

# Modulare Evaluation interaktiver Technik. Entwicklung und Validierung des meCUE Fragebogens zur Messung der User Experience.

Michael Minge<sup>1</sup>, Laura Riedel<sup>1</sup> und Manfred Thüning<sup>2</sup>

*Schlüsselwörter: User Experience, Usability, Emotionen, Evaluation, Fragebogen.*

## Zusammenfassung

Während der menschenzentrierte Gestaltungsprozess noch vor kurzer Zeit vor allem darauf abzielte, eine möglichst hohe Gebrauchstauglichkeit sicherzustellen, geht es mittlerweile darum, auch die Wahrnehmung und das Erleben des Nutzers bei Interaktion mit technischen Geräten zu berücksichtigen und positiv zu beeinflussen. Zur adäquaten Erfassung dieses vielschichtigen Konzepts „User Experience“ wird ein theoriebasierter Fragebogen zur modularen Evaluation von Technik (*meCUE*) vorgestellt. Dieser Beitrag fasst die bisherigen Entwicklungsschritte bei der Konstruktion und Validierung des Fragebogens zusammen. Speziell wird die empirische Überprüfung einer Minimalrevision und einer inhaltlichen Erweiterung behandelt.

## Theoretischer Hintergrund

Die Berücksichtigung des Nutzungserlebens („*User Experience*“) stellt Entwickler und Hersteller von Technik vor die Herausforderung, subjektive Bewertungen aus Nutzerperspektive in einer geeigneten Art und Weise zu erheben und die erzielten Ergebnisse im menschenzentrierten Gestaltungsprozess nachhaltig zu berücksichtigen (DIN EN ISO 9241-2010, 2011). Neben qualitativen Verfahren zur systematischen Ermittlung persönlicher Eindrücke eröffnen Fragebögen die Möglichkeit zur standardisierten und ökonomischen Erfassung subjektiver Bewertungen.

Um relevante Aspekte im Nutzungserleben möglichst ganzheitlich zu erfassen, wurde auf Basis eines weithin etablierten Modells, dem CUE-Modell von Thüning und Mahlke (2007), ein modular aufgebauter Fragebogen entwickelt. In Anlehnung an Hassenzahl (2001) unterscheidet das CUE-Modell zwischen der Wahrnehmung aufgabenbezogener und nicht-aufgabenbezogener Produktqualitäten und berücksichtigt darüber hinaus Nutzeremotionen als einen mediierenden Faktor für die Ausformung bestimmter Konsequenzen, wie z.B. der Akzeptanz (siehe Abb. 1).

Die Zielstruktur des neuen Fragebogens wurde darauf aufbauend entwickelt. Als zentrale Aspekte wurden dabei die separat anwendbaren Module „Produktwahrnehmungen“ (Nützlichkeit, Benutzbarkeit, visuelle Ästhetik, Status, Bindung), „Nutzeremotionen“ (positive und negative Emotionen) und „Konsequenzen“ (Produktloyalität, Nutzungsintention) mit den genannten Subskalen definiert.

---

<sup>1</sup> Technische Universität Berlin, Graduiertenkolleg prometei

<sup>2</sup> Technische Universität Berlin, Fachgebiet Kognitionspsychologie und Kognitive Ergonomie

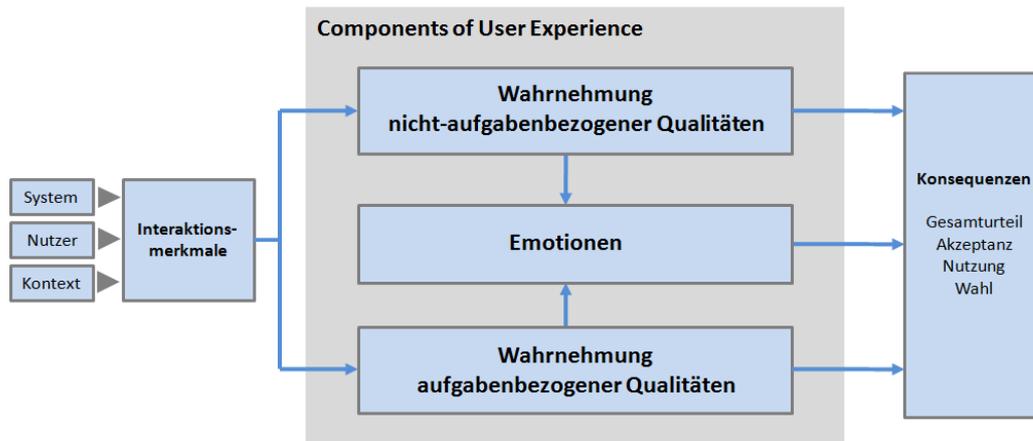


Abb. 1: Modell des Nutzungserlebens (CUE) nach Thüring und Mahlke (2007, S. 262)

## Empirische Vorarbeiten

Die Konstruktion des Fragebogens zur modularen Evaluation zentraler Aspekte der User Experience (*meCUE*) erfolgte auf Basis zweier Online-Datenerhebungen. Der anfängliche Itempool bestand aus 67 Items, die deduktiv entsprechend der Zielstruktur formuliert und den Subskalen theoriegeleitet zugeordnet wurden. Alle Items waren als Aussagen formuliert, deren Zustimmung über ein Likert skaliertes Antwortformat erfasst wird („lehne völlig ab“, „lehne ab“, „lehne eher ab“, „weder noch“, „stimme eher zu“, „stimme zu“, stimme völlig zu“). Beide Stichproben (jeweils  $n = 238$  Probanden) beinhalteten Bewertungen zu technischen Produkten unterschiedlicher Art. Die Itemselektion und Validierung erfolgte durch separate Faktorenanalysen. Im Ergebnis entstand ein Fragebogen, der die Zielstruktur über 33 Items erfasst und solide Werte sowohl hinsichtlich der Varianzaufklärung in den Modulen „Produktwahrnehmungen“ (69,6 %), „Emotionen“ (57,4 %) und „Konsequenzen“ (63,5 %) aufwies als auch hinsichtlich der internen Konsistenz seiner Subskalen mit *Cronbachs Alpha* Werten zwischen .69 und .83 (Minge & Riedel, accepted).

Die faktorielle Struktur konnte mit den Daten aus einer laborexperimentellen Untersuchung stabil repliziert werden (Minge & Riedel, accepted). Hierbei bewerteten  $n = 67$  Probanden jeweils drei interaktive Produkte aus verschiedenen Bereichen (Textbearbeitungssoftware, Audio-Player und Mobiltelefon). Aufgrund größerer Varianzen zwischen den in diesem Fall vorgegebenen Produkten zeigten sich die Anteile erklärter Varianz bei Analyse der vorliegenden 201 Datensätze nochmals erhöht (81,1 % für „Produktwahrnehmungen“, 74,3 % für „Emotionen“ und 74,1 % für „Konsequenzen“). Ebenso unterstützten die Daten erneut die Annahme guter interner Konsistenz aller Subskalen ( $.83 < \text{Cronbachs Alpha} < .94$ ).

Um die Validität des Verfahrens abzuschätzen, wurden im Rahmen der Laborstudie Zusammenhänge zu bestehenden Fragebögen erhoben. Hierbei zeigten sich substantiell hohe Korrelationen vor allem zwischen den Dimensionen „Benutzbarkeit“ (*meCUE*), „pragmatische Qualität“ (*AttrakDiff* nach Hassenzahl, Burmester & Koller, 2008) sowie „Durchschaubarkeit“ und „Vorhersagbarkeit“ (*UEQ* nach

Laugwitz, Schrepp & Held, 2006). Werte der „visuellen Ästhetik“ (*meCUE*) waren hoch verbunden mit den Dimensionen klassischer und expressiver Ästhetik nach Lavie und Tractinsky (2004). Im Bereich Emotionen lagen signifikante Korrelationen zwischen den Dimensionen des *meCUE* und den Skalen des *PANAS* (Watson, Clark & Tellegen, 1988) vor. Erwartungsgemäß korrelierte das Außenkriterium der bearbeiteten Aufgabenmenge mit den subjektiven Angaben zur „Nützlichkeit“ und „Benutzbarkeit“ (*meCUE*), wohingegen zu den nicht-aufgabenbezogenen Dimensionen keine substanziellen Korrelationen vorlagen (Minge & Riedel, accepted).

Zusammenfassend zeigen diese Arbeiten auf, dass die theoriegeleitete Konstruktion eines zuverlässigen und validen Messinstrumentes zur Erhebung zentraler Aspekte des Nutzungserlebens erfolgreich gelungen ist. Hinsichtlich der Formulierung einzelner Items und der angebotenen Module bestand allerdings noch Handlungsbedarf. Dieser wurde durch eine Minimalrevision und eine inhaltliche Erweiterung des *meCUE* Fragebogens abgedeckt und mit der folgenden Studie überprüft.

## Zielsetzung

Das Ziel der Studie bestand einerseits darin, den revidierten Fragebogen hinsichtlich seiner faktoriellen Struktur zu überprüfen und andererseits die Validität eines neuen Moduls zur Erfassung eines Gesamturteils zu bewerten. Die Revision umfasste die Abwandlung zweier Items, wobei ein Item umformuliert wurde, um eine einheitliche Polung in der Subskala „Benutzbarkeit“ zu erzielen und ein weiteres Item, um dessen inhaltliche Aussagekraft zu fokussieren (siehe Tab. 1).

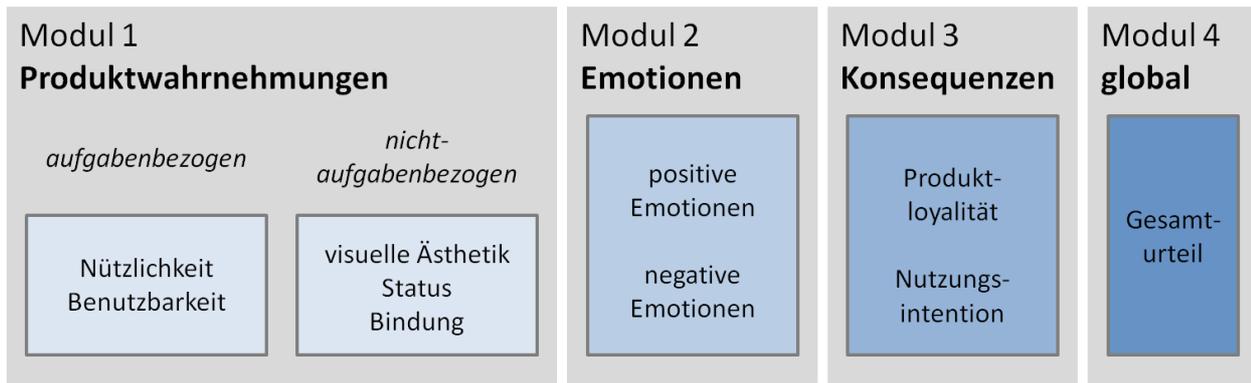
**Tab. 1: umformulierte Items in der minimal revidierten Version**

bisherige Formulierung	revidierte Formulierung
Es dauert zu lange, die Funktionen des Produkts zu erlernen.	Es wird schnell klar, wie das Produkt zu bedienen ist.
Insgesamt halte ich das Produkt für nützlich.	Ich halte das Produkt für absolut nützlich.

Erweitert wurde die modulare Struktur des Fragebogens um die Erhebung eines Gesamturteils (Modul 4). Verfügbare Fragebögen zum Nutzungserleben (z.B. *AttrakDiff* und *UEQ*) erfassen dieses Gesamturteil unter der Bezeichnung „Attraktivität“ als zusätzliche Subskala. In *meCUE* wurde die Skala als feinabgestuftes semantisches Differenzial mit den Polen „schlecht“ und „gut“ realisiert. Die bisherige Umsetzung im Fragebogentool LimeSuvey® gestattet, die Auswahl zwischen den insgesamt 21 Abstufungen über einen Schieberegler vorzunehmen (siehe Abb. 2). Die durch dieses Modul erweiterte Struktur von *meCUE* zeigt Abbildung 3.



**Abb. 2: Realisierung des Moduls 4 als Schieberegler zur Erhebung des Gesamturteils**



**Abb. 3: modulare Struktur der revidierten Version des meCUE Fragebogens**

## Methode

Zur Überprüfung der faktoriellen Struktur des überarbeiteten Fragebogens wurde eine Online-Datenerhebung durchgeführt. Hierzu waren  $n = 237$  Probanden aufgefordert, jeweils ein interaktives technisches Produkt aus ihrem Alltag anhand des minimal revidierten meCUE Fragebogens zu bewerten. Die Stichprobe setzte sich aus 139 Teilnehmerinnen und 98 Teilnehmern zusammen und wies ein Durchschnittsalter von 29,8 Jahren auf. Die vier Module des Fragebogens wurden in der in Abbildung 3 beschriebenen Reihenfolge fest vorgegeben, während die einzelnen Items innerhalb eines jeden Moduls randomisiert präsentiert wurden.

Um die Validität des neu konstruierten Schiebereglers zur Erhebung des globalen Produkturteils (Modul 4) abzuschätzen, wurde zusätzlich zum meCUE auch der AttrakDiff Fragebogen von Hassenzahl, Burmester und Koller (2008) eingesetzt. Entsprechend der theoretischen Vorüberlegungen wurden hierbei substantielle Zusammenhänge zwischen dem Modul 4 des meCUE und der Dimension „Attraktivität“ des AttrakDiff erwartet. Wie sich der Datensatz hinsichtlich Art und Häufigkeit der bewerteten technischen Produkte zusammensetzt, ist in Tabelle 2 aufgeführt.

**Tab. 2: Art und Häufigkeit der bewerteten Produkte in der Stichprobe**

Produktbereich	absolute Häufigkeit	relative Häufigkeit
Mobiltelefon	99	41,8 %
Laptop/ Computer	75	31,6 %
Kaffeemaschine	22	9,3 %
mobiler Audioplayer (MP3)	16	6,8 %
Smartphone App	9	3,8 %
Waschmaschine	7	3,0 %
Digitalkamera	6	2,5 %
Software	3	1,3 %
<b>Summe</b>	<b>237</b>	<b>100,0 %</b>

## Ergebnisse

In einem ersten Schritt erfolgte die modulweise Validierung der revidierten Fragebogenversion durch separate Faktorenanalysen der mehrdimensionalen Module.

Auf Basis des Minimum-Average-Partial-Tests (MAP-Test nach Velicer, 1976) ergab die varianzmaximierende Hauptkomponentenanalyse der zugrundeliegenden 15 Items des Moduls „Produktwahrnehmungen“ die erwarteten fünf Hauptkomponenten. Insgesamt erklären diese fünf Faktoren 73,7 % der aufgetretenen Varianz (siehe Tab. 3). Die ebenfalls in Tabelle 3 dargestellte faktorielle Struktur zeigt, dass die bei der Konstruktion etablierte Struktur stabil repliziert wird und die inhaltliche Bezeichnung der Subskalen weiterhin aufrecht erhalten werden kann. Auf Basis der *Cronbachs Alpha* Koeffizienten ist die interne Konsistenz der Subskalen als ausreichend bis gut zu bezeichnen.

**Tab. 3: Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Produktwahrnehmung“ (dokumentiert sind alle Ladungen > .4)**

<b>Item:</b>	<b>Nützlich- keit</b>	<b>Benutz- barkeit</b>	<b>visuelle Ästhetik</b>	<b>Status</b>	<b>Bindung</b>
Mithilfe des Produkts kann ich meine Ziele erreichen.	.836				
Ich halte das Produkt für absolut nützlich.	.765				
Die Funktionen des Produkts sind genau richtig für meine Ziele.	.654				
Es wird schnell klar, wie das Produkt zu bedienen ist.		.863			
Die Bedienung des Produkts ist verständlich.		.859			
Das Produkt lässt sich einfach benutzen.		.853			
Das Design wirkt attraktiv.			.876		
Das Produkt ist kreativ gestaltet.			.845		
Das Produkt ist stilvoll.			.828		
Durch das Produkt werde ich anders wahrgenommen.				.875	
Das Produkt verleiht mir ein höheres Ansehen.				.830	
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.			.401	.578	
Ohne das Produkt kann ich nicht leben.					.827
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.					.814
Das Produkt ist wie ein Freund für mich.					.756
<b>Cronbachs Alpha</b>	<b>.73</b>	<b>.86</b>	<b>.86</b>	<b>.76</b>	<b>.80</b>
<b>Varianzaufklärung</b>	<b>12,9</b>	<b>16,3</b>	<b>16,0</b>	<b>13,2</b>	<b>15,3</b>

Im Modul „Nutzeremotionen“ ergab die varianzmaximierende Faktorenanalyse der zwölf Items ebenfalls die erwarteten zwei zugrundeliegenden Faktoren für das Erleben positiver und negativer Emotionen. Gemeinsam erklären diese 66,6 % der aufgetretenen Varianz. Zudem liegen für beide Subskalen Cronbachs Alpha-Werte vor, die die Annahme guter interner Konsistenz unterstützen (siehe Tab. 4).

**Tab. 4 : Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Nutzeremotionen“ (dokumentiert sind alle Ladungen > .4)**

<b>Item:</b>	<b>positive Emotionen</b>	<b>negative Emotionen</b>
Das Produkt entspannt mich.	.858	
Durch das Produkt fühle ich mich ausgeglichen.	.846	
Durch das Produkt fühle ich mich fröhlich.	.824	
Das Produkt beruhigt mich.	.822	
Das Produkt beschwingt mich.	.809	
Das Produkt stimmt mich euphorisch.	.805	
Das Produkt frustriert mich.		.866
Das Produkt nervt mich.		.826
Das Produkt macht mich müde.		.818
Das Produkt verärgert mich.		.800
Durch das Produkt fühle ich mich erschöpft.		.781
Durch das Produkt fühle ich mich passiv.		.625
<b>Cronbachs Alpha</b>	<b>.91</b>	<b>.88</b>
<b>Varianzaufklärung</b>	<b>35,0</b>	<b>31,6</b>

Bei Analyse der verbliebenen sechs Items im Modul „Konsequenzen“ wurden unter Bezugnahme auf den MAP-Test zwei Faktoren extrahiert, die zusammen 65,5 % der aufgetretenen Varianz erklären. Die Verteilung der Ladungsmuster auf beide Dimensionen unterstützt die nach Abschluss der Konstruktionsphase formulierte inhaltliche Benennung. Die Cronbachs Alpha-Werte sprechen für eine ausreichende interne Konsistenz der zwei Subskalen (siehe Tab.5).

**Tab. 5: Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Konsequenzen“ (dokumentiert sind alle Ladungen > .4)**

<b>Item:</b>	<b>Produkt-loyalität</b>	<b>Nutzungs-intention</b>
Ich würde das Produkt gegen kein anderes eintauschen.	.873	
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	.855	
Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.	.597	.494
Ich kann es kaum erwarten, das Produkt erneut zu verwenden.		.840
Wenn ich könnte, würde ich das Produkt täglich nutzen.		.755
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.		.711
<b>Cronbachs Alpha</b>	<b>.70</b>	<b>.75</b>
<b>Varianzaufklärung</b>	<b>32,3</b>	<b>33,2</b>

In einem zweiten Schritt wurden die Angaben zum globalen Produkturteil ausgewertet und auf Zusammenhänge untersucht. Einerseits wurde das Gesamturteil über einen Schieberegler im neu konstruierten Modul 4 des *meCUE* erhoben, andererseits über die *AttrakDiff* Dimension „Attraktivität“, die wiederum das arithmetische Mittel aus Antworten auf insgesamt sieben Items darstellt. Die substantielle Korrelation in Höhe von  $r = .559$  zwischen beiden Variablen unterstützt die Annahme konvergenter Validität des Schiebereglers. Statistisch bedeutsame Korrelationen, allerdings von geringerer Höhe, liegen darüber hinaus zu den übrigen Subskalen des *AttrakDiff* vor (siehe Tab. 6).

**Tab. 6: Korrelationen zwischen dem Modul 4 (*meCUE*) und den *AttrakDiff* Dimensionen**  
 Anmerkung: \*  $p < .05$ , \*\*  $p < .01$

Korrelation $r$	<i>AttrakDiff</i>			
	Pragmatik	Stimulation	Identifikation	Attraktivität
<i>meCUE</i> Gesamturteil	.477**	.448**	.439**	.559**

## Diskussion

Die Ergebnisse aus der vorliegenden Studie unterstützen die bisherigen Befunde zur Qualität und Güte des theoriebasiert entwickelten Fragebogens zur modularen Evaluation zentraler Aspekte des Nutzungserlebens, *meCUE*. Insbesondere zeichnet sich der *meCUE* Fragebogen durch eine besonders stabile faktorielle Struktur und eine ausreichende bis gute interne Konsistenz seiner Subskalen aus. Mit der Revision wurde ein konsistentes Antwortformat realisiert, mit dem sich Nutzerbewertungen in kurzer Zeit (ca. 3-5 Minuten) äußerst effizient einholen lassen.

Die Erweiterung des Fragebogens um ein viertes Modul zur Erhebung eines allgemeinen Gesamturteils kann ebenfalls als gelungen bezeichnet werden. Der ermittelte Zusammenhang zu vergleichbaren Skalen ist, insbesondere vor dem Hintergrund der unterschiedlichen Skalierung beider Variablen und der Tatsache, dass es sich bei dem *meCUE* Gesamturteil um eine Single-Item Lösung handelt, substantiell und ausreichend hoch. Für den Fall, dass verschiedene Gestaltungslösungen miteinander verglichen werden sollen, erlaubt die Erhebung eines Gesamturteils aus anwendungsorientierter Sicht bei inferenzstatistischer Überprüfung vorliegender Unterschiede das Aufstellen einer eindeutigen Rangfolge der Untersuchungsobjekte. Aus forschungstheoretischer Sicht eröffnet sich durch die Erhebung des Gesamturteils die Möglichkeit zur Bildung eines Regressionsmodells, um die allgemeine Produktgüte vorherzusagen. Dadurch werden zukünftig nicht nur zusätzliche Aussagen zur strukturellen Güte des Fragebogens ermöglicht, sondern auch zu spezifischen Kausalzusammenhängen zwischen den Komponenten des CUE-Modells.

Um den vorliegenden Fragebogen auch als reine Papier-Bleistift-Variante einsetzen zu können, wird in einer aktuellen Studie die Übereinstimmung der Angaben im Modul 4 zwischen den verschiedenen Eingabeformen überprüft.

## Fazit

Auf Basis eines integrativen Modells zur User Experience (CUE-Modell) wurde der *meCUE* Fragebogen zur modularen Evaluation zentraler Aspekte des Nutzungserlebens erfolgreich konstruiert. Er wurde theoriebasiert entwickelt und ist sowohl bei seiner Konstruktion als auch in zusätzlichen Studien unter Einsatz einer breiten Vielfalt unterschiedlicher interaktiver Produkte empirisch validiert worden (Minge & Riedel, accepted; Doria, Minge & Riedel, accepted; Minge, Riedel & Thüring, accepted). Damit bietet er eine ökonomische, valide und zuverlässige Alternative zu bereits bestehenden Verfahren. Insbesondere die modulare Struktur unter Berücksichtigung von Produktwahrnehmungen, Nutzeremotionen und Konsequenzen erlaubt es, quantitative Vergleiche zwischen verschiedenen Gestaltungslösungen oder zwischen verschiedenen Nutzungszeitpunkten zielorientiert und ganzheitlicher aus nutzerzentrierter Sicht vorzunehmen.

## Literatur

- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- DIN EN ISO 9241-210 (2011). Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.
- Doria, L., Minge, M. & Riedel, L. (accepted). User-centred evaluation of lower-limb orthoses: A new approach. *Dreiländertagung der Deutschen, Schweizerischen und Österreichischen Gesellschaft für Biomedizinische Technik (BMT)*. 19.-21.09.2013, Graz.
- Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, 13(4), 481-499.
- Hassenzahl, M., Burmester, M. & Koller, F. (2008). Der User Experience (UX) auf der Spur: Zum Einsatz von *www.attrak.diff*. In: H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner & K. Rose. (Hrsg.): *Usability Professionals 2008*, 78-82.
- Laugwitz, B., Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In A. M. Heinecke & H. Paul (Hrsg.), *M&C 2006: Mensch und Computer im Strukturwandel* (S. 125-134). München: Oldenbourg.
- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 269-298.
- Minge, M. & Riedel, L. (accepted). *meCUE* – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. *Mensch und Computer*, 08.-11.09.2013, Bremen.
- Minge, M., Riedel, L. & Thüring, M. (accepted). Und ob du wirklich richtig stehst... Zur diskriminativen Validität des User Experience Fragebogens „*meCUE*“. Beitrag im Workshop „Temporale Aspekte des Nutzungserlebens“, *Mensch und Computer*, 08.09.2013, Bremen.
- Thüring, M. & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. *International Journal of Psychology*, 42(4), 253-264.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Watson, D., Clark, A. & Tellegen, A. (1988). Development and Validation of Brief Measure of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.