2 4 1	28	m	1:35	0,92	1,9	0,14	0
b2 5 2	28	m	0:56	4,16	3,89	2,75	6
b3 1 1	21	w	0:38	2,15	1,65	1,25	2
b3 4 1	23	w	0:56	0,67	1,31	1,25	1
b3 5 1	28	m	0:50	1,46	0,73	0,25	1
b4 5 1	42	w	0:40	1,43	4,35	0,63	1
b5 7 1	20	w	0:37	3,4	1,59	0,50	0
b6 3 1	37	w	0:27	2,53	1,57	0,50	2
b7 1 2	30	m	0:39	2,78	4,63	4,00	0
b7 5 2	68	w	1:08	1,68	1	0,25	1
d1 4 2	22	w	0:43	3,35	0,96	0,5	1
d1 6 2	19	w	1:33	1,38	0,95	0,5	1
d2 1 2	40	m	1:02	2,67	2,75	4,5	2
d2 7 1	24	m	0:43	3,67	3,69	3,25	6
d3 1 1	45	m	1:27	2,24	3,01	6	0
d3 1 2	22	w	0:32	2,88	2,78	2,13	0
d4 5 1	31	m	0:54	4,55	4,94	4,63	1
d4 6 2	19	m	0:39	1,95	4,41	1,63	5
d5 1 1	58	m	2:19	4,53	6,2	1,00	0
d5 3 1	30	w	1:18	2,77	0,82	2,75	1
d5 4 2	25	m	1:11	4,9	0,85	0	0
d6 4 1	30	w	1:01	2,4	0,96	0,13	1
m1 2 1	15	m	0:57	3,53	3,4	0,25	6
m1 7 2	26	m	0:31	2,51	3,30	0,88	2
m2 6 1	29	m	0:40	4,62	1,02	0,00	1
m2 6 2	50	w	1:01	2,65	0,5	0	0
m3 1 2	53	w	1:13	0,81	0,81	0	0
m3 2 1	30	m	0:52	3,09	3,98	0,5	5
m3 6 1 1	26	m	0:39	4,72	0,92	0,5	3
m3 6 2	39	w	1:36	1,58	0,74	0	1
m5 2 1	40	w	0:53	2,37	3,54	5,88	0
m5 4 1	43	m	0:56	0,94	0,96	1,14	2
m5 4 2	25	m	0:46	3,28	3,36	1	5
m6 1 2	71	w	0:57	0,58	3,9	0,88	2
m6 4 2 1	49	w	3:06	1,87	4,40	-0,50	0
m6 5 1	48	m	1:22	1,98	0,92	0,86	0
m7 4 1	64	m	1:02	3,09	0,84	0,13	0
p3 6 1	51	m	2:00	3,63	0,75	0,50	1
p3 7 2	33	m	1:27	4	0,9	0,71	0
p6 5 2	22	m	1:03	3,23	2,69	2	5
p6 7 1	48	m	1:06	5,5	4,17	4,29	2
p7 3 1	23	w	1:18	3,25	3,22	5,13	2
p7 4 2	28	m	1:11	3,56	5,02	2,13	5
t1 2 1	20	w	0:50	2,69	2,59	5,38	2
t1 3 1	21	w	1:03	3,43	5,32	4,75	3
t2 1 1	42	w	1:40	1,91	2,71	5,14	1
t2 3 1	20	m	0:46	4,05	5,39	2,25	5

Anmerkung: Die grau hinterlegten Zellen weisen auf ein auffälliges Antwortverhalten der jeweiligen Kategorie hin. Weiblich (w), männlich (m), Standardabweichung (SD).

Tab. E-9.2: Auflistung aller Probanden mit potenziell problematischen Antwortverhalten (Fortsetzung)

Teilnehmercode	Alter	Geschlecht	Bearbeitungsdauer	SD Shop1	SD Shop 2	Usability-reflektiv, konsistentes Antworten	Einkaufsvergnügen, konsistentes Antworten
t4_2_1	22	m	0:53	3,29	3,71	1,13	5
t4_3_2	22	w	0:30	2,25	2,01	-0,63	0
t4_5_1	31	w	2:27	1,81	3,33	1,75	4
t4_5_2	66	m	2:12	2,16	4,15	2,25	4
t5_6_1	29	m	1:00	3,01	2,67	6	0

Anmerkung: Die grau hinterlegten Zellen weisen auf ein auffälliges Antwortverhalten der jeweiligen Kategorie hin. Weiblich (w), männlich (m), Standardabweichung (SD).

Tab. E-10: Mustermatrix der ersten berechneten Hauptachsenanalyse

Faktor	1	2	3	4	5	$h_i^2$	MSA-Wert
Usability ur6	<b>0,756</b>	0,032	0,099	0,010	-0,072	0,704	0,959
Usability ur4	<b>0,721</b>	0,057	0,118	-0,002	-0,069	0,662	0,951
Usability ur1	<b>0,671</b>	0,039	0,167	0,082	-0,168	0,763	0,967
Usability ur7	<b>0,664</b>	0,043	0,136	0,087	-0,245	0,819	0,966
Usability ur2	<b>0,643</b>	0,052	0,033	0,013	-0,125	0,541	0,960
Usability ur9	<b>0,623</b>	0,016	0,134	0,164	-0,163	0,700	0,955
Usability ur3	<b>0,550</b>	0,062	0,080	0,059	-0,253	0,594	0,970
Einkaufsvergnügen se2c	<b>0,478</b>	-0,004	0,151	0,089	<b>-0,311</b>	0,619	0,971
Usability ur8	<b>0,410</b>	-0,020	0,169	0,259	-0,257	0,609	0,951
Produkt Involvement pi1	-0,034	<b>0,936</b>	-0,010	0,028	0,042	0,868	0,631
Produkt Involvement pi3	0,034	<b>0,800</b>	0,013	-0,016	0,027	0,646	0,666
Produkt Involvement pi2	-0,008	<b>0,590</b>	0,031	-0,004	-0,041	0,358*	0,808
Vertrauen t4	0,017	0,028	<b>0,890</b>	0,022	-0,037	0,844	0,905
Vertrauen t2	0,047	0,034	<b>0,830</b>	0,088	0,015	0,765	0,911
Vertrauen t1	0,090	0,004	<b>0,735</b>	0,143	-0,081	0,734	0,940
Vertrauen t3	0,020	0,096	<b>0,693</b>	0,022	-0,153	0,622	0,942
Wahrge. Produktanzahl wa2	0,027	0,017	0,015	<b>0,833</b>	0,013	0,705	0,840
Wahrge. Produktanzahl wa3	-0,023	-0,019	0,043	<b>0,809</b>	-0,112	0,295*	0,892
Wahrge. Produktanzahl wa1	-0,062	-0,001	0,070	<b>0,533</b>	0,005	0,734	0,864
Ästhetik ae1	0,010	-0,008	0,037	0,067	<b>-0,883</b>	0,852	0,905
Ästhetik ae2	-0,043	0,006	0,062	0,019	<b>-0,865</b>	0,760	0,893
Einkaufsvergnügen se1	0,259	0,085	0,122	0,171	<b>-0,544</b>	0,715	0,961
Usability ur5	0,250	0,067	0,081	0,280	<b>-0,431</b>	0,602	0,969
$\Sigma=67,45$	42,70	8,20	7,40	5,69	3,46		
$\lambda_{j=6}=0,66$	10,12	2,20	2,03	1,61	1,06		

Anmerkung: \* Kommunalität liegt unterhalb des kritischen Wertes  $h_i^2=0,5$ . Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_j$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

Tab. E-11: Mustermatrix der zweiten berechneten Hauptachsenanalyse

Faktor	1	2	3	4	5	$h_i^2$	MSA-Wert
Usability ur6	<b>0,759</b>	0,090	0,042	0,018	-0,069	0,959	0,707
Usability ur4	<b>0,722</b>	0,110	0,074	0,004	-0,070	0,952	0,665
Usability ur1	<b>0,670</b>	0,159	0,038	0,109	-0,151	0,967	0,763
Usability ur7	<b>0,658</b>	0,124	0,054	0,129	-0,222	0,967	0,818
Usability ur2	<b>0,638</b>	0,022	0,052	0,062	-0,104	0,962	0,536
Usability ur9	<b>0,613</b>	0,126	0,022	0,196	-0,138	0,955	0,700
Usability ur3	<b>0,551</b>	0,074	0,070	0,067	-0,247	0,970	0,597
Einkaufsvergnügen se2c	<b>0,485</b>	0,146	-0,026	0,121	-0,287	0,973	0,622
Usability ur8	<b>0,392</b>	0,162	-0,008	0,297	-0,226	0,951	0,610
Vertrauen t4	0,023	<b>0,883</b>	0,026	0,034	-0,040	0,906	0,844
Vertrauen t2	0,044	<b>0,824</b>	0,043	0,096	0,012	0,911	0,766
Vertrauen t1	0,088	<b>0,730</b>	0,010	0,145	-0,080	0,941	0,731
Vertrauen t3	0,024	<b>0,685</b>	0,099	0,041	-0,150	0,941	0,621
Produkt Involvement pi3	0,000	-0,003	<b>0,912</b>	-0,032	0,007	0,562	0,829
Produkt Involvement pi1	-0,046	0,002	<b>0,826</b>	0,012	0,027	0,545	0,675
Wahrge. Produktanzahl wa1	-0,082	0,043	-0,018	<b>0,867</b>	-0,061	0,848	0,778
Wahrge. Produktanzahl wa2	-0,030	0,025	0,019	<b>0,840</b>	0,050	0,824	0,679
Ästhetik ae1	0,022	0,030	-0,008	0,101	<b>-0,861</b>	0,905	0,852
Ästhetik ae2	-0,033	0,054	0,012	0,039	<b>-0,860</b>	0,892	0,772
Einkaufsvergnügen se1	0,256	0,115	0,076	0,218	<b>-0,510</b>	0,961	0,713
Usability ur5	0,233	0,076	0,073	<b>0,312</b>	<b>-0,400</b>	0,969	0,601
$\Sigma=70,85$	46,54	7,91	7,23	5,55	3,65		
$\lambda_{i=6}=0,578$	10,06	1,92	1,78	1,43	1,01		

Anmerkung: Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_j$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

Tab. E-12: Mustermatrix der dritten berechneten Hauptachsenanalyse

Faktor	1	2	3	4	5	$h_i^2$	MSA-Wert
Usability ur6	<b>0,758</b>	0,086	0,040	0,017	-0,077	0,707	0,957
Usability ur4	<b>0,722</b>	0,108	0,072	0,003	-0,072	0,662	0,951
Usability ur1	<b>0,676</b>	0,156	0,034	0,106	-0,154	0,762	0,965
Usability ur7	<b>0,664</b>	0,119	0,051	0,127	-0,229	0,820	0,965
Usability ur2	<b>0,641</b>	0,020	0,050	0,053	-0,113	0,538	0,960
Usability ur9	<b>0,621</b>	0,118	0,019	0,196	-0,148	0,704	0,953
Usability ur3	<b>0,556</b>	0,071	0,067	0,068	-0,246	0,594	0,967
Einkaufsvergnügen se2c	<b>0,492</b>	0,142	-0,029	0,118	-0,289	0,687	0,970
Usability ur8	<b>0,404</b>	0,155	-0,012	0,295	-0,231	0,610	0,947
Vertrauen t4	0,026	<b>0,878</b>	0,024	0,039	-0,048	0,844	0,903
Vertrauen t2	0,047	<b>0,817</b>	0,042	0,101	-0,001	0,620	0,908
Vertrauen t1	0,094	<b>0,723</b>	0,007	0,153	-0,087	0,732	0,938
Vertrauen t3	0,031	<b>0,682</b>	0,096	0,047	-0,150	0,765	0,938
Produkt Involvement pi3	0,007	-0,001	<b>0,911</b>	-0,026	0,000	0,831	0,557
Produkt Involvement pi1	-0,038	0,004	<b>0,824</b>	0,019	0,021	0,675	0,542
Wahrge. Produktanzahl wa1	-0,079	0,028	-0,010	<b>0,882</b>	-0,063	0,796	0,828
Wahrge. Produktanzahl wa2	-0,022	0,022	0,020	<b>0,812</b>	0,032	0,647	0,798
Ästhetik ae1	0,018	0,016	-0,007	0,110	<b>-0,881</b>	0,877	0,895
Ästhetik ae2	-0,030	0,044	0,014	0,050	<b>-0,864</b>	0,779	0,882
Einkaufsvergnügen se1	0,278	0,119	0,069	0,204	<b>-0,492</b>	0,622	0,969
$\Sigma=71,36$	45,93	8,30	7,57	5,69	3,87		
$\lambda_{i=6}=0,58$	9,48	1,92	1,77	1,40	1,01		

Anmerkung: Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_j$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

Tab. E-13: Cronbachs-Alpha der reflektiven Usability-Skala ufosV2r für die einzelnen Benutzerschnittstellen

<i>Produktart</i>	<i>Nummer</i>	<i>Online-Shop</i>	<i><math>\alpha</math></i>	<i>N</i>
Bücher	1	Karstadt.de	0,935	22
	2	JPC.de	0,947	22
	3	Libri.de	0,917	24
	4	Booxtra.de	0,943	24
	5	BOL.de	0,934	23
	6	Buchkatalog.de	0,907	23
	7	Buch24.de	0,943	24
	Gesamt		0,933	162
DVD	8	DVD-Laden-24.de	0,911	24
	9	Fröschl.de	0,927	24
	10	inCinea.de	0,898	24
	11	Bücher.de	0,964	24
	12	Avides.com	0,876	24
	13	Delta-Shopping.de	0,951	24
	14	DVD-Galaxis.de	0,900	24
	Gesamt		0,935	168
CD	15	WOM.de	0,957	23
	16	CyberCD.de	0,886	23
	17	AlphaMusic.de	0,949	24
	18	Abella.de	0,952	21
	19	Buch.de	0,897	24
	20	Amazon.de	0,955	24
	21	abc-CD.de	0,859	23
	Gesamt		0,935	162
Druckerpatronen	22	Druckerpatronen.de	0,951	24
	23	PrintPower.net	0,941	23
	24	Express-Toner.de	0,948	23
	25	Alphaco.de	0,949	23
	26	ACI-Druckerpatronen.de	0,928	23
	27	Planet-Shop.de	0,929	23
	28	Alternate.de	0,921	23
	Gesamt		0,944	162
Veranstaltungstickets	29	Eventim.de	0,918	14
	30	Getgo.de	0,825	11
	31	Kartenkaufen.de	0,882	14
	32	Kartenhaus.de	0,918	15
	33	Karten-Online.de	0,965	8
	34	Ticket-Online.de	0,886	11
	35	TicketCenter.de	0,883	9
	Gesamt		0,931	82

Anmerkung: Cronbachs Alpha ( $\alpha$ ), Anzahl der vorliegenden Bewertungen (N).

Tab. E-14: Rangreihe der Shop-Mittelwerte auf der reflektiven Usability-Skala ufosV2r

Produktart	Rang	Shop	M	SD	N
Bücher	1.	BOL.de	4,29	1,39	23
	2.	Booxtra.de	4,02	1,63	24
	3.	Karstadt.de	3,85	1,53	22
	4.	Buch24.de	3,83	1,59	24
	5.	Buchkatalog.de	3,62	1,31	23
	6.	Libri.de	3,59	1,47	24
	7.	JPC.de	3,43	1,66	22
	<i>Gesamt</i>			3,81	1,51
DVDs	1.	inCinea.de	4,78	1,00	24
	2.	Bücher.de	4,17	1,59	24
	3.	Fröschl.de	3,83	1,42	24
	4.	Delta-Shopping.de	3,72	1,55	24
	5.	DVD-Galaxis.de	3,24	1,22	24
	6.	DVD-Laden-24.de	3,19	1,51	24
	7.	Avides.com	2,70	1,18	24
	<i>Gesamt</i>			3,66	1,49
CDs	1.	Amazon.de	4,66	1,26	24
	2.	Buch.de	4,37	0,98	24
	3.	WOM.de	3,92	1,62	23
	4.	abc-CD.de	3,85	1,18	23
	5.	AlphaMusic.de	3,75	1,75	24
	6.	Abella.de	3,72	1,57	21
	7.	CyberCD.de	2,63	1,38	23
	<i>Gesamt</i>			3,85	1,51
Patronen	1.	Druckerpatronen.de	4,43	1,28	24
	2.	Alphaco.de	3,86	1,57	23
	3.	Planet-Shop.de	3,53	1,40	23
	4.	Alternate.de	3,30	1,61	23
	5.	PrintPower.net	2,92	1,50	23
	6.	Express-Toner.de	2,76	1,70	23
	7.	ACI-Druckerpatronen.de	2,73	1,48	23
	<i>Gesamt</i>			3,37	1,59
Tickets	1.	Getgo.de	4,43	0,93	11
	2.	Eventim.de	4,05	1,67	14
	3.	Kartenhaus.de	3,92	1,44	15
	4.	TicketCenter.de	3,36	1,24	9
	5.	Karten-Online.de	2,80	1,92	8
	6.	Kartenkaufen.de	2,40	1,38	14
	7.	Ticket-Online.de	2,28	1,51	11
	<i>Gesamt</i>			3,36	1,62
<i>Gesamt</i>			3,63	1,54	756

Anmerkung: Die reflektive Usability-Skala ufosV2r umfasst die acht Indikatoren ur1, ur2, ur3, ur4, ur6, ur7, ur8 und ur9. Cronbachs Alpha  $\alpha=0,936$ . Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-15: Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zum Vertrauen

Produktart	Rang	Shop	M	SD	N
Bücher	1.	Karstadt.de	4,63	1,35	22
	2.	BOL.de	4,54	1,30	23
	3.	Libri.de	4,42	1,18	24
	4.	Booxtra.de	4,37	1,25	24
	5.	JPC.de	4,36	1,02	22
	6.	Buch24.de	4,23	1,52	24
	7.	Buchkatalog.de	3,99	1,76	23
		<i>Gesamt</i>		4,36	1,35
DVDs	1.	Bücher.de	4,56	1,19	24
	2.	inCinea.de	4,20	1,00	24
	3.	Fröschl.de	3,57	1,54	24
	4.	Avides.com	3,33	1,30	24
	5.	DVD-Galaxis.de	3,28	1,24	24
	6.	Delta-Shopping.de	3,26	1,56	24
	7.	DVD-Laden-24.de	2,98	1,23	24
		<i>Gesamt</i>		3,60	1,39
CDs	1.	Amazon.de	5,17	0,88	24
	2.	WOM.de	4,44	1,17	23
	3.	Buch.de	4,27	1,34	24
	4.	AlphaMusic.de	4,08	1,39	24
	5.	abc-CD.de	3,55	1,42	23
	6.	Abella.de	3,49	1,28	21
	7.	CyberCD.de	2,76	1,61	23
		<i>Gesamt</i>		3,98	1,48
Patronen	1.	Alphaco.de	4,24	1,42	23
	2.	Druckerpatronen.de	4,09	1,26	24
	3.	Alternate.de	3,99	1,36	23
	4.	Planet-Shop.de	3,76	1,26	23
	5.	Express-Toner.de	3,26	1,41	23
	6.	PrintPower.net	3,03	1,33	23
	7.	ACI-Druckerpatronen.de	2,93	1,35	23
		<i>Gesamt</i>		3,62	1,41
Tickets	1.	Eventim.de	4,55	1,39	14
	2.	Getgo.de	4,50	0,93	11
	3.	Kartenhaus.de	4,33	1,34	15
	4.	Karten-Online.de	4,16	1,02	8
	5.	Ticket-Online.de	3,80	1,52	11
	6.	Kartenkaufen.de	3,67	1,12	14
	7.	TicketCenter.de	3,52	1,10	9
		<i>Gesamt</i>		4,10	1,25
<i>Gesamt</i>			3,91	1,42	756

Anmerkung: Die reflektive Vertrauens-Skala umfasst die vier Indikatoren t1, t2, t3 und t4. Cronbachs Alpha  $\alpha=0,913$ . Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-16: Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur wahrgenommenen Produktanzahl

Produktart	Rang	Shop	M	SD	N
Bücher	1.	Karstadt.de	4,95	1,33	22
	2.	BOL.de	4,39	1,64	23
	3.	Libri.de	3,98	1,74	24
	4.	Buch24.de	3,67	1,66	24
	5.	Buchkatalog.de	3,39	2,12	23
	6.	Booxtra.de	3,38	1,69	24
	7.	JPC.de	3,33	1,99	22
	Gesamt			3,86	1,81
DVDs	1.	Bücher.de	4,49	1,67	24
	2.	Fröschl.de	4,02	1,91	24
	3.	inCinea.de	3,90	1,90	24
	4.	DVD-Laden-24.de	3,29	1,59	24
	5.	DVD-Galaxis.de	3,16	1,61	24
	6.	Delta-Shopping.de	2,60	1,75	24
	7.	Avides.com	1,65	1,58	24
	Gesamt			3,30	1,91
CDs	1.	Amazon.de	5,75	0,56	24
	2.	WOM.de	4,50	1,57	23
	3.	Buch.de	4,15	1,43	24
	4.	abc-CD.de	3,09	2,07	23
	5.	AlphaMusic.de	3,01	1,68	24
	6.	Abella.de	2,33	2,01	21
	7.	CyberCD.de	2,11	1,44	23
	Gesamt			3,59	1,98
Patronen	1.	Alternate.de	4,37	1,80	23
	2.	PrintPower.net	3,85	1,60	23
	3.	Alphaco.de	3,20	1,70	23
	4.	Planet-Shop.de	2,63	1,97	23
	5.	Express-Toner.de	2,43	1,56	23
	6.	Druckerpatronen.de	2,23	1,41	24
	7.	ACI-Druckerpatronen.de	2,04	1,73	23
	Gesamt			2,96	1,85
Tickets	1.	Getgo.de	4,50	1,30	11
	2.	Eventim.de	4,32	1,67	14
	3.	Kartenhaus.de	4,20	1,37	15
	4.	Kartenkaufen.de	2,46	1,25	14
	5.	Karten-Online.de	2,44	1,68	8
	6.	Ticket-Online.de	2,36	1,48	11
	7.	TicketCenter.de	1,94	1,65	9
	Gesamt			3,30	1,75
Gesamt			3,41	1,90	756

Anmerkung: Die reflektive Skala zur wahrgenommenen Produktanzahl umfasst die zwei Indikatoren wa1 und wa2. Cronbachs Alpha  $\alpha=0,846$ . Spearman-Brown Korrelation  $r=0,744$ . Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).



Tab. E-17: Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur Ästhetik

Produktart	Rang	Shop	M	SD	N
Bücher	1.	BOL.de	3,89	1,20	23
	2.	Buch24.de	3,35	1,66	24
	3.	Buchkatalog.de	3,22	1,74	23
	4.	Libri.de	2,98	1,72	24
	5.	Karstadt.de	2,84	1,80	22
	6.	Booxtra.de	2,77	1,86	24
	7.	JPC.de	2,00	1,54	22
		<i>Gesamt</i>		3,02	1,71
DVDs	1.	inCinea.de	4,54	1,34	24
	2.	Bücher.de	3,13	2,08	24
	3.	Delta-Shopping.de	2,96	1,81	24
	4.	Fröschl.de	2,77	1,93	24
	5.	DVD-Galaxis.de	2,27	1,72	24
	6.	DVD-Laden-24.de	1,85	1,93	24
	7.	Avides.com	1,75	1,43	24
		<i>Gesamt</i>		2,75	1,95
CDs	1.	Amazon.de	4,00	1,66	24
	2.	Buch.de	3,54	1,57	24
	3.	AlphaMusic.de	3,35	1,70	24
	4.	WOM.de	3,35	1,92	23
	5.	Abella.de	2,87	1,65	21
	6.	abc-CD.de	2,30	1,83	23
	7.	CyberCD.de	2,22	1,57	23
		<i>Gesamt</i>		3,10	1,78
Patronen	1.	Druckerpatronen.de	3,00	1,80	24
	2.	Alphaco.de	3,00	1,85	23
	3.	Alternate.de	2,96	1,81	23
	4.	Planet-Shop.de	2,51	1,96	23
	5.	ACI-Druckerpatronen.de	2,22	1,76	23
	6.	PrintPower.net	1,74	1,51	23
	7.	Express-Toner.de	1,41	1,30	23
		<i>Gesamt</i>		2,41	1,80
Tickets	1.	Getgo.de	4,50	1,30	11
	2.	Eventim.de	4,32	1,67	14
	3.	Kartenhaus.de	4,20	1,37	15
	4.	Kartenkaufen.de	2,46	1,25	14
	5.	Karten-Online.de	2,44	1,68	8
	6.	Ticket-Online.de	2,36	1,48	11
	7.	TicketCenter.de	1,94	1,65	9
		<i>Gesamt</i>		3,30	1,75
<i>Gesamt</i>			2,82	1,81	756

Anmerkung: Die reflektive Ästhetik-Skala umfasst die zwei Indikatoren a1 und a2. Cronbachs Alpha  $\alpha=0,908$ , Spearman-Brown Korrelation  $r=0,837$ , Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-18: Rangreihe der Shop-Mittelwerte über das Item zum wahrgenommenen Preis

Produktart	Rang	Shop	M	SD	N
Bücher	1.	Booxtra.de	4,88	1,30	24
	2.	Buchkatalog.de	4,76	1,72	23
	3.	JPC.de	4,52	1,39	22
	4.	BOL.de	4,40	1,75	23
	5.	Karstadt.de	4,20	1,91	22
	6.	Libri.de	4,13	1,57	24
	7.	Buch24.de	4,00	1,74	24
		<i>Gesamt</i>		4,41	1,63
DVDs	1.	Avides.com	3,46	2,02	24
	2.	inCinea.de	3,37	2,04	24
	3.	Fröschl.de	3,25	1,89	24
	4.	Bücher.de	3,05	1,63	24
	5.	Delta-Shopping.de	2,94	2,05	24
	6.	DVD-Laden-24.de	2,77	2,05	24
	7.	DVD-Galaxis.de	1,42	1,67	24
		<i>Gesamt</i>		2,89	1,99
CDs	1.	Amazon.de	4,25	1,33	24
	2.	AlphaMusic.de	3,92	1,91	24
	3.	Buch.de	3,14	1,79	24
	4.	Abella.de	2,72	2,07	21
	5.	abc-CD.de	2,65	2,19	23
	6.	WOM.de	2,61	2,21	23
	7.	CyberCD.de	1,61	1,53	23
		<i>Gesamt</i>		3,00	2,02
Patronen	1.	Alphaco.de	3,74	1,74	23
	2.	Druckerpatronen.de	3,72	1,76	24
	3.	Alternate.de	3,17	1,65	23
	4.	Planet-Shop.de	2,73	2,04	23
	5.	PrintPower.net	2,61	1,88	23
	6.	Express-Toner.de	2,39	1,34	23
	7.	ACI-Druckerpatronen.de	1,76	1,80	23
		<i>Gesamt</i>		2,88	1,85
Tickets	1.	Karten-Online.de	4,50	1,31	8
	2.	Ticket-Online.de	3,73	1,19	11
	3.	Kartenhaus.de	3,34	2,45	15
	4.	Kartenkaufen.de	3,20	1,27	14
	5.	Getgo.de	2,82	2,04	11
	6.	TicketCenter.de	2,44	1,24	9
	7.	Eventim.de	2,43	1,70	14
		<i>Gesamt</i>		3,16	1,77
<i>Gesamt</i>			3,28	1,96	756

Anmerkung: Der wahrgenommene Preis wurde mit Hilfe des Single-Items subjektiv erfasst. Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-19: Rangreihe der Shop-Mittelwerte auf der reflektiven Skala zur Kaufintention

Produktart	Rang	Shop	M	SD	N
Bücher	1.	Booxtra.de	2,56	1,98	24
	2.	BOL.de	2,25	1,68	23
	3.	Buch24.de	2,16	1,84	24
	4.	Buchkatalog.de	1,98	1,42	23
	5.	Karstadt.de	1,95	1,75	22
	6.	Libri.de	1,88	2,04	24
	7.	JPC.de	1,86	1,74	22
		<i>Gesamt</i>		2,09	1,77
DVDs	1.	inCinea.de	3,14	1,32	24
	2.	Bücher.de	2,85	2,04	24
	3.	Fröschl.de	2,18	1,41	24
	4.	Delta-Shopping.de	1,77	1,55	24
	5.	DVD-Laden-24.de	1,25	1,25	24
	6.	Avides.com	0,90	1,18	24
	7.	DVD-Galaxis.de	0,76	1,02	24
		<i>Gesamt</i>		1,84	1,65
CDs	1.	Amazon.de	4,25	1,31	24
	2.	Buch.de	3,00	1,62	24
	3.	AlphaMusic.de	2,25	2,05	24
	4.	WOM.de	2,10	1,95	23
	5.	Abella.de	1,85	1,94	21
	6.	abc-CD.de	1,35	1,47	23
	7.	CyberCD.de	0,70	1,04	23
		<i>Gesamt</i>		2,24	1,95
Patronen	1.	Druckerpatronen.de	2,90	1,79	24
	2.	Alternate.de	2,30	1,75	23
	3.	Alphaco.de	1,78	1,52	23
	4.	Planet-Shop.de	1,54	1,46	23
	5.	Express-Toner.de	1,33	1,82	23
	6.	PrintPower.net	1,28	1,65	23
	7.	ACI-Druckerpatronen.de	0,75	1,48	23
		<i>Gesamt</i>		1,70	1,75
Tickets	1.	Getgo.de	3,52	0,97	11
	2.	Eventim.de	2,79	1,90	14
	3.	Kartenhaus.de	1,98	1,50	15
	4.	Karten-Online.de	1,92	1,49	8
	5.	Kartenkaufen.de	1,21	1,21	14
	6.	TicketCenter.de	1,07	0,92	9
	7.	Ticket-Online.de	1,06	1,48	11
		<i>Gesamt</i>		1,96	1,61
<i>Gesamt</i>			1,96	1,77	756

Anmerkung: Die reflektive Skala zur Kaufintention umfasst die drei Indikatoren ki1, ki2 und ki4. Cronbachs Alpha  $\alpha=0,911$ , Mittelwert (M), Standardabweichung (SD), Anzahl der vorliegenden Bewertungen (N).

Tab. E-20: Rangreihe Anzahl der im Versuch getätigten Käufe

<i>Produktart</i>	<i>Rang</i>	<i>Shop</i>	<i>N</i>	<i>Absolute Häufigkeit</i>	<i>Relative Häufigkeit</i>
<i>Bücher</i>	1.	BOL.de	23	14	0,61
	2.	Booxtra.de	24	14	0,58
	3.	Buchkatalog.de	23	10	0,43
	4.	Buch24.de	24	9	0,38
	5.	Karstadt.de	22	8	0,36
	6.	Libri.de	24	7	0,29
	7.	JPC.de	22	6	0,27
	<i>Gesamt</i>			168	68
<i>DVDs</i>	1.	inCinea.de	24	18	0,75
	2.	Bücher.de	24	17	0,71
	3.	Fröschl.de	24	15	0,63
	4.	Delta-Shopping.de	24	7	0,29
	5.	DVD-Laden24.de	24	6	0,25
	6.	Avides.com	24	4	0,17
	7.	DVD-Galaxis.de	24	2	0,08
	<i>Gesamt</i>			168	69
<i>CDs</i>	1.	Amazon.de	24	20	0,83
	2.	WOM.de	23	15	0,65
	3.	AlphaMusic.de	24	12	0,50
	4.	Buch.de	24	10	0,42
	5.	Abella.de	21	7	0,33
	6.	abc-CD.de	23	7	0,30
	7.	CyberCD.de	23	2	0,09
	<i>Gesamt</i>			162	73
<i>Patronen</i>	1.	Druckerpatronen.de	24	17	0,71
	2.	Alphaco.de	23	9	0,39
	3.	Express-Toner.de	23	7	0,30
	3.	Planet-Shop.de	23	7	0,30
	5.	PrintPower.net	23	3	0,13
	6.	Alternate.de	23	2	0,09
	7.	ACI-Druckpatronen.de	23	1	0,04
	<i>Gesamt</i>			162	46
<i>Tickets</i>	1.	Eventim.de	14	7	0,50
	1.	Karten-Online.de	8	4	0,50
	3.	Getgo.de	11	5	0,45
	4.	Kartenhaus.de	15	6	0,40
	5.	Ticket-Online.de	11	4	0,36
	6.	TicketCenter.de	9	1	0,11
	7.	Kartenkaufen.de	14	1	0,07
	<i>Gesamt</i>			82	28
<i>Gesamtanzahl Online-Käufe im Versuch</i>				291	
<i>Anzahl Wahl des Gutscheins</i>				87	

Anmerkung: Anzahl der vorliegenden Bewertungen (N).

Tab. E-21.1: Pearson-Korrelationen zwischen den formativen Usability Indikatoren

	uf1	uf4	uf5	uf6	uf7	uf8	uf9	uf10	uf11	uf12	uf13	uf14	uf15	uf16	uf18	uf19	uf20	uf21
uf4	0,38	1																
uf5	0,30	0,36	1															
uf6	0,48	0,43	0,40	1														
uf7	0,43	0,45	0,35	0,60	1													
uf8	0,35	0,27	0,23	0,44	0,44	1												
uf9	0,37	0,62	0,30	0,38	0,39	0,31	1											
uf10	0,52	0,47	0,39	0,53	0,45	0,36	0,40	1										
uf11	0,31	0,31	0,43	0,49	0,43	0,27	0,24	0,43	1									
uf12	0,42	0,40	0,37	0,62	0,45	0,30	0,30	0,49	0,45	1								
uf13	0,39	0,30	0,38	0,47	0,34	0,27	0,26	0,43	0,38	0,47	1							
uf14	0,46	0,40	0,42	0,67	0,52	0,36	0,35	0,57	0,52	0,70	0,51	1						
uf15	0,42	0,34	0,36	0,54	0,41	0,33	0,30	0,50	0,36	0,56	0,42	0,56	1					
uf16	0,20	0,21	0,23	0,21	0,22	0,12	0,13	0,17	0,24	0,27	0,25	0,26	0,13	1				
uf18	0,20	0,23	0,21	0,28	0,24	0,18	0,18	0,27	0,24	0,30	0,22	0,34	0,31	0,24	1			
uf19	0,25	0,25	0,26	0,40	0,31	0,20	0,12	0,33	0,37	0,52	0,33	0,48	0,40	0,35	0,34	1		
uf20	0,48	0,49	0,35	0,54	0,46	0,36	0,44	0,54	0,41	0,49	0,38	0,57	0,48	0,19	0,24	0,31	1	
uf21	0,25	0,19	0,28	0,33	0,28	0,21	0,15	0,30	0,31	0,31	0,32	0,37	0,29	0,26	0,41	0,29	0,30	1
uf22	0,39	0,34	0,36	0,56	0,44	0,33	0,30	0,47	0,41	0,58	0,45	0,58	0,52	0,22	0,24	0,40	0,46	0,35
uf23	0,28	0,25	0,28	0,38	0,32	0,24	0,21	0,38	0,31	0,40	0,27	0,41	0,36	0,09	0,19	0,25	0,30	0,17
uf24	0,36	0,33	0,34	0,53	0,42	0,28	0,28	0,40	0,29	0,54	0,50	0,51	0,51	0,12	0,24	0,35	0,41	0,25
uf25	0,37	0,29	0,32	0,51	0,45	0,31	0,29	0,43	0,34	0,47	0,44	0,51	0,48	0,11	0,20	0,30	0,45	0,27
uf26	0,37	0,33	0,35	0,52	0,40	0,26	0,28	0,40	0,31	0,50	0,48	0,49	0,49	0,14	0,20	0,28	0,40	0,25
uf27	0,33	0,28	0,28	0,47	0,33	0,22	0,27	0,36	0,27	0,48	0,42	0,45	0,53	0,15	0,24	0,28	0,35	0,27
uf28	0,36	0,28	0,28	0,53	0,38	0,28	0,24	0,39	0,33	0,64	0,37	0,60	0,51	0,18	0,27	0,36	0,40	0,25
uf29	0,40	0,28	0,25	0,44	0,35	0,24	0,27	0,39	0,31	0,61	0,36	0,50	0,42	0,16	0,21	0,29	0,38	0,25
uf31	0,26	0,20	0,25	0,34	0,26	0,14	0,16	0,31	0,25	0,41	0,27	0,40	0,54	0,05	0,25	0,27	0,32	0,26
uf32	0,34	0,31	0,32	0,43	0,36	0,28	0,19	0,38	0,32	0,44	0,34	0,45	0,46	0,18	0,23	0,38	0,38	0,25
uf33	0,20	0,11	0,08	0,24	0,19	0,18	0,07	0,22	0,11	0,34	0,17	0,27	0,26	-0,04	0,06	0,18	0,17	0,11
uf34	0,28	0,26	0,27	0,31	0,32	0,29	0,22	0,35	0,22	0,27	0,21	0,31	0,26	0,16	0,21	0,18	0,25	0,29

Tab. E-21.2: Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung)

	uf1	uf4	uf5	uf6	uf7	uf8	uf9	uf10	uf11	uf12	uf13	uf14	uf15	uf16	uf18	uf19	uf20	uf21
uf35	0,23	0,24	0,23	0,22	0,18	0,23	0,20	0,29	0,17	0,21	0,15	0,22	0,27	0,15	0,12	0,13	0,21	0,12
uf36	0,33	0,26	0,21	0,23	0,24	0,22	0,25	0,33	0,23	0,33	0,18	0,31	0,40	0,10	0,18	0,23	0,31	0,17
uf38	0,25	0,28	0,20	0,27	0,25	0,18	0,26	0,30	0,22	0,30	0,20	0,24	0,34	0,16	0,19	0,16	0,29	0,15
uf39	0,21	0,17	0,17	0,16	0,15	0,13	0,09	0,19	0,17	0,13	0,12	0,15	0,22	0,06	0,13	0,16	0,15	0,17
uf40	0,26	0,21	0,13	0,25	0,26	0,23	0,19	0,29	0,22	0,25	0,16	0,28	0,36	0,07	0,17	0,18	0,24	0,19
uf41	0,25	0,19	0,18	0,26	0,21	0,20	0,16	0,28	0,22	0,29	0,16	0,29	0,28	0,11	0,19	0,18	0,23	0,21
uf42	0,35	0,42	0,31	0,38	0,32	0,25	0,39	0,39	0,26	0,38	0,27	0,40	0,36	0,11	0,16	0,26	0,38	0,22
uf43	0,22	0,22	0,19	0,30	0,25	0,18	0,15	0,28	0,21	0,29	0,16	0,26	0,25	0,06	0,12	0,16	0,27	0,10
uf44	0,34	0,37	0,29	0,37	0,34	0,23	0,31	0,42	0,25	0,32	0,32	0,36	0,29	0,17	0,14	0,20	0,38	0,21
uf45	0,40	0,65	0,26	0,35	0,33	0,25	0,50	0,43	0,22	0,32	0,27	0,34	0,30	0,14	0,13	0,20	0,47	0,15
uf46	0,39	0,48	0,32	0,43	0,43	0,28	0,44	0,47	0,30	0,39	0,34	0,43	0,42	0,17	0,21	0,23	0,46	0,23
uf47	0,23	0,21	0,17	0,27	0,21	0,20	0,20	0,27	0,22	0,21	0,19	0,27	0,23	0,08	0,10	0,13	0,19	0,15
uf48	0,17	0,18	0,14	0,24	0,22	0,15	0,20	0,20	0,26	0,20	0,17	0,23	0,18	0,12	0,19	0,14	0,20	0,25
uf49	0,21	0,21	0,21	0,31	0,24	0,14	0,21	0,31	0,22	0,20	0,29	0,28	0,22	0,12	0,19	0,21	0,31	0,16
uf50	0,42	0,45	0,33	0,46	0,40	0,27	0,42	0,51	0,34	0,49	0,37	0,47	0,47	0,19	0,19	0,30	0,44	0,19
uf54	0,10	0,08	0,18	0,13	0,12	0,09	0,06	0,13	0,22	0,15	0,13	0,21	0,09	0,39	0,15	0,29	0,13	0,17
uf55	0,37	0,29	0,26	0,51	0,39	0,36	0,26	0,40	0,37	0,61	0,38	0,53	0,51	0,15	0,26	0,43	0,41	0,21
uf56	0,17	0,15	0,17	0,26	0,21	0,12	0,23	0,20	0,19	0,17	0,19	0,20	0,13	0,11	0,11	0,13	0,30	0,18
uf57	0,25	0,19	0,21	0,31	0,23	0,08	0,20	0,29	0,26	0,30	0,20	0,28	0,51	0,05	0,16	0,21	0,30	0,15
uf58	0,12	0,16	0,21	0,27	0,18	0,15	0,16	0,19	0,16	0,17	0,25	0,21	0,16	0,09	0,17	0,15	0,25	0,13

Tab. E-21.3: Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung)

	uf22	uf23	uf24	uf25	uf26	uf27	uf28	uf29	uf31	uf32	uf33	uf34	uf35	uf36	uf38	uf39	uf40	uf41
uf22	1																	
uf23	0,34	1																
uf24	0,52	0,33	1															
uf25	0,48	0,29	0,58	1														
uf26	0,53	0,32	0,75	0,66	1													
uf27	0,48	0,30	0,68	0,57	0,72	1												
uf28	0,54	0,40	0,54	0,42	0,46	0,49	1											
uf29	0,51	0,37	0,44	0,37	0,41	0,41	0,57	1										
uf31	0,42	0,33	0,42	0,34	0,37	0,44	0,46	0,29	1									
uf32	0,47	0,30	0,39	0,38	0,37	0,37	0,43	0,34	0,40	1								
uf33	0,23	0,20	0,28	0,22	0,25	0,23	0,25	0,30	0,24	0,22	1							
uf34	0,28	0,23	0,24	0,24	0,26	0,25	0,19	0,24	0,13	0,25	0,13	1						
uf35	0,21	0,24	0,19	0,19	0,22	0,16	0,20	0,16	0,20	0,17	0,12	0,22	1					
uf36	0,30	0,26	0,26	0,23	0,24	0,21	0,30	0,26	0,27	0,27	0,21	0,22	0,37	1				
uf38	0,26	0,18	0,27	0,21	0,24	0,21	0,19	0,22	0,21	0,29	0,07	0,25	0,22	0,46	1			
uf39	0,17	0,12	0,14	0,22	0,18	0,11	0,10	0,08	0,10	0,19	0,04	0,24	0,13	0,35	0,18	1		
uf40	0,30	0,21	0,22	0,24	0,20	0,20	0,26	0,23	0,23	0,27	0,13	0,25	0,30	0,54	0,35	0,22	1	
uf41	0,29	0,21	0,17	0,20	0,17	0,14	0,24	0,19	0,22	0,28	0,20	0,22	0,21	0,44	0,32	0,21	0,41	1
uf42	0,33	0,26	0,28	0,28	0,30	0,25	0,32	0,32	0,23	0,29	0,15	0,21	0,19	0,21	0,21	0,10	0,18	0,11
uf43	0,24	0,20	0,17	0,16	0,20	0,15	0,24	0,22	0,23	0,22	0,10	0,14	0,17	0,19	0,15	0,07	0,12	0,11
uf44	0,31	0,13	0,27	0,33	0,31	0,22	0,21	0,27	0,19	0,25	0,11	0,24	0,15	0,20	0,23	0,19	0,17	0,15
uf45	0,24	0,17	0,25	0,26	0,28	0,19	0,22	0,27	0,16	0,25	0,12	0,20	0,17	0,20	0,22	0,13	0,11	0,15
uf46	0,36	0,25	0,38	0,37	0,37	0,28	0,31	0,30	0,27	0,30	0,18	0,26	0,23	0,30	0,35	0,18	0,29	0,22
uf47	0,21	0,19	0,19	0,23	0,22	0,17	0,21	0,15	0,14	0,14	0,14	0,15	0,15	0,25	0,31	0,13	0,17	0,18
uf48	0,21	0,03	0,11	0,12	0,13	0,13	0,11	0,11	0,11	0,15	0,03	0,23	0,07	0,15	0,13	0,18	0,21	0,22
uf49	0,19	0,07	0,21	0,25	0,21	0,23	0,18	0,17	0,15	0,20	0,03	0,19	0,12	0,15	0,18	0,16	0,16	0,11
uf50	0,39	0,26	0,34	0,38	0,36	0,28	0,32	0,33	0,32	0,34	0,22	0,27	0,26	0,36	0,33	0,20	0,33	0,28
uf54	0,15	0,09	0,12	0,09	0,08	0,07	0,11	0,08	0,05	0,15	-0,06	0,09	0,07	0,03	0,09	0,11	0,05	0,07
uf55	0,51	0,33	0,46	0,45	0,46	0,40	0,55	0,47	0,36	0,44	0,31	0,22	0,20	0,28	0,27	0,17	0,20	0,28
uf56	0,15	0,09	0,15	0,18	0,17	0,16	0,12	0,13	0,12	0,15	0,02	0,09	0,05	0,11	0,15	0,08	0,05	0,10
uf57	0,32	0,22	0,26	0,31	0,31	0,35	0,30	0,23	0,50	0,28	0,16	0,15	0,20	0,28	0,25	0,16	0,31	0,26
uf58	0,16	0,04	0,19	0,17	0,22	0,18	0,14	0,16	0,11	0,16	0,06	0,11	0,13	0,09	0,16	0,07	0,09	0,09

Tab. E-21.4: Pearson-Korrelationen zwischen den formativen Usability Indikatoren (Fortsetzung)

	uf42	uf43	uf44	uf45	uf46	uf47	uf48c	uf49	uf50	uf54	uf55	uf56	uf57
uf42	1												
uf43	0,35	1											
uf44	0,36	0,23	1										
uf45	0,46	0,22	0,39	1									
uf46	0,44	0,20	0,42	0,46	1								
uf47	0,25	0,16	0,16	0,16	0,36	1							
uf48	0,16	0,09	0,18	0,15	0,23	0,16	1						
uf49	0,19	0,14	0,24	0,24	0,32	0,07	0,11	1					
uf50	0,41	0,24	0,36	0,39	0,60	0,33	0,33	0,23	1				
uf54	0,10	0,07	0,15	0,05	0,07	0,04	0,21	0,09	0,13	1			
uf55	0,31	0,25	0,26	0,26	0,28	0,22	0,12	0,19	0,36	0,15	1		
uf56	0,18	0,12	0,20	0,17	0,29	0,18	0,07	0,35	0,17	0,06	0,13	1	
uf57	0,24	0,17	0,22	0,14	0,30	0,19	0,15	0,18	0,30	0,08	0,27	0,18	1
uf58	0,15	0,14	0,22	0,20	0,25	0,09	0,04	0,67	0,19	0,06	0,17	0,36	0,11



Tab. E-22: Toleranz- und VIF-Werte der formativen ufosV2f-Indikatoren

Item-Code	Toleranz	VIF
uf1i	0,58	1,73
uf5ci	0,65	1,54
uf7i	0,53	1,88
uf8i	0,67	1,48
uf9i	0,58	1,73
uf10i	0,46	2,19
uf11ci	0,58	1,71
uf12i	0,30	3,28
uf13ci	0,56	1,78
uf14i	0,33	3,04
uf15i	0,39	2,57
uf16ci	0,67	1,49
uf18ci	0,72	1,40
uf19ci	0,56	1,77
uf20i	0,47	2,12
uf21ci	0,65	1,55
uf22i	0,46	2,17
uf23i	0,68	1,47
uf24i	0,37	2,68
uf25i	0,49	2,05
uf27i	0,41	2,45
uf28i	0,42	2,40
uf29i	0,51	1,98
uf31i	0,52	1,91
uf32i	0,61	1,64
uf33i	0,76	1,31
uf34i	0,72	1,39
uf35i	0,77	1,31
uf36i	0,49	2,06
uf38i	0,63	1,59
uf39ci	0,77	1,30
uf40i	0,58	1,71
uf41i	0,68	1,47
uf42i	0,60	1,67
uf43i	0,79	1,27
uf44i	0,67	1,50
uf45i	0,55	1,81
uf46i	0,46	2,17
uf47i	0,74	1,34
uf48ci	0,75	1,33
uf49ci	0,73	1,38
uf50i	0,46	2,19
uf54ci	0,75	1,33
uf55i	0,48	2,10
uf56ci	0,77	1,29
uf57ci	0,58	1,72

Anmerkung: Variance Inflation Factor (VIF).

Tab. E-23: Rotierte Komponentenmatrix zur Erstellung der Multikollinearitätsindizes zur formativen Usability-Skala ufosV2f

Komponente	1	2	3	4	5	6	7	8	9	$h^2$	MSA Wert
uf12	<b>0,76</b>	0,28	0,13	0,10	0,12	0,19	0,05	0,05	-0,02	0,731	0,959
uf28	<b>0,73</b>	0,14	0,10	0,05	0,20	0,07	0,10	0,02	0,07	0,621	0,960
uf55	<b>0,71</b>	0,19	0,16	0,05	0,06	0,10	0,05	0,06	0,02	0,579	0,966
uf29	<b>0,68</b>	0,22	0,09	0,04	0,01	0,01	0,07	0,03	0,03	0,527	0,957
uf14	<b>0,65</b>	<b>0,36</b>	0,10	0,24	0,09	0,19	0,09	0,11	0,03	0,688	0,972
uf22	<b>0,60</b>	0,24	0,15	0,21	0,17	0,13	0,21	0,01	0,04	0,571	0,980
uf24	<b>0,56</b>	0,22	0,08	0,04	0,17	0,04	<b>0,55</b>	0,06	0,07	0,709	0,938
uf33	<b>0,54</b>	0,05	0,16	0,00	-0,03	-0,29	-0,03	-0,07	-0,17	0,441*	0,915
uf19	<b>0,51</b>	0,05	0,09	0,16	0,10	<b>0,46</b>	-0,05	0,15	-0,02	0,548	0,942
uf15	<b>0,49</b>	0,28	0,26	0,15	<b>0,44</b>	0,04	0,20	0,03	0,10	0,654	0,968
uf32	<b>0,47</b>	0,16	0,20	0,17	0,21	0,13	0,08	0,10	0,10	0,406*	0,974
uf23	<b>0,47</b>	0,16	0,14	0,10	0,13	0,02	-0,03	-0,06	<b>0,37</b>	0,429*	0,959
uf13	<b>0,42</b>	0,29	0,01	0,15	0,04	0,21	<b>0,37</b>	0,18	0,02	0,501	0,964
uf11	<b>0,37</b>	0,26	0,10	0,33	0,10	0,29	-0,01	0,11	0,05	0,429*	0,964
uf45	0,15	<b>0,72</b>	0,03	0,01	-0,02	0,02	-0,03	0,10	0,07	0,553	0,932
uf9	0,10	<b>0,67</b>	0,11	0,06	-0,01	-0,01	0,13	0,08	0,06	0,507	0,942
uf46	0,18	<b>0,66</b>	0,26	0,04	0,14	0,05	0,13	0,18	-0,13	0,622	0,952
uf42	0,25	<b>0,63</b>	-0,01	0,04	0,22	0,05	-0,16	0,02	0,09	0,550	0,955
uf50	0,26	<b>0,61</b>	<b>0,32</b>	0,06	0,18	0,13	0,06	-0,04	-0,17	0,624	0,954
uf44	0,11	<b>0,57</b>	0,07	0,13	0,08	0,12	0,11	0,12	0,01	0,403*	0,967
uf10	<b>0,38</b>	<b>0,53</b>	0,20	0,23	0,08	0,05	0,08	0,11	0,17	0,569	0,977
uf20	<b>0,39</b>	<b>0,52</b>	0,11	0,17	0,09	0,06	0,07	0,26	0,10	0,554	0,972
uf1	<b>0,35</b>	<b>0,49</b>	0,21	0,15	-0,04	0,03	0,12	0,06	0,13	0,462*	0,978
uf7	<b>0,40</b>	<b>0,44</b>	0,11	<b>0,30</b>	-0,06	0,04	0,14	0,10	0,10	0,502	0,968
uf47	0,17	<b>0,36</b>	<b>0,32</b>	-0,06	0,06	0,05	0,00	0,01	-0,27	0,342*	0,910
uf5	0,18	<b>0,36</b>	0,03	0,26	0,16	0,27	0,17	0,06	<b>0,31</b>	0,460*	0,962
uf8	<b>0,34</b>	<b>0,34</b>	0,17	0,30	-0,28	-0,08	0,07	-0,02	0,22	0,486*	0,951
uf36	0,22	0,14	<b>0,76</b>	0,02	0,10	0,01	-0,02	0,05	0,15	0,683	0,906
uf40	0,15	0,09	<b>0,68</b>	0,17	0,13	-0,03	0,06	-0,02	0,06	0,552	0,919
uf41	0,26	0,04	<b>0,61</b>	0,22	0,04	-0,01	-0,12	0,03	-0,08	0,512	0,940
uf38	0,13	0,24	<b>0,60</b>	-0,09	0,08	0,16	0,09	0,15	-0,02	0,505	0,916
uf39	-0,05	0,08	<b>0,43</b>	<b>0,31</b>	0,04	0,05	0,17	0,08	0,07	0,332*	0,875
uf21	0,23	0,08	0,05	<b>0,66</b>	0,10	0,16	0,05	0,15	-0,05	0,555	0,924
uf48	-0,03	0,29	0,17	<b>0,53</b>	0,13	0,14	0,01	-0,20	<b>-0,43</b>	0,661	0,885
uf18	0,25	0,00	0,08	<b>0,51</b>	0,16	0,19	-0,06	0,19	0,02	0,435*	0,933
uf34	0,11	0,25	0,25	<b>0,51</b>	-0,11	-0,02	0,18	-0,04	0,21	0,485*	0,945
uf57	0,16	0,16	0,27	0,06	<b>0,70</b>	0,00	0,10	0,09	-0,02	0,644	0,912
uf31	<b>0,40</b>	0,08	0,12	0,13	<b>0,68</b>	-0,05	0,10	0,04	0,09	0,677	0,936
uf16	0,14	0,10	0,08	0,08	-0,12	<b>0,75</b>	0,04	0,08	0,05	0,627	0,841
uf54	0,04	0,06	0,02	0,12	0,04	<b>0,74</b>	0,01	-0,06	-0,02	0,576	0,802
uf27	<b>0,47</b>	0,15	0,03	0,10	<b>0,32</b>	0,00	<b>0,58</b>	0,08	0,06	0,704	0,929
uf25	<b>0,44</b>	0,29	0,10	0,14	0,15	-0,01	<b>0,52</b>	0,11	0,04	0,612	0,965
uf43	0,25	<b>0,33</b>	-0,01	0,07	0,28	-0,01	<b>-0,41</b>	0,06	0,28	0,508	0,938
uf56	0,07	0,21	0,05	0,06	0,05	0,01	-0,03	<b>0,76</b>	-0,09	0,642	0,884
uf49	0,06	0,21	0,11	0,13	0,06	0,06	0,14	<b>0,69</b>	0,07	0,589	0,911
uf35	0,07	0,20	<b>0,42</b>	-0,01	0,10	0,11	0,03	-0,09	<b>0,53</b>	0,532	0,940
$\Sigma=55,0$	14,43	11,27	6,64	4,84	4,18	4,17	3,57	3,28	2,61		
$\lambda_{j=10}=1,00$	13,61	2,23	1,97	1,81	1,34	1,21	1,07	1,04	1,03		

Anmerkung: (\*) Kommunalität liegt unterhalb des kritischen Wertes  $h^2 < 0,5$ . Kommunalität ( $h_i^2$ ), Eigenwert ( $\lambda_i$ ), Measure of Sampling Adequacy (MSA), durch Faktor aufgeklärte Varianz in % ( $\Sigma$ ).

Tab. E-24: Korrelationen der latenten Variablen im PLS-Modell mit reflektiver Usability-Skala ufosV2r mit Quadratwurzel des AVE-Maßes

	Kauf- intention	Usability	Ästhetik	Vertrauen	W. Produkt- anzahl	Wahrg. Preis	Service	Bekannt- heit	Produkt Drucker	Ästhetik x Drucker
Kaufintention	<b>0,922</b>									
Usability	0,618	<b>0,835</b>								
Ästhetik	0,541	0,627	<b>0,958</b>							
Vertrauen	0,469	0,562	0,425	<b>0,892</b>						
W. Produktanz.	0,411	0,393	0,385	0,374	<b>0,931</b>					
W. Preis	0,344	0,229	0,244	0,274	0,172	<b>1</b>				
Service	0,360	0,332	0,229	0,350	0,219	0,328	-			
Bekanntheit	0,350	0,244	0,242	0,299	0,372	0,130	0,181	<b>1</b>		
Produkt Drucker	-0,077	-0,093	-0,111	-0,108	-0,128	-0,108	-0,103	-0,172	<b>1</b>	
Ästhetik x Drucker	0,021	0,001	0,018	0,029	-0,022	0,026	-0,022	0,042	-0,137	<b>0,929</b>

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-25: Korrelationen der latenten Variablen im PLS-Modell mit formativer Usability-Skala ufosV2f mit Quadratwurzel des AVE-Maßes

	Kauf- intention	Usability (formativ)	Ästhetik	Vertrauen	W. Produkt- anzahl	Wahrg. Preis	Service	Bekannt- heit	Produkt Drucker	Ästhetik x Drucker
Kaufintention	<b>0,922</b>									
ufosV2f	0,590	-								
Ästhetik	0,541	0,627	<b>0,958</b>							
Vertrauen	0,469	0,685	0,425	<b>0,892</b>						
W. Produktanz.	0,411	0,429	0,385	0,374	<b>0,931</b>					
W. Preis	0,344	0,269	0,244	0,274	0,172	<b>1</b>				
Service	0,360	0,374	0,229	0,350	0,220	0,328	-			
Bekanntheit	0,350	0,235	0,242	0,298	0,372	0,130	0,181	<b>1</b>		
Produkt Drucker	-0,078	-0,124	-0,118	-0,111	-0,128	-0,108	-0,103	-0,172	<b>1</b>	
Ästhetik xDrucker	0,021	0,034	0,018	0,029	-0,022	0,025	-0,022	0,041	-0,138	<b>0,929</b>

Anmerkung: In der Diagonalen findet sich fettgedruckt die Quadratwurzel der Average Variance Extracted (AVE).

Tab. E-26: Kreuzladungen der Indikatoren im PLS-Modells bei Verwendung der reflektiven Usability-Skala ufosV2r

	Kauf- intention	Usability (reflektiv)	Ästhetik	Vertrauen	W. Produkt- anzahl	Service	Wahrg. Preis	Bekannt- heit	Produkt Drucker	Ästhetik x Drucker
ki1	<b>0,933</b>	0,586	0,522	0,456	0,412	0,355	0,310	0,354	-0,116	0,031
ki2	<b>0,924</b>	0,536	0,472	0,413	0,378	0,338	0,286	0,304	-0,084	0,020
ki4	<b>0,908</b>	0,586	0,499	0,426	0,347	0,302	0,353	0,307	-0,014	0,007
ur1	0,589	<b>0,883</b>	0,549	0,531	0,341	0,315	0,226	0,221	-0,103	0,001
ur2	0,453	<b>0,754</b>	0,435	0,351	0,246	0,205	0,137	0,163	-0,048	0,004
ur3	0,476	<b>0,800</b>	0,551	0,426	0,297	0,240	0,202	0,182	-0,125	0,032
ur4	0,428	<b>0,822</b>	0,456	0,437	0,233	0,258	0,179	0,131	-0,104	-0,005
ur6	0,462	<b>0,841</b>	0,466	0,429	0,245	0,245	0,153	0,170	-0,083	-0,024
ur7	0,600	<b>0,910</b>	0,607	0,529	0,369	0,318	0,237	0,212	-0,056	0,008
ur8	0,534	<b>0,791</b>	0,564	0,514	0,466	0,303	0,200	0,312	-0,071	-0,006
ur9	0,549	<b>0,868</b>	0,527	0,496	0,379	0,308	0,179	0,211	-0,035	-0,007
ae1	0,552	0,631	<b>0,963</b>	0,417	0,400	0,213	0,238	0,236	-0,093	0,030
ae2	0,480	0,568	<b>0,953</b>	0,397	0,334	0,226	0,229	0,228	-0,136	0,003
t1	0,479	0,552	0,425	<b>0,900</b>	0,390	0,327	0,238	0,301	-0,119	0,021
t2	0,384	0,474	0,336	<b>0,901</b>	0,339	0,295	0,230	0,264	-0,086	0,008
t3	0,409	0,481	0,386	<b>0,846</b>	0,282	0,299	0,266	0,223	-0,078	0,035
t4	0,390	0,488	0,362	<b>0,920</b>	0,314	0,326	0,245	0,271	-0,108	0,040
wal	0,401	0,385	0,408	0,370	<b>0,939</b>	0,214	0,171	0,396	-0,130	-0,009
wa2	0,362	0,344	0,303	0,325	<b>0,922</b>	0,193	0,149	0,291	-0,107	-0,034
sv1	0,301	0,310	0,183	0,315	0,154	0,836	0,146	0,138	-0,027	-0,016
sv5c	0,257	0,196	0,174	0,222	0,194	0,714	0,398	0,147	-0,148	-0,017
subpl	0,344	0,229	0,244	0,274	0,172	0,328	1,000	0,130	-0,108	0,026
shop ab	0,350	0,244	0,242	0,299	0,372	0,181	0,130	1,000	-0,172	0,042
druck	-0,077	-0,093	-0,118	-0,111	-0,128	-0,103	-0,108	-0,172	1,000	-0,137
ae1xdruck	0,024	-0,001	0,018	0,033	-0,020	-0,023	0,032	0,048	-0,124	<b>0,995</b>
ae2xdruck	0,005	0,008	0,011	0,002	-0,030	-0,010	-0,008	0,002	-0,182	<b>0,858</b>

Anmerkung: Fett gedruckt dargestellt sind die Ladungen der Indikatoren auf ihren jeweiligen latenten Variablen.

Tab. E-27: Kreuzladungen der Indikaoren im PLS-Modells bei Verwendung der formativen Usability-Skala ufosV2f

	Kauf- intention	Usability (formativ)	Ästhetik	Vertrauen	W. Produkt- anzahl	Service	W. Preis	Bekannt- heit	Produkt Drucker	Ästhetik x Drucker
ki1	<b>0,934</b>	0,564	0,522	0,455	0,412	0,355	0,310	0,354	-0,116	0,031
ki2	<b>0,925</b>	0,517	0,472	0,413	0,378	0,338	0,286	0,304	-0,084	0,020
ki4	<b>0,907</b>	0,547	0,499	0,427	0,347	0,302	0,353	0,307	-0,014	0,007
uf_in1	0,489	0,804	0,612	0,403	0,421	0,190	0,144	0,234	-0,144	0,048
uf_in2	0,219	0,317	0,119	0,272	0,093	0,201	0,112	0,055	0,067	-0,059
uf_in3	0,193	0,368	0,161	0,349	0,135	0,352	0,149	0,127	-0,054	0,039
uf_in4	0,013	0,131	0,091	0,140	0,023	0,019	0,196	-0,025	-0,150	0,075
uf_in5	0,163	0,268	0,178	0,165	0,117	0,060	0,126	0,021	0,005	-0,013
uf_in6	0,037	0,147	0,206	0,019	-0,155	0,016	0,027	-0,169	0,085	-0,048
uf_in7	0,069	0,063	0,000	0,056	0,018	0,087	0,036	0,042	-0,135	0,064
uf_in8	0,023	0,056	0,089	-0,014	-0,042	0,034	-0,037	0,053	0,092	-0,069
uf_in9	-0,015	0,014	0,065	-0,033	0,037	-0,051	-0,033	0,020	-0,059	0,044
ae1	0,552	0,685	<b>0,963</b>	0,417	0,400	0,213	0,238	0,236	-0,093	0,030
ae2	0,480	0,625	<b>0,953</b>	0,397	0,334	0,226	0,229	0,228	-0,136	0,003
t1	0,479	0,584	0,425	<b>0,898</b>	0,390	0,327	0,238	0,301	-0,119	0,021
t2	0,384	0,486	0,336	<b>0,899</b>	0,339	0,295	0,230	0,264	-0,086	0,008
t3	0,409	0,546	0,386	<b>0,849</b>	0,282	0,299	0,266	0,223	-0,078	0,035
t4	0,390	0,537	0,362	<b>0,921</b>	0,314	0,326	0,245	0,271	-0,108	0,040
wa1	0,401	0,427	0,408	0,369	<b>0,939</b>	0,214	0,171	0,396	-0,130	-0,009
wa2	0,362	0,368	0,303	0,324	<b>0,922</b>	0,193	0,149	0,291	-0,107	-0,034
sv1	0,301	0,351	0,183	0,315	0,154	0,836	0,146	0,138	-0,027	-0,016
sv5c	0,257	0,217	0,174	0,222	0,194	0,714	0,398	0,147	-0,148	-0,017
subp1	0,344	0,269	0,244	0,274	0,172	0,328	1,000	0,130	-0,108	0,026
shop_ab	0,350	0,235	0,242	0,298	0,372	0,181	0,130	1,000	-0,172	0,041
druck	-0,078	-0,124	-0,118	-0,111	-0,128	-0,103	-0,108	-0,172	1,000	-0,138
ae1xdruck	0,024	0,031	0,018	0,034	-0,020	-0,023	0,032	0,048	-0,124	<b>0,995</b>
ae2xdruck	0,005	0,042	0,011	0,003	-0,030	-0,010	-0,008	0,002	-0,182	<b>0,858</b>

Anmerkung: Fett gedruckt dargestellt sind die Ladungen der Indikatoren auf ihren jeweiligen latenten Variablen.

Tab. E-28: Ausprägungen der Effektgrößen der verschiedenen Prädiktoren in der PLS-Analyse zum konzeptuellen Erfolgsmodell

Latente exogene Variable	Latente endogene Variable	Modell mit ufosV2r (reflektive Usability)		Modell mit ufosV2f (formative Usability)	
		$f^2$	Beurteilung	$f^2$	Beurteilung
Usability	Vertrauen	0,293	mittel	0,381	hoch
	Kaufintention	0,109	gering	0,038	gering
Ästhetik	Usability	0,647	hoch	0,883	hoch
	Kaufintention	0,041	gering	0,062	gering
Vertrauen	Kaufintention	0,006	keiner	0,013	keiner
Wahrge. Produktanzahl	Vertrauen	0,023	gering	0,010	keiner
	Kaufintention	0,027	gering	0,027	gering
Wahrge. Preis	Kaufintention	0,035	gering	0,029	gering
Service	Kaufintention	0,018	keiner	0,016	keiner
Bekanntheit	Vertrauen	0,022	gering	0,027	gering
Ästhetik x Drucker	Kaufintention	0,002	keiner	0,002	keiner

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-29: Korrelationen der MLR-Prädiktoren

	Alter	Ein- komme	Online- Einkauf	Erheb.- dauer	Prdoukt- Involve.	ufosV2r Shop 1	ufosV2r Shop 2	ufosV2f Shop 1	ufosV2f Shop 2	Kaufint Shop 1
Alter	1									
Einkommen (Eink)	0,582 ***	1								
Online- Einkäufe	-0,027	0,098	1							
Erhebungs- dauer	0,344 ***	0,114 *	-0,158 **	1						
Produkt Involvement	0,152 **	0,080	0,020	0,085	1					
ufosV2r Shop 1	0,150 **	0,079	0,057	0,022	0,161 **	1				
ufosV2r Shop 2	0,049	0,048	0,073	-0,047	0,100	-0,020	1			
ufosV2f Shop 1	0,111 *	0,047	0,037	0,028	0,145 **	0,856 ***	-0,033	1		
ufosV2f Shop 2	0,035	0,066	0,102	-0,028	0,080	0,017	0,881 ***	0,074	1	
Kauf- intention S1	0,209 ***	0,130 *	-0,018	0,128 *	0,153 **	0,544 ***	-0,116 *	0,495 ***	-0,07	1
Kauf- intention S2	0,097	0,120 *	0,093	0,041	0,079	-0,178 ***	0,665 ***	-0,185 ***	0,617 ***	-0,0252

Anmerkung: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$



Tab. E-30: Toleranz und Varance-Inflation-Factors (VIF) der MLR-Prädiktoren

<i>Prädiktor</i>	<i>Schritt 1</i>		<i>Schritt 2a</i>		<i>Schritt 2b</i>		<i>Schritt 2c</i>	
	<i>Tol.</i>	<i>VIF</i>	<i>Tol.</i>	<i>VIF</i>	<i>Tol.</i>	<i>VIF</i>	<i>Tol.</i>	<i>VIF</i>
Alter	0,64	1,56	0,63	1,58	0,64	1,57	0,63	1,60
Einkommen	0,63	1,58	0,63	1,58	0,63	1,59	0,63	1,59
Geschlecht	0,96	1,04	0,96	1,04	0,96	1,04	0,95	1,05
Online-Einkäufe	0,98	1,03	0,97	1,03	0,97	1,04	0,97	1,03
Produkt Involvement	0,98	1,02	0,95	1,06	0,95	1,05	0,95	1,05
ufosV2r Shop 1			0,95	1,06				
ufosV2r Shop 2			0,98	1,02				
ufosV2f Shop 1					0,98	1,02		
ufosV2f Shop 2					0,96	1,04		
Kaufintention Shop 1							0,97	1,03
Kaufintention Shop 2							0,92	1,08
Maximaler Konditionsindex	12,31		15,09		12,54		13,81	

Anmerkung: Toleranz (Tol.), Variance Inflation Factor (VIF).

Tab. E-31.1: Teilnehmer mit standardisierten Residuen  $zRedsid > 1$  in der sequentiellen MLR

<i>Teilnehmer</i>	<i>zRedsid1</i>	<i>zRedsid2</i>	<i>zRedsid3</i>	<i>Teilnehmer</i>	<i>zRedsid1</i>	<i>zRedsid2</i>	<i>zRedsid3</i>
b1_3_1			-1,157	m3_1_2		-1,041	
b1_4_2	-3,425		-1,351	m3_2_1	-1,767		
b2_1_1	-1,466			m3_2_2		1,193	
b2_4_2		1,039		m3_1_2		-1,041	
b2_6_1			1,197	m3_4_2	-3,450		
b2_6_2	-5,013		-1,142	m3_5_1	-1,408		-2,424
b3_1_1			-1,795	m3_6_1_2		3,231	1,167
b3_2_2	-1,110			m3_6_2		-2,214	
b3_4_1		1,208		m3_7_1	-1,255		
b3_7_2	-1,864			m4_3_2		1,175	
b4_1_2	-1,704			m5_3_2		1,170	
b4_2_2	-1,252		-2,353	m5_4_1	-1,734		-1,211
b5_1_2	-1,339			m6_1_1		-2,283	
b5_4_1		-1,473		m6_1_2		1,996	
b5_7_2		-1,083		m7_1_1		1,253	1,099
b6_2_2	-1,371			m7_5_1_1		-1,782	
b6_7_1	1,660			p1_3_2	-1,717		
b7_3_2	-1,939			p1_4_1	1,138	-1,345	
b7_4_2		1,469	1,848	p1_4_2	-1,118		
b7_5_1		-1,933		p1_5_2		-1,027	
b7_6_1		2,224	1,247	p1_7_1	1,466	-2,061	
d1_5_1	-1,037		-1,572	p2_1_2		1,103	1,768
d2_1_2		1,700	1,195	p2_3_1			1,047
d2_6_1		-1,435		p2_3_2		1,740	2,513
d2_6_2		1,192		p2_5_1	1,664		
d2_7_1	-1,191			p2_5_2			-1,334
d3_1_1	-1,006			p2_7_1		-1,250	
d3_2_1	1,182			p2_7_2			-1,416
d3_5_1		-1,084		p3_1_1			1,143
d3_7_2	-1,936			p3_2_1	2,029	-1,038	
d4_2_2		3,471	1,092	p3_5_1		1,899	1,099
d5_2_1		1,016		p3_5_2	1,384		
d5_6_2	-1,189		-1,989	p3_7_2			-1,060
d6_1_2		-1,461		p4_1_2			1,528
d6_3_2		-1,384		p4_3_2	1,345	-1,414	
d6_4_2	-2,906		-2,297	p4_5_2	-1,258		
d6_5_2	-2,816			p4_7_2	-1,037		
d6_7_1	-2,021		-1,486	p5_1_2		1,074	1,265
d7_1_1			-1,170	p5_2_2			-1,072
d7_4_1			-1,947	p5_3_1		3,172	2,315
d7_5_1			-1,633	p5_4_1			2,821
d7_5_2		1,544	1,305	p5_6_1	-1,121		
m1_3_1	1,222	-4,022		p5_7_1	-1,229		-1,358
m1_4_2	-2,091		-1,085	p6_3_2	-1,282		
m2_4_1	-2,250		-1,656	p6_4_1	1,029		
m2_5_1	-1,689		-2,729	p6_4_2	1,377	-1,512	
m2_5_2		-1,762		p6_5_1	-2,026		

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-31.2: Teilnehmer mit standardisierten Residuen  $zRedsid > 1$  in der sequentiellen MLR

p7_1_1	-1,628		-1,564	t3_7_1			1,094
p7_2_1			-1,441	t4_1_1			-1,182
p7_3_1		1,518	1,426	t4_3_1	-1,707		-2,056
p7_5_1	1,150	-1,055		t1_4_1	1,631	-3,869	
p7_5_2	2,991	-1,963		t4_6_1			-1,363
p7_6_1			-2,335	t4_6_2	-1,312		-1,323
t2_1_1	-1,570			t4_6_2			
t2_3_1	1,258			t5_6_1	1,219	-2,613	
t2_3_2	1,250	-1,196		t5_6_1			
t2_6_1		3,453		t6_1_1		5,955	2,311
t3_2_1			1,331	t7_4_1			1,368
t3_2_2	-1,107		-1,522	t7_5_1			-1,460

**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-32: Ausreißer in der MLR

Teilnehmer	$zRedsid1$	$zRedsid2$	$zRedsid3$
b2_4_1	-1,466		
b3_1_1			-1,795
b3_4_1		1,208	
d2_1_2		1,700	1,195
d2_7_1	-1,191		
d3_1_1	-1,006		
m3_1_2		-1,041	
m3_2_1	-1,767		
m3_6_2		-2,214	
m5_4_1	-1,734		-1,211
m6_1_2		1,996	
p3_7_2			-1,060
p7_3_1		1,518	1,426
t2_1_1	-1,570		
t2_3_1	1,258		
t5_6_1	1,219	-2,613	

**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* ( $zRedsid3$ ).

### Schritt 1 der MLR: Hinzufügen der Variablen zur Demographie, Vorerfahrung und Kontrolle

Tab. E-33: Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 1

1. Likelihood-Quotiententest:				
<i>Modell</i>	<i>-2 Log-Likelihood</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
<i>Nur konstanter Term</i>	731,608			
<i>Endgültig</i>	659,804	71,804	22	0,000
2. $\chi^2$ - bzw. Devianz-Statistik:				
	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>	
<i>Pearson</i>	731,104	662	0,025	
<i>Abweichung</i>	659,804	662	0,517	
3. Pseudo-R <sup>2</sup> -Statistiken:				
<i>Cox und Snell</i>	<i>Nagelkerke</i>	<i>McFadden</i>		
0,189	0,214	0,098		

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R<sup>2</sup>).

Tab. E-34: Klassifikationsmatrix – sequentielle MRL Schritt 1

<i>Beobachtet</i>	<i>Vorhersagt</i>			
	<i>Gutschein</i>	<i>Kauf in Shop 1</i>	<i>Kauf in Shop 2</i>	<i>% richtig</i>
<i>Gutschein</i>	40	26	10	52,63
<i>Kauf in Shop 1</i>	15	92	36	64,34
<i>Kauf in Shop 2</i>	18	60	46	37,10
<i>% insgesamt</i>	21,28	51,90	26,82	51,90

Tab. E-35: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 1

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
Konstante	659,804	0,000	0	
Alter	674,871	15,067	2	0,001
Einkommen	674,292	14,487	2	0,001
Geschlecht	662,625	2,821	2	0,244
Anzahl Online-Einkäufe	662,191	2,387	2	0,303
Erhebungsdauer	661,559	1,755	2	0,416
Internetverbindung	661,024	1,219	2	0,544
Produkt Involvement	672,248	12,444	2	0,002
Produktart	693,969	34,164	8	0,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-36: Parameterschätzer – sequentielle MRL Schritt 1

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Kaufen in Shop 2 vs. Gutschein	Konstante	1,835	1,391	1,740	1	0,187			
	Alter	0,070	0,019	13,657	1	0,000	1,072	1,033	1,113
	Einkommen	-0,832	0,236	12,417	1	0,000	0,435	0,274	0,691
	Online-Ein.	0,003	0,002	1,238	1	0,266	1,003	0,998	1,007
	Erhebungsd.	-0,005	0,011	0,170	1	0,680	0,995	0,974	1,017
	Produkt Inv.	-0,360	0,126	8,132	1	0,004	0,698	0,545	0,894
	Internetverb.	-0,337	0,317	1,131	1	0,288	0,714	0,383	1,329
	Geschl.: w	0,132	0,330	0,161	1	0,689	1,142	0,598	2,181
	Geschl.: m	0			0				
	Prod. Buch	-0,864	0,587	2,168	1	0,141	0,421	0,133	1,331
	Prod. DVD	-0,673	0,559	1,450	1	0,229	0,510	0,171	1,526
	Prod. CD	-1,817	0,673	7,293	1	0,007	0,163	0,043	0,608
	Prod. Ticket	0,498	0,532	0,877	1	0,349	1,646	0,580	4,667
	Prod. Druck.	0			0				
	Kaufen in Shop 1 vs. Kaufen in Shop 2	Konstante	-0,034	1,212	0,001	1	0,977		
Alter		0,020	0,016	1,704	1	0,192	1,021	0,990	1,052
Einkommen		-0,407	0,174	5,447	1	0,020	0,665	0,473	0,937
Online-Ein.		-0,001	0,002	0,290	1	0,590	0,999	0,994	1,003
Erhebungsd.		0,008	0,008	0,913	1	0,339	1,008	0,992	1,024
Produkt Inv.		0,049	0,114	0,184	1	0,668	1,050	0,840	1,312
Internetverb.		-0,218	0,277	0,621	1	0,431	0,804	0,468	1,383
Geschl.: w		-0,340	0,263	1,673	1	0,196	0,712	0,425	1,191
Geschl.: m		0			0				
Prod. Buch		0,636	0,503	1,597	1	0,206	1,889	0,704	5,065
Prod. DVD		0,339	0,489	0,482	1	0,488	1,404	0,538	3,662
Prod. CD		0,264	0,500	0,279	1	0,598	1,302	0,489	3,470
Prod. Ticket		0,234	0,521	0,201	1	0,654	1,263	0,455	3,508
Prod. Druck.		0			0				
Kaufen in Shop 1 vs. Gutschein		Konstante	-1,870	1,312	2,031	1	0,154		
	Alter	-0,049	0,018	7,809	1	0,005	0,952	0,919	0,985
	Einkommen	0,425	0,230	3,394	1	0,065	1,529	0,973	2,402
	Online-Ein.	-0,004	0,003	2,345	1	0,126	0,996	0,991	1,001
	Erhebungsd.	0,012	0,011	1,333	1	0,248	1,012	0,992	1,034
	Produkt Inv.	0,408	0,124	10,788	1	0,001	1,504	1,179	1,919
	Internetverb.	0,119	0,283	0,178	1	0,673	1,127	0,647	1,962
	Geschl.: w	-0,472	0,322	2,151	1	0,142	0,624	0,332	1,172
	Geschl.: m	0			0				
	Prod. Buch	1,500	0,583	6,613	1	0,010	4,482	1,429	14,062
	Prod. DVD	1,013	0,567	3,188	1	0,074	2,753	0,906	8,364
	Prod. CD	2,081	0,676	9,479	1	0,002	8,012	2,130	30,133
	Prod. Ticket	-0,264	0,538	0,242	1	0,623	0,768	0,268	2,202
	Prod. Druck.	0			0				

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 MLR: Hinzufügen der ufosV2r-Scores (Usability reflektiv) für beide Shops

Tab. E-37: Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2a

1. Likelihood-Quotiententest:				
<i>Modell</i>	<i>-2 Log-Likelihood</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
<i>Nur konstanter Term</i>	731,608			
<i>Endgültig</i>	535,559	196,049	26	0,000
2. $\chi^2$ - bzw. Devianz-Statistik:				
	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>	
<i>Pearson</i>	729,004	658	0,028	
<i>Abweichung</i>	535,559	658	1,000	
3. Pseudo-R <sup>2</sup> -Statistiken:				
<i>Cox und Snell</i>	<i>Nagelkerke</i>	<i>McFadden</i>		
0,435	0,494	0,268		

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R<sup>2</sup>).

Tab. E-38: Klassifikationsmatrix – sequentielle MRL Schritt 2a

<i>Beobachtet</i>	<i>Vorhersagt</i>			
	<i>Gutschein</i>	<i>Kauf in Shop 1</i>	<i>Kauf in Shop 2</i>	<i>% richtig</i>
<i>Gutschein</i>	31	22	23	40,79
<i>Kauf in Shop 1</i>	8	109	26	76,22
<i>Kauf in Shop 2</i>	14	18	92	74,19
<i>% insgesamt</i>	15,45	43,44	41,11	67,64

Tab. E-39: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2a

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
Konstante	535,559	0,000	0	.
Alter	553,539	17,980	2	0,000
Einkommen	551,564	16,005	2	0,000
Geschlecht	540,443	4,884	2	0,087
Anzahl Online-Einkäufe	537,603	2,044	2	0,360
Erhebungsdauer	537,241	1,682	2	0,431
Internetverbindung	536,488	0,929	2	0,628
Produkt Involvement	543,931	8,372	2	0,015
Produktart	565,976	30,417	8	0,000
Usability ufosV2r Shop 1	580,560	45,001	2	0,000
Usability ufosV2r Shop 2	616,802	81,243	2	0,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-40: Parameterschätzer – sequentielle MRL Schritt 2a

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Kaufen in Shop 2 vs. Gutschein	Konstante	1,466	1,570	0,872	1	0,350			
	Alter	0,079	0,021	14,585	1	0,000	1,082	1,039	1,127
	Einkommen	-0,912	0,252	13,085	1	0,000	0,402	0,245	0,658
	Online-Ein.	0,003	0,003	1,952	1	0,162	1,004	0,999	1,008
	Erhebungsd.	-0,004	0,011	0,148	1	0,701	0,996	0,973	1,018
	Produkt Inv.	-0,313	0,135	5,377	1	0,020	0,731	0,561	0,953
	Internetverb.	-0,282	0,364	0,601	1	0,438	0,754	0,370	1,539
	ufosV2r Sh.1	0,331	0,191	2,997	1	0,083	1,393	0,957	2,026
	ufosV2r Sh.2	-0,962	0,198	23,552	1	0,000	0,382	0,259	0,564
	Geschl.: w	0,099	0,353	0,079	1	0,779	1,104	0,553	2,203
	Geschl.: m	0			0				
	Prod. Buch	-0,893	0,619	2,076	1	0,150	0,410	0,122	1,379
	Prod. DVD	-0,604	0,589	1,054	1	0,305	0,546	0,172	1,733
	Prod. CD	-1,967	0,714	7,581	1	0,006	0,140	0,034	0,567
	Prod. Ticket	0,479	0,569	0,709	1	0,400	1,614	0,529	4,923
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Kaufen in Shop 2	Konstante	-0,675	1,481	0,208	1	0,648			
	Alter	0,026	0,019	1,938	1	0,164	1,027	0,989	1,065
	Einkommen	-0,487	0,211	5,315	1	0,021	0,615	0,406	0,930
	Online-Ein.	0,000	0,003	0,011	1	0,917	1,000	0,995	1,005
	Erhebungsd.	0,008	0,010	0,645	1	0,422	1,008	0,989	1,027
	Produkt Inv.	0,047	0,132	0,125	1	0,724	1,048	0,809	1,357
	Internetverb.	-0,063	0,335	0,035	1	0,852	0,939	0,487	1,811
	ufosV2r Sh.1	1,120	0,184	37,109	1	0,000	3,065	2,138	4,395
	ufosV2r Sh.2	-1,308	0,183	51,101	1	0,000	0,270	0,189	0,387
	Geschl.: w	-0,487	0,314	2,401	1	0,121	0,615	0,332	1,138
	Geschl.: m	0			0				
	Prod. Buch	0,772	0,587	1,728	1	0,189	2,164	0,685	6,843
	Prod. DVD	0,502	0,576	0,758	1	0,384	1,651	0,534	5,107
	Prod. CD	0,245	0,597	0,169	1	0,681	1,278	0,397	4,116
	Prod. Ticket	0,406	0,607	0,447	1	0,504	1,501	0,457	4,934
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Gutschein	Konstante	-2,142	1,378	2,415	1	0,120			
	Alter	-0,053	0,019	7,818	1	0,005	0,949	0,914	0,984
	Einkommen	0,426	0,238	3,190	1	0,074	1,531	0,959	2,442
	Online-Ein.	-0,004	0,003	2,024	1	0,155	0,996	0,991	1,001
	Erhebungsd.	0,012	0,011	1,199	1	0,273	1,012	0,990	1,035
	Produkt Inv.	0,360	0,132	7,473	1	0,006	1,433	1,107	1,855
	Internetverb.	0,219	0,285	0,591	1	0,442	1,245	0,712	2,179
	ufosV2r Sh.1	0,789	0,192	16,911	1	0,000	2,201	1,511	3,205
	ufosV2r Sh.2	-0,346	0,169	4,191	1	0,041	0,708	0,508	0,985
	Geschl.: w	-0,586	0,336	3,048	1	0,081	0,557	0,288	1,075
	Geschl.: m	0			0				
	Prod. Buch	1,665	0,605	7,571	1	0,006	5,284	1,614	17,294
	Prod. DVD	1,106	0,591	3,501	1	0,061	3,023	0,949	9,628
	Prod. CD	2,213	0,705	9,859	1	0,002	9,139	2,297	36,365
	Prod. Ticket	-0,073	0,560	0,017	1	0,896	0,930	0,310	2,785
	Prod. Druck.	0			0				

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 MLR: Hinzufügen der ufosV2f-Scores (Usability formativ) für beide Shops

Tab. E-41: Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2b

1. Likelihood-Quotiententest:				
Modell	-2 Log-Likelihood	$\chi^2$ -Wert	df	Sign.
Nur konstanter Term	731,608			
Endgültig	551,309	180,299	26	0,000
2. $\chi^2$ - bzw. Devianz-Statistik:				
	$\chi^2$ -Wert	df	Sign.	
Pearson	774,044	658	0,001	
Abweichung	551,309	658	0,999	
3. Pseudo-R <sup>2</sup> -Statistiken:				
Cox und Snell	Nagelkerke	McFadden		
0,409	0,464	0,246		

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R<sup>2</sup>).

Tab. E-42: Klassifikationsmatrix – sequentielle MRL Schritt 2b

Beobachtet	Vorhersagt			% richtig
	Gutschein	Kauf in Shop 1	Kauf in Shop 2	
Gutschein	31	23	22	40,79
Kauf in Shop 1	10	102	31	71,33
Kauf in Shop 2	13	22	89	71,77
% insgesamt	15,74	42,86	41,40	64,72

Tab. E-43: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2b

Effekt	-2 Log-Likelihood für reduziertes Modell	$\chi^2$ -Wert	df	Sign.
Konstante	551,309	0	0	
Alter	567,937	16,628	2	0,000
Einkommen	566,151	14,842	2	0,001
Geschlecht	555,258	3,949	2	0,139
Anzahl Online-Einkäufe	553,833	2,525	2	0,283
Erhebungsdauer	552,720	1,412	2	0,494
Internetverbindung	552,082	0,774	2	0,679
Produkt Involvement	560,104	8,795	2	0,012
Produktart	584,623	33,314	8	0,000
Usability ufosV2f Shop 1	598,172	46,863	2	0,000
Usability ufosV2f Shop 2	623,160	71,852	2	0,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).



Tab. E-44: Parameterschätzer – sequentielle MRL Schritt 2b

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Kaufen in Shop 2 vs. Gutschein	Konstante	1,466	1,570	0,872	1	0,350			
	Alter	0,079	0,021	14,585	1	0,000	1,082	1,039	1,127
	Einkommen	-0,912	0,252	13,085	1	0,000	0,402	0,245	0,658
	Online-Ein.	0,003	0,003	1,952	1	0,162	1,004	0,999	1,008
	Erhebungsd.	-0,004	0,011	0,148	1	0,701	0,996	0,973	1,018
	Produkt Inv.	-0,313	0,135	5,377	1	0,020	0,731	0,561	0,953
	Internetverb.	-0,282	0,364	0,601	1	0,438	0,754	0,370	1,539
	ufosV2f Sh.1	0,331	0,191	2,997	1	0,083	1,393	0,957	2,026
	ufosV2f Sh.2	-0,962	0,198	23,552	1	0,000	0,382	0,259	0,564
	Geschl.: w	0,099	0,353	0,079	1	0,779	1,104	0,553	2,203
	Geschl.: m	0			0				
	Prod. Buch	-0,893	0,619	2,076	1	0,150	0,410	0,122	1,379
	Prod. DVD	-0,604	0,589	1,054	1	0,305	0,546	0,172	1,733
	Prod. CD	-1,967	0,714	7,581	1	0,006	0,140	0,034	0,567
	Prod. Ticket	0,479	0,569	0,709	1	0,400	1,614	0,529	4,923
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Kaufen in Shop 2	Konstante	-0,675	1,481	0,208	1	0,648			
	Alter	0,026	0,019	1,938	1	0,164	1,027	0,989	1,065
	Einkommen	-0,487	0,211	5,315	1	0,021	0,615	0,406	0,930
	Online-Ein.	0,000	0,003	0,011	1	0,917	1,000	0,995	1,005
	Erhebungsd.	0,008	0,010	0,645	1	0,422	1,008	0,989	1,027
	Produkt Inv.	0,047	0,132	0,125	1	0,724	1,048	0,809	1,357
	Internetverb.	-0,063	0,335	0,035	1	0,852	0,939	0,487	1,811
	ufosV2f Sh.1	1,120	0,184	37,109	1	0,000	3,065	2,138	4,395
	ufosV2f Sh.2	-1,308	0,183	51,101	1	0,000	0,270	0,189	0,387
	Geschl.: w	-0,487	0,314	2,401	1	0,121	0,615	0,332	1,138
	Geschl.: m	0			0				
	Prod. Buch	0,772	0,587	1,728	1	0,189	2,164	0,685	6,843
	Prod. DVD	0,502	0,576	0,758	1	0,384	1,651	0,534	5,107
	Prod. CD	0,245	0,597	0,169	1	0,681	1,278	0,397	4,116
	Prod. Ticket	0,406	0,607	0,447	1	0,504	1,501	0,457	4,934
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Gutschein	Konstante	-2,142	1,378	2,415	1	0,120			
	Alter	-0,053	0,019	7,818	1	0,005	0,949	0,914	0,984
	Einkommen	0,426	0,238	3,190	1	0,074	1,531	0,959	2,442
	Online-Ein.	-0,004	0,003	2,024	1	0,155	0,996	0,991	1,001
	Erhebungsd.	0,012	0,011	1,199	1	0,273	1,012	0,990	1,035
	Produkt Inv.	0,360	0,132	7,473	1	0,006	1,433	1,107	1,855
	Internetverb.	0,219	0,285	0,591	1	0,442	1,245	0,712	2,179
	ufosV2f Sh.1	0,789	0,192	16,911	1	0,000	2,201	1,511	3,205
	ufosV2f Sh.2	-0,346	0,169	4,191	1	0,041	0,708	0,508	0,985
	Geschl.: w	-0,586	0,336	3,048	1	0,081	0,557	0,288	1,075
	Geschl.: m	0			0				
	Prod. Buch	1,665	0,605	7,571	1	0,006	5,284	1,614	17,294
	Prod. DVD	1,106	0,591	3,501	1	0,061	3,023	0,949	9,628
	Prod. CD	2,213	0,705	9,859	1	0,002	9,139	2,297	36,365
	Prod. Ticket	-0,073	0,560	0,017	1	0,896	0,930	0,310	2,785
	Prod. Druck.	0			0				

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 MLR: Hinzufügen der Variable Kaufintention für beide Shops**

Tab. E-45: Statistiken zur globalen Modellanpassung – sequentielle MRL Schritt 2c

1. Likelihood-Quotiententest:				
<i>Modell</i>	<i>-2 Log-Likelihood</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
<i>Nur konstanter Term</i>	731,608			
<i>Endgültig</i>	489,179	242,430	26	0,000
2. $\chi^2$ - bzw. Devianz-Statistik:				
	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>	
<i>Pearson</i>	763,770	658	0,003	
<i>Abweichung</i>	489,179	658	1,000	
3. Pseudo-R <sup>2</sup> -Statistiken:				
<i>Cox und Snell</i>	<i>Nagelkerke</i>	<i>McFadden</i>		
0,507	0,575	0,331		

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.), Determinationskoeffizient (R<sup>2</sup>).

Tab. E-46: Klassifikationsmatrix – sequentielle MRL Schritt 2c

<i>Beobachtet</i>	<i>Vorhersagt</i>			
	<i>Gutschein</i>	<i>Kauf in Shop 1</i>	<i>Kauf in Shop 2</i>	<i>% richtig</i>
<i>Gutschein</i>	33	30	13	43,42
<i>Kauf in Shop 1</i>	13	113	17	79,02
<i>Kauf in Shop 2</i>	8	20	96	77,42
<i>% insgesamt</i>	15,74	47,52	36,73	70,55

Tab. E-47: Ergebnisse der Likelihood-Quotienten Tests zur Beurteilung des Einflusses der Prädiktoren – sequentielle MRL Schritt 2c

<i>Effekt</i>	<i>-2 Log-Likelihood für reduziertes Modell</i>	$\chi^2$ -Wert	<i>df</i>	<i>Sign.</i>
Konstante	489,179	0	0	
Alter	508,889	19,711	2	0,000
Einkommen	502,929	13,751	2	0,001
Geschlecht	493,748	4,569	2	0,102
Anzahl Online-Einkäufe	492,143	2,965	2	0,227
Erhebungsdauer	490,906	1,728	2	0,422
Internetverbindung	489,928	0,749	2	0,688
Produkt Involvement	497,824	8,645	2	0,013
Produktart	518,876	29,697	8	0,000
Kaufintention Shop 1	620,301	131,123	2	0,000
Kaufintention Shop 2	531,935	42,757	2	0,000

Anmerkung: Differenz zwischen beiden Log-Likelihood-Werten multipliziert mit -2 (-2 Log-Likelihood), Anzahl der Freiheitsgrade (df), Signifikanz (Sign.).

Tab. E-48: Parameterschätzer – sequentielle MRL Schritt 2c

Block	Regressor	B	SE	Wald	df	Sign.	Exp(B)	95 % Konfidenzintervall Exp(B)	
								Untergrenze	Obergrenze
Kaufen in Shop 2 vs. Gutschein	Konstante	1,539	1,610	0,914	1	0,339			
	Alter	0,092	0,022	17,089	1	0,000	1,096	1,049	1,145
	Einkommen	-0,919	0,264	12,104	1	0,001	0,399	0,238	0,669
	Online-Ein.	0,004	0,003	2,567	1	0,109	1,004	0,999	1,010
	Erhebungsd.	0,003	0,013	0,061	1	0,806	1,003	0,979	1,028
	Produkt Inv.	-0,306	0,139	4,852	1	0,028	0,736	0,561	0,967
	Internetverb.	-0,097	0,367	0,070	1	0,792	0,908	0,442	1,863
	Kaufint.Sh.1	-0,748	0,117	40,799	1	0,000	0,473	0,376	0,595
	Kaufint.Sh.2	0,109	0,136	0,640	1	0,424	1,115	0,854	1,457
	Geschl.: w	0,073	0,379	0,037	1	0,847	1,076	0,512	2,262
	Geschl.: m	0			0				
	Prod. Buch	-1,072	0,653	2,696	1	0,101	0,342	0,095	1,231
	Prod. DVD	-0,954	0,627	2,315	1	0,128	0,385	0,113	1,317
	Prod. CD	-1,875	0,734	6,531	1	0,011	0,153	0,036	0,646
	Prod. Ticket	0,402	0,598	0,451	1	0,502	1,494	0,463	4,823
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Kaufen in Shop 2	Konstante	-1,046	1,623	0,415	1	0,519			
	Alter	0,040	0,021	3,452	1	0,063	1,040	0,998	1,085
	Einkommen	-0,594	0,240	6,120	1	0,013	0,552	0,345	0,884
	Online-Ein.	0,001	0,003	0,089	1	0,765	1,001	0,995	1,007
	Erhebungsd.	0,014	0,011	1,482	1	0,223	1,014	0,992	1,037
	Produkt Inv.	0,057	0,144	0,159	1	0,690	1,059	0,799	1,404
	Internetverb.	0,153	0,361	0,179	1	0,672	1,165	0,574	2,366
	Kaufint.Sh.1	-0,975	0,111	76,708	1	0,000	0,377	0,303	0,469
	Kaufint.Sh.2	0,649	0,120	28,992	1	0,000	1,913	1,511	2,422
	Geschl.: w	-0,578	0,355	2,654	1	0,103	0,561	0,280	1,125
	Geschl.: m	0			0				
	Prod. Buch	0,621	0,634	0,959	1	0,327	1,861	0,537	6,448
	Prod. DVD	0,253	0,628	0,162	1	0,687	1,288	0,376	4,406
	Prod. CD	0,286	0,658	0,189	1	0,664	1,331	0,367	4,831
	Prod. Ticket	0,475	0,653	0,528	1	0,467	1,607	0,447	5,777
	Prod. Druck.	0			0				
Kaufen in Shop 1 vs. Gutschein	Konstante	-2,585	1,409	3,367	1	0,067			
	Alter	-0,052	0,019	7,534	1	0,006	0,949	0,915	0,985
	Einkommen	0,325	0,246	1,746	1	0,186	1,385	0,854	2,244
	Online-Ein.	-0,004	0,003	1,699	1	0,192	0,996	0,991	1,002
	Erhebungsd.	0,011	0,012	0,844	1	0,358	1,011	0,988	1,034
	Produkt Inv.	0,363	0,136	7,134	1	0,008	1,438	1,102	1,877
	Internetverb.	0,250	0,289	0,745	1	0,388	1,284	0,728	2,264
	Kaufint.Sh.1	-0,227	0,116	3,817	1	0,051	0,797	0,634	1,001
	Kaufint.Sh.2	0,540	0,123	19,129	1	0,000	1,715	1,347	2,184
	Geschl.: w	-0,651	0,343	3,600	1	0,058	0,521	0,266	1,022
	Geschl.: m	0			0				
	Prod. Buch	1,693	0,627	7,284	1	0,007	5,434	1,589	18,576
	Prod. DVD	1,207	0,614	3,861	1	0,049	3,345	1,003	11,151
	Prod. CD	2,161	0,713	9,200	1	0,002	8,683	2,148	35,093
	Prod. Ticket	0,073	0,585	0,016	1	0,901	1,076	0,342	3,386
	Prod. Druck.	0			0				

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))

# Tabellarischer Lebenslauf

## Persönliche Daten

---

Name: Timo Christophersen

Anschrift: Harmsstr. 54  
24118 Kiel

Geburtsdatum/-ort: 31.01.1976 in Eutin

Familienstand: ledig

Staatsangehörigkeit: deutsch

## Schulbildung

---

1982-1986 Besuch der Grundschule in Eutin

1986-1995 Besuch des Carl-Maria-von-Weber-Gymnasiums in Eutin

06/95 Abschluss mit Abitur

## Hochschulausbildung

---

10/96-05/03 Studium an der Christian-Albrechts-Universität zu Kiel  
Psychologie (Hauptfach)  
Schwerpunkte: AOM-Psychologie, Klinische Psychologie  
Phonetik und digitale Sprachverarbeitung (Nebenfach)  
Musikwissenschaft (freiwilliges Nebenfach)  
Diplomarbeit zum Thema *Usability im Kontext des Online-Shoppings: Einflüsse auf die Kaufentscheidung*

05/03 Studienabschluss als Diplom-Psychologe, Note: sehr gut

10/03-09/06 Stipendiat im DFG-finanzierten Graduiertenkolleg  
„Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business“ der Christian-Albrechts-Universität zu Kiel

07/05-10/05 Forschungsaufenthalt an der University of Maryland in College Park (USA) in der Forschungsgruppe von Prof. Kent Norman

Kiel, den 29.11.06

Timo Christophersen