Entwicklung eines Fragebogens zur Messung des Nutzungserlebens interaktiver Produkte

vorgelegt von

Laura Riedel

Masterarbeit zur Erlangung des akademischen Grades

Master of Science (M.Sc.)

im Studiengang

Human Factors

Technische Universität Berlin

Fakultät V für Verkehrs- und Maschinensysteme

Institut für Psychologie und Arbeitswissenschaft

Erstgutachter: Prof. Dr. Manfred Thüring

Zweitgutachter: Dr. Michael Minge

Berlin, 22.04.2013

Eidesstattliche Erklärung

Die selbständige und eigenhändige Anfertigung versichert an Eides statt. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

……. …….

Ort, Datum Unterschrift

Zusammenfassung

Technische Produkte sind aus dem Alltag nicht mehr wegzudenken. Nachdem der nutzerzentrierte Gestaltungsprozess noch vor kurzer Zeit vor allem darauf abzielte, eine möglichst hohe Usability interaktiver Produkte sicherzustellen, geht der Trend mittlerweile dahin, auch die Wahrnehmung und das Erleben des Nutzers bei der Interaktion mit einem Produkt maßgeblich zu berücksichtigen und zu beeinflussen. Darauf verweist das Konzept *User Experience,* das neben der Usability auch emotionale Bewertungen, ästhetische Erfahrungen sowie die Motivation zur zukünftigen Nutzung eines Produkts integriert. Zur adäquaten Erfassung dieser Aspekte wurde auf Basis eines weithin etablierten und empirisch abgesicherten Modells zur User Experience, dem CUE-Modell von Thüring und Mahlke (2007), ein modular aufgebauter Fragebogen entwickelt. Dafür wurden zunächst bestehende Fragebogenverfahren untersucht und deren Probleme herausgestellt, die sich vor allem auf drei Bereiche beziehen: eine Operationalisierung, der es häufig an einer theoretischen Grundlage mangelt, eine Einschränkung auf einzelne Produkte, wie Software oder Webseiten, und methodische Schwächen bei der Konstruktion. Deshalb fanden diese Bereiche für den Fragebogen besondere Berücksichtigung. Drei Modulgruppen, die sich auf die Wahrnehmung aufgabenbezogener und nicht-aufgabenbezogener *Produkteigenschaften* (*Nützlichkeit*, *Benutzbarkeit*, *visuelle Ästhetik*, *Status* und *Bindung*), auf *Emotionen* und auf *Konsequenzen* der Interaktion (*Produktloyalität* sowie *Nutzungsintention*) beziehen, wurden zur Entwicklung erarbeitet. Die Basis bildeten zwei Online-Studien mit jeweils N = 238 Teilnehmern. Diese wurden instruiert anhand von 67 Items ein selbst gewähltes Produkt aus ihrem Alltag zu beurteilen, wodurch eine Vielzahl von Produktgruppen unterschieden werden konnte. Die gewonnen Daten wurden faktorenanalytisch durch eine Hauptkomponentenanalyse mit Varimax-Rotation untersucht und auf eine Anzahl von Items reduziert, die das Nutzungserleben umfassend beschreiben. Am Ende der Konstruktion enthält der finale Fragebogen 33 Items zur Bewertung der User Experience. Die Ergebnisse stützen dabei die Reliabilitätsannahme der konstruierten Module.

Stichworte: User Experience, Emotionen, Fragebogen, interaktive Produkte

Inhaltsverzeichnis

Zusammenfassung I

Abbildungsverzeichnis V

Tabellenverzeichnis VI

Formelverzeichnis VII

Abkürzungsverzeichnis VIII

1. Einleitung 1

1.1. Ziel der Arbeit 1

1.2. Aufbau der Arbeit 2

2. Theoretischer Hintergrund 3

2.1. Was ist User Experience? 3

2.1.1. User Experience und ihre Abgrenzung zur Usability 3

2.1.2. User Experience Modelle: Holistische und reduktionistische Ansätze 6

2.1.3. Das CUE-Modell von Thüring und Mahlke 8

2.2. Exkurs: Itemanalyse, Faktorenanalyse und Reliabilität 12

2.2.1. Itemanalyse 12

2.2.2. Explorative Faktorenanalyse 13

2.2.3. Reliabilität 16

2.3. Review ausgewählter Usability- und User Experience-Fragebögen 18

2.3.1. AttrakDiff 23

2.3.2. User Experience Questionnaire (UEQ) 26

2.3.3. ISONORM 9241/10 28

2.3.4. Self-Assessment Manikin (SAM) 31

2.4. Zusammenfassung 34

3. Studie 1 35

3.1. Theoretische Überlegungen: UX-Fragebogen auf Basis des CUE-Modells 35

3.2. Generieren des Itempools 37

3.3. Methode ... 40
3.3.1. Material ... 40
3.3.2. Untersuchungsdurchführung ... 40
3.3.3. Stichprobe ... 41
3.4. Ergebnisse ... 43
3.4.1. Vorbetrachtung ... 43
3.4.2. Itemanalyse ... 43
3.4.3. Faktorenanalyse der Module ... 44
3.4.4. Faktorenanalyse der Modulgruppe *Produktbewertungen* ... 50
3.4.5. Reliabilität ... 51
3.5. Diskussion ... 52
4. Studie 2 ... 54
4.1. Methode ... 54
4.1.1. Material ... 54
4.1.2. Untersuchungsdurchführung ... 56
4.1.3. Stichprobe ... 56
4.2. Ergebnisse ... 57
4.2.1. Faktorenanalyse der Modulgruppe *Emotionen* ... 58
4.2.2. Faktorenanalyse der Modulgruppe *Konsequenzen* ... 59
4.2.3. Reliabilität ... 59
4.2.4. Validität ... 60
5. Diskussion ... 63
5.1. Diskussion der Fragebogenstruktur ... 63
5.2. Gütekriterien ... 66
5.3. Qualitativer Vergleich mit weiteren Fragebögen ... 67
5.4. Auswertungsbeispiel ... 68
5.5. Kritik ... 71
5.6. Praktische Relevanz ... 71

5.7. Ausblick 72

Literaturverzeichnis 74

Anhang 83

A. Studie 1 84

B. Studie 2 94

C. Endgültiger Fragebogen 104

Abbildungsverzeichnis

Abbildung 1. Modell zur Definition von User Experience; eigene Darstellung nach DIN EN ISO 9241-210 (2011). 5
Abbildung 2. Hierarchisches User Experience Modell von Kakar (2012); Darstellung aus Kakar (2012, S. 145). 6
Abbildung 3. Hedonisches / Pragmatisches Modell von Hassenzahl (2007); eigene Darstellung nach Hassenzahl (2007). 7
Abbildung 4. Komponentenmodell von Mahlke und Thüring (2007); eigene Darstellung nach Mahlke und Thüring (2007, S. 916). 9
Abbildung 5. Mehrkomponenten-Modell von Scherer (1984); Darstellung aus Minge und Thüring (2011, S. 117). 11
Abbildung 6. Beispiel für ein Screeplot, erstellt mit IBM SPSS 21. 15
Abbildung 7. Self-Assessment Manikin von Lang (1980); Darstellung nach Bradley & Lang (1994, S. 51). 32
Abbildung 8. Komponenten und Module des Fragebogens, abgeleitet aus den Komponenten des CUE-Modells von Thüring & Mahlke (2007); eigene Darstellung. 36
Abbildung 9. Häufigkeitsverteilung und Normalverteilungskurve des Alters der Studien-Teilnehmer aus Studie 1. 41
Abbildung 10. Ladungsdiagramm des Moduls *Identifikation / Wert*. 47
Abbildung 11. Die drei finalen Modulgruppen nach der faktorenanalytischen Betrachtung der einzelnen Module; eigene Darstellung. 50
Abbildung 12. (a) Circumplex-Modell von Russell (1979), (b) Ableitung der vier Emotionsmodule aus dem Circumplex-Modell (1979); Darstellung aus Russell (1980, S. 1167). 53
Abbildung 13. Die acht Emotionsausdrücke des LEMtools, ©SusaGroup. 55
Abbildung 14. Darstellungsform des LEMtools in Studie 2. 55
Abbildung 15. Finale Modulgruppen und Module; eigene Darstellung. 64
Abbildung 16. Darstellungsbeispiel für die Ergebnisse zweier prototypischer Produkte A und B bezüglich der einzelnen Module der Modulgruppe *Produktbewertungen*. 69
Abbildung 17. Darstellungsbeispiel für die Ergebnisse der Modulgruppe *Emotionen* am Beispiel der Mittelwerte der Produktgruppe Smartphone / Handy. 70

Tabellenverzeichnis

Tabelle 1. KMO-Koeffizient und Implikationen für die Durchführbarkeit der Faktorenanalyse (Bühner, 2011, S. 347). 14
Tabelle 2. Bekannte Fragebögen, unterteilt in die Bereiche User Experience, Usability, Emotionen, Sonstige. 20
Tabelle 3. Kenntwerte zur Konstruktion des AttrakDiff-2, Daten basierend auf Hassenzahl, Burmester und Koller (2003). 26
Tabelle 4. Kenntwerte zur Konstruktion des User Experience Questionnaire, Daten basierend auf Laugwitz, Schrepp und Held (2006). 28
Tabelle 5. Grundsätze der Dialoggestaltung (DIN EN ISO 9241-110, 2006), Beispiel-Items aus dem ISONORM 9241/10. 29
Tabelle 6. Kenntwerte zur Konstruktion des ISONORM 9241/10, Daten basierend auf Prümper (1997). 31
Tabelle 7. Beispiel für ein semantisches Differential 38
Tabelle 8. Beispiel für eine Ratingskala 38
Tabelle 9. Von den Probanden gewählte Produktgruppen. 42
Tabelle 10. Nutzungsdauer des Produkts und der Produktgruppe in Monaten, sowie Nutzungshäufigkeit des Produkts pro Woche in Stunden. 42
Tabelle 11. Items mit extremer Schwierigkeit (unter .200 oder über .800). 44
Tabelle 12. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Nützlichkeit* 45
Tabelle 13. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Benutzbarkeit*. 46
Tabelle 14. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *visuelle Ästhetik* 47
Tabelle 15. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Komponente *Emotionen*. 48
Tabelle 16. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Nutzungsintention*. 49
Tabelle 17. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation des Moduls *Globalurteil*. 49
Tabelle 18. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Produktbewertungen* 51

Tabelle 19. Darstellung der Reliabilität der Module der Modulgruppe *Produktbewertungen* mittels Cronbach‘s α. 52
Tabelle 20. Von den Probanden gewählte Produktgruppen. 57
Tabelle 21. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Emotionen.* 58
Tabelle 22. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Konsequenzen* 59
Tabelle 23. α-Koeffizient der vier Module der Modulgruppe *Emotionen.* 59
Tabelle 24. α-Koeffizient der zwei Module der Modulgruppe *Konsequenzen.* 60
Tabelle 25. Korrelation der *Emotionen*-Module mit den Dimensionen des SAM und des LEMtools. 61
Tabelle 26. Nutzungsdauer des gewählten Produkts, aufgegliedert nach Produktgruppen 88
Tabelle 27. Itemanalyse 89
Tabelle 28. Interkorrelation der Items und Cronbach’s α der Modulgruppe Produktbewertung, aufgeteilt nach Modulen. 92
Tabelle 29. Itemanalyse der Roh-Items. 100
Tabelle 30. Interkorrelation der Items und Cronbach’s α der .Modulgruppe *Emotionen.* 102
Tabelle 31. Interkorrelation der Items und Cronbach’s α der Modulgruppe *Konsequenzen.* 103

Formelverzeichnis

Formel 1. Schwierigkeitsindex p (Dahl, 1971); nach Moosbrugger & Kelava (2007, S. 79). 12
Formel 2. Berechnung von Cronbach’s α (Cortina, 1993, S. 99). 18

Abkürzungsverzeichnis

Aufl.	Auflage
bzw.	beziehungsweise
CUE	Components of User Experience
et al.	et alii (lat. und andere)
Hrsg.	Herausgeber
ISO	International Organization for Standardization
KMO	Kaiser-Meyer-Olkin-Koeffizient
KTT	Klassische Testtheorie
M	Mittelwert
MAP	Minimum-Average-Partial
MAX	Maximum
MIN	Minimum
MSA	Measure of Sample Adequacy
N	Stichprobe
PANAS	Positive and Negative Affect Schedule
r	Korrelationskoeffizient
S.	Seite
SAM	Self-Assessment Manikin
SD	Standardabweichung
UEQ	User Experience Questionnaire
Var	Varianz
vgl.	vergleiche
V↑A↑	positive Valenz, hohes Arousal
V↑A↓	positive Valenz, niedriges Arousal
V↓A↑	negative Valenz, hohes Arousal
V↓A↓	negative Valenz, niedriges Arousal
z.B.	zum Beispiel

1. Einleitung

User Experience, welche in der deutschen Literatur häufig auch als *Nutzungserleben* bezeichnet wird (z. B. Hassenzahl, Burmester & Koller, 2008; Minge, 2011), ist ein Begriff, der zurzeit im Bereich der Mensch-Technik-Interaktion sehr populär ist. Ein Beispiel dafür ist die Namensänderung der Usability Professionals Association (UPA) in User Experience Professionals Association (UXPA) im Juni 2012 (Gunther, 2012). Der amtierende Präsident der UXPA, Rich Gunther, begründet dies damit, dass die wachsende User Experience Community eine unterstützende Basis brauche, die sie nun in der UXPA habe. Eine solche Basis ist nötig, da bisher verschiedene Wissenschaftsdisziplinen unter dem Begriff *User Experience* Unterschiedliches verstehen. Die Folge ist eine große Bandbreite an Definitionen und Operationalisierungen dieses Begriffs (Scapin, Senach, Trousse & Pallot, 2012).

Ein wichtiges Instrument der User Experience Forschung sind Fragebögen, da mit deren Hilfe das subjektive, persönliche Erleben des Nutzers adäquat erfasst werden kann. Ein guter Fragebogen basiert zum einen auf einem abgesicherten theoretischen Modell und zum anderen erfüllt er durch Reliabilität, Validität und Objektivität die Hauptgütekriterien der klassischen Testtheorie (KTT) sowie das Nebengütekriterium Ökonomie.

1.1. Ziel der Arbeit

Es gibt einige User Experience Fragebögen, die sowohl in der Forschung als auch in der Praxis Zuspruch finden. Sie haben aber Schwachstellen, die sich entweder durch fehlenden Bezug zu einem Modell oder unzureichend abgesicherte Konstruktion auszeichnen.

Deshalb ist das Anliegen dieser Arbeit, einen Fragebogen vorzustellen, dessen theoretische Grundlage ein etabliertes Modell zur Beschreibung von User Experience bildet.

Da der Fragebogen neben instrumentellen und nicht-instrumentellen User Experience Komponenten auch Emotionen und Konsequenzen des Erlebens abfragt, stellt er ein umfassenderes Instrument als die bisher existierenden dar. Ein weiterer Unterschied zu bisherigen Verfahren ergibt sich aus der Möglichkeit, den Fragebogen für die Untersuchung vieler verschiedener interaktiver Produkte einzusetzen, da verschiedene Produktgruppen bereits gezielt bei der Konstruktion des Fragebogens eingeschlossen wurden. Zudem basiert die Itembetrachtung auf einer großen Stichprobe, was den Fragebogen von bisherigen Verfahren abhebt.

1.2. Aufbau der Arbeit

Ausgehend von einer Definition der User Experience und einer Abgrenzung von dem etablierten Konstrukt der *Usability*, werden wichtige User Experience Modelle vorgestellt. Es gibt zwei Ansätze, die User Experience zu definieren versuchen und dabei verschiedene Sichtweisen annehmen. Die holistische Sichtweise versucht User Experience möglichst umfassend zu erklären, hat aber den Nachteil nicht falsifizierbarer Modelle. Der reduktionistische Ansatz beschränkt sich dagegen auf zentrale Aspekte des Nutzungserlebens und operationalisiert sie (Blythe, Reid, Wright & Geelhoed, 2006). Hierzu zählt das CUE-Modell (Thüring & Mahlke, 2007), das nachfolgend ausführlich beschrieben wird. Daran schließt sich ein kritisches Review über eine Auswahl bisheriger Fragebogenverfahren an, woraus die Anforderungen des vorliegenden Fragebogens abgeleitet werden.

Nach der Operationalisierung des Modells wird die Gestaltung und Durchführung von zwei Online-Studien dargestellt, mit denen die Daten für die Faktorenanalyse zur Konstruktion des Fragebogens gewonnen wurden.

Zum Abschluss wird der Fragebogen vorgestellt. Es erfolgt ein qualitativer Vergleich mit anderen Verfahren und Im Ausblick wird insbesondere auf die Revidierung und Validierung des Tools eingegangen.

2. Theoretischer Hintergrund

In diesem Kapitel werden zunächst die theoretischen Grundlagen zum Konzept der User Experience vorgestellt. Der zweite Teil konzentriert sich auf bekannte Usability- und User Experience-Fragebögen, die kritisch beleuchtet werden.

2.1. Was ist User Experience?

Um eine Vorstellung davon zu gewinnen, was unter User Experience verstanden wird, ist neben dem Versuch einer Definition zunächst eine Abgrenzung zur Usability erforderlich. Anschließend werden mit der holistischen und der reduktionistischen Sicht verschiedene Perspektiven auf User Experience beschrieben und wichtige Vertreter und Modelle der beiden Herangehensweisen vorgestellt.

2.1.1. User Experience und ihre Abgrenzung zur Usability

Die User Experience musste sich lange Zeit neben dem sowohl viel bekannteren als auch schon länger im Fokus der Wissenschaft und Praxis stehenden Konstrukt der Usability behaupten (Hassenzahl & Tractinsky, 2006).

Alben definierte im Jahr 1996 zum ersten Mal *Experience* im Umgang mit interaktiven Produkten:

> By “experience” we mean all the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they’re using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it (Alben, 1996, S. 12).

Alben betonte, dass der Erfolg und die Verbindlichkeit der *Experiences* von dem Wert abhingen, den sie für den Nutzer haben (Alben, 1996).

In den darauf folgenden Jahren gab es eine Vielzahl verschiedener Definitionen und Modelle zu diesem Konzept. Nach Hassenzahl und Tractinsky (2006) behindere diese Uneinheitlichkeit eine breite empirische Erforschung der User Experience.

Im Vergleich zur User Experience war und ist Usability ein gut erforschtes und etabliertes Konstrukt (Hassenzahl, Burmester & Koller, 2003; Nielsen, 1993). Zudem wurde Usability unter der deutschen Bezeichnung *Gebrauchstauglichkeit* bereits 1998 in die ISO-Normen aufgenommen (DIN EN ISO 9241-11, 1998), was schnell zu einer einheitlicheren Definition geführt hat. Usability ist durch eine effektive, effiziente und zufriedenstellende Zielerreichung definiert. Das Ausmaß, in

dem ein System, ein Produkt oder eine Dienstleistung für diese Zielerreichung genutzt werden kann, spiegelt den Grad der Gebrauchstauglichkeit wider (DIN EN ISO 9241-11, 1998). Dies deckt sich auch mit dem Verständnis von Usability als „objektives“ Erhebungsmaß, das wenig von subjektiven Empfindungen des Nutzers beeinflusst wird (Hassenzahl, Burmester & Koller, 2008).

Wie der deutsche Name andeutet, umfasst das Nutzungserleben über die reine Zufriedenheit mit einem Produkt hinaus auch Aspekte, die das subjektiv wahrgenommene Erleben des Benutzers darstellen.

Inzwischen wurde eine einheitliche Definition der User Experience in Form der ISO-Norm DIN EN ISO 9241-210 (2011) herausgegeben. Gemäß Punkt 2.15 umfasst sie alle „Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren“ (DIN EN ISO 9241-210, 2011, S. 7). Drei Anmerkungen ergänzen die Definition und umgrenzen damit das Konstrukt noch genauer:

> ANMERKUNG 1 User Experience umfasst sämtliche Emotionen, Vorstellungen, Vorlieben, Wahrnehmungen, physiologischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen, die sich vor, während und nach der Nutzung ergeben.
>
> ANMERKUNG 2 User Experience ist eine Folge des Markenbilds, der Darstellung, Funktionalität, Systemleistung, des interaktiven Verhaltens und der Unterstützungsmöglichkeiten des interaktiven Systems, des psychischen und physischen Zustands des Benutzers aufgrund seiner Erfahrungen, Einstellungen, Fähigkeiten und seiner Persönlichkeit sowie des Nutzungskontextes.
>
> ANMERKUNG 3 Die Gebrauchstauglichkeit kann, sofern sie unter dem Blickwinkel der persönlichen Ziele des Benutzers interpretiert wird, die Art der typischerweise mit der User Experience verbundenen Wahrnehmungen und emotionalen Aspekte umfassen. Kriterien der Gebrauchstauglichkeit können angewendet werden, um Aspekte der User Experience zu beurteilen.

Diese Definition grenzt die Usability zum einen deutlich von der User Experience ab und schließt sie zum anderen durch Anmerkung 3 als Teilbereich der User Experience ein. Außerdem kann bei dieser Definition von einer ganzheitlichen Sicht auf den Menschen in der Mensch-Technik-Interaktion ausgegangen werden (vgl. Minge, 2011). Ein vereinfachendes Modell dieser Definition ist in Abbildung 1 dargestellt. Hier wird deutlich, dass die Interaktion zwischen dem Nutzer und einem Produkt im Vordergrund steht. Während dieser spielen neben Emotionen und Verhaltensweisen der Person auch Vorerfahrungen mit dem Produkt eine Rolle. Diese Interaktion ist zumeist in einen Nutzungskontext eingebettet, wobei damit nicht nur die eigentliche Situation gemeint ist, in der die Interaktion stattfindet, sondern auch eine zu erledigende Aufgabe gemeint sein kann, die die Art der Interaktion beeinflusst (Minge & Thüring, 2011). Freie Exploration und

das Verfolgen einer spezifischen Aufgabe können die User Experience verschieden beeinflussen (Hassenzahl, Burmester & Koller, 2003; vgl. Fierley & Engl, 2010).

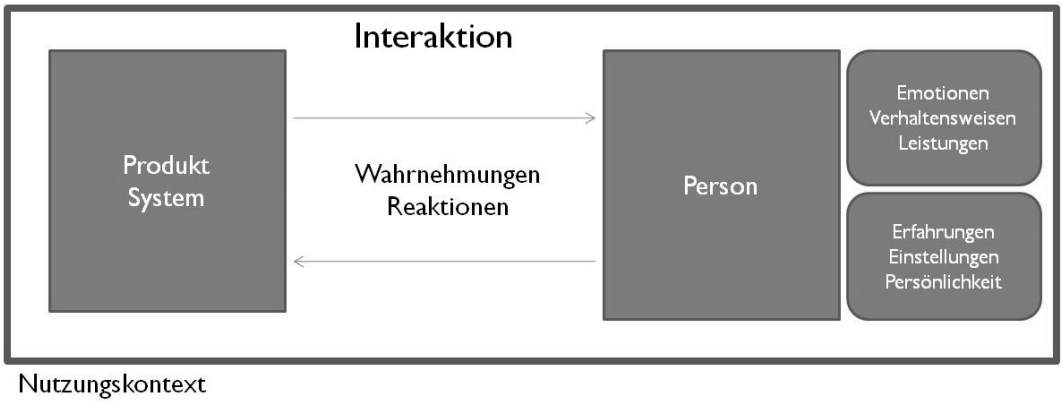


Abbildung 1. Modell zur Definition von User Experience; eigene Darstellung nach DIN EN ISO 9241-210 (2011).

Hassenzahl (2008) weist auf zwei Schwachstellen der englischen Definition der ISO-Norm hin, die aber ebenso auf die deutsche Definition übertragbar sind. Einerseits bemängelt er die Verwendung des Ausdrucks „all aspects“; in der deutschen Fassung synonym der verallgemeinernde Begriff „sämtliche“. Dies sei zwar auf den ersten Blick nachvollziehbar, mindere aber die Eindeutigkeit der Klassifizierung. Anderseits geht er auf den verwendeten, aber nicht definierten Begriff „desirability“ (deutsch „Vorlieben“) ein. Deshalb schlägt er vor, User Experience zu interpretieren „as a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service.” (Hassenzahl, 2008, S. 12).

User Experience kann entsprechend der DIN EN ISO 9241-210 (2011) bei der Nutzung eines Produkts, Systems oder einer Dienstleistung erfasst werden. Der Fokus der Ausführungen in dieser Arbeit liegt auf der User Experience im Zusammenhang mit interaktiven Produkten. Ein interaktives Produkt kann Eingaben von einem Nutzer empfangen und Ausgaben an einen Nutzer übermitteln (vgl. DIN EN ISO 9241-210, 2011). So ist beispielsweise ein Handy ein interaktives Produkt, weil ein Nutzer sowohl Eingaben wie das Eingeben einer Telefonnummer tätigen, als auch Rückmeldungen wie die Anzeige der Telefonnummer auf dem Display durch das System erhalten kann.

Da die Definitionen zur User Experience vielfältig waren und sind, wurden und werden viele verschiedenartige Modelle entwickelt. Diese lassen sich zu zwei Ansätzen zusammenfassen: den holistischen und den reduktionistischen (Blythe et al., 2006).

2.1.2. User Experience Modelle: Holistische und reduktionistische Ansätze

Holistische Ansätze gehen von einer ganzheitlichen Betrachtung eines Konstrukts aus (Gloy, 2006; vgl. Kritzenberger, 2011). Nach Gloy (2006) sei das Ganze mehr als die Summe seiner Teile und so dürfe nicht von der Betrachtung einzelner Aspekte auf ein Ganzes geschlossen werden. Somit verfolgen Holisten eine pragmatische oder phänomenologische Herangehensweise (Blythe et al., 2006; vgl. Minge, 2011). Reduktionistische Ansätze versuchen dagegen das komplexe Konstrukt auf wesentliche und repräsentative Aspekte zu reduzieren. Diese Herangehensweise erhebt den Anspruch, kognitiv und messbar zu sein (Blythe et al., 2006; Liening, 1999; vgl. Minge, 2011).

Beide Ansätze sind nach wie vor Forschungsgegenstand, wie das 2012 veröffentlichte Modell von Kakar zeigt, das zu den holistischen Ansätzen zählt (Kakar, 2012). Den reduktionistischen Modellen können unter anderem das *hedonische / pragmatische Modell der User Experience* von Hassenzahl (2007) und das *CUE-Modell* von Mahlke und Thüring (2007) zugeordnet werden.

Holistische Ansätze

Da die holistischen Ansätze nicht Gegenstand dieser Ausführungen sein sollen, wird stellvertretend kurz auf das aktuelle Modell von Kakar (2012) eingegangen. Weitere holistische Modelle finden sich unter anderem bei Blythe et al. (2006). Kakar (2012) beschreibt eine hierarchische Beziehung utilitarischer und hedonischer Aspekte. Utilitarische Aspekte beziehen sich auf Instrumentalität und Funktionalität eines Produkts. Hedonische Aspekte umfassen Ästhetik und Erleben. Zwischen den kognitiv motivierten utilitarischen und den affektiv motivierten hedonischen Aspekten besteht ein hierarchischer Zusammenhang , da die utlilitarischen Aspekte erfüllt sein müssen um hedonische Bedürfnisse zu wecken (siehe Abbildung 2). Dabei geht Kakar davon aus, dass unzufriedene Nutzer utilitarische Aspekte bevorzugen, während zufriedene hedonische Aspekte bevorzugen, die sie zu „pleasure, joy and fun“ (S. 144) führen.

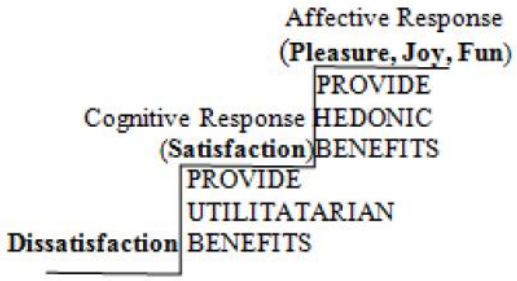


Abbildung 2. Hierarchisches User Experience Modell von Kakar (2012); Darstellung aus Kakar (2012, S. 145).

Holistische Modelle sind von Vorteil, da sie User Experience als ganzheitlichen Prozess begreifen (Blythe et al., 2006; Minge, 2011, S. 11). Sie legen ihren Fokus auf die Einzigartigkeit des Erlebens. Nachteilig ist, dass die Modelle in der Regel keine empirische Absicherung aufweisen und aufgrund ihres ganzheitlichen Charakters kaum falsifizierbar sind (Minge, 2011). Hassenzahl (2008), als Vertreter des reduktionistischen Ansatzes, hält es für kritisch, eine ganzheitliche, von einzigartigem Erleben ausgehende Sichtweise anzunehmen. Es bedürfe dadurch eines ausgesprochen reflektierten „Erlebenden“. Zudem scheint das Erleben von Technik nach seiner Ansicht viel weniger einzigartig zu sein als die Holisten annehmen: „Accounts of according experiences might differ in their quality, the experience itself doesn't.“ (Hassenzahl, 2008, S. 14)

Reduktionistische Ansätze

Reduktionistische Ansätze haben nicht den Anspruch einer ganzheitlichen Betrachtung der User Experience. Vielmehr werden wesentliche Aspekte herausgegriffen, die dabei möglichst eindeutig operationalisiert werden. Dadurch ermöglichen sie eine empirische Absicherung des Modells. Hassenzahl (2008) wehrt sich gegen Kritiken an diesem Ansatz und erklärt, er wolle keine „one-size-fits-all-definition“ der User Experience geben.

Im *hedonischen / pragmatischen Modell der User Experience*, das in Abbildung 3 veranschaulicht ist, unterscheidet Hassenzahl (2007) mithilfe eines zweidimensionalen Ansatzes zwei Produktqualitäten. Deshalb ist es auch unter dem Namen *Zwei-Komponenten-Modell wahrgenommener Produktqualität* bekannt (z.B. Burmester, Hassenzahl & Koller, 2002).

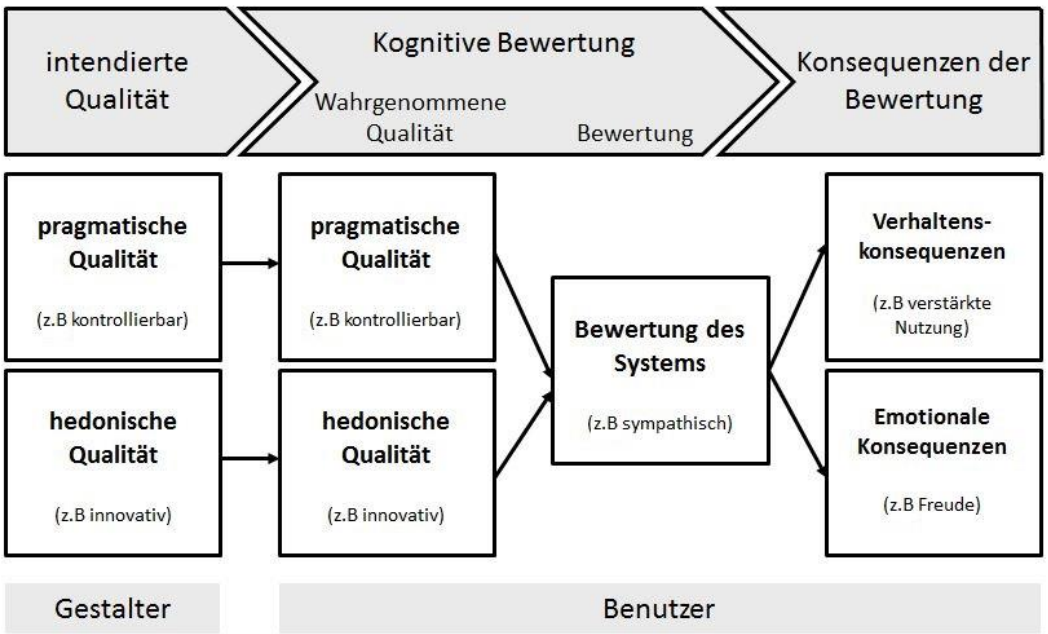


Abbildung 3. Hedonisches / Pragmatisches Modell von Hassenzahl (2007); eigene Darstellung nach Hassenzahl (2007).

Die *pragmatische* Produktqualität, die sich auf die Anforderungen des Produkts bezieht, umfasst zum einen Usability und zum anderen Utility, welche als die Nützlichkeit eines Produkts definiert ist. Die *hedonische* Qualität, die sich auf die Betrachtung der eigenen Person bezieht, wird in *Identität* und *Stimulation* unterteilt (Hassenzahl, Burmester & Koller, 2008). Identität ist geprägt durch die Interaktion mit anderen, im Sinne der Kommunikation der eigenen Werte und durch warmherzige Beziehungen. Stimulation ist der Wunsch, Neues zu entdecken, Veränderung zu erleben oder über sich hinaus zu wachsen. Hassenzahl (2007) geht davon aus, dass die beiden Produktqualitäten unkorreliert sind. Aus der Wahrnehmung der beiden Produktqualitäten entsteht ein Gesamturteil, das Hassenzahl et al. (2008) als *Attraktivität* bezeichnen.

Weiterhin erklärt Hassenzahl (2007), dass *Experience* keine direkte Komponente seines Modells sei, aber in verschiedener Weise mit dem Modell zusammen hänge. Die Produktwahrnehmung sei durch Vorerfahrungen geprägt und beeinflusst. Daraus schließt Hassenzahl (2007), dass die Messung von hedonischen und pragmatischen Qualitäten auch jeweils den Status Quo der Erfahrung enthält. Zudem sei Experience durch eine Zeitkomponente im Modell enthalten. Eine dritte Begründung für den Zusammenhang von Experience und seinem Modell sei die Fähigkeit, zwischen augenblicklicher Wahrnehmung und allgemeiner Attraktivität (z.B. gut – schlecht) zu unterscheiden. Als letzten Punkt führt Hassenzahl (2007) an, dass sich das Modell mit dem momentanen „Modus“ einer Person gegenseitig beeinflussen. Er unterscheidet zwischen dem Action-Modus, in dem eine Person mit dem Produkt aufgrund des Produkts selbst interagiere und „on the fly“ sei, und dem Ziel-Modus, in dem das Produkt für das Erledigen einer Aufgabe genutzt werde. Sei eine Person im Ziel-Modus, stünden bei ihr pragmatische Produkteigenschaften im Vordergrund, sei sie im Action-Modus, habe sie keine Präferenzen für eine bestimmte Produkteigenschaft.

Von dem Modell ausgehend haben Hassenzahl, Burmester und Koller (2003) den Fragebogen AttrakDiff entwickelt, der in Abschnitt 2.3.1 diskutiert wird.

2.1.3. Das CUE-Modell von Thüring und Mahlke

Das zentrale Modell der vorliegenden Abhandlung ist das Komponenten- bzw. CUE-Modell (Components of User Experience) von Thüring und Mahlke (2007). Es beschreibt die Interaktion eines Nutzers mit einem Produkt. Dabei beeinflussen sowohl Faktoren des Systems als auch des Nutzers sowie Kontextfaktoren, wie die Art der Aufgabe, die Interaktion von Nutzer und System. Die Autoren verfolgten den Ansatz, ein Modell aufzustellen, das die wichtigsten Komponenten der User Experience und ihre möglichen Zusammenhänge darstellt, sich empirisch absichern lässt und verschiedene Methoden bereit stellt. Daher differenzieren Mahlke und Thüring (2007) zwischen drei

Komponenten der User Experience, die aufeinander wechselseitig Einfluss nehmen können (siehe Abbildung 4). Diese werden nachfolgend näher betrachtet.

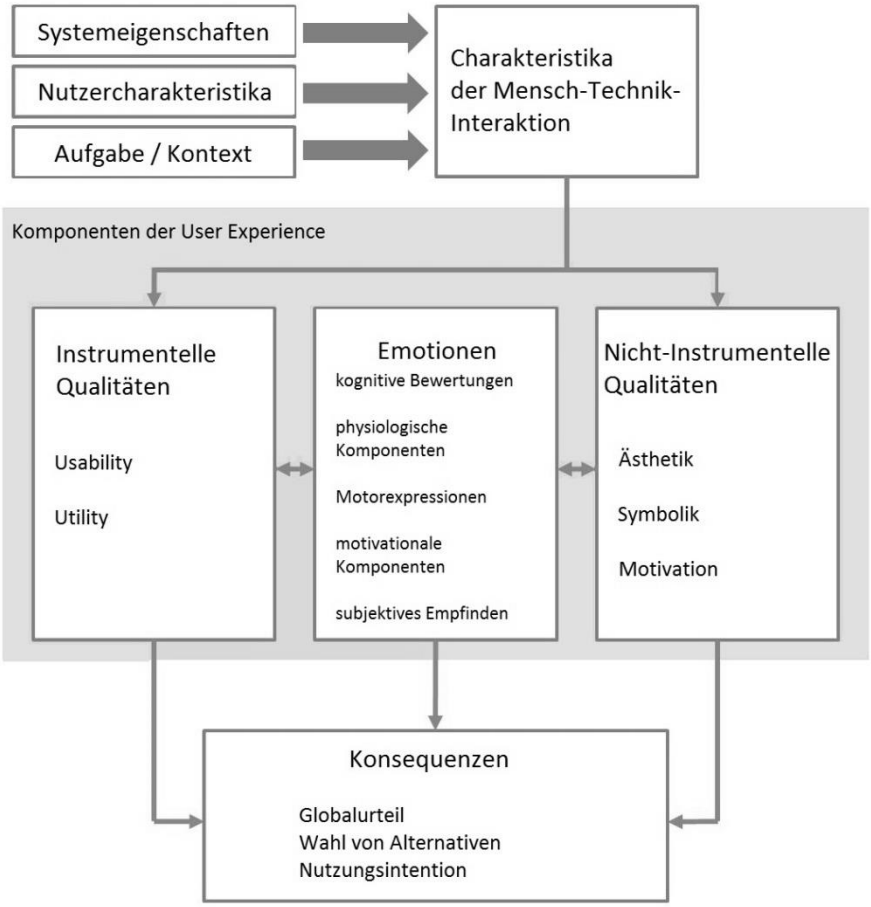


Abbildung 4. Komponentenmodell von Mahlke und Thüring (2007); eigene Darstellung nach Mahlke und Thüring (2007, S. 916).

Mahlke und Thüring (2007) unterscheiden zwei Produktqualitäten, die durch die Art und Weise der Interaktion des Nutzers mit dem System beeinflusst werden. Instrumentelle Qualitäten beziehen sich auf die beiden Bereiche *wahrgenommene Nützlichkeit* (Utility) und *Gebrauchstauglichkeit* (Usability) eines Systems. Mahlke (2008) nennt eine Reihe empirisch relevanter Aspekte der Usability (vgl. Minge, 2011). Effizienz beschreibt das Verhältnis von Genauigkeit und Vollständigkeit zum Aufwand eines Nutzers, um ein bestimmtes Ziel zu erreichen (DIN EN ISO 9241-110, 2006). Weitere wesentliche Aspekte sind die Kontrollierbarkeit, aufzufassen als eine erwartungstreue Systemreaktion, und die Unterstützung durch das System mittels Rückmeldungen Lernförderlichkeit bedeutet eine Unterstützung und Anleitung durch das System bei dem Erlernen von dessen Bedienung (DIN EN ISO 9241-110, 2006). Dies zeigt, dass sich die Aspekte sowohl zum Teil auf die Normen der Grundsätze der Dialoggestaltung (DIN EN ISO 9241-110, 2006) und der Defini-

tion der Gebrauchstauglichkeit (DIN EN ISO 9241-210, 2011) beziehen, als auch aus empirischen Befunden abgeleitet wurden (Minge, 2011).

Nicht-instrumentelle Qualitäten betreffen das „look and feel“ des Produkts. Es werden ästhetische, symbolische und motivationale Eigenschaften eines Produkts unterschieden. Die *Ästhetik* ergibt sich durch die direkte Wahrnehmung eines Produkts mittels visueller, haptisch-somatischer und auditiver Sinnesreize (Lemke, 2007). Dabei nehmen die visuellen Sinnesmodalitäten die wichtigste Rolle bei der Mensch-Technik-Interaktion ein, da von den primären, visuellen Erfahrungen mit einem Produkt auf weitere Eigenschaften geschlossen werden kann (Lemke, 2007; Minge, 2011). Sieht ein Produkt beispielsweise metallisch aus, kann ausgehend von physikalischen Prinzipien angenommen werden, dass es sich kühl anfühlt. Thielsch und Jaron (2012) zeigen, dass Ästhetik neben Usability und Inhalt einen starken Einfluss auf die Bewertung von Webseiten hat (vgl. Jaron & Thielsch, 2009; vgl. Thielsch 2008). Hassenzahl und Monk (2010) weisen einen Zusammenhang von Ästhetik und Usability nach (vgl. Tractinsky, 2011).

Thüring und Mahlke (2007) unterscheiden zwischen den beiden *symbolischen* Eigenschaften assoziative und kommunikative Qualität. Für die Definition von assoziativen Symbolen beziehen sich die Autoren auf Rafaeli und Vilnai-Yavetz (2004), die Symbolismus als durch das Produkt hervorgerufene Bedeutungen oder Assoziationen im Sinne von kognitiven Verknüpfungen mit anderen Situationen bzw. Produkten definieren. Jedes Produkt, selbst ein Stuhl oder Tisch, habe eine Bedeutung, die nicht allein entsprechend der Wünsche des Herstellers hervorgerufen, sondern durch subjektive Bewertungen der Nutzer beeinflusst wird (vgl. Minge, 2011). Kommunikative Symbole stellen den Austausch von Nutzern mit anderen in den Mittelpunkt, z.B. bezüglich Normen und Werten (Minge, 2011), und sind vergleichbar mit der *Identität* im Modell von Hassenzahl (2007).

Motivationale Produkteigenschaften betreffen nach Thüring und Mahlke (2007) insbesondere Aspekte der Selbstregulierung und erhöhter Handlungspotenz. Selbstregulierung bezieht sich auf die Fähigkeit einer Person, sich auf internale oder externale Begebenheiten, wie Gedanken, Gefühle und Handlungen, einzustellen und anzupassen (Blair, Calkins & Kopp, 2010). Effektive und das Selbstwertgefühl stärkende Produkte können eine Erhöhung der Handlungspotenz erreichen (Habermas, 1996), die sich auf die Aneignung von Wissen und Fertigkeiten sowie soziale Kontakte bezieht (Minge, 2011). Hier zeigen sich Überschneidungen mit der *Stimulation* im Modell von Hassenzahl (2007).

Diese beiden Produktqualitäten stehen in einer wechselseitigen Beziehung mit der Komponente *Emotionen* (Thüring & Mahlke, 2007). Emotionen beschreiben Thüring & Mahlke (2007) als

zeitlich begrenzte, subjektive Gefühle, die physiologische und behaviorale Reaktionen hervorrufen. So ist es vorstellbar, dass wenig gebrauchstaugliche Systeme negative Emotionen bei einer Person hervorrufen oder positive Emotionen verhindern können. Emotionen können sich aber auch auf die Bewertung von Produktqualitäten auswirken, indem zum Beispiel ein schlecht gelaunter Nutzer ein Produkt als weniger ästhetisch empfindet. Die Komponenten von Emotionen leiten sie aus dem Mehrkomponenten-Modell von Scherer (1984) ab. Er beschreibt Emotionen als psychologisches Konstrukt mit fünf Komponenten: kognitive Bewertungen, physiologische Komponenten bezogen auf Arousal (Aktiviertheit), Motorexpressionen, motivationale Komponenten einschließlich Verhaltensintention und subjektives Empfinden (Scherer, 1984). Subjektive Empfindungen („Ich fühle mich traurig.“), Motorexpressionen ("Ich ziehe die Mundwinkel nach unten.") und physiologische Komponenten ("Es bilden sich Tränen, ich weine.") bilden dabei eine Emotionstriade, die in Wechselwirkung mit der kognitiven Bewertung und motivationalen Verhaltenskomponenten steht (siehe Abbildung 5; vgl. Minge, 2011; vgl. Minge & Thüring, 2011).

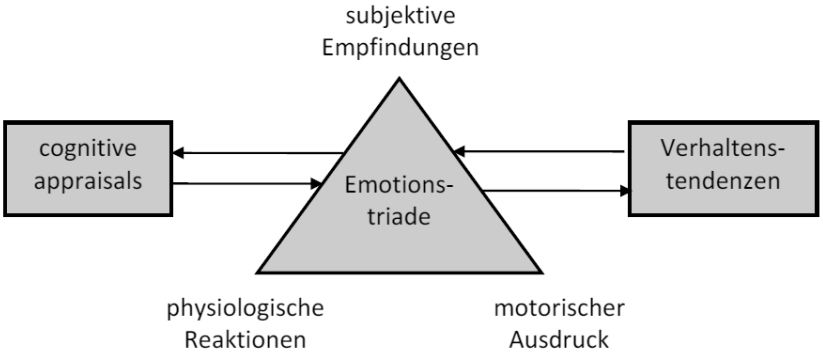


Abbildung 5. Mehrkomponenten-Modell von Scherer (1984); Darstellung aus Minge und Thüring (2011, S. 117).

Sowohl die instrumentellen als auch die nicht-instrumentellen Produktqualitäten und die Emotionen bilden zusammen die zentralen User Experience Komponenten (Mahlke, 2008). Diese Komponenten beeinflussen die *Verhaltenskonsequenzen*. Dabei unterscheidet Mahlke (2008) zwischen der Wahl von Alternativen, dem beobachtbaren Nutzerverhalten bzw. der Nutzungsintention und einem Globalurteil. Verschiedene Studien und Modelle zeigen, dass das Globalurteil mehr als nur die Summe der Einzelantworten repräsentiert und einen zusätzlichen Mehrwert bietet (Harsch, Neumann, Lehmann & Schröder, 2007; vgl. Götze, 2010). Wie dieses allerdings mit den anderen Konsequenzen zusammen hängt, ist bislang weitestgehend ungeklärt.

Zusammenfassend stellt das Komponentenmodell von Mahlke und Thüring (2007) ein Rahmenmodell der User Experience dar, welches die drei Erlebensaspekte instrumentelle und nicht-instrumentelle Produktqualitäten sowie Emotionen integriert (Minge, 2011). Darauf nehmen die

drei Interaktionscharakteristika des Nutzers, des Produkts und des Kontextes Einfluss. Konsequenzen des Erlebens zeigen sich in Globalurteilen und dem Verhalten des Nutzers.

2.2. Exkurs: Itemanalyse, Faktorenanalyse und Reliabilität

Da der nächste Abschnitt verschiedene Usability- und User Experience- Fragebögen vorstellen wird und sich ihre Betrachtung auch auf die statistischen Methoden zur Generierung und Reduktion des Itempools bezieht, soll eine Einführung in die Itemanalyse, Reliabilitätsbetrachtung und explorative Faktorenanalyse gegeben werden.

2.2.1. Itemanalyse

Schwierigkeit, Trennschärfe, Streuung und Varianz sind Kriterien, die bei der Itemanalyse empirisch ermittelt und durch einen Kennwert bezeichnet werden (Lienert & Raatz, 1998).

Kriterium Schwierigkeit

Schwierigkeit sagt aus, wie leicht Personen einem Item zustimmen können. Während Zustimmung zu einem Item zu einer hohen Merkmalsausprägung führt, bedingt das Ablehnen eines Items eine niedrige Merkmalsausprägung. Einem leichten Item wird häufiger zugestimmt. Dagegen wird ein schwieriges Item häufiger abgelehnt. Sehr leichte und sehr schwierige Items sind wenig informativ, da sie keine Bewertungsunterschiede anzeigen können. Die Schwierigkeit beeinflusst somit die Verteilung der Testwerte (Bortz & Döring, 2006). Der Schwierigkeitsindex p kann nach Dahl (1971) für mehrstufige Antwortkategorien bestimmt werden und ergibt Werte zwischen 0 (völlige Ablehnung) und 1 (völlige Zustimmung). Formel 1 veranschaulicht den Schwierigkeitsindex.

Formel 1. Schwierigkeitsindex p (Dahl, 1971); nach Moosbrugger & Kelava (2007, S. 79).

$$p_i = \frac{\sum_{v=1}^{n} x_{vi}}{n(k-1)}$$

p: Schwierigkeitsindex nach Dahl (1971)
x_{vi}: Skalenwert je Person
n: Anzahl der Probanden
k: Anzahl der Skalenstufen

Die Items sollten eine möglichst breite Schwierigkeitsstreuung aufweisen. Um extreme Schwierigkeiten auszuschließen, werden Items mit einem Index zwischen .20 und .80 bevorzugt und andere gegebenenfalls ausgeschlossen (vgl. Bortz & Döring, 2006; vgl. Lienert & Raatz, 1998).

Kriterium Trennschärfe

Die Trennschärfe gibt an, wie gut ein einzelnes Item den Gesamttestwert repräsentiert. Dafür wird die Korrelation jedes Item mit dem Gesamttestwert betrachtet. Je höher die Korrelation ist, desto besser trennt das Item zwischen Probanden mit hoher und niedriger Merkmalsausprägung. Hat ein Item eine Trennschärfe von Null, vermag es nicht zwischen Probanden mit hoher und niedriger Bewertung zu unterscheiden (Lienert & Raatz, 1998). Dadurch, dass das Item selbst auch mit in den Gesamttestwert eingeht, wird die Trennschärfe leicht überschätzt und es kommt zur so genannten Teil-Ganz-Scheinkorrelation (Lienert & Raatz, 1998). Diese kann verhindert werden, indem der Gesamtwert jeweils um das aktuelle Item bereinigt wird. Lienert und Raatz (1998) merken aber an, dass es durch die Trennschärfekorrektur zu einem Mehraufwand komme, der den geringen Nutzen nicht rechtfertige. Die Teil-Ganz-Scheinkorrelation spiele nur eine Rolle, wenn der Test wenige Items enthalte, die eine mittlere Schwierigkeit aufweisen und diese untereinander nicht oder kaum korrelieren.

Kriterien Streuung und Varianz

Bei der Itemanalyse können auch die Betrachtung der Streuung und der Varianz interessant sein. Wird die Streuung betrachtet, ist insbesondere die Verteilung der Bewertungen der Probanden für jedes Item interessant und ob diese Verteilung „überzufällig von einer Normalverteilung abweicht“ (Bortz & Döring, 2006, S. 218). Die Varianz sagt etwas über die Differenzierungsfähigkeit eines Items hinsichtlich der untersuchten Probandenstichprobe aus (Moosbrugger & Kelava, 2007).

2.2.2. Explorative Faktorenanalyse

Die explorative Faktorenanalyse soll helfen, die Anzahl von Items zu reduzieren, indem die Items auf eine geringere Zahl von gemeinsamen Dimensionen (je nach Extraktionsmethode als Faktoren oder Komponenten bezeichnet) zurückgeführt werden. Sie wird angewendet, wenn keine konkreten Hypothesen über die dem Datensatz zugrunde liegenden Faktoren existieren (Moosbrugger & Kelava, 2007).

Damit eine Faktorenanalyse durchgeführt werden kann, muss eine Reihe von Voraussetzungen erfüllt sein. Für die Faktorenanalyse sind Daten von mindestens $N = 100$ Personen nötig, damit sie als ausreichend gelten (Bühner, 2011). Ein $N < 60$ wird als inkompatibel zur Durchführung einer Faktorenanalyse betrachtet. Grundsätzlich müssen zwischen den Items substanzielle Korrelationen vorliegen (Bühner, 2011). Dies kann mithilfe des Kaiser-Meyer-Olkin-Koeffizienten (KMO-Koeffizient) quantifiziert werden. Der Koeffizient kann Werte zwischen Null und Eins annehmen und gibt an, wie geeignet die Korrelationsmatrix für eine Faktorenanalyse ist. Je nach Wert des

Koeffizienten kann abgeleitet werden, ob eine Faktorenanalyse durchgeführt werden sollte oder nicht. Bei einem Wert kleiner .60 ist von einer Faktorenanalyse abzuraten. Tabelle 1 zeigt die Implikationen der Werte (Bühner, 2011).

Tabelle 1. KMO-Koeffizient und Implikationen für die Durchführbarkeit der Faktorenanalyse (Bühner, 2011, S. 347).

KMO- und MSA- Koeffizienten	Durchführbarkeit der Faktorenanalyse
< .50	Nicht durchführbar
.50-.59	Schlecht
.60-.69	Mäßig
.70-.79	Mittel
.80-.89	Gut
> .90	Sehr gut

Mit dem MSA-Koeffizienten (Measure of Sample Adequacy) werden ebenso wie mit dem KMO-Koeffizienten Itemkorrelationen betrachtet. Dies erfolgt jedoch auf Itemebene. Mit diesem Wert kann die Eignung jedes einzelnen Items für die Faktorenanalyse geprüft werden. Dafür werden die gleichen Werte herangezogen wie bei der KMO-Betrachtung (vgl. Tabelle 1). Ein dritter wichtiger Test ist der Bartlett-Test auf Sphärizität. Dieser testet die „globale Nullhypothese, dass alle Korrelationen der Korrelationsmatrix gleich null sind“ (Bühner, 2011, S. 347). Wenn er signifikant ist, sind alle Korrelationen größer Null und die Faktorenanalyse sinnvoll. Der Bartlett-Test erfordert normalverteilte Werte. Nach Bortz und Döring (2006) nähert sich eine Verteilung mit zunehmender Stichprobengröße an eine Normalverteilung an. Ab N = 30 kann von einer Normalverteilung ausgegangen werden. Zur Überprüfung kann auch der Kolmogoroff-Smirnov-Test eingesetzt werden. Sind alle Voraussetzungen erfüllt, kann eine Faktorenanalyse durchgeführt werden.

Je nach Fragestellung werden für die Faktorenanalyse verschiedene Festlegungen getroffen. Zuerst wird eine geeignete Extraktionsmethode gewählt, von denen die bekanntesten die Hauptkomponentenanalyse, die Hauptachsenanalyse und die Maximum Likelihood Faktorenanalyse sind. Am häufigsten wird die Hauptkomponentenanalyse (engl. principle component analysis, PCA) verwendet, die streng genommen keine Faktorenanalyse, sondern ein eigenständiges Verfahren ist. Aber aufgrund ihrer Datenreduktionsfunktion wird sie in der Literatur zusammen mit den Faktorenanalysen genannt (Bühner, 2011; Moosbrugger & Kelava, 2007). Die Hauptkomponentenanalyse zeichnet sich durch ihren Zweck aus, möglichst viel Varianz durch die beobachteten Variablen (also die Items) zu erklären. Dies geschieht durch die Linearkombination der Faktoren, den so genannten Hauptkomponenten (Moosbrugger & Kelava, 2007).

Anschließend wird die zu extrahierende Faktorenzahl festgelegt. Typischerweise wird der Eigenwerteverlauf als Maß herangezogen. Als Eigenwert wird die Höhe der durch jeweils einen Faktor erklärten Varianz aller beobachteten Variablen bezeichnet (Moosbrugger & Kelava, 2007). Entsprechend des Kaiser-Guttman-Kriteriums sind diejenigen Faktoren relevant, die einen Eigenwert größer 1 besitzen. Ein Eigenwert über 1 bedeutet, dass der Faktor mehr Varianz aufklärt als eine Variable allein. Bei einer großen Itemanzahl führt das Verfahren jedoch zu einer Überschätzung der Faktorenanzahl, weshalb häufig auf das zuverlässigere Verfahren der Betrachtung des grafischen Verlaufs des Scree-Plots (Scree-Test nach Cattell, 1966) zurückgegriffen wird (siehe Abbildung 6).

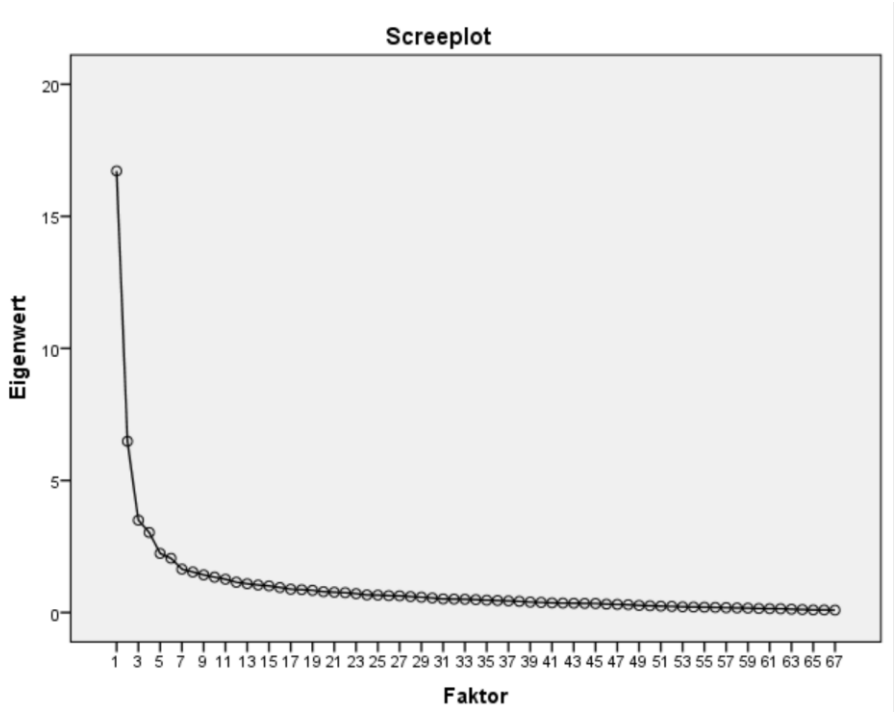


Abbildung 6. Beispiel für ein Screeplot, erstellt mit IBM SPSS 21.

Auf der Abszisse sind Faktoren und auf der Ordinate die Eigenwerte abgetragen. Die nach ihrer Höhe im Plot abgetragenen und zunächst steil abfallenden Eigenwerte der korrelierten Daten zeigen bei einem bestimmten Faktor einen deutlichen Knick, ab dem sich der Graph asymptotisch der Abszisse annähert, d.h. diese Werte stagnieren annähernd auf niedrigem Niveau und sind nicht weiter bedeutsam. Die Faktoren links von dem Knick sind relevant und in der Faktorenanalyse zu extrahieren (Moosbrugger & Kelava, 2007). Die Schwachstellen dieses Verfahrens sind die zum Teil schwierige Ableitung des relevanten Knicks und die häufig stattfindende Überschätzung der Faktorenanzahl (Bühner, 2011). Zwei sehr zuverlässige Methoden zur Bestimmung der Anzahl zu extrahierender Faktoren sind die Parallelanalyse nach Horn (1965), die Eigenwerte aus normalverteilten und nicht normalverteilten Zufallsgrößen bildet, und der Minimum-Average-Partial-Test

von Velicer (MAP-Test; 1976), der mit einer Residualmatrix[1] arbeitet. Für beide Verfahren hat O'Connor (2000) eine SPSS-Syntax veröffentlicht, mit der eine sehr einfache und ökonomische Ableitung der Faktorenanzahl möglich ist (vgl. Bühner, 2011). Da mit all diesen Methoden eine möglichst geringe Faktorenanzahl extrahiert werden soll, kann ein Varianzanteil nicht durch die Faktoren erklärt werden. Darin sind unsystematische Messfehler und systematische Anteile der nicht berücksichtigten Faktoren enthalten (Moosbrugger & Kelava, 2007).

Im letzten Schritt wird die Art der Faktorenrotation gewählt. Diese ist nötig, um die Lösung der Faktorenanalyse interpretierbarer zu machen. Dafür wird ein Ladungsmuster entsprechend einer Einfachstruktur angestrebt. Dabei soll jede Variable möglichst nur auf einem Faktor hoch (Primärladung) und auf den anderen Faktor nicht oder gering laden (Sekundärladung; Moosbrugger & Kelava, 2007). Grob wird zwischen orthogonaler und oblique Rotation unterschieden. Die orthogonale Rotation führt zu unkorrelierten Faktoren, wobei von den verschiedenen Rotationstechniken die Varimax-Rotation am häufigsten Anwendung findet. Im Gegensatz dazu korrelieren bei der oblique Rotation die Faktoren miteinander. Die bedeutendsten Rotationstechniken sind in diesem Fall die Direct-Oblimin- und die Promax-Rotation. Weil die Promax-Rotation zu gut interpretierbaren Lösungen führt, wird sie häufiger verwendet. Für die Entscheidung zwischen den Rotationsmethoden sollte die Faktorkorrelation betrachtet werden. Sind sie gering ($r < .10$) korreliert, kann die orthogonale Rotation durchgeführt werden. Andernfalls wird die oblique Rotation vorgenommen (Moosbrugger & Kelava, 2007).

Bei der Interpretation der Faktoren sollten die Kommunalitäten und die Faktorladungen eine Rolle spielen. Die Faktorladung ist die Korrelation eines Items mit einem Faktor und beschreibt damit deren Beziehung (Backhaus, Erichson, Plinke & Weiber, 2011). Je höher die Korrelation ist, desto eher gehören der Faktor und das Item auch theoretisch zusammen. Items, die sehr niedrig auf einem Faktor oder hoch auf mehreren Faktoren laden, sollten ausgeschlossen werden. Insgesamt sollten aber inhaltliche Überlegungen im Vordergrund stehen (Bühner, 2011). Die Kommunalität wird durch die Summe der quadrierten Faktorladungen eines Items repräsentiert und erklärt, in welchem Maße die Varianz eines Items durch alle Faktoren aufgeklärt wird. Je höher die Kommunalität ist, desto besser wird das Modell durch das Item erklärt (Backhaus et al., 2011).

2.2.3. Reliabilität

Die Reliabilität ist ein Maß zur Validierung eines Messverfahrens und gibt die Zuverlässigkeit des Verfahrens an. Reliabilität betrifft die Genauigkeit der Messung eines Merkmals (Field, 2009; Lie-

[1] Das Residuum beschreibt den Unterschied zwischen einem Wert, der aufgrund einer Regressionsgerade vorhergesagt wurde, und dem tatsächlichen Wert (Field, 2009).

nert & Raatz, 1998). Das bedeutet zum Beispiel einerseits, dass eine Person im selben Test zu zwei Messzeitpunkten ein gleiches bzw. sehr ähnliches Ergebnis erzielen sollte und andererseits, dass zwei Personen, die eine ähnliche Merkmalsausprägung besitzen, auch ähnliche Testwerte aufweisen sollten, damit ein Test als reliabel betrachtet werden kann (Field, 2009). Dies ist unabhängig davon, ob das gemessene Merkmal auch das Konstrukt widerspiegelt, das es messen soll (Lienert & Raatz, 1998). Es existieren verschiedene Methoden, um die Reliabilität eines Tests zu überprüfen.

Die Retest-Reliabilität geht wie im erstgenannten Beispiel davon aus, dass dieselbe Person zu zwei verschiedenen Messzeitpunkten denselben Fragebogen ausfüllt und damit ähnliche bis gleiche Testergebnisse erzielt werden. Diese beiden Werte sollten deshalb hoch korrelieren. Da es sich jedoch als schwierig und zeitlich sowie finanziell aufwendig erweist, einen Probanden zu zwei Testzeitpunkten zu erfassen, wird dieses Verfahren selten angewandt (Lienert & Raatz, 1998).

Bei dem Paralleltest-Verfahren werden zwei vergleichbare Versionen eines Tests erstellt und deren Ergebnisse korreliert (Lienert & Raatz, 1998).

Die Split-Half-Methode (Testhalbierungs-Methode) ähnelt dem Paralleltest-Verfahren, nur dass auf eine Version des Tests zurückgegriffen wird. Der Test wird so in zwei Hälften aufgeteilt, dass eine gleiche Anzahl möglichst ähnlicher Items in beiden Teilen vorliegt. Das Testergebnis der Stichprobe wird für beide Teile gesondert berechnet und anschließend korreliert. Die beiden Hälften sollten möglichst hoch bis perfekt korrelieren (Field, 2009; Lienert & Raatz, 1998). Problematisch sind die vielen verschiedenen Möglichkeiten der Halbierung des Tests, sodass unterschiedliche Reliabilitäts-Werte zustande kommen können (Lienert & Raatz, 1998).

Der am häufigsten verwendete Reliabilitäts-Koeffizient ist Cronbach's α. Damit lässt sich die interne Konsistenz, d.h. der wechselseitige Zusammenhang der Items, bestimmen (Cortina, 1993; Cronbach, 1951; Field, 2009). Cronbach's α hängt mit der Split-Half-Reliabilität zusammen und stellt den Mittelwert aller Split-Half-Reliabilitäten dar (Cronbach, 1951). Die Maßzahl Cronbach's α ist stabiler als die anderen Testhalbierungs-Reliabilitäten, da diese von der Art und Weise der Halbierung beeinflusst sind (Cortina, 1993). Cronbach's α lässt sich mit Formel 2 berechnen, die Unterschiede der Standardabweichungen der Items berücksichtigt.

Formel 2. Berechnung von Cronbach's α (Cortina, 1993, S. 99).

$$\frac{N^2 \times \overline{cov}}{\sum \frac{var}{cov}}$$

N^2: quadrierte Anzahl der Items der Skala

$\overline{cov}$: mittlere Interitem-Kovarianz

$\sum \frac{var}{cov}$: Summe aller Elemente in der Varianz-Kovarianz-Matrix

Der Koeffizient wird von der Testlänge, d.h. der Anzahl der Items, beeinflusst. Mit zunehmender Testlänge nimmt die Reliabilität zu (Schmitt, 1996). Auf Grund dieser Beeinflussbarkeit und des Fehlens eines einheitlichen Cut-Off-Wertes für die Gültigkeit eines Tests als reliabel, sollte die Interpretation von Cronbach's α laut Schmitt (1996) immer zusammen mit weiteren Angaben zu den Items, wie Interkorrelationen, erfolgen.

2.3. Review ausgewählter Usability- und User Experience-Fragebögen

Modelle stellen eine theoretische, auf das Wesentliche reduzierte Betrachtung der Wirklichkeit dar (Brockhaus, 2006). Wie in Abschnitt 2.1.2 erwähnt, erweist sich die empirische Absicherung holistischer Modelle wegen ihres ganzheitlichen Charakters als schwierig bis unmöglich. Reduktionistische Modelle dagegen erheben den Anspruch einer empirischen Verankerung. In mehreren Studien wurde bisher ein Unterschied zwischen *instrumentellen / pragmatischen* und *nicht-instrumentellen / hedonischen Produktqualitäten* gefunden (Hassenzahl, Kekez & Burmester, 2002; Mahlke, 2008; Mahlke & Thüring, 2007; Thüring und Mahlke, 2007; vgl. Scapin et al., 2012). Für empirische Betrachtungen ist es üblich, aus dem Modell einen Fragebogen abzuleiten. Damit lassen sich quantitative Aussagen über das zu erhebende Konstrukt und das Modell bewertbar machen (Hassenzahl, Burmester & Koller, 2008).

Anders herum betrachtet sollte jeder entwickelte Fragebogen auf einem Modell basieren, das die erhobenen Daten erklärt. In diesem Abschnitt werden Fragebögen vorgestellt, die gegenwärtig in der Forschung und in der Praxis zum Einsatz kommen. Dabei wird das jeweils zugrunde liegende Modell kritisch untersucht, sofern eine theoretische Grundlage überhaupt vorliegt, und eventuelle Probleme des Fragebogens aufgezeigt. Tabelle 2 liefert eine Übersicht über verschiedene User Experience-, Usability-, Emotions- und sonstige relevante Fragebögen. Die Aufstellung ist nicht abschließend. Fragebögen, die entweder nur einen sehr eingeschränkten Anwendungskontext besitzen, nur aus einem einzigen Item bestehen (zum Beispiel die Beanspruchungsskala SEA von Eilers, Nachreiner & Hänecke, 1986) oder bei denen die empirische Absicherung fehlt oder fraglich ist, sind nicht aufgeführt.

Im Folgenden wird auf zwei User Experience-Fragebögen näher eingegangen, die sowohl in der Forschung als auch in der Praxis große Relevanz erlangt haben. Dies sind der AttrakDiff-2 (Hassenzahl, Burmester & Koller, 2003) und der User Experience Questionnaire (UEQ; Laugwitz, Schrepp & Held, 2006). Stellvertretend für die Gruppe der Usability-Fragebögen wird der in Deutschland häufig angewandte ISONORM 9241/10 (Prümper & Anft, 1993) vorgestellt. Darüber hinaus kommen bei der Bewertung von Produkten häufig Emotions-Fragebögen zum Einsatz. In Deutschland werden der Self-Assessment Manikin (SAM; Lang, 1980) und der Positive and Negative Affect Schedule (PANAS; Tellegen, Watson & Clark, 1988; deutsche Übersetzung Krohne, Egloff, Kohlmann, Tausch, 1996) sehr häufig verwendet, wovon der SAM (Lang, 1980) vorgestellt wird.

Tabelle 2. Bekannte Fragebögen, unterteilt in die Bereiche User Experience, Usability, Emotionen, Sonstige.

	Fragebogen	Modell	Skala	Dimensionen	Anzahl Items	Probleme / Besonderheiten
User Experience	AttrakDiff-2 (Hassenzahl, Burmester & Koller, 2003)	Hedonisches / pragmatisches Modell (Hassenzahl, 2007)	Semantisches Differential	Pragmatische Qualität Hedonische Qualitäten - Stimulation - Identität Attraktivität	28	Faktorenanalyse mit 22 Personen; Nur Websites untersucht
	User Experience Fragebogen (Mahlke, 2005)	Vorentwicklung zum CUE-Modell (Thüring & Mahlke, 2007)	Ratingskala	Nützlichkeit Ease of Use Hedonische Qualität Visuelle Attraktivität Nutzungsintention	22	Nur Websites
	User Experience Questionnaire (UEQ; Laugwitz, Schrepp und Held, 2006)	Hedonisches / pragmatisches Modell (Hassenzahl; 2008)	Semantisches Differential	Attraktivität Durchschaubarkeit Effizienz Verlässlichkeit Stimulation Originalität	26	Nur Software; Kaum Modellbeschreibung
Usability / v.A. instrumentelle Qualitäten	IsoMetrics (Willumeit, Gediga & Hamborg, 1996)	DIN EN ISO NORM 9241-110	Ratingskala	Aufgabenangemessenheit Selbstbeschreibungsfähigkeit Steuerbarkeit Erwartungskonformität Fehlertoleranz Individualisierbarkeit Lernförderlichkeit	75	Kurz- (summative Evaluation) und Langform (formative Evaluation)
	ISONORM 9241/10 (Prümper & Anft, 1993)	DIN EN ISO NORM 9241-110 (2006)	Semantisches Differential	Aufgabenangemessenheit Selbstbeschreibungsfähigkeit Steuerbarkeit Erwartungskonformität Fehlertoleranz Individualisierbarkeit Lernförderlichkeit	35	Nur Software-Usability; Kurzform vorhanden

	Fragebogen	Modell	Skala	Dimensionen	Anzahl Items	Probleme / Besonderheiten
Usability / v.A. instrumentelle Qualitäten	Questionnaire for User Interaction Satisfaction (QUIS; Chin, Diehl & Norman, 1988)	DIN EN ISO 9241-110 (2006) , DIN EN ISO 9241-12 (1998)	Semantisches Differential	screen factor terminology and system feedback learning factors system capabilities technical manuals on-line tutorials multimedia teleconferencing software installation	27	
	Software Usability Measurement Inventory (SUMI; Kirakowski, 1994)	Nicht vorhanden	Ratingskala	Effizienz Affekt Hilfe und Unterstützung Kontrollierbarkeit Erlernbarkeit	50	Nur eine Skala zu Emotionen
	System Usability Scale (SUS; Brooke, 1986)	Nicht vorhanden	Ratingskala	Globalbewertung von Usability	10	5 positive und 5 negative Usability-Items
Emotionen	Affect Grid (Russell, 1989)	Circumplex Modell (Russell, 1979)	Gitter-Struktur	Arousal Valenz	1	
	LEMtool (Capota, van Hout & van der Geest, 2007)	Nicht vorhanden	Semantisches Differential	Liking: Joy, Sadness Ästhetik: Desire, Disgust Ästhetik / Usability: Fascination, Boredom Usability: Satisfaction, Dissatisfaction	8	Nur Web und Media; Rein piktoral
	Positive and Negative Affect Schedule (PANAS; Tellegen, Watson & Clark, 1988; deutsche Übersetzung Krohne, Egloff, Kohlmann, Tausch, 1996)	Positive Affect-/Negative Affect-Model (Watson & Tellegen, 1985)	Ratingskala	Positiver Affekt Negativer Affekt	20	
	Self-Assessment Manikin (SAM; Lang, 1980)	Drei-Faktoren-Theorie der Emotionen (Russell & Mehrabian, 1977)	Ratingskala	Arousal Valenz Dominanz	3	Rein piktoral
	Semantic Differential Scale (Mehrabian & Russell, 1974)	Drei-Faktoren-Theorie der Emotionen (Russell & Mehrabian, 1977)	Semantisches Differential	Arousal Valenz Dominanz	18	

	Fragebogen	Modell	Skala	Dimensionen	Anzahl Items	Probleme / Besonderheiten
Sonstige	Centrality of visual product aesthetics (CVPA; Bloch, Brunel & Arnold, 2003)	Nicht vorhanden	Ratingskala	Value acumen response determinancy	11	Visuelle Ästhetik von Produkten
	INTUI (Ullrich & Diefenbach, 2010)	Nicht vorhanden	Semantisches Differential	Mühelosigkeit Bauchgefühl Magisches Erleben Verbalisierungsfähigkeit	17	Erfassung intuitiver Nutzung
	Questionnaire for Intuitive Use (Quesi; Hurtienne & Naumann, 2010)	Image schemas and design for intuitive use (Hurtienne, 2010)	Ratingskala	Subjective mental workload Perceived achievement of goals Perceived effort of learning Familiarity Perceived error rate	14	Basis Usability Erfassung intuitiver Nutzung
	Visual aesthetics of websites inventory (VisAWI; Moshagen & Thielsch, 2011)	Nicht vorhanden	Ratingskala	Einfachheit Vielseitigkeit Farbigkeit Kunstfertigkeit	18	Ästhetik von Webseiten Kurzversion verfügbar

2.3.1. AttrakDiff

Der von Hassenzahl, Burmester und Koller (2003) entwickelte AttrakDiff-2 ist die Weiterentwicklung des AttrakDiff-1 (Hassenzahl, 2002) und basiert auf dem Modell hedonischer und pragmatischer Qualitäten von Hassenzahl (2007), das in Abschnitt 2.1.2 beschrieben wurde.

Der AttrakDiff-1 war in Form eines semantischen Differentials konstruiert und bestand aus 23 siebenstufigen Adjektivpaaren sowie den drei Skalen pragmatische Qualität, hedonische Qualität und Attraktivität. Hassenzahl et al. (2003) zeigen, dass pragmatische und hedonische Qualitäten konsistent sind und unabhängig voneinander wahrgenommen werden. Nachteilig am AttrakDiff-1 war den Autoren zufolge die fehlende Trennbarkeit der hedonischen Qualitäten Stimulation und Identifikation. Den Autoren zufolge wurden die Aspekte ursprünglich gemeinsam definiert, wobei sie keinen Wert auf die Trennschärfe der Adjektivpaare legten. Da die Trennschärfe aber angibt, wie gut ein einzelnes Item den Gesamtwert eines Fragebogens vorhersagt (Bortz & Döring, 2006), ist die Trennschärfeberechnung kein Maß für die Unterscheidung von Skalen. Es ist somit nicht nachvollziehbar, was die Autoren in diesem Zusammenhang mit „Trennschärfe“ meinen. Um Simulation und Identität in getrennten Skalen darstellen zu können, wurde der AttrakDiff-2 entwickelt.

Auch der AttrakDiff-2 ist in Form eines semantischen Differentials mit siebenstufigen Adjektivpaaren aufgebaut, wobei jeweils sieben Items eine Skala bilden. Somit umfasst der AttrakDiff-2 insgesamt 28 Items. Die pragmatische Qualität bezieht sich auf nützliche und benutzbare Funktionen eines Produkts. Sie unterscheidet damit im Vergleich zum theoretischen Modell von Hassenzahl (2007) nicht zwischen Utility und Usability. Hedonische Qualitäten sind nun in Stimulation und Identität geteilt. Stimulation bezieht sich auf Produkteigenschaften, die mit der Fertigkeitsverbesserung zusammen hängen. Identität bezieht sich auf die Kommunikation mit „relevanten Anderen“ (vgl. Abschnitt 2.1.2; Hassenzahl, Burmester & Koller, 2008). Die vierte Skala Attraktivität beschreibt globale positiv - negativ Bewertungen des Produkts. Durch die Abfrage abstrakter Attribute, wie z.B. harmlos - herausfordernd, ist den Autoren zufolge ein breiter Anwendungskontext gegeben. Hinzu kommt, dass der Fragebogen im Internet[2] kostenlos zur Anwendung zur Verfügung steht, und der Fragebogen bei einem Nutzertest auch online ausgefüllt wird (zum Zeitpunkt der Erstellung dieser Abhandlung). Dadurch verfügen die Autoren über einen großen Datenpool (mehr als 2300 Beurteilungen in 302 Projekten, Stand 2007; Hassenzahl, Burmester & Koller, 2008).

[2] http://www.attrakdiff.de [06.03.2013]

Der AttrakDiff-2 wurde in drei Schritten entwickelt. In einem Expertenworkshop wurden zuerst Items in einer freien Explorationsphase gesammelt, von denen während der anschließenden Bewertungsphase 50 passende Items ausgewählt wurden. Diese bildeten das Ausgangsmaterial für den Fragebogen. Der zweite Schritt umfasste eine empirische Erhebung von Daten mittels der gewählten Items, die Hassenzahl und Kollegen (2008) „Pilotanwendung“ nennen. Die Ergebnisse sind in Tabelle 3 zusammengefasst. 22 Personen, die in zwei Gruppen eingeteilt wurden, sollten drei verschiedene reale Websites von Alkoholherstellern testen. Eine Gruppe sollte unter einer bestimmter Aufgabenstellung und die andere ohne Aufgabe auf den Webseiten navigieren. Anschließend waren sie aufgefordert die 50 Items auf ihre Erfahrung mit der jeweiligen Webseite hin zu beantworten. Mit den erhobenen Daten wurde eine Hauptkomponentenanalyse durchgeführt. Die sieben Items der Skala Attraktivität wurden aus nicht genannten Gründen aus der Analyse ausgeschlossen und ohne weitere Überprüfung aus dem AttrakDiff-1 in den neuen Fragebogen übernommen. Da die Items für die Attraktivität somit durch eine andere Stichprobe und abhängig von den Items aus dem AttraDiff-1 gewonnen wurden, ist dieses Vorgehen als kritisch zu betrachten. Für die übrigen Items extrahierten die Autoren drei Komponenten, die varimax-rotiert wurden. Diese Art der Rotation wendeten sie an, weil der Fragebogen die drei Skalen möglichst unabhängig messen sollte (Hassenzahl et al., 2003). Wie hoch die Komponenten tatsächlich korrelieren, wird in der ersten Studie berichtet, die mit dem neuen Fragebogen durchgeführt wurde. Die beiden hedonischen Skalen korrelieren mit .55 deutlich. Die Autoren erklären dies mit deren inhaltlichen Verbundenheit.

Anschließend wurden Items entfernt, die inhaltlich oder nach statistischen Gesichtspunkten nicht in die jeweiligen Komponenten passten. Die 21 verbliebenen Items klären 72% der aufgetretenen Varianz auf. Da die Attraktivität für diese Analysen ausgeschlossen wurde, bleibt offen, ob dieser Faktor einen Einfluss auf die aufgeklärte Gesamtvarianz hat. Es wird ebenfalls nicht berichtet, wie viel Varianz der Faktor Attraktivität einzeln betrachtet aufklärt.

Der dritte Schritt der Konstruktion bestand aus einer Studie mit dem neu entwickelten AttrakDiff-2. Hierfür wurden 33 Psychologiestudenten zu einer Software befragt, die Musik abspielen kann. Für die Benutzeroberfläche dieser Software können verschiedene vom Hersteller bereitgestellte Designs ausgewählt werden. Die Autoren haben vier verschiedene Designs vorgegeben. Die Stichprobe unterscheidet sich deutlich von der ersten, indem sie wesentlich homogener ist. Zum einen handelt es sich um eine sehr junge und rein akademische Gruppe (M = 24 gegenüber zuvor M = 38), außerdem ist das Geschlecht nicht ausgewogen (28 Frauen, 5 Männer). Ebenfalls kritisch ist das erneute Fehlen der Betrachtung der Skala Attraktivität.

Mit den Daten aus dieser Studie wurde die interne Konsistenz mittels Cronbach's α bestimmt. Dafür wurde ein von der Norm abweichendes Vorgehen gewählt. Hassenzahl et al. (2003) berichten, pro Skala Cronbach's α für jedes der vier Designs ermittelt zu haben und daraus einen durch den höchsten und niedrigsten Wert abgegrenzten Bereich angegeben zu haben. Die Reliabilitäten liegen zwischen .73 und .90.

Da von den Autoren die fehlende Trennschärfe-Berechnung für den Attrak-Diff-1 kritisiert wurde, ist kritisch anzumerken, dass die Trennschärfen für den AttrakDiff-2 nicht erwähnt wurden.

Der entscheidendste Kritikpunkt an diesem Vorgehen ist aus empirischer Sicht in der Anzahl der Versuchspersonen für die Faktorenanalyse zu sehen. Nach Bühner (2011) sollten mindestens Daten von $N = 100$ Personen erhoben werden um als ausreichend zu gelten. Ein $N < 60$ wird als inkompatibel zur Durchführung einer Faktorenanalyse betrachtet (vgl. Abschnitt 2.2.2.). Hassenzahl et al. (2003) verwendeten jedoch als Grundlage die Daten von 22 Versuchspersonen. Zudem wurden keine Kommunalitäten berichtet, die nach Bühner (2011) insbesondere bei kleinen Stichproben angegeben werden sollten.

Mit der Attraktivität, der pragmatischen und der hedonischen Qualität haben die Autoren einen Teil von Hassenzahls (2008) Grundmodell zur pragmatischen und hedonischen Qualität im AttrakDiff-2 umgesetzt (vgl. Abbildung 3). Items zu emotionalen und Verhaltens-Konsequenzen wurden nicht betrachtet. Diefenbach und Hassenzahl (2010) erklären den Mehrwert des Einsatzes des Emotionsfragebogens PANAS (Tellegen, Watson & Clark, 1988) zusätzlich zum AttrakDiff-2, weil er mediierende Faktoren von Emotionen auf die Produktbewertung aufdecken und beschreiben könne. Eine Integration einer eigenen Emotionsskala in den AttrakDiff-2 schlagen sie jedoch nicht vor.

Seit der Entwicklung des Attrak-Diff-2 im Jahr 2003 gab es keine Revision. 2008 geben Hassenzahl, Burmester und Koller an, dass sie eine Kurzform des AttrakDiff-2 mit nur 10 Items planen. Studien, in denen der „AttrakDiff mini" eingesetzt wurde, zeigen zwar, dass diese Kurzform existiert (z.B. Drexler, 2012; Sieger, Kirschnick & Möller, 2010), doch bisher fand der Fragebogen nur in einer Studie zur Untersuchung des Zusammenhangs von „beauty" und Usability unter Verwendung einer US-amerikanischen Stichprobe durch den Autor Hassenzahl selbst Verwendung (Hassenzahl & Monk, 2010). Die verwendete Kurzform besteht aus 8 Items und wird in englischer Übersetzung angewendet. Interessant wäre eine erneute Prüfung der Langversion und der Kurzversion des Fragebogens mit der vermutlich großen Anzahl an online erhobenen Daten.

Tabelle 3. Kenntwerte zur Konstruktion des AttrakDiff-2, Daten basierend auf Hassenzahl, Burmester und Koller (2003).

Versuchs-teilnehmer	Alter	Stimulus	Itemkennwerte	Faktorenanalyse	Aufgeklärte Gesamt-varianz	Cronbach's Alpha
22 (9 Frauen, 13 Männer) Aufgeteilt in zwei Gruppen mit unterschiedlichen Aufgaben	M = 38 MIN: 23, MAX: 59	Drei Websites verschiedener Alkoholmarken	Nicht berichtet	Hauptkomponenten-analyse, drei Kompo-nenten extrahiert, varimax-rotiert	72%	Skalen-weise

2.3.2. User Experience Questionnaire (UEQ)

Laugwitz, Schrepp und Held (2006; vgl. Laugwitz, Held & Schrepp, 2008) wollten einen Fragebogen entwickeln, in dem der Nutzer eine Software und die Interaktion mit ihr beurteilt. Den Autoren zufolge reicht die Messung von Usability nicht aus, weshalb sie eine Erweiterung durch Kriterien des Nutzungserlebens in Betracht ziehen. Daraus leitet sich das Ziel „die schnelle Erhebung eines umfassenden Gesamteindrucks der User Experience“ ab (Laugwitz et al., 2006, S. 126). Die Autoren kritisieren an dieser Stelle einerseits den AttrakDiff (Hassenzahl et al., 2003), der ihrer Meinung nach zu viel Wert auf hedonische Aspekte legt und andererseits den SUMI (Kirakowski, 1994), der zu wenig Wert auf hedonische Aspekte legt, da nur eine seiner fünf Skalen emotionale Aspekte erfasst. Der Nutzer soll unmittelbar und auf einfache Weise seine Rückmeldung mittels des Fragebogens geben, der zudem nicht als einzige Informationsquelle dienen soll. Im Normalfall soll der Fragebogen Nutzertests ergänzen, weshalb der Fokus auf subjektiven Aspekten liegen könne.

Laugwitz et al. (2006) streben einen ähnlichen Fragebogen wie den AttrakDiff an, weshalb sie auch ihr methodisches Vorgehen von Hassenzahl et al. (2003; siehe Abschnitt 2.3.1) ableiten. Dabei erklären sie aber weder, welchen Mehrwert ihr eigener Fragebogen gegenüber dem von Hassenzahl et al. (2003) bietet, noch gehen sie ad-hoc auf ein Modell ein, das ihrer Empirie zugrunde liegt. Erst post-hoc ordnen Laugwitz und Kollegen ihren Fragebogen in das Rahmenmodell von Hassenzahl (2008) ein.

Zur Konstruktion des Fragebogens haben die Autoren mithilfe von Expertenbefragungen zunächst einen Itempool generiert, der aus für User Experience charakteristischen Begriffen besteht. Daraus wurden 80 Adjektive identifiziert und im Format eines siebenstufigen semantischen Differentials zusammengetragen. Aus dieser vorläufigen Form wurden zwei Fragebögen mit verschiedenen Itemreihenfolgen erstellt (Versionen A und B), mit denen sechs verschiedene Unter-

suchungen zur Datengewinnung durchgeführt wurden. An den Untersuchungen nahmen insgesamt 153 Versuchspersonen teil. Über die Probanden wird wenig berichtet. Die Autoren schreiben lediglich, dass es sich überwiegend um studentische Stichproben handelt, die in verschiedenen Studien mithilfe des entwickelten Fragebogens Software-Anwendungen bewerten sollten.

Zur Itemreduktion wurden die Items zunächst in zwei Gruppen aufgeteilt. Die erste Gruppe bestand aus 14 Items, die die Zustimmung oder Ablehnung von Attraktivitätsaspekten messen, dabei aber keine inhaltlichen Aspekte beachten. Die zweite, mit 66 Probanden wesentlich größere Gruppe, enthielt „alle restlichen Items“ (Laugwitz et al., 2006). Mit beiden Gruppen wurde eine nicht näher beschriebene Faktorenanalyse durchgeführt (vgl. Tabelle 4). Für die erste Gruppe ergab sich ein Faktor, auf dem alle Items luden, und mit dem 60% Varianz aufgeklärt werden konnte. Die Autoren wählten sechs Items aus und nannten den Faktor *Attraktivität*. Die Faktorenanalyse der zweiten Gruppe ergab fünf Faktoren, die 53% der Varianz aufklärten. Die Faktorbezeichnungen richteten sich nach den am höchsten ladenden Items. Diese ordnen die Autoren aufgabenorientierten und nicht aufgabenorientierten Aspekten zu. Zu den aufgabenorientierten Aspekten zählen Durchschaubarkeit, Verlässlichkeit, Effizienz und zu den nicht-aufgabenorientierten Aspekten Originalität und Stimulation. Je Faktor wurden vier Items ausgewählt. Die Autoren erklären, dass die Items des Faktors Attraktivität erwartungsgemäß hoch auf allen anderen Faktoren laden. Die Erwartung wurde jedoch vorher nicht genannt und ist auch nicht durch Daten veranschaulicht. Kritisch ist zudem die zum AttrakDiff-2 äquivalente Bezeichnung der Dimensionen Attraktivität und Stimulation. Positiv ist die Differenzierung der zu Hassenzahls, Burmesters und Kollers (2003) pragmatischen Qualität äquivalenten aufgabenorientierten Aspekte in einzelne Dimensionen, was eine gezieltere Betrachtung als bei Hassenzahl et al. (2003) ermöglicht.

Die Validierung des Fragebogens erfolgte mittels zweier Usability-Studien. In der ersten Studie mit 13 Teilnehmern wurden Zusammenhänge zwischen den Skalen und Bearbeitungszeiten sowie die Reliabilität mittels Cronbach’s α untersucht. Die α-Werte wurden skalenweise angegeben und sind zwar mit Werten zwischen .65 und .89 in einem aussagefähigen Bereich, jedoch ist die Aussagekraft der Studie durch die geringe Anzahl von nur 13 Versuchsteilnehmern wiederum eingeschränkt.

Mithilfe der zweiten Usability-Studie mit 15 Versuchsteilnehmern haben die Autoren die Validität des UEQ geprüft. Die Validität als Gütekriterium gibt die Genauigkeit an, mit der das Verfahren tatsächlich das misst, was es zu messen vorgibt (Field, 2009; Lienert & Raatz, 1998). Es werden verschiedene Validitätsarten unterschieden. Bei der äußeren kriterienbezogenen Validität wird das Testergebnis mit einem so genannten unabhängigen Außenkriterium verglichen. Bei ausrei-

chend hohen Korrelationswerten kann von einem validen Testverfahren gesprochen werden (Lienert & Raatz, 1998). Für die innere kriterienbezogene Validität wird das Messverfahren mit einem bereits existierenden Fragebogen für ein identisches Konstrukt verglichen. Laugwitz et al. (2006) haben als Vergleichsverfahren den AttrakDiff-2 (Hassenzahl et al. 2003) verwendet. Entsprechend der Hypothese der Autoren korrelieren die Attraktivitätsskalen beider Fragebögen mit .75 hoch. Die pragmatischen Skalen des AttrakDiff-2 und des UEQ korrelierten ebenso mäßig bis hoch miteinander und eher gering mit den hedonischen Skalen. Die hedonischen Skalen korrelieren zum Teil nur mäßig. Insgesamt leiten die Autoren aber ab, dass der UEQ valide sei.

Der UEQ wurde in den letzten Jahren in einer englischen (Laugwitz, Held & Schrepp, 2008) und einer spanischen Version (Rauschenberger, Schrepp, Olschner, Thomaschewski & Cota, 2012) veröffentlicht. Zudem liegt eine Version des Fragebogens in vereinfachter Sprache, also ohne Fachbegriffe und Fremdworte, vor, mit der User Experience insbesondere bei Jugendlichen erfasst werden kann (Hinderks, Schrepp, Rauschenberger, Olschner & Thomaschewski, 2012).

Tabelle 4. Kenntwerte zur Konstruktion des User Experience Questionnaire, Daten basierend auf Laugwitz, Schrepp und Held (2006).

Versuchsteilnehmer	Alter	Stimulus	Itemkennwerte	Faktorenanalyse	Aufgeklärte Gesamtvarianz	Cronbach's Alpha
Version 1: 76 Version 2: 77	Nicht berichtet	Sechs verschiedene Software-Produkte	Nicht berichtet	Nicht berichtet	„Attraktivität“:60% Restliche Faktoren: 53%	skalenweise

2.3.3. ISONORM 9241/10

Neben den vorgestellten User Experience Fragebögen existieren auch Fragebögen, die nur wahrgenommene Usability erfassen. Stellvertretend wird im Rahmen dieser Ausführungen der ISONORM 9241/10, erstmals veröffentlicht von Prümper und Anft (1993), vorgestellt, der zur Ergonomiebewertung von Software und Softwareprototypen und als Grundlage für Untersuchungen in Nutzergruppen entwickelt wurde. Als theoretische Basis diente die ISO Norm 9241-10 (DIN EN ISO 9241-10, 1998), die inzwischen in überarbeiteter Form als DIN EN ISO 9241-110 (2006) vorliegt, und „Grundsätze der Dialoggestaltung“ beschreibt. Sie hängt mit der DIN EN ISO 9241-11 zusammen, in der Gebrauchstauglichkeit definiert ist. Die Prinzipien der Dialoggestaltung stellen eine Grundlage der Gebrauchstauglichkeit dar (DIN EN ISO 9241-110, 2006). Prümper und Anft (1993) fassen verschiedene Probleme bisheriger Fragebögen zusammen. Am gravierendsten ist das „Defizit an wissenschaftlich fundierten und gleichzeitig praxisgerechten Evaluationsverfahren“ (Prümper & Anft, 1993, S. 146) einzustufen, weshalb sie ihren Fragebogen anhand der ISO Norm

9241-10 (1998) operationalisieren. Zusätzlich kritisieren sie die fehlende Absicherung durch Gütekriterien, fehlende Robustheit und hohe Preise für die Anschaffung von Fragebögen.

In der Norm werden die sieben Grundsätze der Dialoggestaltung, vorgestellt, beschrieben und mit Beispielen versehen (DIN EN ISO 9241-110, 2006). Diese Grundsätze sind: Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit und Lernförderlichkeit. Zu jedem Gestaltungsziel haben Prümper und Anft (1993) fünf Items in Form semantischer Differentiale mit sieben Skalenstufen entwickelt, die einen Handlungscharakter aufweisen. Die Stufen kodieren sie von --- über *0* (null) bis *+++*. Die Grundsätze mit je einem Beispiel aus dem ISONORM 9241/10 sind in Tabelle 5 dargestellt.

Tabelle 5. Grundsätze der Dialoggestaltung (DIN EN ISO 9241-110, 2006), Beispiel-Items aus dem ISONORM 9241/10.

Grundsatz	**Beispiel aus dem ISONORM 9241/10**		
	Die Software …		*Die Software …*
Aufgabenangemessenheit	ist kompliziert zu bedienen.	-	ist unkompliziert zu bedienen.
Selbstbeschreibungsfähigkeit	bietet einen schlechten Überblick über ihr Funktionsangebot.	-	bietet einen guten Überblick über ihr Funktionsangebot.
Steuerbarkeit	erzwingt eine unnötig starre Einhaltung von Bearbeitungsschritten.	-	erzwingt keine unnötig starre Einhaltung von Bearbeitungsschritten.
Erwartungskonformität	erschwert die Orientierung, durch eine uneinheitliche Gestaltung.	-	erleichtert die Orientierung, durch eine einheitliche Gestaltung.
Fehlertoleranz	ist so gestaltet, daß kleine Fehler schwerwiegende Folgen haben können.	-	ist so gestaltet, daß kleine Fehler keine schwerwiegende Folgen haben können.
Individualisierbarkeit	läßt sich von dem Benutzer schwer erweitern, wenn für ihn neue Aufgaben entstehen.	-	läßt sich von dem Benutzer leicht erweitern, wenn für ihn neue Aufgaben entstehen.
Lernförderlichkeit	erfordert viel Zeit zum Erlernen.	-	erfordert wenig Zeit zum Erlernen.

Die konkreten Schritte der Konstruktion des ISONORM 9241/10 werden durch die Autoren nicht beschrieben. Prümper (1997) stellt aber die Reliabilität, Validität und Normierung vor. Zu ihrer Bestimmung wurden die Daten von 1265 Personen, die 178 verschiedene Software-Produkte beurteilt haben, ausgewertet. Alle Probanden hatten bereits Erfahrung mit dem zu beurteilenden Produkt, im Mittel 25 Monate. Der Autor hat zwei Reliabilitätsverfahren eingesetzt. Die Retest-Reliabilität (vgl. Abschnitt 2.2.3) spiegelt sich darin wider, wie hoch zwei Messzeitpunkte korrelieren. Problematisch ist dieses Verfahren allerdings, wenn instabile oder zeitabhängige Merkmalen gemessen werden. Da Einschätzungen zur Dialoggestaltung, wie Usability und User Experience, mindestens durch den Nutzungskontext, aber auch durch Emotionen oder Erfahrungen, beein-

flusst sein können, kann eine niedrige Retest-Reliabilität auch auf diese äußeren Faktoren zurückzuführen sein und muss nicht zwangsläufig einen Mangel im Fragebogen selbst darstellen. Prümper (1997) hat die Retest-Reliabilität anhand einer Stichprobe von N = 49 Nutzern bestimmt und mit Werten zwischen .59 (Lernförderlichkeit) und .68 (Fehlertoleranz) akzeptable Zusammenhänge gefunden. Die innere Konsistenz wurde mit Cronbach's α bestimmt (vgl. Abschnitt 2.2.3). Die Überprüfung erfolgte skalenweise und ergab aussagekräftige Werte zwischen .81 (Aufgabenangemessenheit) und .89 (Individualisierbarkeit). Figl (2009) findet bei einer erneuten Betrachtung der Reliabilität ebenfalls annehmbare bis gute Werte zwischen .54 und .91 für die einzelnen Skalen.

Prümper (1997) setzte zur Prüfung der inneren kriterienbezogenen Validität die Usability-Fragebögen QUIS (Chin, Diehl & Norman, 1988) und BBD (Spinas, 1987) ein und verglich deren Gesamtwerte mit dem Gesamtwert des ISONORM 9241/10, weil zu diesem Zeitpunkt noch keine weiteren Fragebögen auf Grundlage der ISO-Norm 9241-10 (1998) existierten. Über die Gesamtscores wurde nicht berichtet. Datengrundlage des Vergleichs war die Bewertung von Software durch 31 Probanden mit allen drei Fragebögen. Es ergaben sich hohe Korrelationen (QUIS: r = .73, BBD: r = .71), was der Autor als positiv einschätzt. Figl (2009) vergleicht den ISONORM 9241/10 mit dem IsoMetrics (Willumeit, Gediga & Hamborg, 1996), der ebenfalls auf der Grundlage der ISO-Norm 9241-10 (1998) konstruiert wurde, und findet dabei keine Unterschiede in ihrer Vorhersagekraft für die Usability. Für die äußere kriterienbezogene Validität wird oft ein objektives Kriterium wie eine Expertenbewertung herangezogen und mit dem Ergebnis des Testverfahrens korreliert (Lienert & Raatz, 1998). Hierfür wurden Beurteilungen eines Software-Ergonomen und eines erfahrenen Nutzers mit den ISONORM 9241/10-Nutzerurteilen von N = 383 Personen anhand von 13 verschiedenen Software-Anwendungen verglichen. Die Urteile korrelieren zu r = .59 für Prümper (1997) ausreichend hoch.

Die Konstruktvalidität überprüft, ob der Test hinreichend gut das zugrunde liegende Konstrukt oder Modell erfasst (Lienert & Raatz, 1998). Ein Usability-Fragebogen verfügt über hinreichende Konstruktvalidität, wenn die Komponenten des Fragebogens mit dem Konstrukt „Usability" ausreichend übereinstimmen. Zur Untersuchung der Konstruktvalidität verglich Prümper die Bewertungen im ISONOM 9241/10 bezüglich zweier Versionen von MS-Word (4.0 und 5.0) und gab an, dass der Fragebogen deutlich zwischen diesen beiden unterscheiden können sollte. Insbesondere hinsichtlich der Grundsätze Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Fehlertoleranz und Individualisierbarkeit wurden Unterschiede gefunden, die für die bessere Usability von MS-Word 5.0 sprechen. Auch der Gesamtwert des Fragebogens unterscheidet sich signifikant für die beiden Versionen.

Tabelle 6. Kenntwerte zur Konstruktion des ISONORM 9241/10, Daten basierend auf Prümper (1997).

Versuchsteilnehmer	Alter	Stimulus	Itemkennwerte	Faktorenanalyse	Aufgeklärte Gesamtvarianz	Cronbach's Alpha
1265	Mittelwert: 34	178 verschiedene Software-Produkte	Nicht berichtet	Nicht berichtet	Nicht berichtet	Skalenweise, .81 - .89

2.3.4. Self-Assessment Manikin (SAM)

Emotions-Fragebögen stellen eine wichtige Ergänzung des Methoden-Inventars zur Untersuchung der User Experience interaktiver Produkte dar. Emotionen sind sowohl in der ISO-Norm 9241-210 (2011) als auch in den beiden bekannten, zuvor vorgestellten User Experience Modellen von Hassenzahl (2007) sowie Thüring und Mahlke (2007) eine zentrale Komponente, die die Interaktion zwischen Mensch und Produkt beeinflusst. Obwohl Emotionen für das Nutzungserleben eine wichtige Rolle spielen, existiert bisher kein Fragebogen, der emotionale Aspekte des Erlebens interaktiver Produkte erfasst. Deshalb wird sich in der Forschung oft mit etablierten, jedoch nicht für den Einsatz mit Produkten entwickelten, Emotions-Fragebögen beholfen. Diefenbach und Hassenzahl (2010) schlagen den PANAS (Tellegen, Watson & Clark, 1988) als ergänzenden Fragebogen vor, während viele andere Forscher, wie beispielsweise Mahlke (2006), Mahlke und Lindgaard (2007), Minge (2008) und Schleicher und Trösterer (2009), den Self-Assessment Manikin (SAM; Lang, 1980) bevorzugen, der im Folgenden vorgestellt wird.

Der von Lang (1980) entwickelte Fragebogen basiert auf der Drei-Faktoren-Theorie der Emotionen von Russell und Mehrabian (1977; vgl. Mehrabian, 1970). Laut diesem Modell existieren die drei voneinander unabhängigen bipolaren Dimensionen *Valenz* (pleasure – displeasure), *Arousal* (degree of arousal) und *Dominanz* (dominance – submissiveness). Jede Dimension gibt ein Kontinuum von Ausprägungen vor, die eine Person annehmen kann. Arousal verläuft zum Beispiel von dem extremen Pol *Schlaf* über *Schläfrigkeit* und *Wachheit* bis zu dem anderen Extrem *Aufregung* (Russell & Mehrabian, 1977). Aus diesem Modell hat Lang (1980) für den Fragebogen die drei dazu äquivalenten Dimensionen *Valenz* (happy – unhappy), *Arousal* (excited – calm) und *Dominanz* (control – controlled) entnommen. Aus Gründen der Kultur-Unabhängigkeit des Verfahrens, handelt es sich um ein piktorales Instrument, wie in Abbildung 7 veranschaulicht.

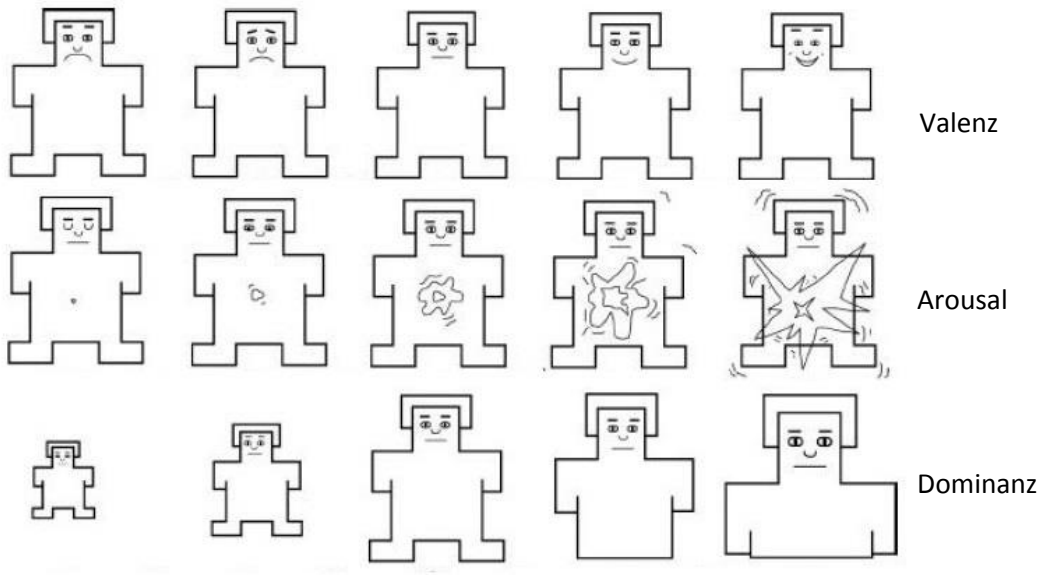


Abbildung 7. Self-Assessment Manikin von Lang (1980); Darstellung nach Bradley & Lang (1994, S. 51).

Die drei Skalen sind jeweils 9-fach gestuft, wobei links die geringste Ausprägung und rechts die stärkste Ausprägung dargestellt ist.

In der Studie von Bradley und Lang (1994) wurde der SAM mit der Semantic Defferential Scale von Mehrabian und Russell (1974; vgl. Tabelle 2) verglichen, die, ebenso wie der SAM, auf der Drei-Faktoren-Theorie der Emotionen basiert (Russell & Mehrabian 1977). Sie besteht aus 18 bipolaren, neun-fach gestuften Items. Je sechs Items bilden eine der drei Dimensionen Valenz, Arousal oder Dominanz. Auf Basis von 21 Bildern des International Affective Picture Systems (IAPS; Lang, Öhman & Vaitl, 1988) wurden in der Studie 78 Personen befragt, ihre subjektiven Emotionen mit der Semantic Differential Scale anzugeben. Die SAM-Ratings stammen aus zwei von den Autoren zu einem früheren Zeitpunkt durchgeführten Studien, wobei in der einen die Paper-and-Pencil-Version (PP) eingesetzt wurde und in der anderen die Computer-Version. Weil über die Durchführung und die Stichprobe der beiden SAM-Studien durch Bradley und Lang (1994) wenig berichtet wird, ist unklar, ob eine Vergleichbarkeit der Daten gegeben ist.

Die Ratings der Semantic Defferential Scale wurden faktorenanalytisch untersucht, indem eine Hauptkomponentenanalyse mit Faktorenextraktion entsprechend des Kaiser-Guttmann-Kriteriums durchgeführt wurde. Es ergaben sich drei Faktoren, auf denen jeweils die Items hoch laden, die auch theoretisch der jeweiligen Dimension entsprechen. Die aufgeklärte Varianz jedes Faktors (Valenz: 24.6%, Arousal: 23.12%, Dominanz: 12.18%) ist ebenfalls sehr ähnlich zu den Dimensionen der Drei-Faktoren-Theorie der Emotionen (Russell & Mehrabian, 1977; Valenz: 27%, Arousal: 23%, Dominanz: 14%). Anschließend wurde jeweils der Mittelwert jedes Faktors mit dem Mittelwert der SAM-Ratings für die drei Dimensionen korreliert. Zusätzlich haben Bradley und

Lang (1994) die PP- und die Computerversion des SAM verglichen. Für die Dimensionen Valenz und Arousal zeigen sich hohe Zusammenhänge zwischen allen untersuchten Verfahren. Für die Valenz-Dimension korreliert die Semantic Differential Scale mit der PP-Version des SAM zu .96 und mit der Computer-Version zu .97. Für Arousal ergeben sich ebenso hohe Zusammenhänge. Die Ergebnisse zeigen, dass alle Verfahren bezüglich Valenz und Arousal das Gleiche messen, was auf innere Kriteriumsvalidität hindeutet.

Für die Dominanz-Skala zeigt sich dagegen ein anderes Bild. Die PP- und die Computer-Version des SAM korrelieren zwar mit .79 sehr hoch, doch die Zusammenhänge zwischen den SAM-Skalen und der Semantic Differential Scale sind mit .23 und .18 gering und nicht signifikant. Die Analyse der beurteilten Bilder zeigt, dass insbesondere für die Dominanzskalen Unterschiede zwischen den Bewertungen der Bilder bestehen. Je nachdem, welches Verfahren die Probanden zur Beurteilung der Bilder verwendeten, haben sie verschiedene Bilder als mehr oder weniger dominant erlebt. Bradley und Lang (1994) stellen die Hypothese auf, dass bei der Beurteilung mit dem SAM Aspekte der *subjektiven Einschätzung von Kontrolle* Ausschlag gebend seien, wohingegen mit der Semantic Differential Scale die *in das Bild interpretierte Kontrolle* beurteilt wurde.

Aus der Studie leiten die Autoren ab, dass es sich beim SAM um ein sehr ökonomisches Messinstrument handelt, weil es mit seinen drei Skalen sehr schnell auszufüllen sei, insbesondere im Vergleich zu Verfahren mit vielen Einzel-Ratings wie der Semantic Differential Scale. Vor allem die Dimensionen Valenz und Arousal führen Bradley und Lang (1994) zufolge immer wieder zu aussagekräftigen Ergebnissen. Die Dominanz-Dimension zeige dagegen mehr Varianz und habe auch in vorherigen Untersuchungen nicht eindeutige Ergebnisse geliefert.

In Studien zum Nutzungserleben werden zumeist nur die beiden Skalen Valenz und Arousal verwendet (z.B. Mahlke, 2006; Mahlke & Lindgaard, 2007). Diese beiden Dimensionen sind, wie auch im Circumplex-Modell von Russell (1979) belegt, zwei unabhängige und hinreichende Dimensionen. Während die Valenz-Dimension zumeist gute Ergebnisse liefert, treten vor allem beim Arousal Probleme auf. Oft ist diese Skala nicht in der Lage zwischen hoch und niedrig aktivierenden Aufgaben oder Produkten zu unterscheiden (z.B. Mahlke, 2006; Mahlke & Lindgaard, 2007), was dadurch begründet sein könnte, dass das Instrument nicht für interaktive Produkte entwickelt wurde. Außerdem ist es fraglich, ob Personen in der Lage sind ihren Aktivierungsgrad adäquat widerzugeben, zumal die Beurteilung häufig retrospektiv erfolgt und ob die Dimensionen *Valenz* und *Arousal* tatsächlich unabhängig voneinander beurteilt werden.

Diese Problematik macht es erforderlich, emotionales Erleben bereits während der Entwicklung eines Messinstruments zum Nutzungserleben interaktiver Produkte aktiv einzubeziehen.

2.4. Zusammenfassung

In diesem Kapitel erfolgte ein Review über etablierte Verfahren zur Erfassung von User Experience, Usability oder Emotionen. Die hohe Anwendungsrate zeigt ihren Wert für die Bewertung von interaktiven Produkten. Es konnten aber auch Mängel aufgezeigt werden.

Zusammenfassend ergeben sich drei problematische Bereiche. Ein Problem besteht in einer unzureichenden theoretischen Grundlage, die die Operationalisierung des Fragebogens fragwürdig erscheinen lässt. Außerdem vermögen viele Fragebögen nur Aussagen über eine bestimmte Produktgruppe zu machen. Gründe dafür sind entweder methodische Schwächen oder der Fokus des Fragebogens auf diesen einen Bereich. Das dritte Problem sind methodische Mängel bei der Durchführung der Faktorenanalyse. Hinzu kommt, dass die Fragebögen entweder einen Fokus auf instrumentelle oder nicht-instrumentelle Qualitäten legen und Emotionen sowie Nutzungskonsequenzen bisher nicht berücksichtigt wurden. Ziel dieser Ausführungen ist deshalb die Entwicklung eines Instruments, welches User Experience im Umgang mit interaktiven Produkten zuverlässig misst und dabei großen Wert auf methodisch adäquates Vorgehen legt.

3. Studie 1

Das folgende Kapitel beschreibt das Vorgehen bei der Entwicklung des Fragebogens zur Messung des Nutzungserlebens, der insbesondere die in Abschnitt 2.4 zusammengefassten Schwachstellen bisheriger User Experience Fragebögen thematisiert und zu vermeiden versucht.

Zunächst werden die als wichtig erachteten Komponenten des CUE-Modells von Thüring und Mahlke (2007) dargestellt. Davon ausgehend erfolgt die Ableitung relevanter Komponenten und Items zur Gestaltung des Fragebogens. Der Abschnitt zur Methode beschreibt die Durchführung der Online-Erhebung sowie die Untersuchungsstichprobe. Die Itemreduktion erfolgte auf Basis faktorenanalytischer Betrachtungen. Die Analyseschritte werden im Ergebnisteil präsentiert. Als finales Produkt entstand ein Fragebogen mit 33 Items. Zum Abschluss des Kapitels beleuchtet die Diskussion Probleme des Fragebogens und leitet Implikationen für Studie 2 ab, die insbesondere die Erfassung von Emotionen und der Nutzungsintention in den Mittelpunkt rückt.

3.1. Theoretische Überlegungen: UX-Fragebogen auf Basis des CUE-Modells

Als Grundlage für die Ableitung der Items des Fragebogens wurde das CUE-Modell von Thüring und Mahlke (2007) gewählt, da das Modell sowohl den Rahmen als auch eine Erklärung für das Konzept des Nutzungserlebens, das mit dem Fragebogen erhoben wird, bietet. Auf Grund der geplanten Anwendung des Fragebogens für verschiedene interaktive Produktgruppen ist es unabdingbar, nur besonders relevante und allgemeingültige Komponenten des CUE-Modells auszuwählen.

Zuerst wurde das CUE-Modell darauf untersucht, welche Komponenten besonders bedeutsam für ein Fragebogenverfahren sind. Im Fokus steht das Nutzungserleben einer Person beim Umgang mit einem interaktiven Produkt, weshalb die Kontextfaktoren der Mensch-Maschine-Interaktion, wie beispielsweise Personencharakteristika, die im CUE-Modell dem Erleben voran gehen, nicht direkt in den Fragebogen aufgenommen werden. Ergänzend sei jedoch erwähnt, dass Fragebögen zur Beschreibung von Personencharakteristika existieren, wie beispielsweise der Fragebogen TA-EG, der die Technikaffinität eines Nutzers bezüglich elektronischer Geräte erfasst (Karrer, Glaser, Clemens & Bruder, 2009).

Es wurde festgelegt, dass lediglich soziodemografische Basisdaten, wie Alter, Geschlecht und Bildungsstand der Person sowie Herstellerangaben und die Nutzungsdauer des Produkts abgefragt werden.

Aus dem CUE-Modell wurden vier Komponenten übernommen, die jeweils Subaspekte des Nutzungserlebens beinhalten. Sie werden auf Grund der Itemselektion unter faktorenanalytischer Einzelbetrachtung nachfolgend als *Module* bezeichnet (vgl. Abbildung 8).

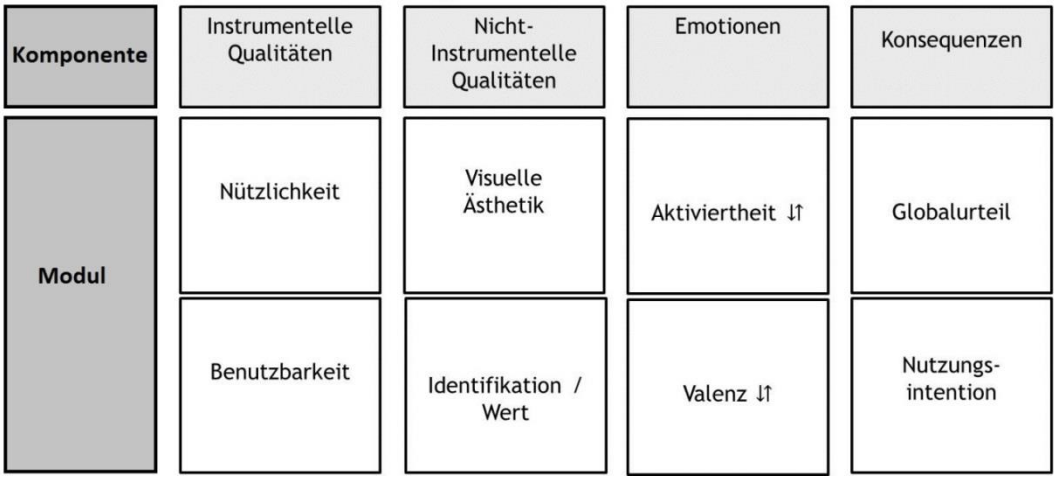


Abbildung 8. Komponenten und Module des Fragebogens, abgeleitet aus den Komponenten des CUE-Modells von Thüring & Mahlke (2007); eigene Darstellung.

Instrumentelle Qualitäten

Die Komponente *Instrumentelle Qualitäten* enthält die auch im CUE-Modell zentralen Aspekte *Utility* und *Usability*. Diese Unterscheidung geht aus dem Technology Acceptance Model (TAM) von Davis (1993) hervor, der die beiden Begriffe *ease of use* (hier: Usability) und *usefulness* (hier: Utility) verwendet. Utility ist definiert als die Nützlichkeit eines Produkts für das Erreichen eines Ziels (Nielsen, 1993). Dieses Ziel kann zum Beispiel eine bestimmte Aufgabe sein, die eine Person mit ihrem Produkt lösen möchte. Im Fragebogen wird deshalb die deutsche Bezeichnung *Nützlichkeit* verwendet. Weil Usability ein eigenes Konstrukt darstellt, das, wie die ISO-Normen 9241-11 (1998) und 9241-110 (2006) zeigen, verschiedene Aspekte beinhaltet, die mit dem Fragebogen nicht gänzlich abgedeckt werden, wird nachfolgend die Bezeichnung *Benutzbarkeit* verwendet.

Nicht-Instrumentelle Qualitäten

Bei den *nicht-instrumentellen Qualitäten* wird die *visuelle Ästhetik* als erstes Modul gewählt. Visuelle Ästhetik hat einen hohen Stellenwert beim Erleben von Produkten (Norman, 2004; Tractinsky, 2011). Zudem ist sie in der Lage, andere Aspekte der Ästhetik, wie Haptik und Akustik, zu kompensieren (Lemke, 2007; Minge, 2011). Deshalb wird es als legitim betrachtet, Aspekte haptischer und akustischer Ästhetik zunächst auszublenden, da der Fragebogen sowohl im Bereich Software als auch für die Bewertung von Prototypen eingesetzt werden soll und Aussagen über diese Aspekte nur zum Teil oder nicht valide getroffen werden können. Das zweite Modul der Komponente *Nicht-instrumentellen Qualitäten* wird als *Identifikation / Wert* bezeichnet und ist aus den symbolischen Qualitäten des CUE-Modells abgeleitet. Motivationale Aspekte sowie Assoziationen und

Erinnerungen finden keinen Eingang in den Fragebogen. Assoziationen und Erinnerungen sind rein subjektiv und müssen sich nicht direkt auf das Produkt beziehen (Beispielsweise auf ein Handy, das der Nutzer irgendwann einmal besessen hat.). Die Operationalsierung, worauf sich die Erinnerung des Nutzers tatsächlich bezieht, ist somit problematisch.

Emotion

im Bereich der Emotionsforschung werden zwei Ansätze unterschieden. Bei kategorialen Emotionsmodellen spielen voneinander getrennte Basisemotionen, wie Freude, Trauer oder Überraschung, die zentrale Rolle (Ekman, 1991). Dagegen existiert bei dimensionalen Emotionsmodellen ein Kontinuum von Ausprägungen einer Emotion, die eine Person annehmen kann (Russel, 1979). Weil dimensionale Faktormodelle eine klare empirische Basis haben und die vielen Emotionsbegriffe auf wenige, eindeutig operationalisierte Dimensionen reduzieren (Minge, 2011), wird eines dieser Modelle für die Entwicklung des Fragebogens genutzt. Das Circumplex-Modell (Russel, 1979) beschreibt die beiden unabhängigen Dimensionen *Arousal* und *Pleasure*. Arousal gibt den Grad der Aktivierung einer Person an, die von geringer bis sehr hoher Aktivierung reichen kann. In der vorliegenden Arbeit wird in Anlehnung an Lang (1980) statt des Begriffs Pleasure der Ausdruck Valenz verwendet, was die Stimmung einer Person widergibt. Negative Valenzen haben zum Beispiel die Emotionen „traurig“ und „wütend“. Positive Valenzen sind beispielsweise „erfreut“ und „glücklich“. Aufgrund der Entwicklung konsistenter Antwortformate für den Fragebogen, scheint eine verbale Variante mit verschiedenen Items besser geeignet als ein non-verbales Format, wie beispielsweise im Self-Assessment Manikin verwendet (Lang, 1980; vgl. Abschnitt 2.3.4).

Konsequenzen

Die *Konsequenzen* stellen die vierte Komponente dar. Sie umfassen als Module das *Globalurteil*, also ein „Alles-in-Allem-Gesamturteil“, sowie die *Nutzungsintention*, die als Konsequenz der übrigen Komponenten entsteht und letztendlich darin resultiert, ob das Produkt wieder verwendet oder ein anderes bevorzugt wird.

3.2. Generieren des Itempools

Zuerst wurde ein geeignetes Antwortformat festgelegt. Ein besonders häufig verwendetes Format ist die Ratingskala, bei der eine Antwort vorgegeben wird und mittels einer bestimmten Anzahl von Abstufungen die Ausprägung der Zustimmung oder Ablehnung vom Probanden festgelegt werden kann (Bortz & Döring, 2006; siehe Tabelle 8). Eine Sonderform bildet dabei das semantische Differential, bei dem zwei extreme Antwortalternativen die äußeren Pole bilden und dazwischen verschiedene Abstufungen der Antwort möglich sind (Bortz & Döring, 2006; siehe Tabelle 7). Semantische Differentiale haben den Nachteil, dass die Pole oft nicht die beiden Extreme einer

Aussage darstellen, weil keine geeigneten verbalen Gegensatzpaare vorliegen, sodass die Verteilung der Antworten durch Nutzer schiefwinklig wird (Bradley & Lang, 1994). Entsprechend müsste die Angemessenheit von semantischen Items untersucht werden. Bei der Wahl des Antwortformats sind zudem die Anzahl der Antwort-Abstufungen und deren Beschriftung wichtig. Nach Rohrmann (1978) sind zwischen fünf und neun Antwortstufen angemessen. Spatz (2008) zeigt, dass bei sieben-stufigen im Vergleich zu fünf-stufigen Formaten die Reliabilität höher, aber die Validität geringer ist. Letztendlich sei aber gemäß diesem Autor das Ziel der Erhebung ausschlaggebend. Weniger Abstufungen sind nicht ratsam, weil dadurch feine Unterschiede nicht mehr sichtbar werden (vgl. Rohrmann, 1978). Bei einer ungeraden Anzahl von Skalenstufen haben die Probanden die Möglichkeit, eine neutrale Antwort zu geben und sind nicht gezwungen, eine eindeutige Entscheidung zu jedem Item abzugeben. Rohrmann (1978) erklärt, dass „nur angemessen skalierte Variablen theoretische Modelle prüfbar machen“ (S. 223) und schlägt verschiedene Beschriftungen für die Skalenstufen vor, die er empirisch getestet hat. Rein numerische Skalenstufen seien zu abstrakt, insbesondere für Probanden, die nicht mit Fragebögen vertraut seien. Deshalb sei ein qualitatives, verbalisiertes Format, insbesondere in Kombination mit numerischen Entsprechungen, zu bevorzugen.

Aus diesen Gründen wird für die Studie eine sieben-stufige Ratingskala gewählt mit den Abstufungen von *Lehne völlig ab* über *Weder / noch* bis *Stimme völlig zu* entsprechend Tabelle 8.

Tabelle 7. Beispiel für ein semantisches Differential

Das Design ist schön.	○	○	○	○	○	○	○	Das Design ist hässlich.

Tabelle 8. Beispiel für eine Ratingskala

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Das Design ist hässlich.	○	○	○	○	○	○	○

Zur Itemgewinnung wurde zum einen eine Analyse bestehender Fragebogenverfahren und zum anderen ein Brainstorming über Items durchgeführt, die ein entsprechendes Modul repräsentieren.

Grundlage bildeten Usability- und User Experience Fragebögen, die ähnliche Komponenten wie die für den Fragebogen abgeleiteten Module aufweisen. Der Usability-Fragebogen ISONORM 9241/10, der sich auf die ISO-Norm zu den Grundlagen der Dialoggestaltung bezieht, bot sich als Ausgangspunkt für die Entwicklung von Items der instrumentellen Qualitäten an (vgl. Abschnitt

2.3.3). Insbesondere Items zur *Benutzbarkeit* sind von den Grundsätzen Aufgabenangemessenheit, Erwartungskonformität, Fehlertoleranz und Lernförderlichkeit beeinflusst (vgl. DIN EN ISO 9241-110, 2006). Da zur *Nützlichkeit* bisher kein validierter deutscher Fragebogen existiert, wurden die Items aus dem User Experience Fragebogen von Mahlke (2005) und aus dem Fragebogen von Davis (1989) zur wahrgenommenen Nützlichkeit, Benutzbarkeit und Akzeptanz als Vorlage verwendet. Der VisAWI (Moshagen & Thielsch, 2011), der die Ästhetik von Webseiten erfasst, diente den Items der *visuellen Ästhetik* als Vorbild. Für das Modul *Identifikation / Wert* wurden aus den existierenden Fragebögen AttrakDiff-2 und UEQ jeweils die Dimensionen, die sich mit hedonischen Aspekten befassen (AttrakDiff-2: Stimulation und Identifikation, UEQ: Stimulation, Originalität), als Anhaltspunkt für die Itemgewinnung herangezogen. Für die Emotionen wurden die Fragebögen SAM (Lang, 1980) und PANAS (Tellegen, Watson & Clark, 1988) als Ausgangspunkt gewählt. Da im Bereich Nutzungsintention und Globalurteil keine ausführlichen, validierten Fragebogenverfahren verfügbar waren, erfolgte die Itemgenerierung für diese Dimensionen ausschließlich schöpferisch über die Methode des Brainstormings.

Zur Itemgewinnung erfolgten für alle abgeleiteten Module zusätzlich Brainstormings, in denen aus konkreten Überlegungen zu den zu beurteilenden Produkten Probleme, Schwachstellen, aber auch positive Aspekte gewonnen wurden. Daraus wurden im nächsten Schritt Fragen und Items abgeleitet. Für die Gestaltung der Emotionsitems kamen insbesondere Synonyme-Lexika zum Einsatz (z.B. Chr. Agricola & E. Agricola, 1982).

Im Ergebnis entstand ein Itempool von etwa 100 Items, von denen aufgrund inhaltlicher Überlegungen zur Verständlichkeit und Eindeutigkeit letztendlich 67 für die erste Rohversion Fragebogens ausgewählt wurden.

Bei der Itemreduktion werden drei Items, die inhaltlich möglichst weit gefächert sein sollen, je Modul angestrebt, damit einerseits die Items des Fragebogens durch Nutzer schnell bearbeitet werden können, was ihn zu einem ökonomischen Instrument macht, und andererseits die Aussagefähigkeit des Fragebogens trotz der Reduktion erhalten bleibt.

3.3. Methode

Auf der Grundlage der 67 entwickelten Items wurde eine Online-Befragung durchgeführt, die im Folgenden näher beschrieben wird.

3.3.1. Material

Der Fragebogen wurde als Online-Fragebogen mittels der Open Source Software *LimeSurvey*[3] (Version 2.0) erstellt. Auf sieben Seiten wurden je neun bis zehn Items randomisiert dargeboten. Der Online-Fragebogen ist in vollständig dargestellt.

Der Fragebogen startet mit einer Willkommen-Seite, die über den Untersuchungshintergrund und die Zeitdauer informiert (vgl. Anhang A.1).

Nachdem sich die Probanden ein Alltagsprodukt entschieden haben, sollen sie es einer der vorgegebenen Produktgruppen zuordnen: Smartphone bzw. Handy, Laptop bzw. Computer, Kaffeemaschine, Digitalkamera, Waschmaschine, Fernseher, Smartphone Apps, Microsoft Office. Zusätzlich wird die Bezeichnung des Produkts (z.B. Philips SENSEO), die bisherige Nutzungsdauer für das konkrete Produkt und für die Produktgruppe, jeweils in Jahren und Monaten, abgefragt sowie die Nutzung des Produkts pro Woche in Stunden.

Zusätzlich werden die Probanden gebeten, ihre demografische Daten Alter, Geschlecht und Bildungsstand einzugeben, um Informationen zu der Stichprobe zu erhalten. Die Probanden hatten die Möglichkeit, eine optional Rückmeldung über ihre Meinung zum Fragebogen zu geben. Bei allen übrigen Eingaben im Fragebogen wurden die Probanden auf eventuell ausgelassene Angaben mittels visuellen Hinweisen aufmerksam gemacht.

3.3.2. Untersuchungsdurchführung

Die Studie wurde in einem Zeitraum von zwei Wochen im November 2012 durchgeführt. Auf die Studie wurde zum einen durch eine direkte Einladung von Personen über das Probandentool des Graduiertenkollegs *prometei* und zum anderen über universitäre Mailinglisten und soziale Netzwerke aufmerksam gemacht. Um einen möglichst großen Personenkreis anzusprechen, galt als einzige Teilnahmevoraussetzung, dass es sich bei den Probanden um Personen mit der Muttersprache *Deutsch* handelt. In der Einladung wurde das Thema der Studie beschrieben. Unter allen Probanden wurden fünfmal 20 Euro verlost. Dafür konnten die Probanden ihre Email-Adresse

[3] http://www.limesurvey.org [06.03.2013]

angeben, die unabhängig von den Fragebogendaten gespeichert und nach Beendigung der Verlosung gelöscht wurden. Die Durchführung der Befragung dauerte circa 15 bis 20 Minuten.

3.3.3. Stichprobe

Für die Untersuchung wurde eine Teilnehmerzahl von mindestens 200 angestrebt, da dieser Stichprobenumfang als ausreichend für aussagekräftige Ergebnisse einer Faktorenanalyse eingeschätzt wird (Bühner, 2011). Die Online-Studie haben insgesamt 370 Personen aufgerufen, von denen 238 Personen die Studie erfolgreich abgeschlossen haben. Es kam somit zu einem Drop-Out von 36%. Die nicht erfolgreich abgeschlossenen Befragungen sind auf vorzeitige Abbrüche zurückzuführen, von denen die Mehrzahl (*N* = 100) direkt nach der Betrachtung der Willkommensseite aufgetreten sind, was nach Thielsch und Weltzin (2009) unproblematische Abbrüche seien. Spätere Abbrüche, die auf inhaltliche Probleme der Studie bzw. des Fragebogens hinweisen könnten, traten mit 8,6% nur selten auf (*N* = 32).

An der Studie nahmen 127 Frauen und 111 Männer erfolgreich teil. Im Mittel waren die Probanden 28.67 Jahre alt (*SD* = 8.23). Der jüngste Teilnehmer war 19, der älteste 61 Jahre. Das recht niedrige mittlere Alter[4] (siehe Abbildung 9) ist auf die vornehmlich aus Studenten bestehende Stichprobe zurückzuführen.

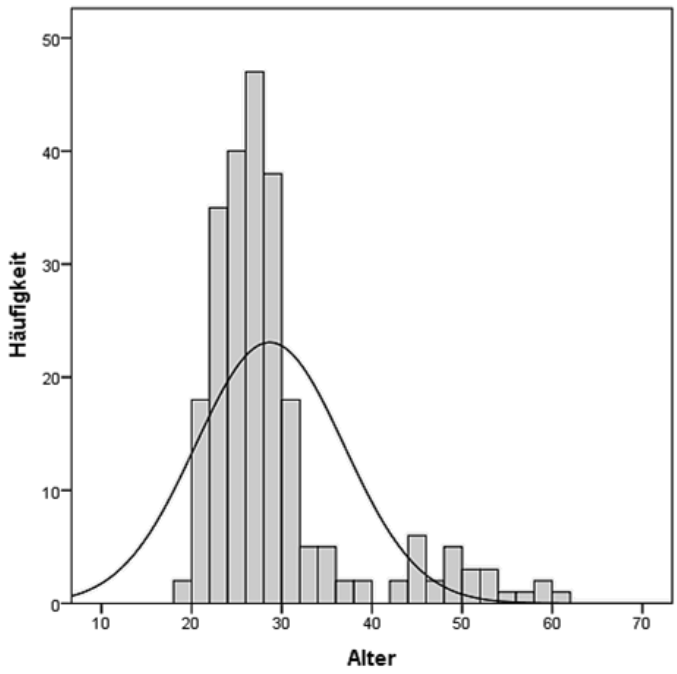


Abbildung 9. Häufigkeitsverteilung und Normalverteilungskurve des Alters der Studien-Teilnehmer aus Studie 1.

[4] Das Durchschnittsalter der Gesamtbevölkerung Deutschlands betrug 2011 44.9 Jahre, http://www.welt-auf-einen-blick.de/bevoelkerung/durchschnittsalter.php [05.03.2013]

Dadurch lässt sich auch erklären, dass mit 47.9% fast die Hälfte der Teilnehmer bereits einen (ersten) Hochschulabschluss hat. Die zweitgrößte Gruppe sind Abiturienten mit 42.9%. Andere Bildungsabschlüsse sind entsprechend gering vertreten (Realschule: 7,6%, Hauptschule: 1,7%).

In Tabelle 9 sind die Produktgruppen angegeben, für die sich die Probanden entschieden haben, wenn sie nach einem interaktiven Alltagsprodukt gefragt wurden. Die meisten Probanden entschieden sich für ihr Handy bzw. Smartphone (*N* = 100 bzw. 42%) oder ihren Laptop bzw. Computer (*N* = 75 bzw. 31.5%). Die anderen Produktgruppen wurden wesentlich seltener gewählt.

Tabelle 9. Von den Probanden gewählte Produktgruppen.

Produktgruppe	*N*	Anteil an Gesamtstichprobe in %
Handy / Smartphone	100	42.0
Laptop / Computer	75	31.5
Digitalkamera	16	6.7
Waschmaschine	13	5.5
Fernseher	12	5.0
Smartphone App	2	0.8
Microsoft Office	1	0.4

Die Nutzungsdauer der von den Teilnehmern gewählten Produkte variiert zwischen einem Monat und 12 Jahren (siehe Tabelle 10). Im Mittel werden die Produkte seit 2 Jahren und 2 Monaten genutzt (*SD* = 21.37 Monate). Die zum gewählten Produkt gehörige Produktgruppe hat eine durchschnittliche Nutzungsdauer von circa 7 Jahren (*SD* = 77.10 Monate). Die Nutzungsdauer pro Woche variiert ebenfalls deutlich. Das Minimum beträgt eine Stunde, das Maximum 147 Stunden pro Woche (*M* = 21.95, *SD* = 24.19). Wie Tabelle 26 **Tabelle 26**(Anhang A.3) zeigt, sind die Nutzungsdauer und die Wochennutzung von dem jeweils bewerteten Produkt bzw. der Produktgruppe abhängig.

Tabelle 10. Nutzungsdauer des Produkts und der Produktgruppe in Monaten, sowie Nutzungshäufigkeit des Produkts pro Woche in Stunden.

	N	MIN	MAX	*M*	*SD*
Wie lange nutzen Sie das Produkt bereits? *	233	1	144	26.28	21.37
Wie viele Stunden pro Woche nutzen Sie das Produkt? **	214	1	147	21.94	24.19
Wie lange nutzen Sie Produkte dieser Produktgruppe bereits? *	227	1	456	87.66	77.10

* Angabe in Monaten ** Angabe in Stunden

3.4. Ergebnisse

Zur Analyse der Daten wurde eine Hauptkomponentenanalyse mit Varimax-Rotation eingesetzt. Zu deren nachfolgenden Erläuterung wird zuerst auf die Vorbetrachtung zur Durchführbarkeit der Faktorenanalyse eingegangen, gefolgt von den konkreten Schritten der Itemreduktion.

3.4.1. Vorbetrachtung

Wichtige Verfahren der Dateneignung für die Faktorenanalyse wurden bereits in Abschnitt 2.2.2 vorgestellt. Darauf Bezug nehmend stehen die Ergebnisse im Vordergrund.

Kaiser-Meyer-Olkin-Koeffizient (KMO) und Measure of Sample Adequacy (MSA)

Der KMO-Koeffizient, der Auskunft darüber gibt, ob die Daten ausreichend hoch korrelieren, hat einen Wert von .886. Entsprechend Tabelle 1 kann daraus eine gesicherte Dateneignung abgeleitet werden. Der MSA-Koeffizient der einzelnen Items ergibt Werte zwischen .671 („Ich kann mit dem Produkt viele verschiedene Aufgaben erledigen.“) und .930 („Das Produkt ist hässlich.“). Die meisten Items weisen einen MSA-Koeffizienten über .800 auf. Somit korrelieren die Items mäßig bis sehr gut miteinander.

Kolmogorov-Smirnov (K-S)-Test und Bartlett-Test auf Sphärizität

Die Ergebnisse des K-S-Tests zeigen, dass keine Normalverteilung der Werte vorliegt (K-S-Werte zwischen 2.270 und 5.567, $p < .001$). Deshalb sind die Ergebnisse des Bartlett-Tests auf Sphärizität nur unter Vorbehalt zu interpretieren. Da der Bartlett-Test jedoch signifikant wird ($p < .001$), kann die Eignung der Daten für eine Faktorenanalyse gestützt werden.

Aus den Ergebnissen der Dateneignungsprüfung lässt sich somit schlussfolgern, dass eine Faktorenanalyse sinnvoll durchführbar ist.

3.4.2. Itemanalyse

Itemschwierigkeit

Wie in Abschnitt 2.2.1 ausgeführt, gibt die Itemschwierigkeit an, wie leicht Personen einem Item eher zustimmen (Wert gegen 1) oder es ablehnen (Wert gegen 0). Über die gesamte Stichprobe betrachtet, bietet eine Schwierigkeit zwischen .2 und .8 den größten Informationswert. Items mit extremen Schwierigkeiten sollten ausgeschlossen werden, sofern sie keine inhaltlichen Relevanzen aufweisen. In Tabelle 11 sind die Items mit Schwierigkeiten unter .2 und über .8 aufgeführt. Items, die mit einem Stern (*) markiert sind, wurden ausgeschlossen. Die Itemschwierigkeit für alle Items befindet sich in Tabelle 27. Itemanalyse (Anhang A.4).

Tabelle 11. Items mit extremer Schwierigkeit (unter .200 oder über .800).

Items mit besonders hoher oder niedriger Schwierigkeit	*p* (von Dahl)
Insgesamt halte ich das Produkt für nützlich.	.882
Das Produkt erfüllt seine Bestimmung.	.861*
Alles in allem ist das Produkt gut.	.835*
Das Produkt lässt sich einfach benutzen.	.832
Die Bedienung des Produkts ist verständlich.	.815
Das Produkt ist hässlich.	.183*
Ich schäme mich, wenn andere mich mit dem Produkt sehen.	.125*
Ich würde das Produkt auf keinen Fall wieder benutzen.	.095*

* ausgeschlossene Items

Trennschärfe

Die Trennschärfe stellt die Summe eines Items korreliert mit der Gesamtsumme eines Tests dar. Bei einem mehrdimensionalen Ansatz wird die Gesamtsumme skalenweise bestimmt, weil der Unterschied zwischen den Skalen von Interesse ist (vgl. Bortz & Döring, 2006). Deshalb wurde die Trennschärfe im Rahmen dieser Untersuchung für jedes Modul einzeln bestimmt (siehe Tabelle 27, Anhang A.4). Mit Ausnahme des Items „Ich schäme mich, wenn mich andere mit dem Produkt sehen“, das mit einer Trennschärfe von r = .048 die Gesamtsumme des Moduls *Identifikation / Wert* nicht repräsentiert, werden Korrelationen von mindestens .300 erreicht. Somit sind 66 der 67 Items trennscharf und vermögen zwischen Personen mit hoher und niedriger Ausprägung zu unterscheiden (Lienert & Raatz, 1998).

3.4.3. Faktorenanalyse der Module

Unter der Bedingung der Erfüllung der Voraussetzungen, wurde für jedes Modul zuerst getrennt eine Faktorenanalyse unter Einschluss der theoretisch für das Modul gedachten Items durchgeführt. Geprüft wurde, welche Items sich sinnvoll zu einem Faktor verbinden lassen. Dies ermöglicht zum einen eine gezielte Reduktion von Items. Zum anderen findet dadurch eine separate Validierung einzelner Module statt, sodass diese für spätere Untersuchungen theoretisch auch getrennt eingesetzt werden können.

Anschließend wurde die Dateneignung der Items pro Modul geprüft. Durch die je sehr geringe Anzahl an Items (jeweils acht bis zehn pro Modul), verschlechtern sich die KMO- und MSA-Werte, jedoch unter Einhaltung der Voraussetzungen zur Durchführung einer Faktorenanalyse. Der Bartlett-Test auf Sphärizität wurde ebenfalls signifikant (p < .001).

Für jedes Modul wurde des Weiteren eine Hauptkomponentenanalyse mit Varimax-Rotation mit dem Ziel durchgeführt, eine Einfachstruktur, die mit einer selegierten Menge an Items beschrieben werden kann, zu erhalten. Die Selektion erfolgt auf Basis der Faktorladungen und Kommunalitäten sowie aufgrund inhaltlicher Überlegungen.

Instrumentelle Qualitäten

Die Analyse der instrumentellen Qualitäten erwies sich als unproblematisch. Sowohl für das Modul Nützlichkeit als auch für Benutzbarkeit zeigten sich drei Items als empirisch besonders zusammengehörig. In Tabelle 12 ist die Faktorstruktur der Faktorenanalyse des Moduls Nützlichkeit dargestellt. Es ergibt sich zunächst eine zweifaktorielle Struktur, doch liegen nur zwei Items auf dem zweiten Faktor, die zusätzlich zusammen mit den restlichen Items auch auf dem ersten Faktor laden. Deshalb wurden drei Items aus dem ersten Faktor gewählt, die sowohl empirisch als auch inhaltlich das Modul Nützlichkeit im besonderen Maße widerspiegeln.

Tabelle 12. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Nützlichkeit*.

	1	2	Kommunalität
Das Produkt ist hilfreich für das Erreichen meiner Ziele.	.812		.670
Mithilfe des Produkts kann ich meine Ziele erreichen.*	.777		.604
Die Funktionen des Produkts sind genau richtig für meine Ziele.*	.742		.564
Insgesamt halte ich das Produkt für nützlich.*	.703		.582
Mit dem Produkt lassen sich meine geplanten Aufgaben nicht lösen.	.593		.352
Das Produkt erfüllt seine Bestimmung.	.539	-.510	.551
Das Produkt ist nicht zweckmäßig.	.425		.185
Das Produkt ist sinnvoll erweiterbar, falls sich meine Bedürfnisse ändern.	.408	.689	.642
Ich kann mit dem Produkt viele verschiedene Aufgaben erledigen.	.451	.674	.658

* ausgewählte Items

Die Faktorladungen des Moduls Benutzbarkeit zeigen ein ähnliches Bild wie das der Nützlichkeit. Hier ergibt sich zunächst eine dreifakorielle Struktur (siehe Tabelle 13). Die meisten Items laden hoch auf Faktor 1. Alle Items beziehen sich in irgendeiner Weise auf die Einfachheit der Bedienung des Produktes. Drei Items laden auf dem zweiten Faktor, wovon allerdings zwei Items eine zu große inhaltliche Nähe haben („Manchmal verstehe ich nicht, was das Produkt tut.“, „Manchmal verstehe ich nicht, was das Produkt mir sagen will.“, r = .74), sodass dieser Faktor nicht alle Facetten der Benutzbarkeit abdecken könnte. Auf dem dritten Faktor laden zwei Items, die sich mit der Anstrengung bei der Bedienung des Produktes befassen. Es wurden deshalb drei

Items aus Faktor 1 und Faktor 3 gewählt, da dadurch ein breites Spektrum der inhaltlichen Bedeutung des Moduls abgedeckt wird.

Tabelle 13. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Benutzbarkeit*.

	1	2	3	Kommunalität
Die Bedienung des Produkts ist verständlich.*	.773			.663
Es dauert zu lange die Funktionen des Produkts zu erlernen.*	-.757			.612
Das Produkt ist kompliziert zu bedienen.	-.736			.622
Ich muss mich kaum auf die Benutzung des Produkts konzentrieren.	.498			.385
Es ist immer eindeutig, was das Produkt von mir will.	.465		.454	.472
Von der Benutzung des Produkts bekomme ich manchmal Kopfschmerzen.		.774		.676
Manchmal verstehe ich nicht, was das Produkt tut.	.442	.737		.739
Manchmal verstehe ich nicht, was das Produkt mir sagen will.	.499	.709		.755
Die Benutzung des Produkts ist nicht anstrengend.			.845	.746
Das Produkt lässt sich einfach benutzen.*	.415		.731	.710

* ausgewählte Items

Nicht-Instrumentelle Qualitäten

Die Faktorenanalyse der nicht-instrumentellen Qualität *visuelle Ästhetik* ergab, wie Tabelle 14 zeigt, eine zweifaktorielle Struktur, wobei positiv gepolte Items auf dem ersten Faktor und negativ gepolte Items auf dem anderen hoch laden. Inhaltlich fiel die Wahl auf den Faktor, auf dem die Items mit den positiven Itemformulierungen laden. Drei Items wurden ausgewählt, die eine hohe Kommunalität aufweisen und sowohl die visuelle „Schönheit“ als auch die kreative Gestaltung eines Produkts erfassen.

Tabelle 14. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *visuelle Ästhetik*.

	1	2	Kommunalität
Das Produkt sieht minderwertig aus.	.759		.686
Das Design harmonisiert nicht mit dem Inhalt.	.758		.579
Das Produkt ist hässlich.	.722		.661
Das Design ist altbacken.	.616		.527
Hier passt alles zusammen.	.462		.364
Die einzelnen Komponenten passen zueinander.	-.429		.239
Das Produkt ist originell.		.852	.726
Das Produkt ist kreativ gestaltet.*		.779	.685
Das Produkt ist stilvoll.*	.458	.751	.774
Das Design wirkt attraktiv.*	.534	.696	.770

* ausgewählte Items

Für das Modul *Identifikation / Wert* zeigte sich eine zweifaktorielle Struktur, wobei sich thematisch ähnliche Items auf einem Faktor zusammenordneten (siehe Abbildung 10).

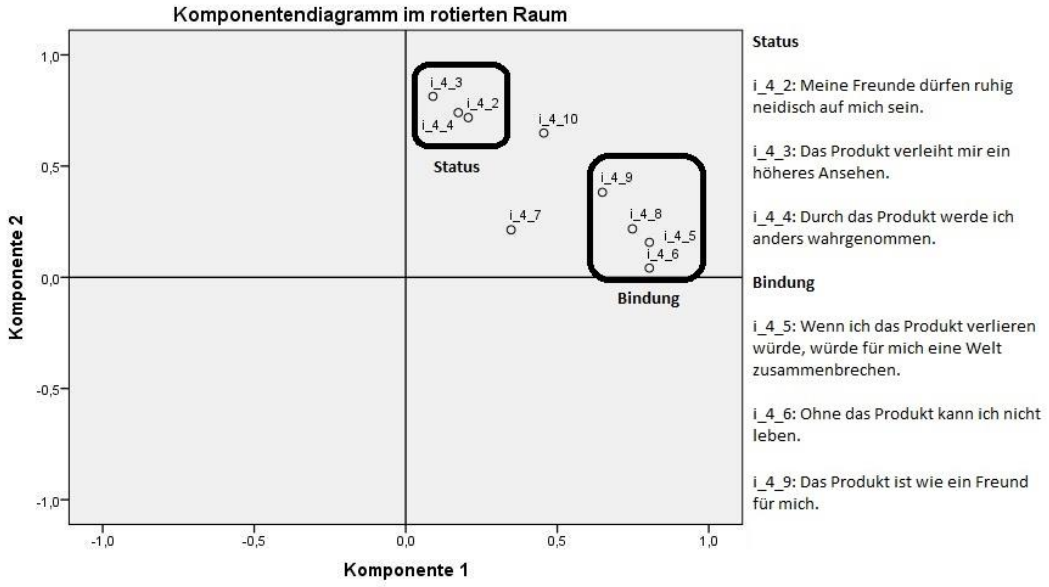


Abbildung 10. Ladungsdiagramm des Moduls *Identifikation / Wert*.

Daraus wurden zwei separate Module abgeleitet. Das erste, der *Status*, bezieht sich auf die soziale Identität einer Person. Es beschreibt, inwiefern das Produkt zum Ansehen der Person innerhalb einer sozialen Struktur beiträgt. O'Cass und Choy (2008) zeigen zum Beispiel, dass Nutzer bereit sind, mehr für ein Produkt zu zahlen, wenn es einen hohen Status vermittelt (vgl. O'Cass &

Frost, 2002). Die *Bindung* stellt als zweites Modul die Beziehung der Person zu (s)einem Produkt in den Mittelpunkt. Die Items beziehen sich darauf, wie wichtig das Produkt der Person ist bzw. wäre, wenn sie es besäße. Die ausgewählten Items sind in Abbildung 10 rechts aufgelistet.

Emotionen

Die Komponente *Emotionen* wurde in die Module *Valenz* und *Arousal* aufgeteilt und einzeln faktorenanalytisch betrachtet. Dabei spannte sich jeweils ein Raum mit zwei Faktoren auf, wobei ein Faktor Items mit hoher Ausprägung (positive Valenz, hohes Arousal) und der andere Faktor Items mit niedriger Ausprägung (niedrige Valenz, niedriges Arousal) zusammenfasste.

Weil mit dieser Lösung keine Itemreduktion vorgenommen werden kann, da die Items sowohl durch ihre Faktorladung als auch inhaltlich eine gute Struktur bilden, wurden alle Emotionsitems gemeinsam faktorenanalytisch betrachtet. Wie Tabelle 15 zeigt, ergab sich eine dreifaktorielle Struktur, wobei auf dem ersten Faktor Items mit positiver Valenz sowie Items, die ein niedriges Arousal ausdrücken, laden. Der zweite Faktor enthält die Items mit negativer Valenz und auf dem dritten Faktor laden die Items mit niedrigem Arousal.

Tabelle 15. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Komponente *Emotionen*.

	1	2	3	Kommunalität
Das Produkt macht mich glücklich.	.756			.651
Durch das Produkt fühle ich mich energiegeladen.	.750			.566
Das Produkt macht mich hellwach.	.701			.521
Das Produkt erfreut mich.	.699	- .396		.709
Das Produkt beruhigt mich.	- .682			.519
Das Produkt begeistert mich.	.665	-.329		.591
Das Produkt macht mich zufrieden.	.658	-.465		.661
Das Produkt verärgert mich.		.851		.781
Das Produkt frustriert mich.		.851		.766
Das Produkt nervt mich.		.776		.721
Das Produkt macht mich müde.			.812	.707
Das Produkt macht mich träge.			.810	.731
Durch das Produkt fühle ich mich passiv.			.724	.530
Das Produkt langweilt mich.	- .424		.571	.539
Aufgeklärte Varianz	**26.468**	**19.934**	**17.686**	**Gesamt: 64.088**

Konsequenzen

Die Betrachtung der *Nutzungsintention* ergab zunächst eine dreifaktorielle Struktur. Hier wäre eine Reduktion auf eine einfaktorielle Struktur nicht ratsam, da dadurch wichtige Verhaltenskonsequenzen nicht erfasst werden würden. Das Faktormodell in Tabelle 16 deutet bereits an, dass sich die Nutzungsintention in zwei Bereiche aufteilen lässt. Die Faktoren 1 und 3 sind inhaltlich identisch, mit dem Unterschied, dass auf dem einen Faktor die Items der Zustimmung und auf dem anderen Faktor die der Ablehnung hoch laden. So wird die tatsächliche Nutzungsintention (Faktor 1 und 3), ausgedrückt über Items, die das Verhalten dem Produkt gegenüber darstellen („Wenn ich könnte, würde ich das Produkt täglich nutzen."), unterschieden von der Produktloyalität (Faktor 2), die sich auf die Wahl zwischen Produktalternativen bezieht („Ich würde das Produkt gegen kein anderes eintauschen.").

Tabelle 16. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items des Moduls *Nutzungsintention*.

	1	2	3	Kommunalität
Wenn ich könnte, würde ich das Produkt täglich nutzen.	.818			.674
Ich kann es kaum erwarten, das Produkt erneut zu benutzen.	.769			.662
Ich würde dieses Produkt einem anderen, ähnlichen Produkt vorziehen.		.755		.594
Ich würde das Produkt meinen Freunden weiterempfehlen.		.684		.632
Ich würde das Produkt gegen kein anderes eintauschen.		.661		.619
Das Produkt motiviert mich nicht, es weiterhin zu nutzen.			.864	.763
Ich würde das Produkt auf keinen Fall wieder benutzen.			.624	.607
Aufgeklärte Varianz	**24.228**	**22.455**	**18.335**	**Gesamt: 65.019**

Die Analyse der fünf Items des Moduls *Globalurteil* ergab eine Einfachstruktur, wie sie in Tabelle 17 veranschaulicht ist. Alle Items weisen eine hohe Faktorladung und eine akzeptable Kommunalität auf. Weil alle fünf Items gute Werte erzielen, wurde zunächst keines ausgeschlossen.

Tabelle 17. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation des Moduls *Globalurteil*.

	1	Kommunalität
Das Produkt entspricht vollkommen meinen Erwartungen.	.825	.681
Ich bin von dem Produkt begeistert.	.822	.676
Alles in allem ist das Produkt gut.	.766	.586
Ich habe mir von dem Produkt mehr erwartet.	-.715	.511
Das Produkt befriedigt meine Bedürfnisse.	.670	.681

Zusammenfassend lassen sich die Module zu drei Modulgruppen ordnen. Die Modulgruppe *Produktbewertungen* umfasst die instrumentellen (*Nützlichkeit* und *Benutzbarkeit*) und nicht-instrumentellen Module (*Ästhetik*, *Status* und *Bindung*). *Emotionen* bilden ebenso wie die *Konsequenzen* eine eigene Modulgruppe. In Abbildung 11 ist dies veranschaulicht.

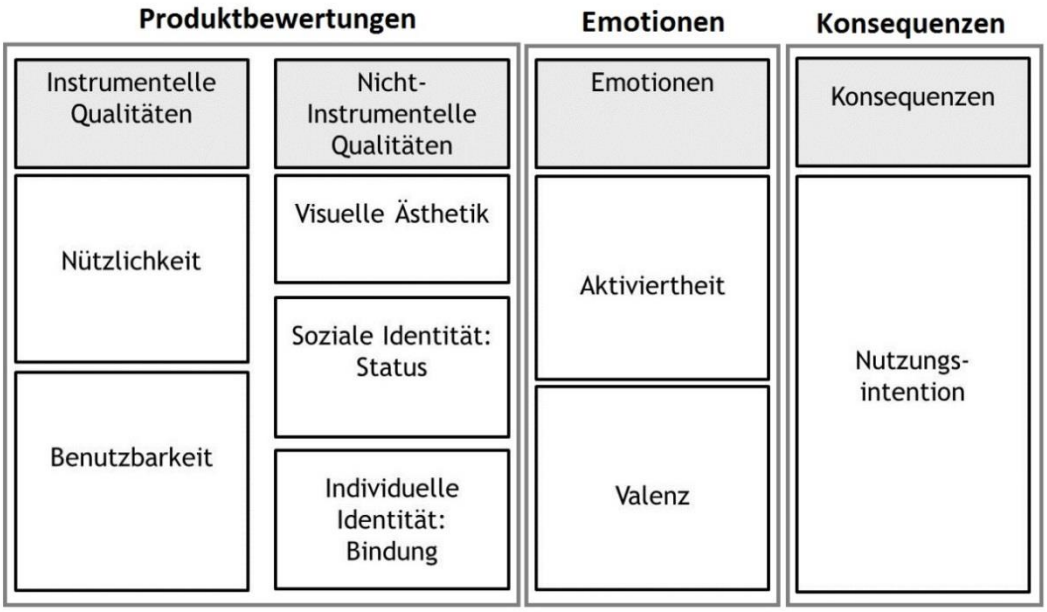


Abbildung 11. Die drei finalen Modulgruppen nach der faktorenanalytischen Betrachtung der einzelnen Module; eigene Darstellung.

3.4.4. Faktorenanalyse der Modulgruppe *Produktbewertungen*

Jede Modulgruppe wurde gesondert faktorenanalytisch betrachtet. Im ersten Schritt wurde erneut die Dateneignung geprüft. Für die Produktbewertungen zeigt sich durch die KMO-Werte mit .781 eine ausreichende Eignung. Mit Werten zwischen .632 („Es dauert zu lange die Funktionen des Produkts zu erlernen.“) und .889 („Meine Freunde dürfen ruhig neidisch auf mich sein.“) des MSA-Koeffizienten ergibt sich eine mäßige bis gute Eignung der Daten. Der Bartlett-Test ist signifikant ($p < .001$).

Sowohl der MAP-Test als auch die Parallelanalyse (vgl. Abschnitt 2.2.2) zur Prüfung der nötigen Anzahl der Faktoren, ergeben eine dem aufgestellten theoretischen Modell (vgl. Abbildung 11) entsprechende Anzahl von fünf Faktoren. Wie Tabelle 18 zeigt, laden jeweils die Items hoch auf einem Faktor, die bei Erstellung des Itempools als zusammengehörig erachtet wurden. Zusätzlich sind in der Tabelle die Kommunalitäten der Items angegeben. Je höher die Kommunalität ist, desto besser wird das Modell durch das Item erklärt (Backhaus et al., 2011). Das Modell klärt 69.43% der aufgetretenen Varianz auf.

Tabelle 18. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Produktbewertungen*.

	1	2	3	4	5	Kommunalität
Das Design wirkt attraktiv.	.895					.857
Das Produkt ist stilvoll.	.869					.667
Das Produkt ist kreativ gestaltet.	.783					.838
Durch das Produkt werde ich anders wahrgenommen.		.791				.675
Das Produkt verleiht mir ein höheres Ansehen.		.779				.662
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.		.657				.566
Es dauert zu lange, um die Funktionen des Produkts zu erlernen.			-.856			.739
Die Bedienung des Produkts ist verständlich.			.828			.712
Das Produkt lässt sich einfach benutzen.			.668			.512
Ohne das Produkt kann ich nicht leben.				.879		.791
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.				.783		.683
Das Produkt ist wie ein Freund für mich.				.650		.625
Mithilfe des Produkts kann ich meine Ziele erreichen.					.833	.730
Die Funktionen des Produkts sind genau richtig für meine Ziele.					.791	.706
Insgesamt halte ich das Produkt für nützlich.					.707	.644
Aufgeklärte Varianz (%)	**16.446**	**13.438**	**13.339**	**13.424**	**12.964**	**Gesamt: 69.430**

3.4.5. Reliabilität

Im Fokus der Reliabilitätsbetrachtung stand die Modulgruppe *Produktbewertungen*. Reliabilität beschreibt die Genauigkeit, mit der eine Skala ein Merkmal misst (Lienert & Raatz, 1998; vgl. Abschnitt 2.2.3) und wurde im Rahmen dieser Untersuchung durch Cronbach's α-Koeffizienten (1951), der den wechselseitigen Zusammenhang der Items darstellt, bestimmt.

Die Reliabilität der Produktbewertung wurde modulweise berechnet. Nach Nunnally und Bernstein (1994) kann ab einem Wert von .7 von einer befriedigenden und ab einem Wert von .8 von einer guten Reliabilität ausgegangen werden. Da entsprechend Schmitt (1996) die Reliabilität

eines Verfahrens immer im Zusammenhang mit weiteren Maßen interpretiert werden soll, wurden zusätzlich Interkorrelationen zwischen den Items betrachtet. Eine vollständige Darstellung der Reliabilitäten und Interkorrelationen kann Tabelle 28 (Anhang A.5) entnommen werden. Die Reliabilitäten sind in Tabelle 19 zusammen gefasst.

Tabelle 19. Darstellung der Reliabilität der Module der Modulgruppe *Produktbewertungen* mittels Cronbach's α.

Modulgruppe	Cronbach's α
Nützlichkeit	.714
Benutzbarkeit	.701
visuelle Ästhetik	.870
Soziale Identität: Status	.755
Individuelle Identität: Bindung	.706

Die Werte des α-Koeffizienten und die Korrelationen der Items innerhalb der Module weisen auf eine hinreichende Reliabilität der Modulgruppe *Produktbewertungen* hin.

Emotionen und *Konsequenzen* wurden aufgrund der nicht eindeutigen faktoriellen Struktur in Studie 1 hinsichtlich ihrer internen Konsistenz nicht untersucht.

3.5. Diskussion

Die Ergebnisse zeigen, dass die Modulgruppe *Produktbewertungen* eine sehr klare, gut abgegrenzte Struktur aufweist. Sie beschreibt das Produkt über die Dimensionen der instrumentellen und nicht-instrumentellen Qualitäten durch vielfältige Aspekte. Dagegen zeigt sich für das *Globalurteil*, die *Emotionen* und die *Nutzungsintention* ein weniger eindeutiges Bild.

Die Interpretation der faktorenanalytisch ermittelten, empirisch „guten" Daten des Moduls *Globalurteil* erwies sich als schwierig. Die Überlegung, das Globalurteil aus mehreren Items aufzubauen, wurde aus den Fragebögen AttrakDiff (Hassenzahl et al., 2003) und UEQ (Laugwitz et al., 2006) abgeleitet, die jeweils eine Dimension enthalten, die die Autoren als *Attraktivität* bezeichnen und die globale positiv-negativ Bewertungen über das Produkts abgibt. Jedoch wurde bei diesen beiden Fragebögen nicht geprüft, inwiefern die Attraktivität durch die anderen Dimensionen vorhergesagt wird und ob diese Darstellungsweise deshalb gerechtfertigt ist. Bei dem *Globalurteil* handelt es sich um eine das Produkt im Ganzen betrachtende Beurteilung (Götze, 2010). Deshalb erscheint nach der Prüfung der Daten und Betrachtung der anderen Fragebogenverfahren eine Vorgehensweise mit vielen Items zur Beschreibung des *Globalurteils* nicht ratsam zu sein. Denkbar wäre dagegen eine Single-Item-Variante, in der der Nutzer zum Beispiel mit einem Schie-

beregler auf einem Kontinuum von schlecht bis gut das Produkt insgesamt einschätzt. Aus diesem Grund wurde das *Globalurteil* vorerst aus der weiteren Betrachtung ausgeschlossen.

Für die *Emotionen* wurde auf das Circumplex-Modell von Russel (1979) zurückgegriffen, das die beiden unabhängigen Dimensionen Valenz und Arousal beschreibt (siehe Abbildung 12a). Die daraus abgeleiteten Items sollten entweder Arousal oder Valenz ausdrücken. Die faktorenanalytische Untersuchung zeigte jedoch, dass sich die Items auf andere Weise zusammen ordnen.

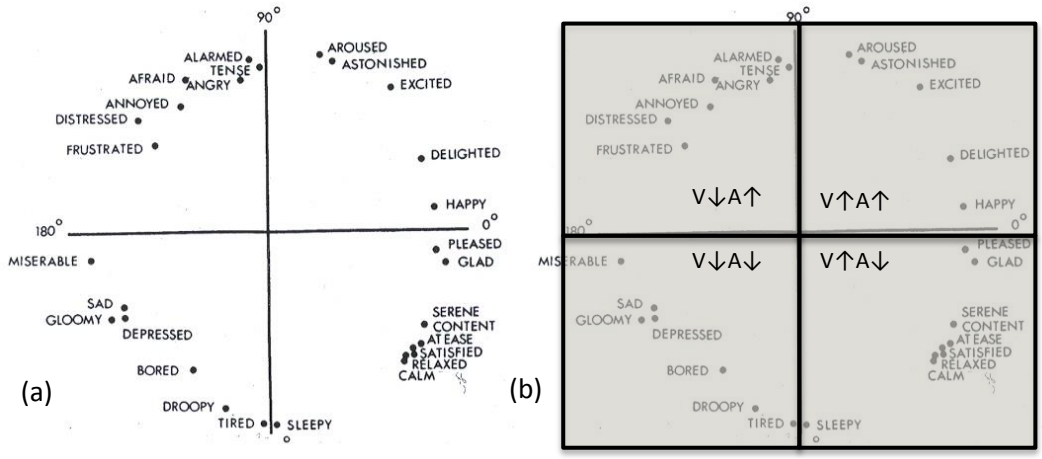


Abbildung 12. (a) Circumplex-Modell von Russell (1979), (b) Ableitung der vier Emotionsmodule aus dem Circumplex-Modell (1979); Darstellung aus Russell (1980, S. 1167).

Die Struktur der Modulgruppe Emotionen lässt vermuten, dass darin vier Module enthalten sind: (1) positive Valenz, hohes Arousal (V↑A↑), (2) positive Valenz, niedriges Arousal (V↑A↓), (3) negative Valenz, hohes Arousal (V↓A↑), (4) negative Valenz, niedriges Arousal (V↓A↓; siehe Abbildung 12b). Für die negative Valenz zeigt sich bereits eine Unterscheidung zwischen den beiden Arousal-Ausprägungen. Für die positive Valenz laden bisher sowohl hoch als auch niedrig aktivierte Items auf einem Faktor. Zur weiteren Untersuchung dieser Struktur wurde eine zweite Studie durchgeführt, in der die bereits hoch korrelierenden Items durch neue Items ergänzt wurden und bei der anhand einer Online-Erhebung eine faktorenanalytische Untersuchung der Daten erfolgte. Da das Ergebnis der Betrachtung der Nutzungsintention ebenfalls nicht eindeutig ist, wurden auch in der zweiten Studie weitere Items auf der Basis der beiden aus der Faktorstruktur in Tabelle 16 abgeleiteten Module *Nutzungsintention* und *Produktloyalität* hinzugefügt und untersucht.

4. Studie 2

Die zweite Studie konzentrierte sich ausschließlich auf die beiden Modulgruppen *Emotionen* und *Konsequenzen*, die in der ersten Rohversion des Fragebogens zu empirisch nicht eindeutigen Ergebnissen geführt haben.

Da die faktorenanalytische Betrachtung und Reliabilitätsprüfung der Modulgruppe *Produktbewertungen* eindeutige Ergebnisse geliefert hat, kann sie als vorerst erfolgreich abgeschlossen betrachtet werden.

Für die vier angestrebten Module in der Modulgruppe *Emotionen* (vgl. Abbildung 12b) wurden mittels Brainstormings neue Items entwickelt, die in Studie 2 getestet und faktorenanalytisch betrachtet werden sollen. Um Aussagen zur Gültigkeit der Module machen zu können, wurden in Studie 2 zur Abschätzung der inneren Kriteriumsvalidität die etablierten Emotionsfragebögen SAM (Lang, 1980) und LEMtool (Capota, van Hout & van der Geest, 2007) eingesetzt.

Auch für die Modulgruppe *Konsequenzen* wurden neue Items hinzugenommen, um die beiden Faktoren *Nutzungsintention* und *Produktloyalität* zukünftig messbar zu machen.

4.1. Methode

Das methodische Vorgehen in Studie 2 ist äquivalent zu Studie 1. Die Probanden wurden ebenfalls online befragt. Die Unterschiede in den Erhebungen beziehen sich vor allem auf das verwendete Material.

4.1.1. Material

Für Studie 2 wurde, analog zu Studie 1, die freie Applikation für Online-Befragungen *LimeSurvey* verwendet. Die Instruktionen entsprechen ebenfalls denen der ersten Studie (siehe Anhang B.1, vgl. Anhang A.2). Die faktorenanalytisch ermittelten Emotionsitems wurden beibehalten (vgl. Tabelle 15). In mehreren Brainstorming-Sitzungen wurden weitere Items entwickelt, die den jeweiligen Modulen (1) positive Valenz, hohes Arousal, (2) positive Valenz, niedriges Arousal, (3) negative Valenz, hohes Arousal oder (4) negative Valenz, niedriges Arousal zugeordnet sind (vgl. Tabelle 29, Anhang B.2).

Die aus dem ursprünglichen Modul *Nutzungsintention* abgeleiteten neuen Module *Nutzungsintention* und *Produktloyalität* wurden jeweils um weitere Items ergänzt (vgl. Tabelle 29, Anhang B.2).

Zur Überprüfung der inneren Kriteriumsvalidität (vgl. Abschnitt 2.3.2) der Module kamen in der Befragung zusätzlich zwei etablierte Emotionsfragebögen zum Einsatz. Zum einen erfolgte ein Vergleich mit den Dimensionen Valenz und Arousal des SAM (Lang, 1980), der in Abschnitt 2.3.4 ausführlich vorgestellt wurde. Zum anderen kam das LEMtool (Capota et al., 2007; vgl. Tabelle 2) zum Einsatz, das Emotionen kategorial erfasst und für die Beurteilung von Webseiten entwickelt wurde (siehe Abbildung 13).

Abbildung 13. Die acht Emotionsausdrücke des LEMtools, ©SusaGroup.

Aufgrund der eingeschränkten technischen Möglichkeiten von *LimeSurvey* konnte das LEMtool nicht in der originalen Art und Weise, sondern als semantisches Differential dargestellt werden. Es ergaben sich acht drei-stufige Items, wobei das linke Ende des Differentials als neutrales Gesicht und das rechte Ende als Gesicht mit einer prototypischen Basisemotion gezeigt wurde (vgl. Abbildung 14).

Abbildung 14. Darstellungsform des LEMtools in Studie 2.

Die Darstellung des SAM und LEMtools erfolgte randomisiert auf einer jeweils eigenen Befragungsseite.

Die Probanden wurden analog zur ersten Studie aufgefordert, sich ein Alltagsprodukt zu überlegen und in die vorgegebenen Produktgruppen einzuordnen. Die Gruppen entsprechen fast denen aus Studie 1. Es wurde lediglich der Fernseher auf Grund der Annahme, dass die Interaktion

mit dem Fernseher indirekter als bei den anderen Produkten geschieht, durch einen MP3-Player ersetzt.

Zusätzlich wurden Art und Bezeichnung des Produkts (z.B. Philips SENSEO) sowie die Nutzungsdauer in Jahren und Monaten erfasst. Als neues Element wurde die Frage nach der konkreten Situation, in der die Probanden das Produkt zuletzt verwendet haben, aufgenommen. Diese Frage soll insbesondere in Bezug auf die Beurteilung des subjektiven Emotionserlebens das Bewusstsein für das *konkrete* Produkt stärken und eine Bewertung der Produktgruppe im Allgemeinen ausschließen.

Die Erfragung von demografischen Daten wie Alter, Geschlecht und Bildungsstand erfolgte analog zu Studie 1. Die Probanden hatten auch bei dieser Studie die Möglichkeit an einem Gewinnspiel teilzunehmen, in dem fünf zufällig gewählte Personen 20 Euro erhielten. Dafür konnten sie ihre Email-Adresse angeben. Diese Angabe war freiwillig.

4.1.2. Untersuchungsdurchführung

Die Studie wurde in der ersten Februarwoche durchgeführt. Die Einladung erfolgte über das Probandentool des Graduiertenkollegs *prometei*, über das eine E-Mail an alle potentiellen Probanden verschickt wurde. Darin waren das Thema der Studie sowie die Verlosung beschrieben. Wie in der vorherigen Studie galt als einzige Anforderung, dass es sich bei den Probanden um deutsche Muttersprachler handelt. Die Teilnahme dauerte circa 10 bis 15 Minuten.

4.1.3. Stichprobe

Nach einer Woche hatten mit N = 238 bereits ausreichend Personen an der Befragung teilgenommen. Insgesamt begannen 323 Personen die Studie, die meisten Abbrüche erfolgten direkt auf der ersten Seite. Der Drop-Out ist mit 26% etwas geringer als in Studie 1 (36%).

Etwas mehr Frauen (N = 134) als Männer (N = 104) nahmen an der Befragung teil. Die Teilnehmer waren im Mittel 28.42 Jahre alt (SD = 9.12), wobei die beiden jüngsten 17 Jahre alt waren und der älteste 65 Jahre war. Das geringe Durchschnittsalter und die Verteilung des Bildungsabschlusses sind auch bei dieser Studie auf die Art der Probandengewinnung zurückzuführen. Die Hälfte der Probanden hat Abitur. Die zweitgrößte Gruppe umfasst die Hochschulabsolventen (39.5%). Einen Realschul- (8.8%) oder Hauptschulabschluss (1.7%) haben vergleichsweise wenig Probanden.

Die Teilnehmer nutzen das bewertete Produkt im Durchschnitt seit 2 Jahren und einem Monat (SD = 28.88). Das Minimum liegt bei 3 Tagen und das Maximum bei 29 Jahren.

Die Verteilung der Produktgruppen ähnelt der aus Studie 1. Fast die Hälfte der Probanden hat ihr Handy / Smartphone bewertet (*N* = 106, 44.5%), gefolgt von der Gruppe der Laptops / Computer (*N* = 75, 31.5%). Die übrigen Produkte wurden wieder wesentlich seltener gewählt (siehe Tabelle 20).

Tabelle 20. Von den Probanden gewählte Produktgruppen.

Produktgruppe	*N*	Anteil an Gesamtstichprobe in %
Handy / Smartphone	106	44.5
Laptop / Computer	75	31.5
MP3-Player	20	8.4
Waschmaschine	16	6.7
Kaffeemaschine	8	3.4
Smartphone App	5	2.1
Microsoft Office	5	2.1

4.2. Ergebnisse

Zuerst wurde die Dateneignung der Items der Modulgruppen *Emotionen* und *Nutzungskonsequenzen* geprüft. Für die *Emotionen* ergibt sich ein KMO-Koeffizient von .912, was auf eine sehr gute Dateneignung hinweist (vgl. Tabelle 1). Die MSA-Werte liegen zwischen .823 („Durch das Produkt fühle ich mich erschöpft.“) und .957 („Das Produkt erfreut mich.“), wobei die meisten Werte über .900 liegen und die einzelnen Items somit alle gut bis sehr gut geeignet sind. Der Bartlett-Test auf Sphärizität wurde signifikant ($p < .001$). Somit weisen die Items der Modulgruppe *Emotionen* eine sehr gute Eignung auf.

Die Voraussetzungen für eine Faktorenanalyse über die Items der *Konsequenzen* sind ebenfalls erfüllt. Der KMO-Koeffizient nimmt einen aussagekräftigen Wert von .857 an. Die MSA-Werte fallen mit .742 („Ich verwende das Produkt nur, wenn es unbedingt sein muss.“) bis .890 („Ich möchte das Produkt nicht mehr hergeben.“) geringer aus als die Werte der Items der *Emotionen*. Dennoch genügen die Items den notwendigen Anforderungen. Der Bartlett-Test wurde für diese Items ebenfalls signifikant ($p < .001$).

Wie in Studie 1 wurden auch die Items aus Studie 2 bezüglich ihrer Schwierigkeit und Trennschärfe analysiert. Die Itemschwierigkeit nimmt nur für das Item „Das Produkt macht mich zufrieden“ einen extremen Wert von .805 an, der jedoch nicht deutlich von dem gesetzten Maximalwert von .800 abweicht. Alle p-Werte (p von Dahl, 1971) der übrigen Items liegen zwischen .200 und .800 (vgl. Tabelle 29, Anhang B.2). Die Trennschärfe wurde wie in der vorherigen Studie modulweise berechnet. Alle Items erweisen sich als trennscharf (vgl. Tabelle 29, Anhang B.2).

4.2.1. Faktorenanalyse der Modulgruppe *Emotionen*

Wie bei den Modulen der *Produktbewertungen* wurde bei den Items der Modulgruppe *Emotionen* eine Hauptkomponentenanalyse mit Varimax-Rotation durchgeführt. Entsprechend der Empfehlung des MAP-Tests wurden vier Faktoren extrahiert. Die Items wurden durch Ausschluss aufgrund geringer Faktorladung oder Kommunalität auf drei je Faktor reduziert. Die endgültige Struktur zeigt Tabelle 21.

Tabelle 21. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Emotionen.*

	1	2	3	4	Kommunalität
Das Produkt nervt mich.	.865				.813
Das Produkt frustriert mich.	.851				.792
Das Produkt verärgert mich.	.831				.782
Das Produkt beruhigt mich.		.887			.794
Das Produkt entspannt mich.		.819			.758
Durch das Produkt fühle ich mich ausgeglichen.		.784			.671
Das Produkt begeistert mich.			.857		.779
Das Produkt macht mich glücklich.			.808		.754
Das Produkt erfreut mich.			.727		.677
Durch das Produkt fühle ich mich passiv.				.793	.641
Das Produkt macht mich müde.				.781	.692
Durch das Produkt fühle ich mich erschöpft.				.761	.712
Aufgeklärte Varianz	**20.485**	**18.869**	**17.956**	**16.583**	**Gesamt: 73.866**

Wie bereits aus Studie 1 erkennbar (vgl. Tabelle 15), unterscheiden sich die vier Module entsprechend ihrer Valenz und ihrer Aktivierung. Das erste Modul drückt eine negative Valenz mit gleichzeitiger hoher Aktivierung aus (Beispiel: „Das Produkt verärgert mich.“). Das zweite Modul fasst Items mit wenig Aktivierung und positiver Valenz zusammen (Beispiel: „Das Produkt beruhigt mich.“). Im dritten Modul kommen hohe Aktivierung und positive Valenz zum Ausdruck (Beispiel: „Das Produkt begeistert mich“). Das vierte Modul umfasst negative Valenz mit geringer Aktivierung (Beispiel: „Durch das Produkt fühle ich mich passiv.“). Mit diesem Modell können etwa 73.9 % der insgesamt aufgetretenen Varianz erklärt werden.

4.2.2. Faktorenanalyse der Modulgruppe *Konsequenzen*

Bezüglich der Ergebnisse zur *Nutzungsintention* in Studie 1 wurde die These aufgestellt, dass sich diese in die beiden Subdimensionen *Nutzungsintention* und *Produktloyalität* unterteilen lässt. Die Faktorenanalyse erzielte zwei Faktoren, deren inhaltliche Struktur zu den beiden erdachten Dimensionen passt. Erneut kam es zum Ausschluss von Items mit geringer Faktorladung oder Kommunalität, sodass jeweils drei Items einen Faktor bilden. Das Modell klärt 60.64% der aufgetretenen Varianz auf. Tabelle 22 stellt die Faktorstruktur dar.

Tabelle 22. Faktorstruktur nach der Hauptkomponentenanalyse mit Varimax-Rotation der ausgewählten Items der Modulgruppe *Konsequenzen*.

	1	2	Kommunalität
Ich würde das Produkt gegen kein anderes eintauschen.	.819		.671
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	.790		.642
Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.	.763		.611
Wenn ich könnte, würde ich das Produkt täglich nutzen.		.769	.510
Ich verwende das Produkt nur, wenn es unbedingt sein muss.		-.761	.579
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.		.704	.625
Aufgeklärte Varianz	**32.070**	**28.570**	**Gesamt: 60.639**

4.2.3. Reliabilität

Wie in Studie 1 wurde auch in Studie 2 die Reliabilität modulweise mittels Cronbach's α geprüft. In Tabelle 23 sind die modulweise bestimmten Werte der *Emotionen* dargestellt. Diese befriedigenden bis guten Werten des α-Koeffizienten (vgl. Nunnally & Bernstein, 1994) in Kombination mit den mittleren bis hohen Interkorrelationen der Items (r = .418 bis r = .715; vgl. Tabelle 30, Anhang B.3) zeigt, dass die Module und daraus schließend auch die Modulgruppe *Emotionen* insgesamt reliabel ist.

Tabelle 23. α-Koeffizient der vier Module der Modulgruppe *Emotionen*.

Module	Cronbach's α
Positive Valenz, aktiviert	.809
Positive Valenz, deaktiviert	.815
Negative Valenz, aktiviert	.872
Negative Valenz, deaktiviert	.744

Für die Modulgruppe *Konsequenzen* wurde die Reliabilität ebenfalls modulweise bestimmt (siehe Tabelle 24). Die Items der *Nutzungsintention* korrelieren nur mäßig (r = .283 bis r = .388; vgl. Tabelle 31, Anhang B.3), wohingegen die Items der *Produktloyalität* einen deutlich stärkeren Zusammenhang zeigen (r = .448 bis r = .471). Der Reliabilitätskoeffizient nimmt für die *Produktloyalität* einen befriedigenden Wert an, weshalb dieses Modul als hinreichend reliabel bezeichnet werden kann. Der α-Wert sowie die Interitemkorrelationen der *Nutzungsintention* sind zu gering als dass das Modul als reliabel bezeichnet werden kann (vgl. Nunnally & Bernstein, 1994).

Tabelle 24. α-Koeffizient der zwei Module der Modulgruppe *Konsequenzen.*

Module	Cronbach's α
Nutzungsintention	.617
Produktloyalität	.714

4.2.4. Validität

Im Rahmen der zweiten Studie kamen zur ersten Validitätsabschätzung der Modulgruppe *Emotionen* existierende Emotions-Fragebögen zum Einsatz. Die Validität gibt Auskunft darüber, ob das Verfahren tatsächlich das misst, was es zu messen vorgibt (Bortz & Döring, 2006). Allgemein lassen sich verschiedene Konzepte unterscheiden, wobei einige im Zusammenhang mit dem UEQ (Laugwitz et al., 2006) vorgestellt wurden (vgl. Abschnitt 2.3.2). Für die Gruppe der Emotionen wurde die innere kriteriumsbezogene Validität untersucht. Dazu wurde für jedes Emotionen-Modul ein Mittelwert berechnet, um diesen anschließend mit den Dimensionen des SAM und denen des LEMtools zu korrelieren (vgl. Bortz & Döring, 2006). Da für Ratingskalen eine Intervallskalierung angenommen werden kann (Bortz & Döring, 2006), wurde die Pearson-Korrelation angewendet. Tabelle 25 veranschaulicht die Korrelationskoeffizienten. Signifikant korrelierende Paare sind mit Sternen (*) markiert.

Tabelle 25. Korrelation der *Emotionen*-Module mit den Dimensionen des SAM und des LEMtools.

		V ↑A↑	V ↑A↓	V ↓A↑	V ↓A↓
SAM	Valenz	.548***	.294***	-.381***	-.362***
	Arousal	-.024	.047	.135*	.183**
LEMtool	Desire	.547***	.250***	-.312***	-.267***
	Fascination	.436***	.216**	-.242***	-.135*
	Satisfaction	.482***	.246***	-.328***	-.307***
	Joy	.505***	.330***	-.259***	-.306***
	Disgust	-.296***	-.261***	.275***	.219**
	Sadness	-.242***	-.139*	.276***	.317***
	Dissatisfaction	-.254***	-.164*	.417***	.264***
	Boredom	-.224***	-.152**	.271***	.349***

*** p < .001
** p < .01
* p < .05

Zwischen den *Emotionen*-Modulen und der *Valenz*-Dimension des SAM zeigen sich die erwarteten Korrelationen, die zwar nur mäßig stark, jedoch alle hoch signifikant, sind. Dies ist durch die große Stichprobe (*N* = 238) begründet, bei der die bereits kleine Zusammenhänge zu signifikanten Ergebnissen führen (Bortz, 2005). Die Korrelation mit der *Arousal*-Dimension zeigt kein eindeutiges Ergebnis. Insbesondere für die Module mit positiver Valenz sind die Korrelationen mit $r_{\text{SAM_Arousal, V↑A↑}}$ = -.024 und $r_{\text{SAM_Arousal, V↑A↓}}$ = .047 sehr gering. In Abschnitt 2.3.4 wurde bereits erläutert, dass die Arousal-Dimension des SAM in vielen Untersuchungen keine annehmbaren Ergebnisse geliefert hat. Der SAM wurde im Gegensatz zu den vorliegenden Modulen nicht anhand von interaktiven Produkten im Bereich Mensch-Maschine-Kontext, sondern ursprünglich für den Einsatz im klinischen Bereich entwickelt (Lang, 1980). Deshalb ist seine Aussagekraft in dem hier vorliegenden Bereich grundsätzlich fraglich.

Die Dimensionen des LEMtools zeigen ein eindeutigeres Korrelationsmuster mit den *Emotionen*-Modulen. Die positive Emotionen darstellenden Dimensionen des LEMtools korrelieren positiv mit den beiden Modulen positiver Valenz und negativ mit den beiden Modulen negativer Valenz. Bei den negativen Dimensionen des LEMtools ist es erwartungsgemäß genau umgekehrt.

Die positiven LEMtool-Emotionen (*Desire, Fascination, Satisfaction, Joy*) korrelieren besonders hoch mit dem Modul *positiver Valenz und hoher Aktivierung*. Die schwächsten Korrelationen liegen zwischen den negativen LEMtool-Emotionen (*Disgust, Sadness, Dissatisfaction, Boredom*) und dem Modul *hoher Valenz und geringer Aktivierung* vor. Aus den Dimensionen des LEMtools ist jedoch nicht der Grad der Aktiviertheit herauszulesen, sodass dieser nicht in die Analyse der Korrelationen einbezogen werden kann.

Zusammenfassend zeigt insbesondere das Modul *positiver Valenz und hoher Aktivierung* starke Korrelationen mit den Dimensionen der beiden anderen Vergleichsfragebögen. Die anderen Module korrelieren ebenfalls wie erwartet mit den Dimensionen der Emotions-Fragebögen, jedoch tendenziell schwächer. Aufgrund dieser Ergebnisse kann daher angenommen werden, dass eine ausreichende Kriteriumsvalidität für das Modul *Emotionen* vorliegt.

5. Diskussion

Abschließend werden die Ergebnisse der Entwicklung eines Fragebogens zur Messung der User Experience diskutiert. Dazu werden die wesentlichen Resultate aus Studie 1 und 2 zunächst zusammengefasst, wobei insbesondere auf das vorläufige Endmodell des Fragebogens eingegangen wird. Die Gütekriterien der Klassischen Testtheorie (KTT) werden zusammenhängend berichtet. Da die Validitätsbestimmung des Fragebogens kein zentraler Gegenstand dieser Abhandlung ist, erfolgt ein hauptsächlich qualitativer Vergleich mit weiteren Fragebogenverfahren. Abschließend wird die praktische Relevanz des entwickelten Fragebogens dargestellt sowie ein Ausblick auf weitere Entwicklungsschritte des Fragebogens gegeben.

5.1. Diskussion der Fragebogenstruktur

Das Ziel der ersten Studie bestand zunächst in der Auswahl geeigneter Items für die aus dem CUE-Modell abgeleiteten Module. Dafür wurden die Items modulweise faktorenanalytisch mittels Hauptkomponentenanalysen untersucht und so reduziert, dass jeweils drei Items pro Modul verblieben.

Die *instrumentellen Qualitäten Benutzbarkeit* und *Nützlichkeit* konnten problemlos von einer mehrfaktoriellen auf eine einfaktorielle Struktur mit drei Items reduziert werden. Diese drei Items beschreiben verschiedene Aspekte der operationalisierten Komponenten aus dem CUE-Modell (Thüring & Mahlke, 2007).

Für die *visuelle Ästhetik* ergab sich zunächst eine zweifaktorielle Struktur, wobei positiv formulierte Items auf dem einen und negativ formulierte Items auf dem anderen Faktor luden. Es wurden nur positiv formulierte Items gewählt, die das Erleben von Schönheit und Kreativität des Produktes darstellen.

Identifikation / Wert war das zunächst zweite Modul der Komponente *nicht-instrumentelle Qualitäten*. Die faktorenanalytische Betrachtung ergab zwei unabhängige Faktoren, aus denen die beiden Module *Status* und *Bindung* abgeleitet wurden. Sie erheben, inwiefern das Produkt zur individuellen und sozialen Identität des Nutzers beiträgt.

Die beiden Module *Valenz* und *Arousal* der *Emotionen* wurde zunächst aus dem Circumplex-Modell von Russell (1979) abgeleitet. Die faktorenanalytische Betrachtung der Daten aus Studie 1 zeigte jedoch, dass, anders als im Circumplex-Modell, nicht die zwei unabhängigen Dimensionen *Valenz* und *Arousal* nachgewiesen werden können, sondern zunächst drei Faktoren entstanden, die jeweils verschiedene Ausprägungen von *Valenz* und *Arousal* ausdrückten. Die erneute empiri-

sche Betrachtung mit neu entwickelten Items in Studie 2, führte zu einer vierfaktoriellen Struktur mit einer jeweiligen Kombination hoch oder niedrig ausgeprägter Valenz und Aktivierung.

Für die Module *Nutzungsintention* und *Globalurteil* ergaben sich in Studie 1 keine eindeutigen Ergebnisse. Die Faktorenanalyse der *Nutzungsintention* ergab eine zweifaktorielle Struktur, die zwei eigenständige Bereiche von Konsequenzen beschreibt: *Nutzungsintention*, im Sinne einer Verhaltensabsicht, und *Produktloyalität*, im Sinne einer Wahl zwischen Produktalternativen. Diese Struktur wurde in der zweiten Studie näher untersucht, bei der ein ergänzter Itempool für beide Modulgruppen getestet wurde. Die Prüfung der Dimensionalität der Items der *Nutzungsintention* ergab erneut eine zweifaktorielle Lösung, die durch die zu den Items aus Studie 1 hinzugefügten Items eine klarere Struktur aufweist. Analog zu den Produktbewertungen wurden die Items auf drei pro Skala reduziert. Das *Globalurteil* wurde vollständig aus dem Fragebogen-Modell entfernt, da die Mehr-Item-basierte Form der Abfrage als nicht sinnvoll betrachtet wurde (vgl. Abschnitt 3.5). Stattdessen wird eine Single-Item-Variante präferiert, auf die in Abschnitt 5.7 eingegangen wird.

Modulgruppen

Im Zuge der Fragebogenkonstruktion konnten drei unabhängige Modulgruppen für die vorgestellten Module identifiziert werden: *Produktbewertungen*, *Emotionen* und *Konsequenzen* (siehe Abbildung 15).

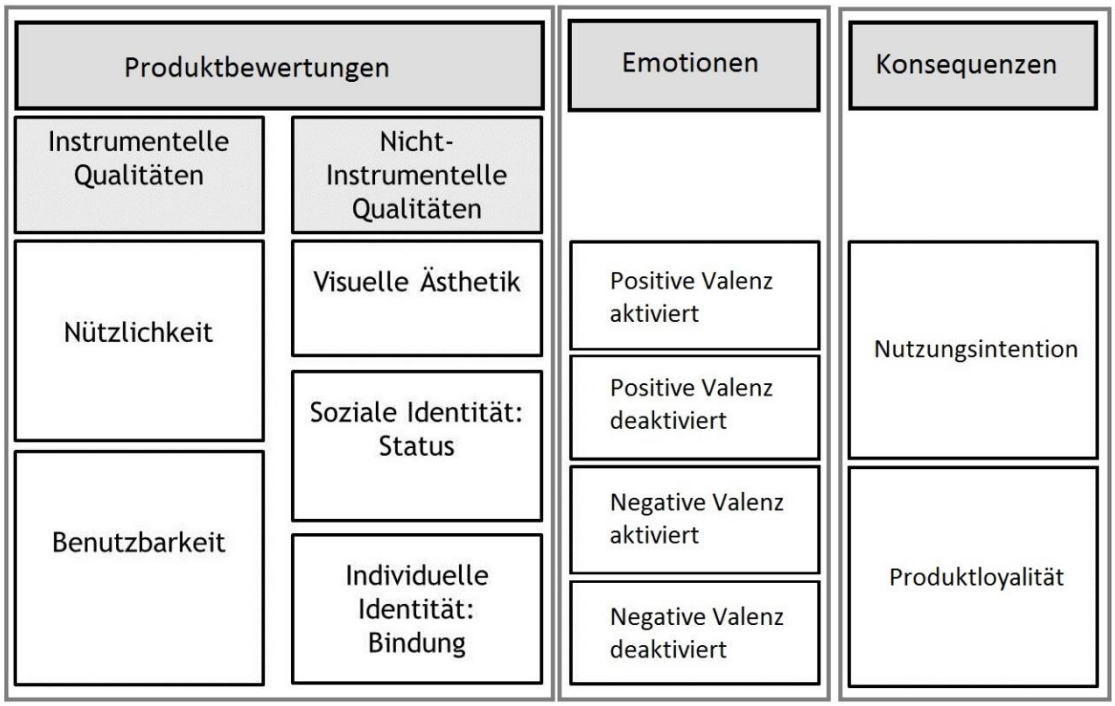


Abbildung 15. Finale Modulgruppen und Module; eigene Darstellung.

Die Modulgruppe *Produktbewertungen* bezieht sich auf die *instrumentellen* und *nicht-instrumentellen Qualitäten* aus dem CUE-Modell (Thüring & Mahlke, 2007). Sie beinhaltet fünf faktorenanalytisch ermittelte Skalen, die verschiedene Charakteristika eines interaktiven Produktes abdecken, anhand derer ein Nutzer ein Produkt erlebt. Den größten Anteil an der Bewertung hat das Modul *visuelle Ästhetik* (16.45% erklärter Varianzanteil). Es weist zudem mit $\alpha = .870$ die größte interne Konsistenz auf. Dieses Ergebnis deckt sich mit Norman (2004), der darlegt, dass das äußere Erscheinungsbild eines interaktiven Produktes das Erste sei, was ein Nutzer wahrnimmt und dass das visuelle Erleben bereits Schlüsse auf andere Qualitäten, wie *Benutzbarkeit* und *Nützlichkeit*, zulasse. Die übrigen Module werden erst relevant, wenn der Nutzer bereits mit dem Produkt interagiert und gehen entsprechend ihres aufgeklärten Varianzanteils in etwa gleich stark in die Produktbewertung ein. Die *Benutzbarkeit* und die *Nützlichkeit* zeigen auf, wie effektiv und effizient der Nutzer das interaktive Produkt zum Erreichen eines Zieles einschätzt. Die Identität, die das Produkt einer Person verleiht, wird dadurch deutlich, wie sehr sich die Person mit einem Produkt verbunden fühlt (*Bindung*) und welchen sozialen *Status* über das Produkt vermittelt wird. Die Modulgruppe *Produktbewertungen* umfasst 15 Items. Die Module wurden mittels Cronbach's α auf ihre interne Konsistenz hin untersucht und zeigten sich alle hinreichend reliabel.

Die Modulgruppe *Emotionen* wurde zunächst aus dem Circumplex-Modell von Russell (1979) abgeleitet. Die faktorenanalytische Betrachtung zeigte jedoch, dass anders als im Circumplex-Modell, nicht die zwei unabhängigen Dimensionen Valenz und Arousal nachgewiesen werden können, sondern vier Module entstehen, die jeweils verschiedene Ausprägungen von Valenz und Arousal ausdrücken. Das Modul, das *negative Valenz und hohes Arousal* umfasst, hat den größten Einfluss auf das emotionale Erleben eines interaktiven Produkts (20.49% erklärter Varianzanteil). Alle vier Emotionsausprägungen gehen in die Modulgruppe Emotionen ein und stellen den affektiven Aspekt des Nutzungserlebens umfassend dar. Die Validitätsabschätzung der Modulgruppe *Emotionen* aus Basis von Korrelationen zu anderen Emotionsfragebögen (SAM, Lang, 1980; LEM-tool, Capota et al., 2007) deutet darauf hin, dass die Items das emotionale Erleben valide erfassen.

Die *Konsequenzen* bilden die dritte Modulgruppe, die sich, entsprechend dem CUE-Modell, aus den Aspekten des Erlebens ergibt. Die *Nutzungsintention* bildet final ein Modul neben einem zweiten, der *Produktloyalität*. Die *Produktloyalität* trägt etwas mehr zur Gesamtvarianz bei als die *Nutzungsintention* (32.07% gegenüber 28.57% erklärtem Varianzanteil). Während das Modul *Produktloyalität* hinreichend reliabel ist, ergab sich für das Modul *Nutzungsintention* keine ausreichende Reliabilität. Es ist denkbar, dass die unterschiedliche Polung der Items das Urteil der Nut-

zer beeinflusst. Deshalb sollten die Items dieses Moduls entsprechend angepasst und eine erneute Reliabilitätsprüfung vorgenommen werden.

Die vorläufige Endversion des Fragebogens besteht aus 33 Items, die jeweils in eine der drei Modulgruppen eingeordnet sind (siehe Abhang C).

5.2. Gütekriterien

Die guten bis sehr guten Werte von Cronbach's α als Maß für die Reliabilität zeigen zusammen mit den substantiellen Korrelationen der Items innerhalb eines Moduls (siehe Anhang A.5, B.3), dass die unabhängigen Module eine hinreichende Reliabilität aufweisen.

Im Rahmen der vorliegenden Arbeit wurde die Validität zunächst nur für die Modulgruppe Emotionen überprüft. Es zeigt sich ein deutlicher Zusammenhang mit der Skala *Valenz* des SAM (Lang, 1980) und mit den Skalen des LEMtools (Capota et al., 2007), was darauf hinweist, dass es sich bei dem neu entwickelten *Emotion*smodul um ein valides Messinstrument handelt.

Neben der Reliabilität und Validität ist die Objektivität ein weiteres Hauptgütekriterium der KTT. Ein Verfahren gilt dann als objektiv, wenn das Testergebnis vom Testleiter und -auswerter unbeeinflusst ist (Bortz & Döring, 2006). Bei der Objektivität werden drei Arten unterschieden: Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität.

Die Durchführungsobjektivität gibt den Grad der Beeinflussung des Probanden durch den Testleiter an (Bortz & Döring, 2006). Durch standardisierte Instruktionen und Rückmeldungen kann eine hohe Objektivität dieser Art erreicht werden. Da der Fragebogen in den beiden vorgestellten Studien als standardisierte Online-Befragung vorgegeben wurde, kann er hinsichtlich der Durchführung grundsätzlich als objektiv betrachtet werden. Einschränkend wirkt die fehlende Standardisierung der Untersuchungssituation. Die Probanden können verschiedene, das Antwortverhalten beeinflussende, Verhaltensweisen während der Bearbeitung des Fragebogens gezeigt haben, wie Unterbrechungen oder Nebentätigkeiten. Mit einer Laboruntersuchung könnte das Verhalten der Nutzer kontrolliert werden.

Bei einer gelungenen Auswertungsobjektivität darf die Höhe des Testwertes nicht vom Auswerter beeinflusst sein (Bortz & Döring, 2006). Die Skalenstufen sind eindeutig bestimmt und werden jedem Probanden gleich vorgegeben. Der Auswerter polt die Items mit negativer Formulierung (z.B. „Es dauert zu lange, um die Funktionen des Produkts zu erlernen.“) um, sodass mit steigendem Testwert die Stärke der Zustimmung zunimmt. Anschließend wird modulweise der Mittelwert gebildet. Aufgrund dieses standardisierten Vorgehens ist die Auswertungsobjektivität hier ebenfalls gegeben.

Aufgrund des aktuellen Entwicklungsstandes des Fragebogens, kann eine Interpretation der ermittelten Daten in Bezug auf eine Vergleichsstichprobe (Normierung, Bortz & Döring, 2006) noch nicht erfolgen. Deshalb ist die Interpretationsobjektivität des Verfahrens nicht relevant.

Die Items wurden von anfänglich 67 auf eine wesentlich geringere Anzahl von 33 reduziert. Dadurch kann von einer kürzeren Bearbeitungszeit ausgegangen werden. Zudem handelt es sich um einen Fragebogen, der am Computer bearbeitet wird. Dadurch werden die durch die Nutzer abgegebenen Antworten zum einen direkt gespeichert und müssen nicht extra, wie es bei einem Paper-and-Pencil-Verfahren nötig ist, in den Computer übertragen werden. Zum anderen werden Nutzer auf fehlende Antworten aufmerksam gemacht. Dies führt zu einer hohen Ökonomie (Nebengütekriterium laut Lienert & Raatz, 1998) sowohl für den Nutzer als auch für den Testleiter und den Auswerter.

Aus der Betrachtung der Reliabilität, Validität, Objektivität und Ökonomie kann zusammenfassend geschlossen werden, dass es sich bei dem Fragebogen um ein qualitativ hinreichend geprüftes Instrument zur validen Erfassung des Nutzungserlebens interaktiver Produkte handelt.

5.3. Qualitativer Vergleich mit weiteren Fragebögen

Der Vergleich zwischen dem entwickelten Fragebogen und etablierten User Experience Fragebögen kann nur qualitativ erfolgen, da bei den beiden Online-Studien die Erhebung von Daten für die Konstruktion und Reliabilitätsprüfung des eigenen Fragebogens im Vordergrund stand. Der qualitative Vergleich erfolgt mit den Veröffentlichungen zur Konstruktion der beiden meist zitierten User Experience Fragebögen AttrakDiff (Hassenzahl et al., 2003) und UEQ (Laugwitz et al., 2006). Die hier präsentierten Anmerkungen können nur Hinweis über die Validität des Fragebogens geben.

Dimensionen

Der vorliegende Fragebogen beinhaltet, ebenso wie der AttrakDiff (Hassenzahl et al., 2003) und der UEQ (Laugwitz et al., 2006), Skalen, die instrumentelle (pragmatische) Qualitäten und nicht-instrumentelle (hedonische) Qualitäten beschreiben. Während der AttrakDiff nur eine Skala zur pragmatischen Qualität erfasst, liegen bei dem UEQ drei und bei dem hier entwickelten Fragebogen zwei Skalen vor. Dadurch können differenziertere Aussagen über aufgabenbezogene Aspekte, wie die erlebte Benutzbarkeit eines Produktes, getroffen werden. Laugwitz et al. (2006) kritisieren die zu starke Fokussierung des AttrakDiff auf die Hedonik, die mit den beiden Skalen Stimulation und Identifikation erhoben wird. Der UEQ enthält ebenfalls zwei Skalen zur Erfassung hedonischer Aspekte des Erlebens. Die hedonischen Skalen der beiden Fragebögen korrelieren nur mäßig, was daran liegen kann, dass die hedonischen Skalen des UEQ stärker auf die Designqualität fokussie-

ren, wohingegen die Skalen des AttrakDiff, ebenso wie die beiden Module *Status* und *Bindung* des entwickelten Fragebogens, jeweils die soziale Kommunikation mithilfe des Produkts bzw. die Bindung zum Produkt erfassen. Es ist denkbar, dass die hedonischen Skalen des UEQ hoch mit dem Modul *visuelle Ästhetik* korrelieren. Die Modulgruppe *Emotionen* wurde erstmals in einen User Experience Fragebogen integriert. Deshalb war hier eine Abschätzung der Validität besonders wichtig. Die Überprüfung mit dem SAM (Lang, 1980) und dem LEMtool (Capota et al., 2007) deuten auf ausreichende innere kriteriumsbezogene Validität hin. Die Modulgruppe *Konsequenzen*, die aus den Modulen *Nutzungsintention* und *Produktloyalität* besteht, ist ebenfalls neuartig und kann daher nicht unmittelbar mit dem AttrakDiff und dem UEQ verglichen werden. Diese beiden Fragebögen enthalten jeweils zusätzlich eine Skala, die die Attraktivität des Produktes insgesamt betrachtet erfasst. Laugwitz et al. (2006) stellen fest, dass die beiden Skalen hoch korrelieren. Eine solche Skala, die ein Globalurteil über das interaktive Produkt erfasst, könnte zukünftig in den entwickelten Fragebogen als Single-Item-Variante eingehen.

Stichprobe

Bei allen drei Fragebögen ist kritisch anzumerken, dass die Fragebogenentwicklung auf Daten von überwiegend studentischen Stichproben basiert. Dies hat ein sehr niedriges Durchschnittsalter zur Folge, welches nicht repräsentativ für die Bevölkerung Deutschlands ist. Über Hochschulverteiler können, gerade für Online-Erhebungen, schnell viele Probanden gewonnen werden, was aber ältere Probandengruppen nahezu ausschließt. Bei der Revidierung des Fragebogens sollte deshalb auf eine ausgewogene Altersverteilung geachtet werden.

Reliabilität

Alle drei Fragebögen wurden mittels Cronbach's α skalen- bzw. modulweise auf ihre interne Konsistenz hin untersucht. Obwohl im AttrakDiff und UEQ keine Interkorrelationen zwischen den Items berichtet werden, weisen die α-Werte auf eine hinreichende Reliabilität aller Verfahren hin.

5.4. Auswertungsbeispiel

Der Fragebogen soll zukünftig in der Lage sein, zwischen verschiedenen Gestaltungslösungen zu unterscheiden (diskriminative Validität, Bortz & Döring, 2006), zum Beispiel Lösungen, bei denen instrumentelle und / oder nicht-instrumentelle Qualitäten unterschiedlich hoch ausgeprägt sind.

Abbildung 16 zeigt, wie die *Produktbewertungen* mittels einer Spinnennetz-Darstellung veranschaulicht werden können, indem jedes der fünf Module eine Ecke des Netzes darstellt. Dadurch sind alle Module auf einen Blick zu erfassen und können schnell augenscheinlich interpretiert werden.

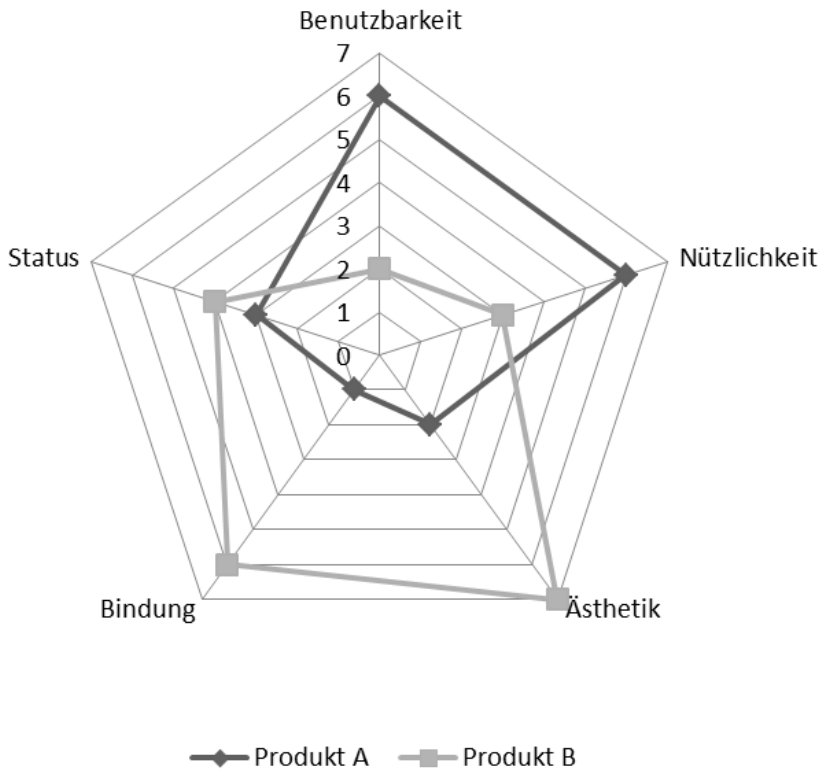


Abbildung 16. Darstellungsbeispiel für die Ergebnisse zweier prototypischer Produkte A und B bezüglich der einzelnen Module der Modulgruppe *Produktbewertungen*.

Für die Auswertung je Modul sind die Skalierungen von innen nach außen entsprechend der Antwortskalen von 1 (*lehne völlig ab*) bis 7 (*stimme völlig zu*) abgetragen.

Die beispielhafte Veranschaulichung erfolgt mit zwei prototypischen Produkten, die unterschiedliche Aspekte der Produktbewertung in den Mittelpunkt.

Tatsächliche Bewertungen aus Studie 1 wurden nicht betrachtet, weil es sich um eine Post-hoc-Betrachtung handelt und deshalb keine Hypothesen über die Unterschiede formuliert werden können. Außerdem unterscheiden sich die Teilstichproben deutlich (z.B. Handy / Smartphone: N = 100, Waschmaschine: N = 16), was die Aussagekraft ebenfalls einschränkt.

Für den Vergleich sind modulweise die Mittelwerte dargestellt. Produkt A zeigt hohe Werte bezüglich der *Benutzbarkeit* und der *Nützlichkeit* und dagegen niedrige Werte für die Produkt*bindung* und die *visuelle Ästhetik*. *Status* ist mit einem Mittelwert von 4.0 mäßig ausgeprägt. Das Produkt scheint den Nutzern als reiner Gebrauchsgegenstand zum Erledigen von Aufgaben zu dienen, der vor allem auf die *instrumentellen Qualitäten* abzielt. Dagegen wird Produkt B als sehr *visuell ästhetisch* eingeschätzt und es zeigt sich eine starke *Bindung* des Nutzers zum Produkt. Die *instrumentellen Qualitäten Benutzbarkeit* und *Nützlichkeit* sind deutlich geringer. Um tatsächliche Unterschiede zwischen den Produkten A und B aufzuzeigen, müsste eine Signifikanzprüfung vorgenommen werden.

Die *Emotionen* können über eine Vier-Felder-Tafel (siehe Abbildung 17) dargestellt werden. Dabei deckt jedes der vier Module ein Feld in der Tabelle ab, sodass auf einen Blick ein Urteil über das Emotionserleben abgeleitet werden kann.

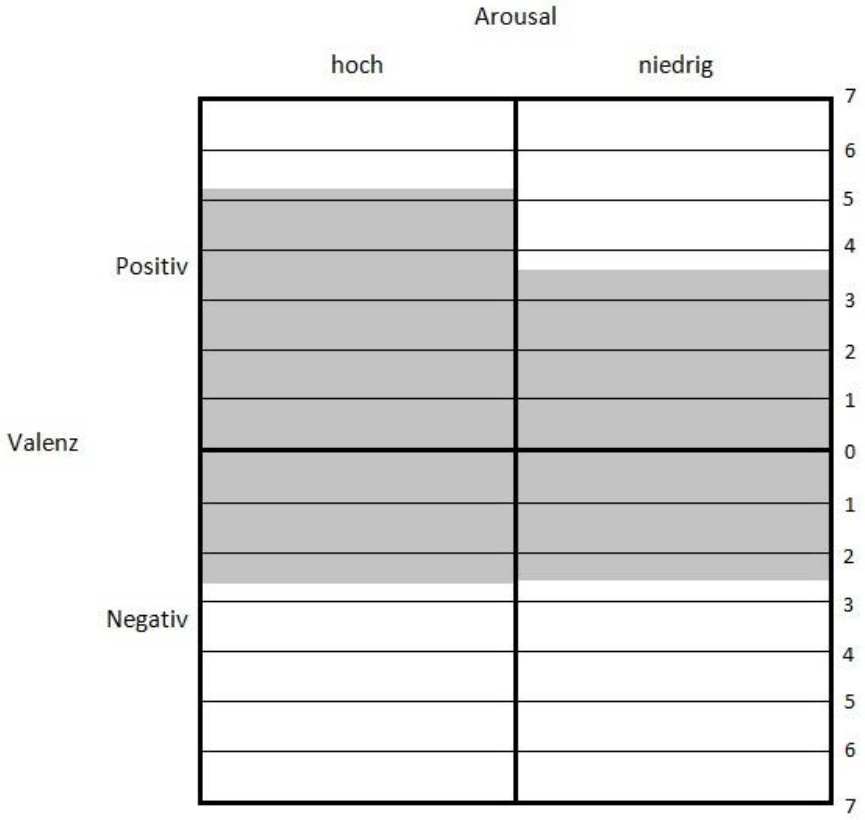


Abbildung 17. Darstellungsbeispiel für die Ergebnisse der Modulgruppe *Emotionen* am Beispiel der Mittelwerte der Produktgruppe Smartphone / Handy.

Jedes Feld ist siebenfach unterteilt, was die Skalenstufen widerspiegelt. Es wird jeweils der Mittelwert jedes Moduls je Feld abgetragen, indem die Zeilen bis zum Wert des Mittelwertes farbig gefüllt werden. In dem in Abbildung 17Abbildung 17. Darstellungsbeispiel für die Ergebnisse der Modulgruppe *Emotionen* am Beispiel der Mittelwerte der Produktgruppe Smartphone / Handy. dargestellten Beispiel sind die Mittelwerte der Bewertungen von Nutzern bezüglich ihres Handys bzw. Smartphones abgetragen. Es zeigt sich, dass der Mittelwert des Moduls *positiver Valenz und hohen Arousals* am höchsten ausfällt. Auch hier müssten die Unterschiede zwischen den Modulen auf ihre Signifikanz geprüft werden. Augenscheinlich kann jedoch aus den Daten geschlossen werden, dass die Probanden positive, anregende Emotionen bei der Interaktion mit ihrem Handy erleben.

5.5. Kritik

Im Rahmen dieser Abhandlung wurden einige Kritikpunkte, wie die fehlende Interpretationsobjektivität (siehe Abschnitt 5.2) und die überwiegend aus Studenten bestehende Stichprobe (siehe Abschnitt 5.3), bereits genannt.

Eine weitere mögliche Kritik ist das in beiden Online-Studien gewählte methodische Vorgehen Probanden aufzufordern, ein eigenes Produkt zu beurteilen. Insbesondere bei Produkten, wie einem Handy oder Laptop, ist davon auszugehen, dass die bereits getätigte Kaufentscheidung die Beurteilung des Produkts nachträglich beeinflusst. Dieses Phänomen kann mit dem Konstrukt der kognitiven Dissonanz und einer daraus resultierenden Dissonanzreduktion erklärt werden (Festinger, 1957; Piesbergen & Sarubin, 2011; Schuchard-Ficher, 1979). Kognitive Dissonanz beschreibt ein unangenehmes Gefühl als Resultat sich widersprechender Kognitionen. Im Fall der Produktbewertung kann Dissonanz zum Beispiel dann auftreten, wenn ein Nutzer ein Produkt gekauft hat und anschließend feststellt, dass es seine Erwartungen nicht erfüllt. Während der Dissonanzreduktion werden beispielsweise positive Aspekte des Produktes in den Vordergrund gestellt, die zuvor unwichtig waren oder das Produkt wird positiver bewertet wird um den Kauf zu rechtfertigen (Schuchard-Ficher, 1979). Solche Kognitionen können sich wiederum auf die Bewertung in einem Fragebogen auswirken. Piesbergen und Sarubin (2011) haben dazu die Münchener Dissonanzskala entwickelt, einen Fragebogen, der Dissonanz nach einer Kaufentscheidung erfasst. Um solche Effekte auszuschließen, ist deshalb eine faktorenanalytische Überprüfung von Daten nötig, die mit den Probanden zuvor unbekannten Produkten gewonnen wurden, zum Beispiel in Form eines Laborexperimentes mit vorgegebenen Produkten.

Die Probanden wurden in Studie 1 nicht darum gebeten, sich an eine konkrete Situation, in der sie das Produkt zuletzt verwendet haben, zu erinnern und zu beschreiben. Dadurch kann nicht sichergestellt werden, dass die Personen sich tatsächlich und möglichst ausschließlich auf das Produkt konzentriert haben. In Studie 2 wurde daher die Frage nach der konkreten Situation, in der sie das Produkt verwendet haben, eingeführt. In zukünftigen Untersuchungen könnten konkrete Aufgaben vorgegeben werden, die die Probanden mit dem Produkt lösen sollen. Diese Aufgaben sollten darauf abzielen, dass sich der Proband sowohl über instrumentelle als auch über nicht-instrumentelle Aspekte des Produkts bewusst ist und diese Erfahrungen in die Beurteilung mittels des Fragebogens einfließen lässt.

5.6. Praktische Relevanz

Auf Basis des integrativen CUE-Modells von Thüring und Mahlke (2007) wurde ein modularer Fragebogen zur Evaluation von Komponenten des Nutzungserlebens konstruiert. Durch die theorie-

geleitete Entwicklung und die empirische Validierung unter Einsatz unterschiedlicher interaktiver Produkte wird sowohl Personen aus der Forschung als auch aus der Praxis ein Instrument an die Hand gegeben, welches Nutzungserleben bei der Interaktion mit einem Produkt auf verschiedenen Stufen des Entwicklungsprozesses zu erfassen vermag. Dadurch bietet der Fragebogen eine ökonomische, flexible und zuverlässige Alternative zu bereits existierenden Verfahren. Insbesondere die modulare Struktur unter Berücksichtigung von *Produktwahrnehmungen*, *Emotionen* und *Konsequenzen* erlaubt es, quantitative Vergleiche zwischen verschiedenen Gestaltungslösungen oder verschiedenen Nutzungszeitpunkten analytisch und ein Stück weit ganzheitlicher aus nutzerzentrierter Sicht vorzunehmen.

5.7. Ausblick

Der Fragebogen zeichnet sich durch eine stabile faktorielle Struktur und eine ausreichende bis gute interne Konsistenz seiner Module aus. Die dadurch als erfüllt zu betrachtende Reliabiliät stellt eine notwendige Voraussetzung für seine Validität dar. Im Rahmen einer weiteren Studie ist nun die Validität aller Modulgruppen zu untersuchen. Dies sollte über die Prüfung von Zusammenhängen mit Skalen anderer Fragebögen, insbesondere des AttrakDiff (Hassenzahl et al., 2003) und des UEQ (Laugwitz et al., 2006) erfolgen. Dadurch würde die innere Kriteriumsvalidität bestimmt. Zusätzlich besteht die Möglichkeit, den Zusammenhang zu einem Außenkriterium zu untersuchen. Bei einer aufgabengeleiteten Studie könnte das Kriterium beispielsweise die Bearbeitungsdauer oder Fehlerrate sein.

In zukünftigen Studien sollten den Probanden zudem interaktive Produkte vorgegeben werden. Diese können sowohl aus verschiedenen Produktgruppen stammen als auch unterschiedliche Gestaltungslösungen darstellen. Dadurch wird zum einen die Stabilität der Faktorstruktur der Modulgruppen in Abhängigkeit von selbst oder nicht selbst gewählten Produkten überprüft, zum anderen kann das Vermögen des Fragebogens verschiedene Gestaltungslösungen zu differenzieren, validiert werden (diskriminative Validität; Bortz & Döring, 2006). Eine Laboruntersuchung würde zudem dazu beitragen, dass die Probanden unter kontrollierten Bedingungen agieren.

Weiterhin ist es denkbar, dass weitere Module zum Fragebogen hinzukommen. Die Haptik und Akustik eines interaktiven Produktes werden zum Beispiel bisher noch nicht betrachtet. Diese Module könnten so gestaltet werden, dass sie optional in den Fragebogen einbezogen werden können.

Eine andere Erweiterung des Fragebogens bezieht sich auf die Entwicklung und Überprüfung eines geeigneten Formats für die Erhebung der globalen Attraktivität eines Produkts. Denkbar wäre ein Single-Item-Format in Form eines Schiebereglers, mit dem ein Urteil zwischen *schlecht*

und *gut* differenziert dargestellt werden könnte. Forschungstheoretisch eröffnet sich durch die Erhebung des Globalurteils zusätzlich die Möglichkeit zur Bildung eines Regressionsmodells, mit dem die Produktgüte vorhergesagt werden kann. Davon ausgehend besteht die Möglichkeit, weitere Untersuchungen über Kausalzusammenhänge zwischen den einzelnen Komponenten innerhalb des CUE-Modells selbst zu vollziehen.

Zusätzlich ist eine Normierung für verschiedene interaktive Produktgruppen denkbar. Dadurch eröffnet sich die Möglichkeit, untersuchte Produkte mit der Referenzgruppe zu vergleichen und entsprechende Handlungsschritte abzuleiten (Bortz & Döring, 2006). Ein solches Vorgehen würde zu einer Interpretationsobjektivität beitragen und der Fragebogen wäre auch durch nicht psychologisch geschulte Personen anwendbar.

Abschließend wäre eine Übersetzung des Fragebogens denkbar, da bisher nur wenige User Experience Fragebögen für den angelsächsischen Sprachraum existieren. Dadurch könnten auch kulturelle Unterschiede im Nutzungserleben interaktiver Produkte untersucht werden. Zum Beispiel könnten Hinweise darauf abgeleitet werden, weshalb im Bereich Smartphones in den Vereinigten Staaten Apple Marktführer ist, wohingegen in Europa die Android-Geräte den Markt anführen (vgl. Matthey, 2013).

Literaturverzeichnis

Agricola, Chr. & Agricola, E. (1982). Wörter und Gegenwörter. Antonyme der deutschen Sprache. Leipzig: VEB Bibliographisches Institut

Alben, L. (1996). Quality of experience: defining the criteria for effective interaction design. *Interactions*, 3, 11 – 15.

Backhaus, K., Erichson, B., Plinke, W. &Weiber, R. (2006). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (11. Aufl.). Berlin: Springer.

Blair, C., Calkins, S., Kopp, L. (2010). Self-Regulation as the Interface of Emotional and Cognitive Development. In Rick H. Hoyle (Hrsg.), *Handbook of Personality and Self-Regulaltion* (S. 64-90). Willey-Blackwell.

Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual differences in the centrality of visual product aesthetics: concept and measurement. *Journal of consumer research*, *29*(4), 551-565.

Blyhte, M., Reid, J., Wright, P. & Geelhoed, E. (2006). Interdisciplinary criticism: Analyzing the experience of riot! A location-sensitive digital narrative. *Behaviour & Technology, 25,* 127-139.

Bortz, J. (2005). *Statistik für Sozial-und Humanwissenschaftler* (6. Aufl.). Heidelberg: Springer Medizin Verlag.

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation* (4. Aufl.). Berlin: Springer.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Brockhaus (2006). Modell. In F.A. Brockhaus GmbH (Hrsg.*), Brockhaus, Enzyklopädie, Band 18* (21. Aufl.), Leipzig: F.A. Brockhaus GmbH.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*, 194.

Bühner, M. (2008). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson Studium.

Burmester, M., Hassenzahl, M., & Koller, F. (2002). Usability ist nicht alles–Wege zu attraktiven Produkten (Beyond Usability–Appeal of interactive Products). *I-com*, *1*(1/2002), 32-40.

Capota, K., van Hout, M., & van der Geest, T. (2007). Measuring the emotional impact of websites: a study on combining a dimensional and discrete emotion approach in measuring visual appeal of university websites. *Proceedings of the 2007 conference on Designing pleasurable products and interfaces*, 135-147.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*, 245–276.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 213-218.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*, 98-98.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Dahl, G. (1971). Zur Berechnung des Schwierigkeitsindex bei quantitative abgestufter Aufgabenbewertung. *Diagnostika*, 17, 139-142.

Davis, F. D. (1989) Perceived usefuness, 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*(3), 475-487.

Diefenbach, S. & Hassenzahl, M. (2010). *Handbuch zur Fun-ni Toolbox*. Unveröffentlichtes Manuskript. Folkwang Universität der Künste.

DIN EN ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.* Geneva: International Standardiziation Organization (ISO).

DIN EN ISO 9241-110 (2006). *Ergonomic requirements for office work with visual display terminals – Part 110: Dialogue principles.* Geneva: International Standardiziation Organization (ISO).

DIN EN ISO 9241-210 (2010). *Ergonomics of human-computer interaction – Part 210: Human centered design process for interactive systems.* Geneva: International Standardiziation Organization (ISO).

Drexler, D. (2012). *Methoden zur Exploration von großen Zusammenhängenden Datenmengen und deren Evaluation am Beispiel von LASAD*, Diplomarbeit, Technische Universität Clausthal.

Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft 4*, 214-224.

Ekman, P. (1991). *Gesichtsausdruck und Gefühl.* Paderborn: Jungfermann Verlag.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford: Stanford University Press.

Field, A. (2009). *Discovering statistics using SPSS.* London: Sage Publications Limited.

Fierley, R., & Engl, S. (2010). User Experience Methoden und Games: Erkenntnisse aus der Praxis. *Mensch & Computer 2010 Entertainment Interfaces Track*, 2-8.

Figl, K. (2009). Usability-Fragebögen im Vergleich. *Mensch & Computer 2009,*143-152.

Gloy, K. (2006). *Grundlagen der Gegenwartsphilosophie. Eine Einführung.* Paderborn: UTB.

Götze, F. (2010). *Innovationsakzeptanz von Smartphones bei chinesischen Konsumenten: eine Analyseder Einflussfaktoren*. Berlin: Springer.

Gunther, R. (2012). *An Open Letter to The User Experience Community.* http://www.usabilityprofessionals.org/uxpa/about/an-open-letter-to-the-user-experience-community/ [04.04.2013].

Habermas, T. (1996). *Geliebte Objekte: Symbole und Instrumente der Identitätsbildung* (19. Aufl.). Berlin: Walter de Gruyter.

Harsch, C., Neumann, A., Lehmann, R., & Schröder, K. (2007). Schreibfähigkeit. *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International*, 42-62.

Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction,13*, S. 479-497.

Hassenzahl, M. (2007). The hedonic/pragmatic model of user experience. In Law, E., Vermeeren, A. P. O. S., Hassenzahl, M., & Blythe, M. (2007), *Towards a UX manifesto.* Proc. of HCI 2008: People and Computers XXI: HCI...but not as we know it - Volume 2, Swinton, UK.

Hassenzahl, M. (2008). User Experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme Machine,* 11-15.

Hassenzahl, M., Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. *Mensch & Computer 2003: Interaktion in Bewegung,* 187- 196.

Hassenzahl, M., Burmester, M., & Koller, F. (2008). Der User Experience (UX) auf der Spur: Zum Einsatz von attrakdiff. de. *Usability Professionals 08*, 73-82.

Hassenzahl, M., Kekez, R. & Burmester, M. (2002). The importance of a software's pragmatic quality depends on usage modes. *Proceedings of the 6th International conference on work with display units 2002,* 275-276.

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human–Computer Interaction*, *25*(3), 235-260.

Hassenzahl, M. & Tractinsky, N. (2006). User Experience – a research agenda. *Behavior & Information Technology, 25*(2), 91 -97.

Hinderks, A., Schrepp, M., Rauschenberger, M., Olschner, S., Thomaschewski, J. (2012). Konstruktion eines Fragebogens für jugendliche Personen zur Messung der User Experience. *Usability Professionals 2012,* 78 – 83.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

Hurtienne, J., & Naumann, A. (2010). QUESI-A questionnaire for measuring the subjective consequences of intuitive use. *Interdisciplinary College*, 536.

Jaron, R., & Thielsch, M. T. (2009). Die dritte Dimension–Der Einfluss der Ästhetik auf die Bewertung von Websites. *Planung & Analyse*, *1*, 22-25.

Kakar, A. K. (2012). Engendering Holistic User Experiences: A Hierarchical Model for Balancing the Utilitarian and Hedonic features of IS Products. *Proceedings of the Southern Association for Information Systems Conference,* 143-147.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen - Der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel & C. Clemens (Eds.), *Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme*, 196-201.

Kirakowski, J. (1994). The use of questionnaire methods for usability assessment. *http://sumi.ucc.ie/sumipapp.html* [20.04.2013].

Kritzenberger, H. (2011). On the Role of User Experience in a Holistic Evaluation Model for Serious Games. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, *1,* 1353-1359.

Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Investigations with a German version of the positive and negative affect schedule (PANAS). *Diagnostica*, *42*(2), 139-156.

Lang, P. J. (1980). Behavioral Treatment and Bio-behavioral Assessment: Computer Applications. In J. B. Sidowski, J. H. Johnson, T. A. Willams (Hrsg.), *Technology in mental health care delivery systems.* Norwood: ABLEX Publishing Corporation.

Lang, P. J., Ohman, A., & Vaitl, D. (1988). The international affective picture system [photographic slides]. *Gainesville, FL: Center for Research in Psychophysiology, University of Florida*.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *HCI and Usability for Education and Work*, 63-76.

Laugwitz, B., Schrepp, M., & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. *Mensch und Computer 2006: Mensch und Computer im Strukturwandel*, 125-134.

Lemke, I. (2007). *"Beyond usability", Ästhetische und Symbolische Aspekte des Nutzererlebens*. Diplomarbeit. Technische Universität Berlin.

Liening, A. (1999). *Komplexe Systeme zwischen Ordnung und Chaos. Neuere Entwicklungen in der Theorie nicht-linearer dynamischer Systeme und die Bedeutung für die Wirtschaftswissenschaft und ihre Didaktik*. Münster: LIT Verlag.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.

Mahlke, S. (2005). An integrative model on web user experience. *Proceeding of ICWI2005, 2*, 91-95.

Mahlke, S. (2006). Studying user experience with digital audio players. In R. Harper, M. Rauterberg & M. Combetto (Hrsg.), *Entertainment Computing*, ICEC 2006, LNCS 4161, 358-361. Berlin: Springer.

Mahlke, S. (2008). Visual aesthetics and the user experience. *Proceedings of the study of visual aesthetics in HCI*, 1-6.

Mahlke, S. & Lindgaard, G. (2007). Emotional Experiences and Quality Perceptions of Interactive Products. *Interaction Design and Usability*, 164-173.

Mahlke, S. & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 915-918.

Matthey, F. (2013). USA: Apple baut Führung vor Samsung auf Smartphone-Markt aus. *http://www.giga.de/smartphones/iphone-5/news/usa-apple-baut-fuhrung-vor-samsung-auf-smartphone-markt-aus/* [16.04.2013].

Moshagen, M. & Thielsch, M. T. (2011, in press). *A short version of the visual aesthetics of web-sites inventory. Behaviour & Information Technology.*

Mehrabian, A. (1970). The development and validation of affiliative tenendy and sensitivity to reection. *Educational and Psychological Measurement, 30*, 417-428.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology* (Vol. 11). Cambridge, MA: MIT press.

Minge, M. (2008). Dynamics of User Experience. *Positionspapier im Workshop „Research Goals and Strategies for Studying User Experience and Emotion" auf der NordiCHI 2008*, Lund, Schweden.

Minge, M. (2011). *Dynamische Aspekte des Nutzungserlebens der Interaktion mit technischen Systemen.* Dissertation. Technische Universität Berlin.

Minge, M. & Thüring, M. (2011). Der Einfluss von Gebrauchstauglichkeit und Gestaltung auf Kognition und Emotion. In C. Lutsch & F. Adler (Hrsg.), *Forum für Entwerfen – Der kurze Weg zum Glück*. Forum für Entwerfen e.V., Ulm, 109-119.

Moosbrugger, H., & Kelava, A. (2011). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer.

Nielsen, J. (1993). *Usability Engineering.* Academic Press, San Diego, CA.

Norman, D.A., (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3. Aufl.). New York: McGraw-Hill.

O'Cass, A., & Choy, E. (2008). Studying Chinese generation Y consumers' involvement in fashion clothing and perceived brand status. *Journal of Product & Brand Management*, *17*(5), 341-352.

O'Cass, A., & Frost, H. (2002). Status brands: examining the effects of non-product-related brand associations on status and conspicuous consumption. *Journal of Product & Brand Management*, *11*(2), 67-88.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32,* 396-402.

Piesbergen, C., & Sarubin, N. (2011). *Erfassung von kognitiver Dissonanz nach Kaufentscheidungen.* Unveröffentlichtes Manuskript. LMU München.

Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. *Software-Ergonomie, 97*, 253-262.

Prümper, J., & Anft, M. (1993). Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - Ein Fallbeispiel. *Software-Ergonomie*, *93*, 145-156.

Rafaeli, A., & Vilnai-Yavetz, I. (2004). Instrumentality, aesthetics and symbolism of physical artifacts as triggers of emotion. *Theoretical Issues in Ergonomics Science*, *5*(1), 91-112.

Rauschenberger, M., Schrepp, M., Olschner, S., Thomaschewski, J. & Cota, M.P. (2012). Measurement of User Experience. A Spanish Language Version of the User Experience Questionnaire (UEQ). In: Rocha, A., Calvo-Manzano, J.A., Reis, L.P. & Cota, M.P. (Hrsg.), *Sistemas y Tecnologías de Información – Actas de la 7a Conferencia Ibérica de Sistemas y Tecnologías de Información – Madrid, Espana, 20 al 23 de Junio de 2012*.

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, *9*, 222-245.

Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality andSocial Psychology, 37,* 345-356.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161-1178.

Russell, J. A. (1989). Measures of emotion. In R. Plutchik & H. Kellerman (Hrsg.), *Emotion: Theory, research, and experience* (S. 83-111). San Diego: Academic Press

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11,* 273-294.

Scapin, D. L., Senach, B., Trousse, B., & Pallot, M. (2012). User Experience: Buzzword or New Paradigm? *ACHI 2012, The Fifth International Conference on Advances in Computer-Human Interactions*, 336–341.

Scherer, K.R. (1984). On the nature and function of emotion: A component process approach. In K.R. Scherer & P. Ekman (Hrsg.). *Approaches of emotion* (S. 293-317). Hillsdale: Erlbaum.

Schleicher, R., & Trösterer, S. (2009). The 'Joy-of-Use'-Button: Recording Pleasant Moments While Using a PC. *Human-Computer Interaction–INTERACT 2009*, 630-633.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, *8*(4), 350-353.

Schuchard-Ficher, C. (1979). Ein Ansatz zur *Messung* von Nachkauf-*Dissona*nz. *Vertriebswirtschaftliche Abhandlungen, 21,* Duncker & Humbolt Berlin.

Sieger, H., Kirschnick, N., Möller, S. (2010). Poster: User preferences for biometric authentication methods and graded security on mobile phones. *Symposium on Usable Privacy and Security (SOUPS)*.

Spatz, A. (2008). *Zur Messung im Wirtschafts- und Sozialwissenschaftlichen Umfeld. Eine empirische Studie zur Anzahl der Antwortkategorien bei einer Ratingskala.* Dissertation. Universität Augsburg.

Spinas, P. (1987). *Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen.* ADAG. Zürich.

Tellegen, A., Watson, D., & Clark, L. A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of personality and social psychology, 54(6),* 1063-1070.

Thielsch, M. T. (2008). Inhalt, Usability und Ästhetik in der Bewertung durch Webnutzer. *Mensch & Computer* 2008, 441-444.

Thielsch, M. T., & Jaron, R. (2012). Das Zusammenspiel von Website-Inhalten, Usability und Ästhetik. *Mensch & Computer 2012: interaktiv informiert–allgegenwärtig und allumfassend!?*, 123-132.

Thielsch, M. T., & Weltzin, S. (2009). Online-Befragungen in der Praxis. *Praxis der Wirtschaftspsychologie: Themen und Fallbeispiele für Studium und Praxis*, 69-85.

Thüring, M. & Mahlke, S. (2007). Usability, aesthetics, and emotions in humantechnology interacttion. *International Journal of Psychology, 42* (4), 253-264.

Tractinsky, N. (2011). Visual aesthetics in human-computer interaction: Justification and findings. In Mads Bødker, Ather Nawaz, Gitte Skou Petersen (Hrsg*). Proceedings of the 11th Danish Human-Computer Interaction Research Symposium (DHRS2011),* 2-9.

Ullrich, D., & Diefenbach, S. (2011). Erlebnis intuitive Interaktion - ein phänomenlogischer Ansatz. *i-com*, *10*(3), 63-68.

Velicer, W.F. (1976). Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika*, 41(3), 321-327.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219.

Willumeit, H., Gediga, G. & Hamborg, K.-C. (1996). IsoMetrics: Ein Verfahren zur formativen Evaluation von Software nach ISO 9241/10. *Ergonomie & Informatik, 27*, 5-12.

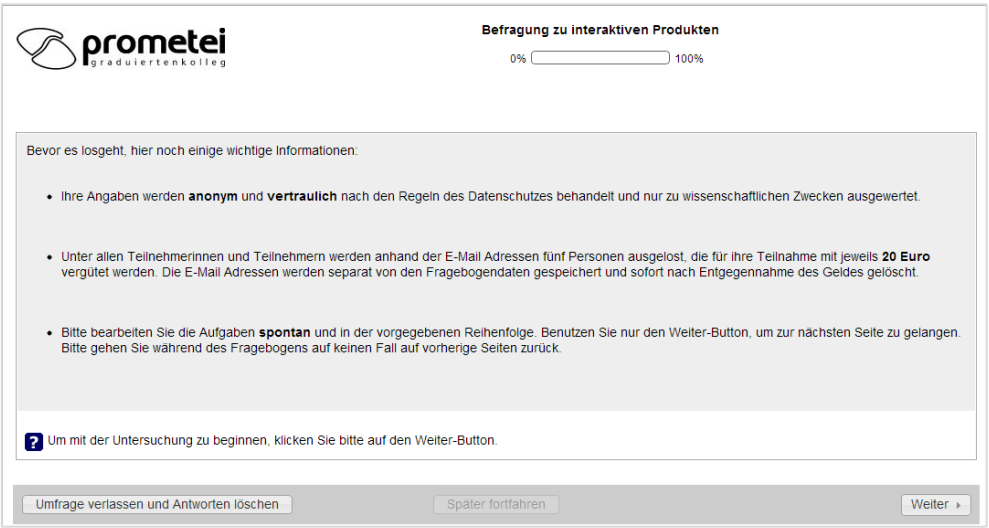
Anhang

A. Studie 1

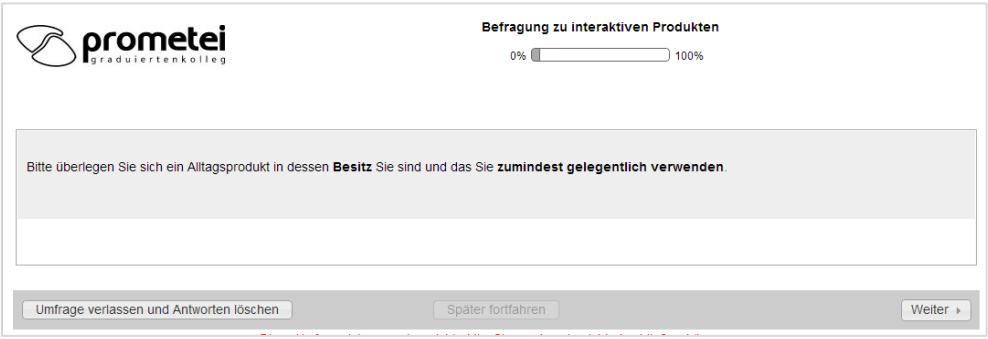
A.1. Online-Befragung mit LimeSurvey

Seite 1: Willkommen-Bildschirm

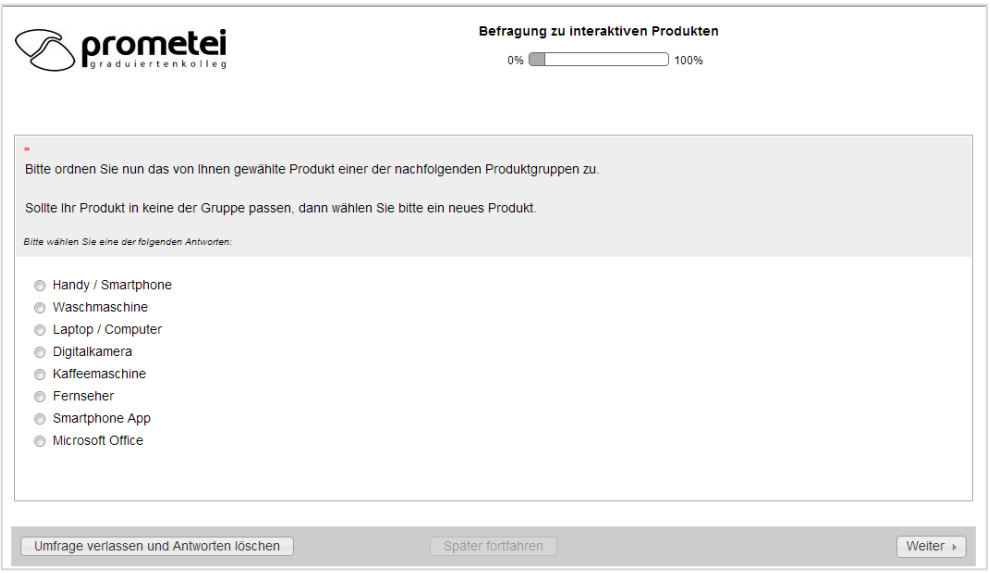
Seite 2: Einleitung



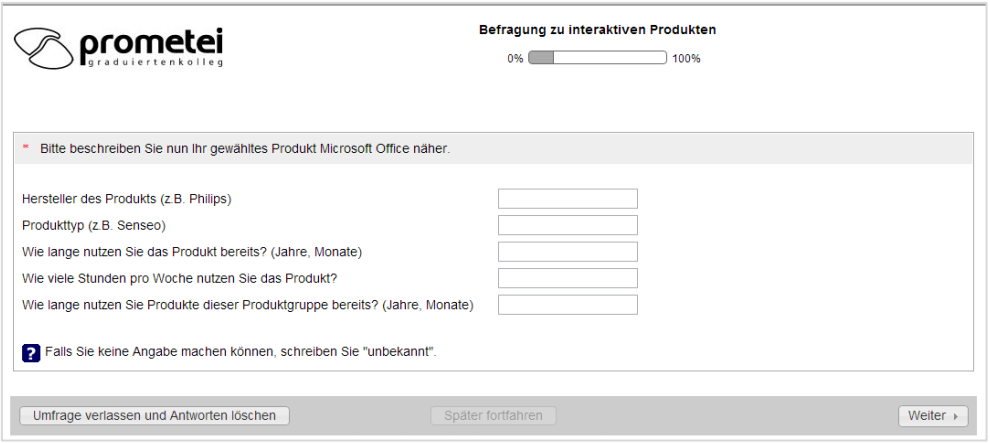
Seite 3: Der Proband soll sich ein interaktives Produkt überlegen.



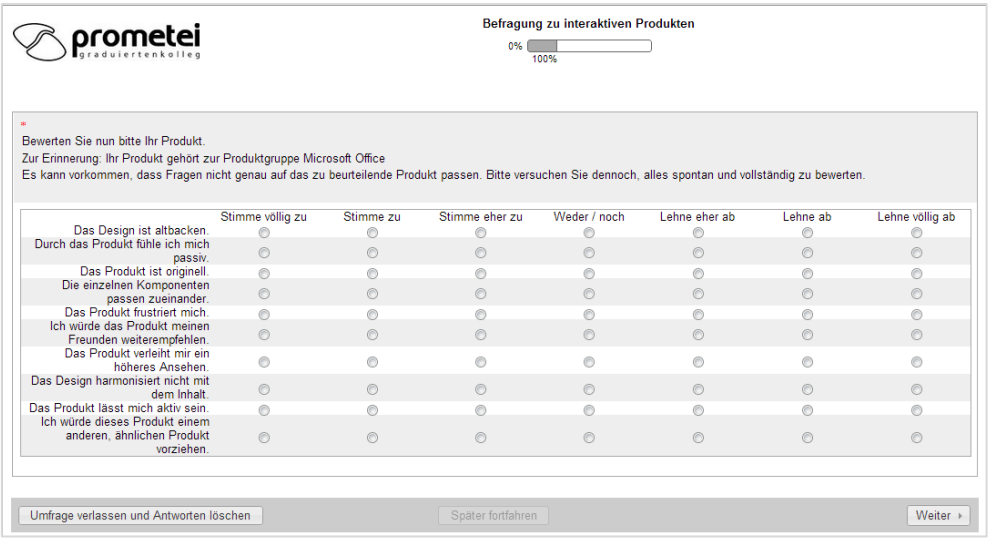
Seite 4: Der Proband soll sein Produkt in die Liste der Produktgruppen einordnen.



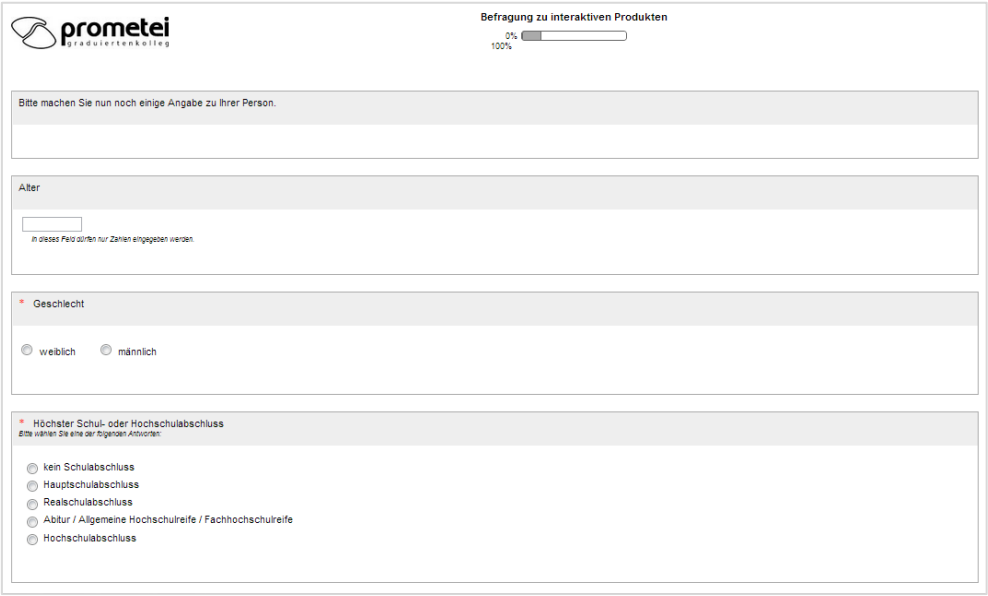
Seite 5: Fragen zum gewählten Produkt.



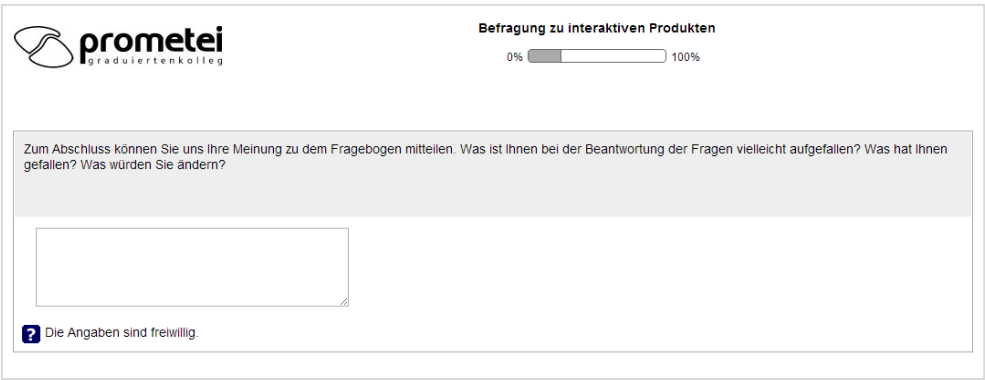
Seite 6 – 11: 67 Items, je 9 bis 10 Items pro Seite.



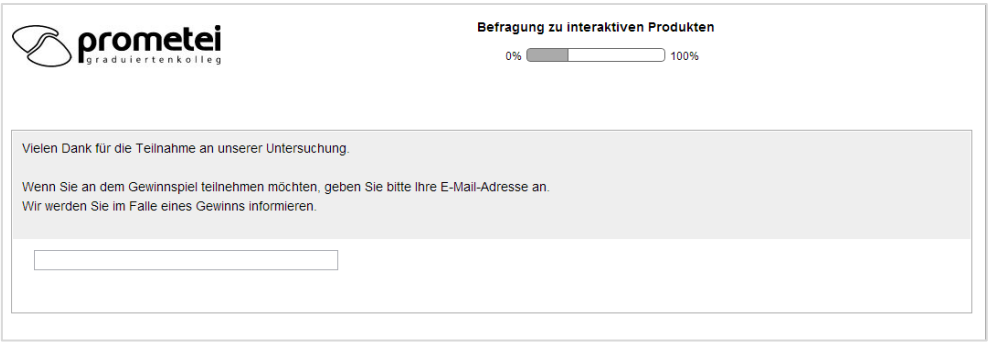
Seite 12: Nachbefragung (demographische Angaben)



Seite 13: Verbesserungsvorschläge



Seite 14: Gewinnspiel



Seite 15: Abschluss

Ihre Daten wurden gespeichert. Sie können das Fenster jetzt schließen.

Vielen Dank für die Teilname!

Bei Anregungen und Fragen können Sie sich wenden an:

Dr. Michael Minge
Technische Universität Berlin
GRK prometei
michael.minge(at)zmms.tu-berlin.de

Laura Riedel
laura.riedel(at)googlemail.com

A.2. Instruktion

An der TU Berlin untersuchen wir, wie Menschen **interaktive technische Produkte** beurteilen. Interaktiv bedeutet, dass es sich um Technik handelt, die einerseits vom Nutzer bestimmte Eingaben erfordert und andererseits bestimmte Ergebnisse zurückmeldet.

Dies kann zum Beispiel ein Handy sein, das einerseits die Eingabe einer Telefonnummer benötigt und andererseits darstellt, welche Ziffer man soeben eingegeben hat.

Zu Beginn werden wir Sie bitten, sich für ein **konkretes interaktives Produkt** aus ihrem Alltag zu entscheiden und genau dieses Produkt mit den folgenden Fragen zu beurteilen.

Die Gesamtdauer der Untersuchung beträgt **ca. 20 Minuten**.

Mit Ihrer Arbeit unterstützen Sie die wissenschaftliche Arbeit an der TU Berlin. Für Ihre Teilnahme und Ihr Interesse bedanken wir uns vorab.

A.3. Nutzungsdauer

Tabelle 26. Nutzungsdauer des gewählten Produkts, aufgegliedert nach Produktgruppen

	Wie lange nutzen Sie das Produkt bereits?		**Wie lange nutzen Sie Produkte dieser Produktgruppe bereits?**		**Wie viele Stunden pro Woche nutzen Sie das Produkt?**	
	M	*SD*	*M*	*SD*	*M*	*SD*
Handy / Smartphone	19.74	19.18	63.78	52.84	23.24	26.48
Laptop / Computer	30.12	19.82	95.49	66.86	29.80	24.87
Kaffeemaschine	29.28	21.13	76.94	60.76	3.03	3.74
Digitalkamera	27.63	19.12	88.80	42.01	4.27	3.44
Waschmaschine	38.85	23.44	143.17	131.92	4.18	4.45
Fernseher	30.67	21.64	223.90	140.61	22.50	7.23
Smartphone App	9.00	4.24	3.00	4.24	21.00	-
Microsoft Office	120.00	-	120	-	20.00	-

A.4. Itemanalyse der 67 Roh-Items

Tabelle 27. Itemanalyse

Dimension	Item	MIN	MAX	*M*	*SD*	Var	∑	∑ (0-6)*	*p* von Dahl	Trenn-schär-fe**
Nützlich-keit	Mit dem Produkt lassen sich meine geplanten Aufgaben nicht lösen.	1	7	2.58	1.384	1.915	615	377	.264	.577
	Insgesamt halte ich das Produkt für nützlich.	2	7	6.29	.830	.690	1498	1260	.882	.610
	Das Produkt ist hilfreich für das Erreichen meiner Ziele.	1	7	5.58	1.363	1.857	1327	1089	.763	.766
	Das Produkt ist sinnvoll erweiter-bar, falls sich meine Bedürfnisse ändern.	1	7	4.38	1.835	3.368	1043	805	.564	.553
	Ich kann mit dem Produkt viele verschiedene Aufgaben erledigen.	1	7	5.64	1.677	2.813	1343	1105	.774	.589
	Das Produkt ist nicht zweckmäßig.	1	7	2.34	1.580	2.495	557	319	.223	.426
	Das Produkt erfüllt seine Bestim-mung.	1	7	6.16	.906	.821	1467	1229	.861	.460
	Mithilfe des Produkts kann ich meine Ziele erreichen.	1	7	5.44	1.303	1.699	1295	1057	.740	.709
	Die Funktionen des Produkts sind genau richtig für meine Ziele.	1	7	5.57	1.122	1.259	1326	1088	.762	.669
Benutzbar-keit	Die Benutzung des Produkts ist nicht anstrengend.	1	7	5.65	1.293	1.671	1345	1107	.775	.505
	Ich muss mich kaum auf die Benut-zung des Produkts konzentrieren.	1	7	4.89	1.578	2.489	1163	925	.648	.468
	Die Bedienung des Produkts ist verständlich.	1	7	5.89	.874	.764	1402	1164	.815	.688
	Das Produkt lässt sich einfach benutzen.	2	7	5.99	.923	.852	1426	1188	.832	.571
	Das Produkt ist kompliziert zu bedienen.	1	7	2.32	1.248	1.559	551	313	.219	.701
	Manchmal verstehe ich nicht, was das Produkt tut.	1	7	2.98	1.566	2.451	709	471	.330	.711
	Manchmal verstehe ich nicht, was das Produkt mir sagen will.	1	7	2.79	1.463	2.141	664	426	.298	.701
	Es dauert zu lange die Funktionen des Produkts zu erlernen.	1	7	2.39	1.332	1.774	568	330	.231	.673
	Von der Benutzung des Produkts bekomme ich manchmal Kopf-schmerzen.	1	7	2.53	1.640	2.689	603	365	.256	.442
	Es ist immer eindeutig. was das Produkt von mir will.	1	7	5.34	1.288	1.659	1270	1032	.723	.644
Visuelle Ästhetik	Das Produkt ist originell.	1	7	4.49	1.572	2.470	1069	831	.582	.629
	Das Produkt ist hässlich.	1	7	2.10	1.307	1.708	499	261	.183	.760
	Das Design wirkt attraktiv.	1	7	5.05	1.415	2.001	1203	965	.676	.848
	Hier passt alles zusammen.	1	7	5.43	1.184	1.402	1292	1054	.738	.601
	Das Produkt ist stilvoll.	1	7	4.96	1.348	1.817	1181	943	.660	.831
	Das Produkt sieht minderwertig aus.	1	6	2.41	1.377	1.897	574	336	.235	.761
	Das Produkt ist kreativ gestaltet.	1	7	4.61	1.400	1.960	1098	860	.602	.743
	Das Design harmonisiert nicht mit dem Inhalt.	1	7	2.67	1.634	2.670	635	397	.278	.543
	Das Design ist altbacken.	1	7	2.68	1.572	2.470	639	401	.281	.714
	Die einzelnen Komponenten pas-sen zueinander.	1	7	5.74	1.111	1.233	1365	1127	.789	.486
Identifika-tion / Wert	Ich schäme mich. wenn mich ande-	1	7	1.75	1.192	1.421	416	178	.125	.048

	re mit dem Produkt sehen.									
	Meine Freunde dürfen ruhig neidisch auf mich sein.	1	7	3.78	1.693	2.866	899	661	.463	.630
	Das Produkt verleiht mir ein höheres Ansehen.	1	7	3.57	1.586	2.516	850	612	.429	.612
	Durch das Produkt werde ich anders wahrgenommen.	1	7	3.15	1.631	2.661	750	512	.359	.596
	Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.	1	7	3.63	1.928	3.718	865	627	.439	.703
	Ohne das Produkt kann ich nicht leben.	1	7	3.62	1.874	3.510	862	624	.437	.618
	Wenn ich mit dem Produkt arbeite, vergesse ich schon mal die Zeit.	1	7	4.72	1.779	3.165	1123	885	.620	.452
	Das Produkt ist mir wichtig.	1	7	5.51	1.190	1.416	1311	1073	.751	.697
	Das Produkt ist wie ein Freund für mich.	1	7	3.25	1.732	2.999	774	536	.375	.718
	Ich bin stolz dieses Produkt zu besitzen.	1	7	4.56	1.494	2.231	1085	847	.593	.758
Valenz	Das Produkt macht mich zufrieden.	1	7	5.24	1.218	1.485	1248	1010	.707	.774
	Das Produkt macht mich glücklich.	1	7	4.87	1.330	1.768	1159	921	.645	.714
	Das Produkt begeistert mich.	1	7	5.34	1.248	1.557	1270	1032	.723	.747
	Das Produkt langweilt mich.	1	6	2.33	1.274	1.624	555	317	.222	.635
	Das Produkt erfreut mich.	1	7	5.42	1.137	1.292	1291	1053	.737	.810
	Das Produkt nervt mich.	1	7	2.62	1.393	1.942	623	385	.270	.751
	Das Produkt verärgert mich.	1	6	2.59	1.419	2.015	617	379	.265	.724
	Das Produkt frustriert mich.	1	7	2.34	1.311	1.719	557	319	.223	.723
Arousal	Durch das Produkt fühle ich mich energiegeladen.	1	7	3.71	1.368	1.871	882	644	.451	.472
	Durch das Produkt fühle ich mich träge.	1	7	2.56	1.488	2.214	609	371	.260	.668
	Das Produkt macht mich hellwach.	1	7	3.79	1.443	2.082	902	664	.465	.492
	Das Produkt macht mich müde.	1	7	2.64	1.430	2.046	628	390	.273	.640
	Das Produkt beruhigt mich.	1	7	4.13	1.306	1.706	982	744	.521	.414
	Das Produkt macht mich nervös.	1	7	2.25	1.348	1.818	536	298	.209	.503
	Das Produkt lässt mich aktiv sein.	1	7	5.05	1.540	2.373	1203	965	.676	.569
	Durch das Produkt fühle ich mich passiv.	1	7	2.68	1.449	2.100	638	400	.280	.568
Nutzungs-intention	Ich würde das Produkt meinen Freunden weiterempfehlen.	1	7	5.67	1.293	1.673	1350	1112	.779	.670
	Ich würde dieses Produkt einem anderen, ähnlichen Produkt vorziehen.	1	7	4.73	1.702	2.898	1126	888	.622	.519
	Wenn ich könnte, würde ich das Produkt täglich nutzen.	1	7	5.44	1.608	2.585	1294	1056	.739	.565
	Das Produkt motiviert mich nicht es weiterhin zu nutzen.	1	7	3.12	1.886	3.556	742	504	.353	.345
	Ich kann es kaum erwarten das Produkt erneut zu benutzen.	1	7	4.32	1.480	2.191	1027	789	.553	.634
	Ich würde das Produkt auf keinen Fall wieder benutzen.	1	7	1.57	.942	.888	373	135	.095	.521
	Ich würde das Produkt gegen kein anderes eintauschen.	1	7	3.97	1.803	3.252	945	707	.495	.573
Global-urteil	Alles in allem ist das Produkt gut.	1	7	6.01	1.041	1.084	1431	1193	.835	.741
	Ich bin von dem Produkt begeistert.	2	7	5.32	1.229	1.511	1267	1029	.721	.806
	Das Produkt befriedigt meine Bedürfnisse.	1	7	5.36	1.220	1.489	1276	1038	.727	.683
	Ich habe mir von dem Produkt	1	7	2.92	1.459	2.128	694	456	.319	.747

	mehr erwartet.								
	Das Produkt entspricht vollkommen meinen Erwartungen.	1	7	5.41	1.259	1.584	1287	1049	.735	.818

* Summe umgerechnet in Abstufung von 0 bis 6

** Trennschärfe modulweise berechnet

A.5. Reliabilitäten und Interkorrelationen

Tabelle 28. Interkorrelation der Items und Cronbach's α der Modulgruppe Produktbewertung, aufgeteilt nach Modulen.

	Items A ↓	**Items B →**		
Visuelle Ästhetik		Das Design wirkt attraktiv.	Das Produkt ist stilvoll.	Das Produkt ist kreativ gestaltet.
	Das Design wirkt attraktiv.	-	.835	.637
	Das Produkt ist stilvoll.	.835	-	.603
	Das Produkt ist kreativ gestaltet.	.637	.603	-
			(α = .870)	
Status		Durch das Produkt werde ich anders wahrgenommen.	Das Produkt verleiht mir ein höheres Ansehen.	Meine Freunde dürfen ruhig neidisch auf das Produkt sein.
	Durch das Produkt werde ich anders wahrgenommen.	-	.447	.417
	Das Produkt verleiht mir ein höheres Ansehen.	.447	-	.444
	Meine Freunde dürfen ruhig neidisch auf das Produkt sein.	.417	.444	-
			(α = .755)	
Benutzbarkeit		Es dauert zu lange, um die Funktionen des Produkts zu erlernen.	Die Bedienung des Produkts ist verständlich.	Das Produkt lässt sich einfach benutzen.
	Es dauert zu lange, um die Funktionen des Produkts zu erlernen.	-	-.561	-.406
	Die Bedienung des Produkts ist verständlich.	-.561	-	.417
	Das Produkt lässt sich einfach benutzen.	-.406	.417	-
			(α = .701)	

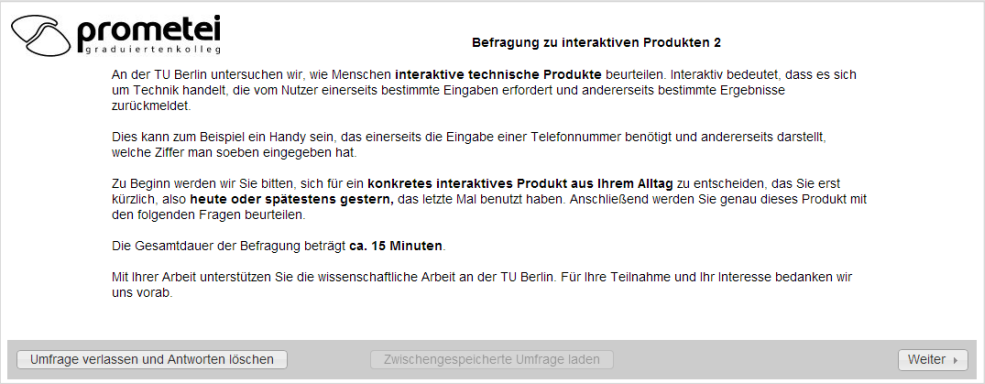
	Items A ↓	**Items B →**		
Bindung		Ohne das Produkt kann ich nicht leben.	Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.	Das Produkt ist wie ein Freund für mich.
	Ohne das Produkt kann ich nicht leben.	-	.565	.508
	Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.	.565	-	.446
	Das Produkt ist wie ein Freund für mich.	.508	.446	-
			(α =.706)	
Nützlichkeit		Mithilfe des Produkts kann ich meine Ziele erreichen.	Die Funktionen des Produkts sind genau richtig für meine Ziele.	Insgesamt halte ich das Produkt für nützlich.
	Mithilfe des Produkts kann ich meine Ziele erreichen.	-	.493	.449
	Die Funktionen des Produkts sind genau richtig für meine Ziele.	.493	-	.485
	Insgesamt halte ich das Produkt für nützlich.	.449	.485	-
			(α = .714)	

Anmerkung. Der α-Koeffizient steht in Klammern unter den Korrelationen der Items jedes Moduls.

B. Studie 2

B.1. Online-Befragung mit LimeSurvey

Seite 1: Willkommen-Bildschirm



prometei
graduiertenkolleg

Befragung zu interaktiven Produkten 2

An der TU Berlin untersuchen wir, wie Menschen **interaktive technische Produkte** beurteilen. Interaktiv bedeutet, dass es sich um Technik handelt, die vom Nutzer einerseits bestimmte Eingaben erfordert und andererseits bestimmte Ergebnisse zurückmeldet.

Dies kann zum Beispiel ein Handy sein, das einerseits die Eingabe einer Telefonnummer benötigt und andererseits darstellt, welche Ziffer man soeben eingegeben hat.

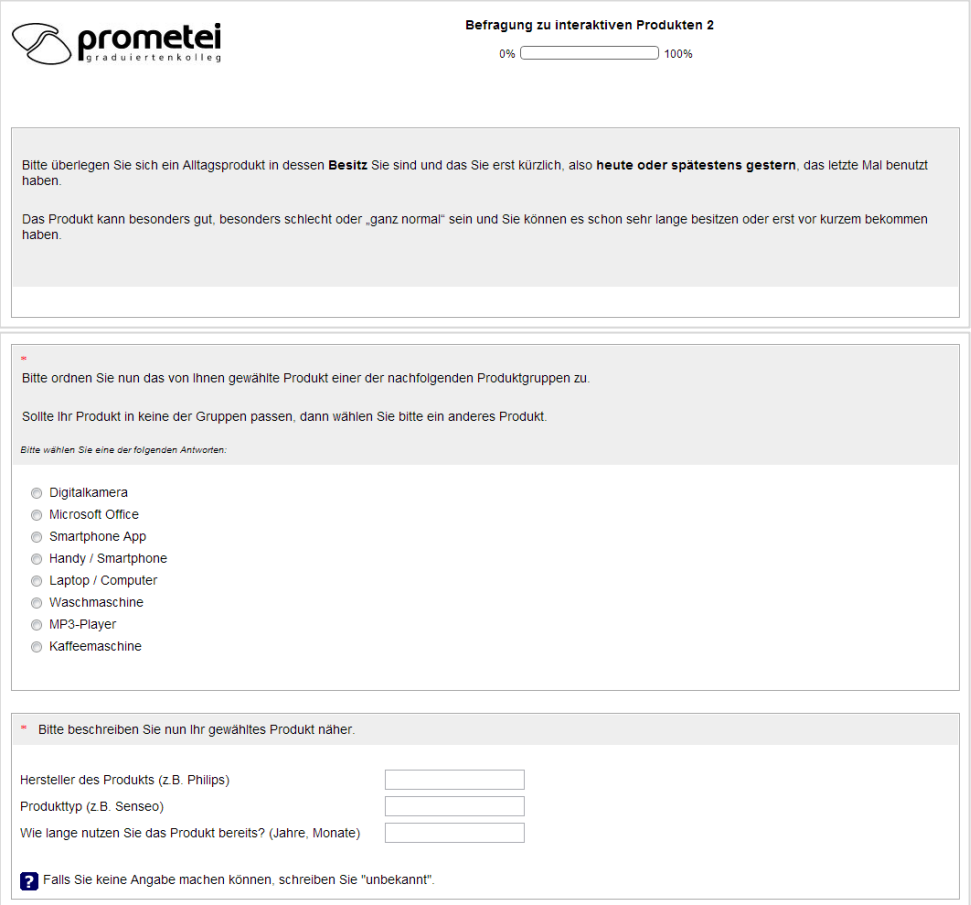
Zu Beginn werden wir Sie bitten, sich für ein **konkretes interaktives Produkt aus Ihrem Alltag** zu entscheiden, das Sie erst kürzlich, also **heute oder spätestens gestern,** das letzte Mal benutzt haben. Anschließend werden Sie genau dieses Produkt mit den folgenden Fragen beurteilen.

Die Gesamtdauer der Befragung beträgt **ca. 15 Minuten**.

Mit Ihrer Arbeit unterstützen Sie die wissenschaftliche Arbeit an der TU Berlin. Für Ihre Teilnahme und Ihr Interesse bedanken wir uns vorab.

Umfrage verlassen und Antworten löschen | Zwischengespeicherte Umfrage laden | Weiter

Seite 2: Der Proband soll sich ein Produkt überlegen, es in die Liste der Produktgruppen einordnen und Angaben dazu machen.



prometei
graduiertenkolleg

Befragung zu interaktiven Produkten 2

0% 100%

Bitte überlegen Sie sich ein Alltagsprodukt in dessen **Besitz** Sie sind und das Sie erst kürzlich, also **heute oder spätestens gestern**, das letzte Mal benutzt haben.

Das Produkt kann besonders gut, besonders schlecht oder „ganz normal" sein und Sie können es schon sehr lange besitzen oder erst vor kurzem bekommen haben.

*
Bitte ordnen Sie nun das von Ihnen gewählte Produkt einer der nachfolgenden Produktgruppen zu.

Sollte Ihr Produkt in keine der Gruppen passen, dann wählen Sie bitte ein anderes Produkt.

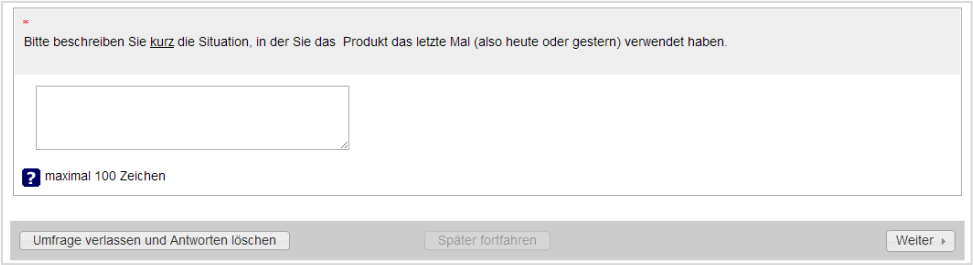
Bitte wählen Sie eine der folgenden Antworten:

- Digitalkamera
- Microsoft Office
- Smartphone App
- Handy / Smartphone
- Laptop / Computer
- Waschmaschine
- MP3-Player
- Kaffeemaschine

* Bitte beschreiben Sie nun Ihr gewähltes Produkt näher.

Hersteller des Produkts (z.B. Philips)
Produkttyp (z.B. Senseo)
Wie lange nutzen Sie das Produkt bereits? (Jahre, Monate)

? Falls Sie keine Angabe machen können, schreiben Sie "unbekannt".



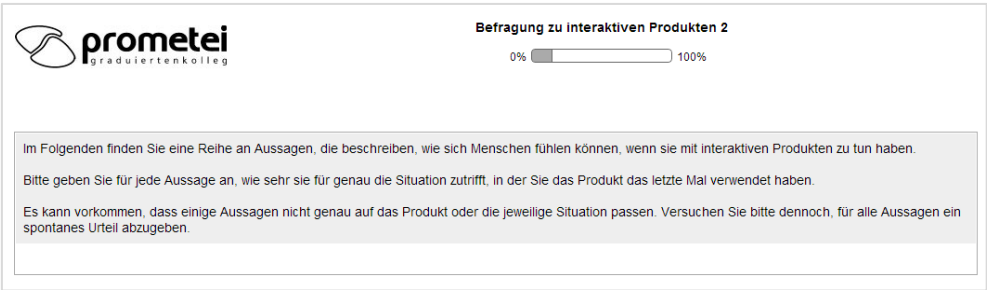
*

Bitte beschreiben Sie kurz die Situation, in der Sie das Produkt das letzte Mal (also heute oder gestern) verwendet haben.

maximal 100 Zeichen

Umfrage verlassen und Antworten löschen | Später fortfahren | Weiter

Seite 3: Items der Modulgruppe *Emotionen*



prometei graduiertenkolleg

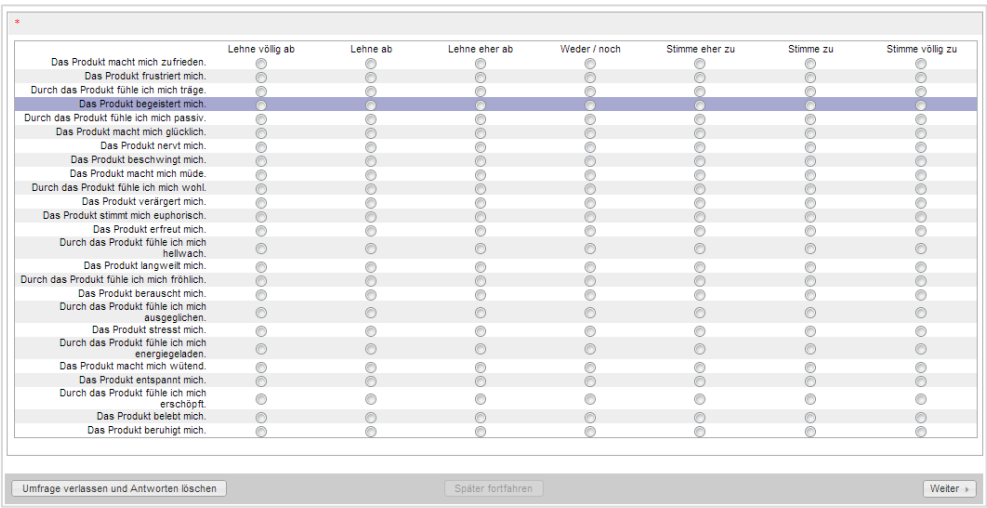
Befragung zu interaktiven Produkten 2

0% 100%

Im Folgenden finden Sie eine Reihe an Aussagen, die beschreiben, wie sich Menschen fühlen können, wenn sie mit interaktiven Produkten zu tun haben.

Bitte geben Sie für jede Aussage an, wie sehr sie für genau die Situation zutrifft, in der Sie das Produkt das letzte Mal verwendet haben.

Es kann vorkommen, dass einige Aussagen nicht genau auf das Produkt oder die jeweilige Situation passen. Versuchen Sie bitte dennoch, für alle Aussagen ein spontanes Urteil abzugeben.

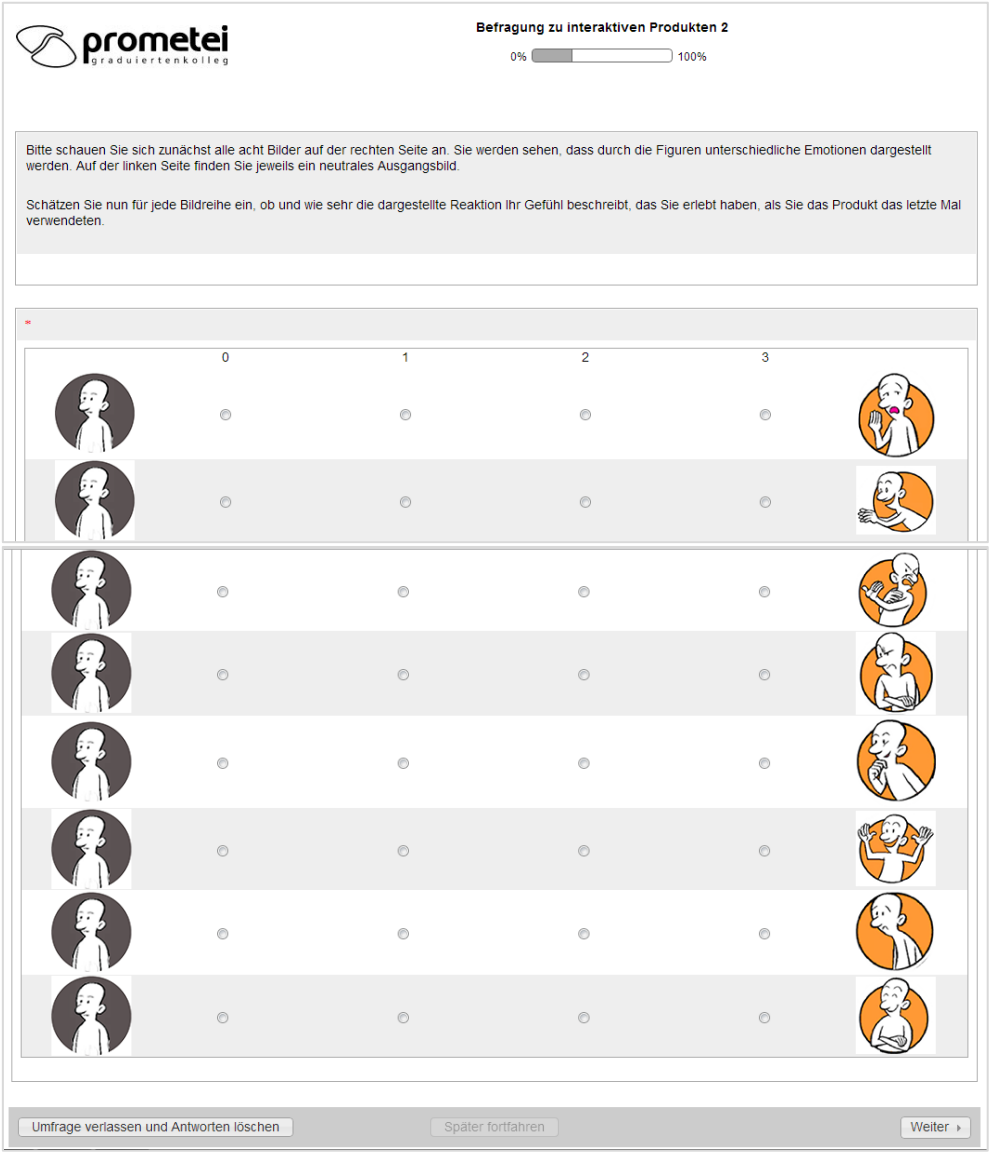


*

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Das Produkt macht mich zufrieden.							
Das Produkt frustriert mich.							
Durch das Produkt fühle ich mich träge.							
Das Produkt begeistert mich.							
Durch das Produkt fühle ich mich passiv.							
Das Produkt macht mich glücklich.							
Das Produkt nervt mich.							
Das Produkt beschwingt mich.							
Das Produkt macht mich müde.							
Durch das Produkt fühle ich mich wohl.							
Das Produkt verärgert mich.							
Das Produkt stimmt mich euphorisch.							
Das Produkt erfreut mich.							
Durch das Produkt fühle ich mich hellwach.							
Das Produkt langweilt mich.							
Durch das Produkt fühle ich mich fröhlich.							
Das Produkt berauscht mich.							
Durch das Produkt fühle ich mich ausgeglichen.							
Das Produkt stresst mich.							
Durch das Produkt fühle ich mich energiegeladen.							
Das Produkt macht mich wütend.							
Das Produkt entspannt mich.							
Durch das Produkt fühle ich mich erschöpft.							
Das Produkt belebt mich.							
Das Produkt beruhigt mich.							

Umfrage verlassen und Antworten löschen | Später fortfahren | Weiter

Seite 4: LEMtool



prometei graduiertenkolleg

Befragung zu interaktiven Produkten 2

0% 100%

Bitte schauen Sie sich zunächst alle acht Bilder auf der rechten Seite an. Sie werden sehen, dass durch die Figuren unterschiedliche Emotionen dargestellt werden. Auf der linken Seite finden Sie jeweils ein neutrales Ausgangsbild.

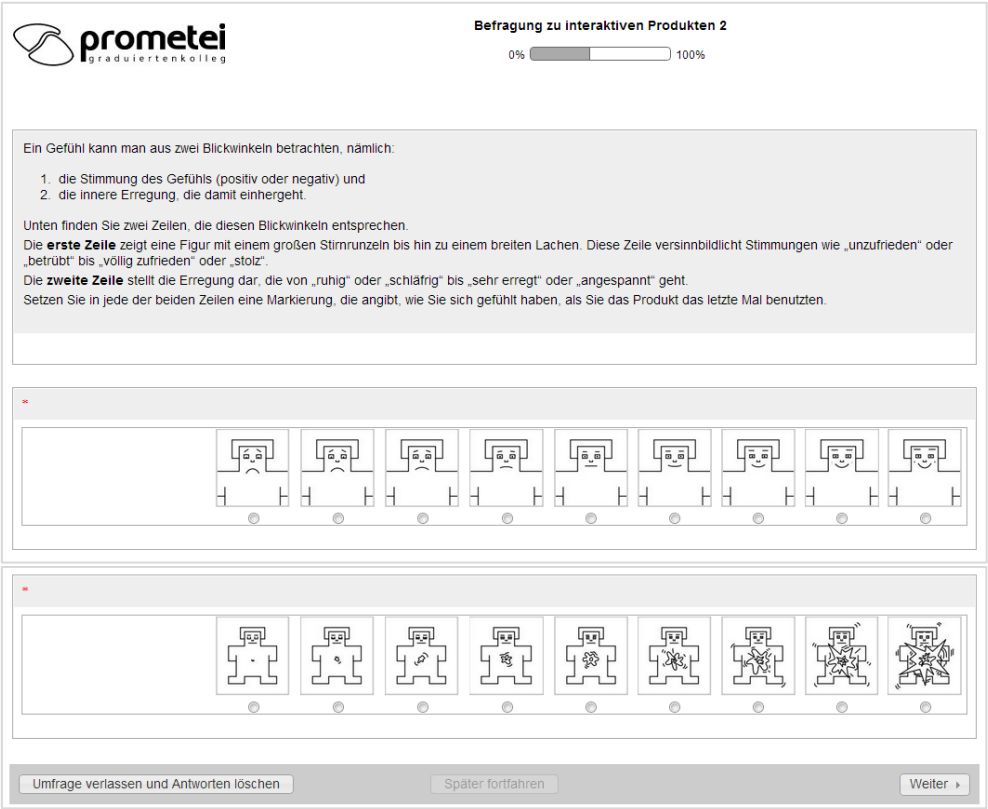
Schätzen Sie nun für jede Bildreihe ein, ob und wie sehr die dargestellte Reaktion Ihr Gefühl beschreibt, das Sie erlebt haben, als Sie das Produkt das letzte Mal verwendeten.

*

	0	1	2	3	

Umfrage verlassen und Antworten löschen | Später fortfahren | Weiter

Seite 5: SAM



prometei graduiertenkolleg

Befragung zu interaktiven Produkten 2

0% 100%

Ein Gefühl kann man aus zwei Blickwinkeln betrachten, nämlich:

1. die Stimmung des Gefühls (positiv oder negativ) und
2. die innere Erregung, die damit einhergeht.

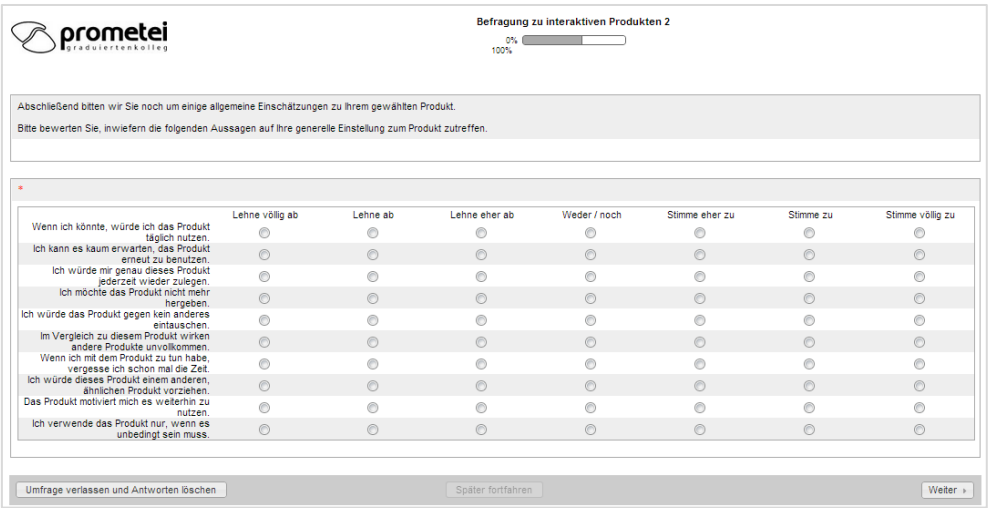
Unten finden Sie zwei Zeilen, die diesen Blickwinkeln entsprechen.
Die **erste Zeile** zeigt eine Figur mit einem großen Stirnrunzeln bis hin zu einem breiten Lachen. Diese Zeile versinnbildlicht Stimmungen wie „unzufrieden" oder „betrübt" bis „völlig zufrieden" oder „stolz".
Die **zweite Zeile** stellt die Erregung dar, die von „ruhig" oder „schläfrig" bis „sehr erregt" oder „angespannt" geht.
Setzen Sie in jede der beiden Zeilen eine Markierung, die angibt, wie Sie sich gefühlt haben, als Sie das Produkt das letzte Mal benutzten.

*

*

Umfrage verlassen und Antworten löschen | Später fortfahren | Weiter

Seite 6: Items der Modulgruppe *Konsequenzen*



prometei graduiertenkolleg

Befragung zu interaktiven Produkten 2

0% 100%

Abschließend bitten wir Sie noch um einige allgemeine Einschätzungen zu Ihrem gewählten Produkt.

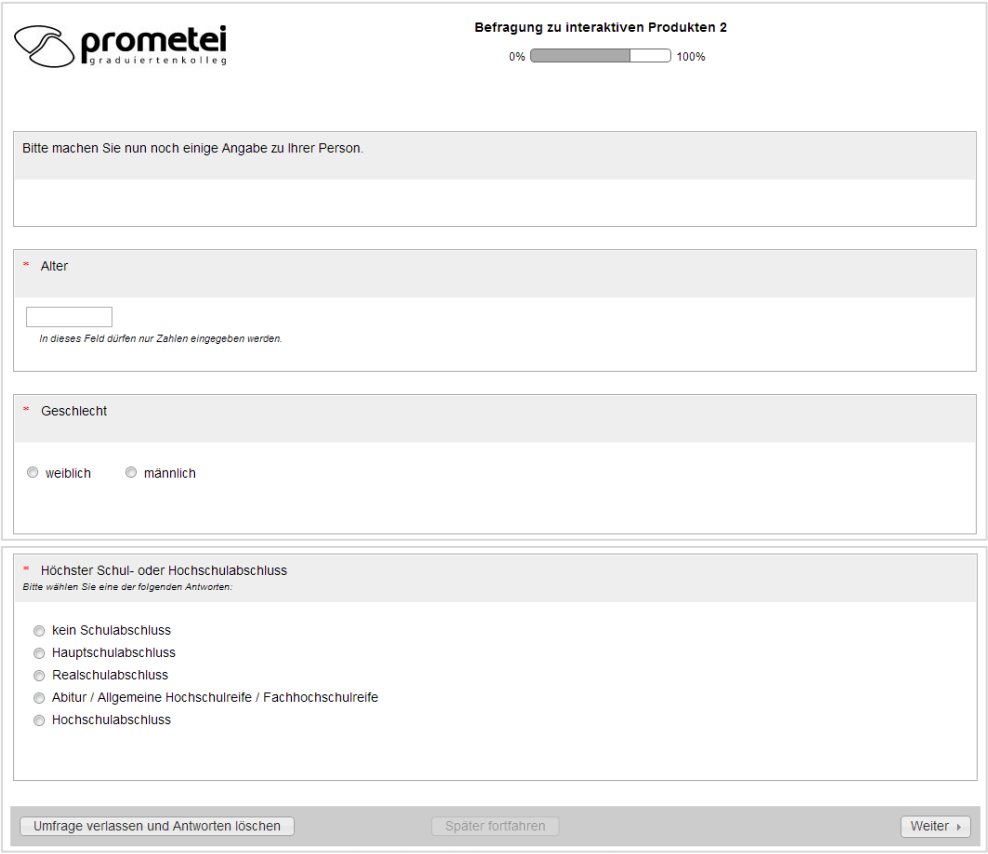
Bitte bewerten Sie, inwiefern die folgenden Aussagen auf Ihre generelle Einstellung zum Produkt zutreffen.

*

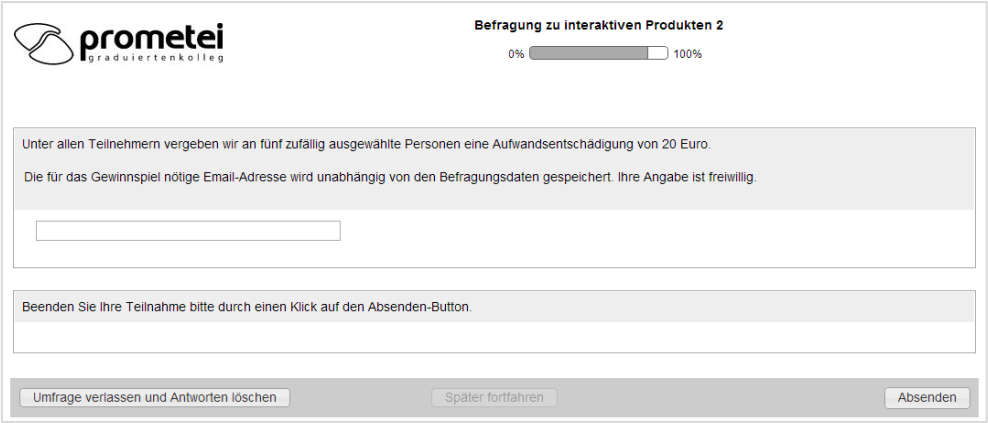
	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Wenn ich könnte, würde ich das Produkt täglich nutzen.							
Ich kann es kaum erwarten, das Produkt erneut zu benutzen.							
Ich würde mir genau dieses Produkt jederzeit wieder zulegen.							
Ich möchte das Produkt nicht mehr hergeben.							
Ich würde das Produkt gegen kein anderes eintauschen.							
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.							
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.							
Ich würde dieses Produkt einem anderen, ähnlichen Produkt vorziehen.							
Das Produkt motiviert mich es weiterhin zu nutzen.							
Ich verwende das Produkt nur, wenn es unbedingt sein muss.							

Umfrage verlassen und Antworten löschen | Später fortfahren | Weiter

Seite 7: Angaben zur Person (demographische Daten)



Seite 8: Gewinnspiel



Seite 9: Abschluss

Ihre Daten wurden gespeichert. Sie können das Fenster jetzt schließen.

Vielen Dank für die Teilname!

Bei Anregungen und Fragen können Sie sich wenden an:

Dr. Michael Minge
Technische Universität Berlin
GRK prometei
michael.minge(at)zmms.tu-berlin.de

Laura Riedel
laura.riedel(at)googlemail.com

B.2. Itemanalyse

Tabelle 29. Itemanalyse der Roh-Items.

Dimension	Item	MIN	MAX	*M*	*SD*	Var	∑	∑ (0-6)*	*p* von Dahl	Trenn-schär-fe**
Emotionen	Das Produkt macht mich zufrieden.	1	7	5.83	1.079	1.164	1387	1149	.805	.590
	Das Produkt frustriert mich.	1	7	2.52	1.383	1.913	600	362	.254	.555
	Durch das Produkt fühle ich mich träge.	1	7	2.66	1.512	2.286	632	394	.276	.451
	Das Produkt begeistert mich.	1	7	5.26	1.215	1.476	1252	1014	.710	.632
	Durch das Produkt fühle ich mich passiv.	1	7	2.78	1.483	2.199	661	423	.296	.332
	Das Produkt macht mich glücklich.	1	7	4.92	1.267	1.605	1170	932	.653	.702
	Das Produkt nervt mich.	1	7	2.75	1.511	2.282	654	416	.291	.557
	Das Produkt beschwingt mich.	1	7	4.03	1.315	1.729	958	720	.504	.606
	Das Produkt macht mich müde.	1	7	2.82	1.510	2.281	672	434	.304	.418
	Durch das Produkt fühle ich mich wohl.	1	7	4.89	1.270	1.613	1165	927	.649	.692
	Das Produkt verärgert mich.	1	7	2.74	1.456	2.120	651	413	.289	.551
	Das Produkt stimmt mich euphorisch.	1	7	3.71	1.459	2.129	884	646	.452	.596
	Das Produkt erfreut mich.	1	7	5.10	1.225	1.500	1214	976	.683	.703
	Durch das Produkt fühle ich mich hellwach.	1	7	3.64	1.412	1.994	867	629	.440	.467
	Das Produkt langweilt mich.	1	6	2.61	1.288	1.658	620	382	.268	.478
	Durch das Produkt fühle ich mich fröhlich.	1	7	4.29	1.327	1.761	1022	784	.549	.701
	Das Produkt berauscht mich.	1	7	3.17	1.508	2.275	754	516	.361	.488
	Durch das Produkt fühle ich mich ausgeglichen.	1	7	3.68	1.380	1.904	877	639	.447	.491
	Das Produkt stresst mich.	1	7	2.89	1.543	2.381	689	451	.316	.501
	Durch das Produkt fühle ich mich energiegeladen.	1	7	3.39	1.400	1.960	806	568	.398	.494
	Das Produkt macht mich wütend.	1	7	2.42	1.481	2.195	577	339	.237	.554
	Das Produkt entspannt mich.	1	7	3.96	1.440	2.074	942	704	.493	.580
	Durch das Produkt fühle ich mich erschöpft.	1	7	2.67	1.376	1.892	636	398	.279	.449
	Das Produkt belebt mich.	1	7	3.86	1.361	1.853	918	680	.476	.579
	Das Produkt beruhigt mich.	1	7	3.95	1.400	1.959	939	701	.491	.477
Nutzungs-intention	Wenn ich könnte, würde ich das Produkt täglich nutzen.	1	7	5.40	1.668	2.782	1286	1048	.733	.631
	Das Produkt motiviert mich es weiterhin zu nutzen.	1	7	5.31	1.310	1.716	1263	1025	.717	.696
	Ich kann es kaum erwarten, das Produkt erneut zu benutzen.	1	7	4.16	1.541	2.374	991	753	.527	.676
	Ich verwende das Produkt nur, wenn es unbedingt sein muss.	1	7	2.57	1.557	2.423	612	374	.262	.438
	Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.	1	7	4.76	1.688	2.850	1133	895	.627	.535
Loyalität	Ich würde dieses Produkt einem anderen, ähnlichen Produkt vorziehen.	1	7	4.55	1.652	2.730	1082	844	.591	.552
	Ich würde das Produkt gegen kein anderes eintauschen.	1	7	3.83	1.779	3.164	911	673	.471	.584
	Ich würde mir genau dieses Produkt jederzeit wieder zulegen.	1	7	5.04	1.545	2.386	1200	962	.674	.681

	Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	1	7	3.91	1.648	2.717	930	692	.485	.636
	Ich möchte das Produkt nicht mehr hergeben.	1	7	4.95	1.664	2.769	1177	939	.658	.749

* Summe umgerechnet in Abstufung von 0 bis 6

** Trennschärfe modulweise berechnet

B.3. Reliabilität und Interkorrelationen

Tabelle 30. Interkorrelation der Items und Cronbach's α der .Modulgruppe *Emotionen.*

	Items A ↓	**Items B →**		
Negative Valenz, deaktiviert		Durch das Produkt fühle ich mich passiv.	Das Produkt macht mich müde.	Durch das Produkt fühle ich mich erschöpft.
	Durch das Produkt fühle ich mich passiv.	-	.418	.454
	Das Produkt macht mich müde.	.418	-	.614
	Durch das Produkt fühle ich mich erschöpft.	.454	.614	-
			(α = .744)	
Negative Valenz, aktiviert		Das Produkt nervt mich.	Das Produkt frustriert mich.	Das Produkt verärgert mich.
	Das Produkt nervt mich.	-	.715	.706
	Das Produkt frustriert mich.	.715	-	.666
	Das Produkt verärgert mich.	.706	.666	-
			(α = .872)	
Positive Valenz, aktiviert		Das Produkt begeistert mich.	Das Produkt macht mich glücklich.	Das Produkt erfreut mich.
	Das Produkt begeistert mich.	-	.623	.544
	Das Produkt macht mich glücklich.	.623	-	.587
	Das Produkt erfreut mich.	.544	.587	-
			(α = .809)	
Positive Valenz, deaktiviert		Das Produkt beruhigt mich.	Das Produkt entspannt mich.	Durch das Produkt fühle ich mich ausgeglichen.
	Das Produkt beruhigt mich.	-	.669	.559
	Das Produkt entspannt mich.	.669	-	.554
	Durch das Produkt fühle ich mich ausgeglichen.	.559	.554	-
			(α =.815)	

Tabelle 31. Interkorrelation der Items und Cronbach's α der Modulgruppe *Konsequenzen.*

	Items A ↓	**Items B →**		
Nutzungs-intention		Wenn ich könnte, würde ich das Produkt täglich nutzen.	Ich verwende das Produkt nur, wenn es unbedingt sein muss.	Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.
	Wenn ich könnte, würde ich das Produkt täglich nutzen.	-	.388	.378
	Ich verwende das Produkt nur, wenn es unbedingt sein muss.	.388	-	.283
	Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.	.378	.283	-
			(α = .617)	
Loyalität		Ich würde das Produkt gegen kein anderes eintauschen.	Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.
	Ich würde das Produkt gegen kein anderes eintauschen.	-	.471	.448
	Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	.471	-	.451
	Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.	.448	.451	-
			(α = .714)	

C. Endgültiger Fragebogen

Nutzungserleben interaktiver Produkte

Wir untersuchen, wie Menschen interaktive technische Produkte beurteilen.

Interaktiv bedeutet, dass es sich um Technik handelt, die vom Nutzer einerseits bestimmte Eingaben erfordert und andererseits bestimmte Ergebnisse zurückmeldet. Dies kann zum Beispiel ein Handy sein, das einerseits die Eingabe einer Telefonnummer benötigt und andererseits darstellt, welche Ziffer man soeben eingegeben hat.

Mit Ihrer Arbeit unterstützen Sie die wissenschaftliche Arbeit an der TU Berlin. Für Ihre Teilnahme und Ihr Interesse bedanken wir uns vorab.

Angaben zur Person

Alter	_____	Jahre
Schulabschluss	○	Hauptschulabschluss
	○	Realschulabschluss
	○	Abitur / Allgemeine Hochschulreife / Fachhochschulreife
	○	Hochschulabschluss

Angaben zum Produkt

Wie lange nutzen Sie dieses Produkt bereits?	_____	Jahre / Monate (Unzutreffendes streichen)
Ordnen Sie das Produkt bitte einer der folgenden Produktgruppen zu.	○	Smartphone / Handy
	○	Laptop / Computer
	○	Mp3-Player
	○	Waschmaschine
	○	Software

Beispiel

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Hier passt alles zusammen.	○	○	●	○	○	○	○

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Das Design wirkt attraktiv.							
Das Produkt ist stilvoll.							
Das Produkt ist kreativ gestaltet.							
Durch das Produkt werde ich anders wahrgenommen.							
Das Produkt verleiht mir ein höheres Ansehen.							
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.							
Es dauert zu lange, um die Funktionen des Produkts zu erlernen.							
Die Bedienung des Produkts ist verständlich.							
Das Produkt lässt sich einfach benutzen.							
Ohne das Produkt kann ich nicht leben.							
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.							
Das Produkt ist wie ein Freund für mich.							
Mithilfe des Produkts kann ich meine Ziele erreichen.							
Die Funktionen des Produkts sind genau richtig für meine Ziele.							

Insgesamt halte ich das Produkt für nützlich.	○	○	○	○	○	○	○

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Durch das Produkt fühle ich mich passiv.	○	○	○	○	○	○	○
Das Produkt macht mich müde.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich erschöpft.	○	○	○	○	○	○	○
Das Produkt nervt mich.	○	○	○	○	○	○	○
Das Produkt frustriert mich.	○	○	○	○	○	○	○
Das Produkt verärgert mich.	○	○	○	○	○	○	○
Das Produkt begeistert mich.	○	○	○	○	○	○	○
Das Produkt macht mich glücklich.	○	○	○	○	○	○	○
Das Produkt erfreut mich.	○	○	○	○	○	○	○
Das Produkt beruhigt mich.	○	○	○	○	○	○	○
Das Produkt entspannt mich.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich ausgeglichen.	○	○	○	○	○	○	○

	Lehne völlig ab	Lehne ab	Lehne eher ab	Weder / noch	Stimme eher zu	Stimme zu	Stimme völlig zu
Wenn ich könnte, würde ich das Produkt täglich nutzen.							
Ich verwende das Produkt nur, wenn es unbedingt sein muss.							
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.							
Ich würde das Produkt gegen kein anderes eintauschen.							
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.							
Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.							

Vielen Danke für die Teilnahme.