

# Und ob du wirklich richtig stehst... Zur diskriminativen Validität des User Experience Fragebogens „meCUE“

Michael Minge<sup>1</sup>, Laura Riedel<sup>1</sup> & Manfred Thüring<sup>2</sup>

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

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

## **Zusammenfassung**

In diesem Beitrag werden die bisherigen Entwicklungsschritte bei der Konstruktion eines modularen Fragebogens zur Messung des Interaktionserlebens zusammengefasst. Inhaltlich orientiert sich die Fragebogenstruktur am Komponentenmodell des Nutzungserlebens, *CUE*, nach Thüring und Mahlke (2007). Die Validierung fand sowohl im Rahmen mehrerer Datenerhebungen im Feld als auch unter laborexperimentellen Bedingungen statt. Speziell geht dieser Beitrag auf erste Befunde zur Bestimmung der diskriminativen Validität ein, also zur Fähigkeit des Fragebogens, vorliegende Unterschiede im Erleben interaktiver Produkte zu identifizieren.

## 1 Einleitung

Bei der Evaluation von Technik eröffnen Fragebögen die Möglichkeit, subjektive Bewertungen aus nutzerzentrierter Perspektive standardisiert und ökonomisch zu erheben. Fragebögen sind insbesondere dann zweckmäßig, wenn das Ziel besteht, verschiedene Gestaltungslösungen relativ zueinander zu vergleichen oder bei einem technischen Produkt Veränderungen im Interaktionserleben über einen längeren Nutzungszeitraum zu erfassen. Entsprechend des „User Experience Lifecycle Models“ (*ContinUE*) erfolgt die Bewertung in der Regel in retrospektiver Weise in der *Past Use Phase* (Pohlmeyer, Hecht & Blessing, 2009).

Zur möglichst ganzheitlichen Erfassung des Nutzungserlebens wurde auf Basis eines weithin etablierten und empirisch abgesicherten Modells zur User Experience, dem *CUE*-Modell von Thüring und Mahlke (2007), der modular aufgebaute Fragebogen *meCUE* entwickelt.

Zur Erfassung der a priori abgeleiteten Dimensionen wurden zunächst 67 Items deduktiv vorgeschlagen und in einem siebenfach gestuften Likert-Skalenformat formuliert. Die Itemselektion erfolgte auf Basis zweier Datenerhebungen im Feld und führte zu einer Rohversion mit 33 Items. Neben einer Validierung des Gesamtfragebogens erfolgte eine Zuordnung der Dimensionen zu drei einzeln einsetzbaren Modulen, die separat validiert wurden. Das erste Modul bezieht sich auf die Wahrnehmung aufgabenbezogener und nicht-aufgabenbezogener Produktqualitäten (Nützlichkeit, Benutzbarkeit, visuelle Ästhetik, Status, Bindung), das zweite Modul auf Nutzeremotionen (positive sowie negative Emotionen) und das dritte Modul auf Konsequenzen der Interaktion (Produktloyalität, Nutzungsintention).

Die faktorielle Struktur konnte im Rahmen einer laborexperimentellen Studie bereits erfolgreich repliziert werden (Minge & Riedel, submitted). Ebenso konnte aufgezeigt werden, dass alle Skalen eine ausreichende bis sehr gute interne Konsistenz aufweisen ( $.76 < \text{Cronbachs Alpha} < .94$ ), und dass erwartungsgemäße Korrelationen sowohl zu inhaltlich vergleichbaren Skalen anderer Fragebögen als auch zu objektiven Außenkriterien, wie der bearbeiteten Aufgabenmenge in einem vorgegebenen Zeitrahmen, vorliegen (ebenda).

Neben einer empirischen Absicherung der faktoriellen Struktur, der Reliabilität und der Konstruktvalidität des konstruierten Verfahrens bezog sich eine wesentliche Fragestellung in dieser laborexperimentellen Studie bereits auf eine erste Bestimmung der diskriminativen Validität bzw. Sensitivität des Fragebogens und damit auf die Fähigkeit, vorliegende Unterschiede im Erleben zwischen verschiedenen interaktiven Systemen identifizieren zu können.

## 2 Bestimmung der diskriminativen Validität

### 2.1 Methode

Die Fähigkeit des Fragebogens, vorliegende Unterschiede im Interaktionserleben zu identifizieren, wurde laborexperimentell geprüft. Hierzu waren Probanden aufgefordert, die Interaktion mit Produkten unterschiedlicher Art (Audio-Player und Anwendungssoftware) zu bewerten. Aus beiden Produktbereichen kamen jeweils zwei verschiedene Gestaltungsvarianten zum Einsatz. Die Varianten unterschieden sich in erster Linie hinsichtlich ihres Bekanntheitsgrades und ihrer aktuellen Marktpräsenz. So handelte es sich bei der Textbearbeitungssoftware um eine gängige kostenpflichtige Standardanwendung einerseits und einer frei verfügbaren, quelloffenen Software andererseits. Bei den digitalen Audio-Playern wurde das Produkt eines Marktführers gegen das Produkt eines Außenseiters getestet. Bei allen vier Varianten handelte es sich also um reale Produkte, wobei in diesem Beitrag allerdings sowohl die Namen der Hersteller als auch die konkrete Produktbezeichnung ungenannt bleiben.

Alle Probanden bekamen Systeme aus beiden Bereichen vorgelegt (Messwiederholung). Innerhalb der Produktarten bewertete jeder Proband jeweils nur eine Variante (Zwischensubjektfaktor). Die Reihenfolge der Produktgruppen wurde über alle Teilnehmer ausbalanciert, die Zuordnung der Gestaltungsvarianten wurde randomisiert.

Die Beurteilung der Produkte erfolgte nach einer jeweils angeleiteten fünfminütigen Interaktionsphase. Bei der Textbearbeitung wurden die Probanden instruiert, einen vorgelegten Brief inklusive aller Formatierungen abzutippen. Beim mobilen Audioplayer ging es darum, Musikstücke anzuhören und abschließend einen persönlichen Favoriten zu nennen. Hierzu konnten die Teilnehmer unter einer kontrollierten Auswahl zur Verfügung gestellter Musikstücke selbstständig entscheiden, welche Lieder sie wie lange bzw. wie oft anhören wollten.

Neben dem neu konstruierten Instrument kamen als weitere Fragebogenverfahren im Bereich Nutzungserleben zusätzlich der AttrakDiff (Hassenzahl, Burmester & Koller 2008) und der User Experience Questionnaire UEQ (Laugwitz, Schrepp & Held, 2006) zum Einsatz.

An der Studie nahmen 67 Personen mit einem Durchschnittsalter von 28.8 Jahren ( $s = 8.5$ ) teil, darunter 34 Frauen und 33 Männer. Eine Untersuchung dauerte ungefähr 50 Minuten.

## 2.2 Ergebnisse zur Produktgruppe „Mobiler Audio-Player“

Zur Analyse von Bewertungsunterschieden zwischen den Audio-Playern wurde eine multivariate Varianzanalyse mit dem zweifachgestuften Zwischensubjektfaktor Produktvariante (A bzw. B) gerechnet. Die Ergebnisse des *meCUE* Fragebogens sind in Abbildung 1, diejenigen des AttrakDiff und des UEQ in Abbildung 2 dargestellt. Die durchweg positiveren Bewertungen für das Produkt A erweisen sich auf den erhobenen Dimensionen aller Fragebögen als statistisch bedeutsam mit Werten für die Effektstärke  $d$  im mindestens mittleren bis hohen Bereich (Cohen, 1988). Auffällig erscheint, dass Urteile auf den Fragebögen AttrakDiff und UEQ mit wenig Variation über die Dimensionen hinweg entweder im überdurchschnittlichen (Produkt A) oder im unterdurchschnittlichen Bereich (Produkt B) liegen, wohingegen beim *meCUE* Fragebogen eine höhere Variation zwischen den Dimensionen beobachtet werden kann. So fallen beispielsweise auch für das Produkt A Bewertungen zum Status, zur Bindung und zur Loyalität trotz erkannter relativer Vorteile tendenziell unterdurchschnittlich aus.

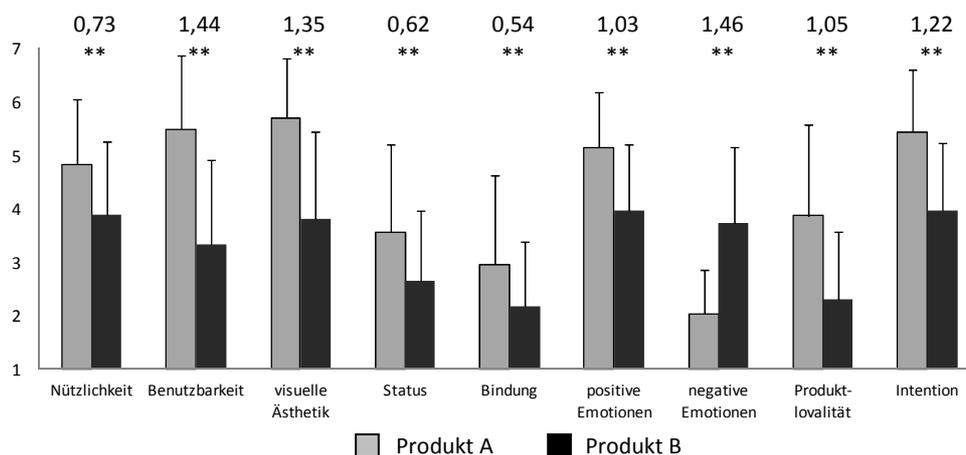


Abbildung 1: Ergebnisse des *meCUE* Fragebogens zur Produktgruppe „Mobiler Audio-Player“, als Werte angegeben werden Effektstärken  $d$  nach Cohen (1988); \*\*  $p < .01$ .

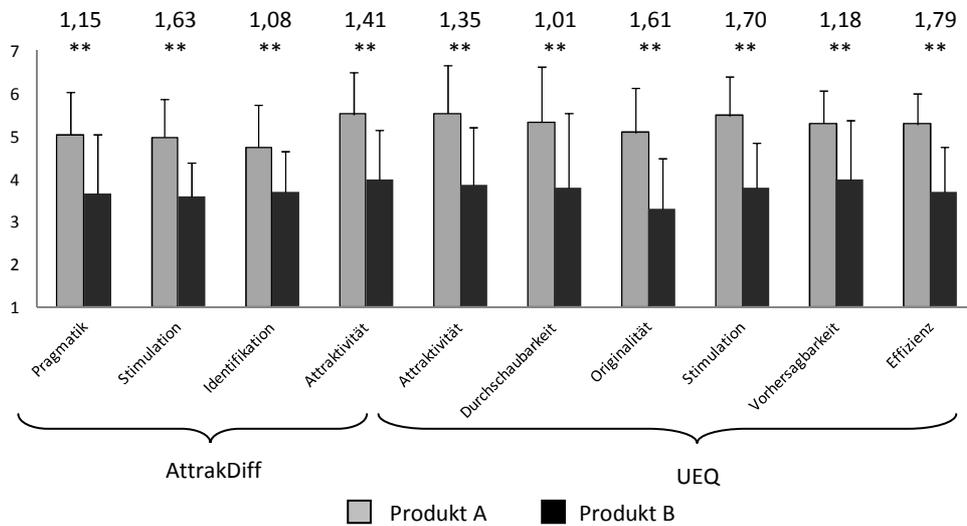


Abbildung 2: Ergebnisse der Fragebögen AttrakDiff und UEQ zur Produktgruppe „Mobiler Audio-Player“, als Werte angegeben werden Effektstärken d nach Cohen (1988); \*\* p < .01.

### 2.3 Ergebnisse zur Produktgruppe „Textbearbeitungssoftware“

Die Bewertungen der beiden Textbearbeitungsprogramme wurden ebenfalls mit einer multivariaten Varianzanalyse und dem zweifachgestuften Zwischensubjektfaktor Produktvariante (C bzw. D) ausgewertet. Hierbei erwiesen sich die Unterschiede auf den Dimensionen des meCUE und des AttrakDiff als statistisch nicht bedeutsam (siehe Abb. 3 und 4). Allerdings konnten mit dem UEQ die auch mit dem meCUE und AttrakDiff tendenziell vorliegenden Vorteile bei der Bewertung pragmatischer Produktqualitäten (Durchschaubarkeit, Vorhersagbarkeit, Effizienz) für das Produkt C inferenzstatistisch gestützt werden (siehe Abb. 4).

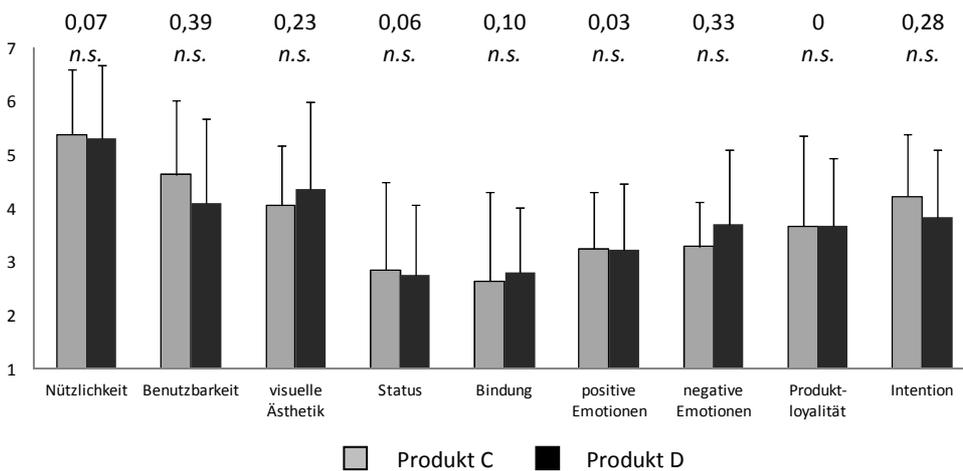


Abbildung 3: Ergebnisse des meCUE Fragebogens zur Produktgruppe „Textbearbeitungssoftware“, als Werte angegeben werden Effektstärken d nach Cohen (1988).

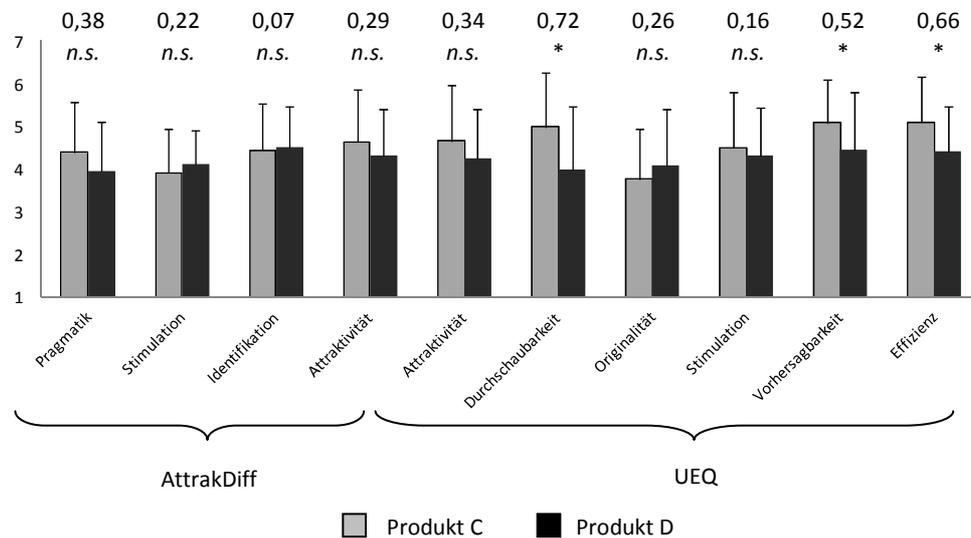


Abbildung 4: Ergebnisse der Fragebögen AttrakDiff und UEQ zur Produktgruppe „Textbearbeitungssoftware“, als Werte angegeben werden Effektstärken  $d$  nach Cohen (1988); \*  $p < .05$ .

### 3 Diskussion

Zusammenfassend kann die Annahme diskriminativer Validität des *meCUE* Fragebogens auf Basis der vorliegenden Ergebnisse vorläufig gestützt werden. Neben der theoriegeleiteten Konstruktion des Fragebogens und der empirisch abgesicherten inneren und äußeren kriteriumsbezogenen Validität liegt ein wesentlicher Vorteil und der Mehrwert, der durch die Verwendung des neu entwickelten Messinstrumentes im Vergleich zu den bereits verfügbaren Fragebögen AttrakDiff und UEQ erzielt wird, in seiner mehrdimensionalen und modularen Fragenbogenstruktur. Diese Struktur ermöglicht es, zusätzliche Aspekte beim Erleben von Technik zu erheben, welche über die reine Wahrnehmung und Bewertung von Produkteigenschaften hinaus gehen, wie z.B. Emotionen, Ästhetik und Nutzungsintention.

Einschränkend muss auf Basis der vorliegenden Befunde jedoch erwähnt werden, dass zumindest für den Bereich Anwendungssoftware der UEQ, vermutlich auch aufgrund seiner facettenreicheren Substruktur, sensitiver erscheint, um die Wahrnehmung pragmatischer Produktqualitäten zu erfassen. Die identifizierten Differenzen stehen einerseits in Zusammenhang mit den tendenziell vorliegenden Unterschieden im *meCUE* und im *AttrakDiff*, andererseits können sie auch durch Maße der Performanz zusätzlich gestützt werden. So zeigte sich für das Produkt C, welches mit höheren Werten im aufgabenbezogenen Bereich beurteilt wurde, dass damit im zeitlich begrenzten Zeitrahmen auch eine signifikant höhere Bearbeitungsleistung erreicht werden kann.

In diesem Zusammenhang wird es für zukünftige Studien noch stärker erforderlich sein, eine systematische Vorgehensweise bei der Auswahl von Testprodukten vorzunehmen und die

Analyse von Produktunterschieden zusätzlich durch andere methodische Verfahren im Vorfeld sicherzustellen (z.B. Heuristische Evaluation, Expertenbefragung, Fokusgruppen).

Mit der Erweiterung des Fragebogens um ein viertes Modul zur Erhebung eines globalen Produkturteils wurde jüngst eine revidierte Endfassung des Fragebogens erstellt und überprüft (Minge, Riedel & Thüring, accepted). Eine Übersicht aller Items dieser Endfassung findet sich im Anhang dieses Beitrages.

### **Danksagung**

Wir danken der Deutschen Forschungsgemeinschaft (DFG) für ihre Förderung im Rahmen des Graduiertenkollegs 1013/2: „Prospektive Gestaltung von Mensch-Technik-Interaktion“ (prometei).

### **Literaturverzeichnis**

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2. Auflage. Hillsdale: Erlbaum Associates.
- 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.
- Minge, M. & Riedel, L. (accepted). meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. *Mensch und Computer 2013*, 08.-11.09.2013, Bremen.
- Minge, M., Riedel, L. & Thüring, M. (accepted). Modulare Evaluation von Technik. Entwicklung und Validierung des meCUE Fragebogens zur Messung der User Experience. *10. Berliner Werkstatt Mensch-Maschine-Systeme*, 10.-12.10.2013, Berlin.
- Pohlmeyer, A.E., Hecht, M. & Blessing, L. (2009). User Experience Lifecycle Model ContinUE [Continuous User Experience]. In A. Lichtenstein, C. Stöbel & C. Clemens (Hrsg.), *Der Mensch im Mittelpunkt technischer Systeme. Fortschritt-Berichte VDI Reihe 22 Nr. 29* (S. 314-317). Düsseldorf: VDI-Verlag.
- Thüring, M. & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. *International Journal of Psychology*, 42(4), 253-264.

### **Kontaktinformationen**

Dr. Michael Minge / Laura Riedel / Prof. Dr. Manfred Thüring  
Mail: [mmi@zmms.tu-berlin.de](mailto:mmi@zmms.tu-berlin.de) / [lriedel@zmms.tu-berlin.de](mailto:lriedel@zmms.tu-berlin.de) / [manfred.thuering@tu-berlin.de](mailto:manfred.thuering@tu-berlin.de)

## **Anhang**

### Modul I: Produktbewertungen

#### *Nützlichkeit*

Insgesamt halte ich das Produkt für absolut nützlich.  
Mithilfe des Produkts kann ich meine Ziele erreichen.  
Die Funktionen des Produkts sind genau richtig für meine Ziele.

#### *Benutzbarkeit*

Die Bedienung des Produkts ist verständlich.  
Das Produkt lässt sich einfach benutzen.  
Es wird schnell klar, wie man das Produkt bedienen muss.

#### *Visuelle Ästhetik*

Das Design wirkt attraktiv.  
Das Produkt ist stilvoll.  
Das Produkt ist kreativ gestaltet.

#### *Status*

Das Produkt verleiht mir ein höheres Ansehen.  
Durch das Produkt werde ich anders wahrgenommen.  
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.

#### *Bindung*

Ohne das Produkt kann ich nicht leben.  
Das Produkt ist wie ein Freund für mich.  
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.

### Modul II: Nutzeremotionen

#### *Positive Emotionen*

Durch das Produkt fühle ich mich ausgeglichen.  
Das Produkt beruhigt mich.  
Das Produkt entspannt mich.  
Das Produkt stimmt mich euphorisch.  
Das Produkt beschwingt mich.  
Durch das Produkt fühle ich mich fröhlich.

#### *Negative Emotionen*

Das Produkt macht mich müde.  
Durch das Produkt fühle ich mich erschöpft.  
Das Produkt entspannt mich.  
Durch das Produkt fühle ich mich passiv.  
Das Produkt nervt mich.  
Das Produkt verärgert mich.  
Das Produkt frustriert mich.

Modul III:

*Nutzungsintention*

Wenn ich könnte, würde ich das Produkt täglich nutzen.

Ich kann es kaum erwarten, das Produkt erneut zu verwenden.

Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.

*Produktloyalität*

Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.

Ich würde dieses Produkt gegen kein anderes eintauschen.

Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.

Modul IV:

*Globales Produkturteil*

Geben Sie bitte abschließend an, wie Sie das Produkt insgesamt bewerten:

schlecht – gut (\*)

Alle Items sind siebenstufig Likert-skaliert mit den folgenden Antwortvorgaben: lehne völlig ab / lehne ab / lehne eher ab / weder noch / stimme eher zu / stimme zu / stimme völlig zu;

(\*) beim letzten Item handelt es sich um ein semantisches Differenzial mit 21 Stufen.