

# Messung der User Experience von Services mithilfe des UEQ: Validierung des User Experience Questionnaire zur Anwendung auf Dienstleistungen

Catherine Alberola<sup>1</sup>, Henning Brau<sup>1</sup>

Corporate Strategic Marketing – User Experience, BSH Hausgeräte GmbH<sup>1</sup>

## Zusammenfassung

Die vorliegende Studie soll dem kooperierenden Unternehmen der Hausgeräteindustrie eine Einschätzung geben, ob sich der User Experience Questionnaire (UEQ) auch zur Messung der UX von Dienstleistungen sowie zur Festlegung von UX-Zielwerte im Rahmen des Entwicklungsprozesses eignet. Für die Validierung wird ein Service-Prototyp entwickelt und anschließend von Experten und potentiellen Nutzern mit dem Fragebogen und zusätzlichen Items zur Nutzerzufriedenheit bewertet. Die mittels Korrelationen und Faktorenanalyse dargestellten Ergebnisse lassen die Validität des UEQ zur Anwendung auf Services als zweifelhaft erscheinen: Weder das Modell, dem der UEQ zugrundeliegt, kann vollständig durch die Faktoren abgebildet werden noch die Korrelationswerte auf Skalenebene sind ausreichend. Es wird folglich zur Durchführung weiterer Studien sowie zur servicespezifischen Anpassung der Fragebogenitems geraten.

## 1 Einleitung

Ein häufig verwendetes quantitatives Instrument zur Messung der User Experience (UX) ist der User Experience Questionnaire (UEQ; Laugwitz et al., 2006). Die Autoren des Fragebogens versichern, dass dieser nicht einzig als Messinstrument der UX von materiellen Produkten sondern auch zur Messung der UX von Dienstleistungen verwendet werden kann (Cota et al. 2014: 498). Bislang wurden hierzu allerdings keine Studien veröffentlicht. Es erscheint allerdings möglich, dass der Nutzer manche UEQ-Items anders interpretiert, als er dies im Produktkontext tut und diese daher ihre eigentliche Bedeutung verlieren.

Die vorliegende Arbeit wurde in Kooperation mit einem Unternehmen der Hausgeräteindustrie verfasst, welches die Erzeugung einer positiven UX künftig auch in den Entwicklungsprozess von Services integrieren möchte. Im Rahmen einer Validierungsstudie sollte überprüft werden, ob bei der Bewertung von Dienstleistungen alle Skalen und Dimensionen durch die erhobenen UEQ-Werte abgebildet werden und somit die Modellstruktur, welcher der UEQ zugrunde liegt, weiterhin besteht.

## 2 Der User Experience Questionnaire (UEQ)

Der UEQ wurde 2006 von Laugwitz, Held und Schrepp entwickelt und definiert die UX als ein Konstrukt, das aus der durch den Nutzer wahrgenommenen pragmatischen (PQ) und hedonischen Qualität (HQ) sowie einer allgemeinen Bewertung des Produkts oder Services (ATT) besteht. Die beiden Qualitäten bestehen unabhängig voneinander (Hassenzahl, 2008). Insgesamt umfasst der UEQ 26 Items in Form von semantischen Differentialen, wobei diese zu sechs Skalen zusammengefasst werden (Laugwitz et al., 2006): Attraktivität (ATT), Effizienz (EFF), Steuerbarkeit oder Vorhersagbarkeit (VOR), Durchschaubarkeit (DUR), Stimulation (STI) und Originalität (ORI). Jede der erwähnten Skalen umfasst vier miteinander hoch korrelierende Items – mit Ausnahme der Skala ATT, die über sechs Items verfügt. EFF, DUR und VOR zählen zur PQ oder aufgabenbezogenen Qualität. STI und ORI zählen zur HQ oder nicht-aufgabenbezogenen Qualität (Laugwitz et al., 2006).

Sowohl Hassenzahl et al. (2008) als auch Bevan (2008) erwähnen den Zusammenhang zwischen einer hohen Zufriedenheit und einer positiven UX. Folglich kann die Validität des UEQ zur Messung von Services über die Berechnung der Zusammenhänge zwischen den Dimensionen/Skalen des Fragebogens und den Aspekten der Nutzerzufriedenheit geprüft werden. Derartige Aspekte sind beispielsweise die globale Zufriedenheit, die Absicht den Service selbst zu nutzen oder aber die gezielte Weiterempfehlung der Servicedienstleistung.

Das Ziel dieser Arbeit ist der Nachweis, dass der Fragebogen als valides Messinstrument der UX für Servicedienstleistungen eingesetzt werden kann. Die praktische Relevanz erwächst aus der künftigen Messbarkeit der UX von Services beziehungsweise aus der Identifizierung von möglichen Verbesserungsaspekten.

## 3 Methode

Die Validität des UEQ zur Anwendung auf Servicedienstleistungen wurde mithilfe eines Messenger Service Prototyps für Smartphones im Jahr 2016 überprüft. Das getestete Nutzungsszenario hat eine defekte Waschmaschine zum Inhalt, deren Schäden durch den Versand von Instruktionen eines Technikers über den Messenger Service behoben werden sollten. Testpersonen bewerteten den Service mithilfe des UEQ und beantworteten zusätzlich Fragen zum Gesamteindruck des Prototyps und zur Zufriedenheit (Nutzerzufriedenheit, Weiterempfehlung und Verwendung im Alltag) mit der Problemlösung.

Es wurden insgesamt zwei Versionen des Prototyps getestet, welche sich hinsichtlich ihres Interaktions-Grades unterschieden (rezeptiv – interaktiv). Bei der rezeptiven Version des Prototyps gehen die Teilnehmer einen vorgegebenen Chat-Verlauf durch. Die interaktivere Version des Prototyps soll durch das Einfügen zusätzlicher Screens mit Antwortoptionen das Nutzererlebnis realistischer gestalten. Um weiterhin die Validität des UEQ zur Anwendung auf Servicedienstleistungen zu überprüfen, wurde bei der interaktiven Prototyp-Version mit einer Kontrollgruppe gearbeitet (Hussy et al., 2010).

### 3.1 Stichprobe

Die Stichprobe umfasst insgesamt 85 Fälle, welche in Experten und potentielle Nutzer der Servicedienstleistung unterteilt wurden. Die erwähnte rezeptive Prototyp-Version wurde ausschließlich von 31 Experten mit einem beruflichen Service-Hintergrund getestet (17 ♀, 14 ♂). Die interaktivere Version wurde von potentiellen Nutzern der Servicedienstleistung bewertet. Diese Testpersonen wurden wiederum in eine Experimentalgruppe (n= 28; 11 ♀, 17 ♂) und eine Kontrollgruppe (n= 26; 10 ♀, 16 ♂) unterteilt.

### 3.2 Analyseverfahren

Bei der zunächst durchgeführten Reliabilitätsanalyse wurde ein Zielwert von  $\alpha \geq .70$  definiert (George & Mallery, 2003). Die Experimentalgruppe erzielt im Vergleich zur Expertengruppe sowohl auf Skalen- als auch auf Dimensionsebene höhere  $\alpha$ -Werte. In der Gesamtauswertung sind die Werte der Dimensionen ATT und PQ höher als die der Dimension HQ. Des Weiteren erzielen die Skalen DUR und ORI die schwächsten Werte und liegen nicht über einem Wert von .70. Schwache  $\alpha$ -Werte weisen darauf hin, dass die Items der jeweiligen Skalen das zugrundeliegende Konstrukt nicht umfassend messen (Schrepp, 2015). Der Fragebogen wird daher nur als teilweise reliabel angesehen.

Zudem wird eine explorative Faktorenanalyse gerechnet, um zu überprüfen, ob die Skalen- und Dimensionsstruktur stets vorhanden sind und die Items der einzelnen Skalen diese möglichst eindeutig abbilden (Bühl, 2014). Es erfolgt eine Orientierung an der Konstruktion des UEQ, wobei die Skala ATT separat von den restlichen Items in einer zweiten Analyse untersucht wird. Bei der ersten Analyse wird die Varimax-Rotation ausgewählt.

Zur Überprüfung der Kriteriumsvalidität wurden die Dimensionen des UEQ mit Zufriedenheitsitems (Nutzerzufriedenheit, Weiterempfehlung und Verwendung im Alltag) korreliert (vgl. Punkt 2). Als Richtwert sollte  $r$  mindestens einen Wert von .50 erzielen, sodass von einem starken Zusammenhang ausgegangen werden kann (Bühner & Ziegler, 2009). Das Item „Weiterempfehlung“ wurde bei der Expertengruppe nicht abgefragt.

## 4 Ergebnisse

Bei der ersten explorativen Faktorenanalyse (Items aller Skalen außer ATT) wurden entsprechend der Annahme fünf Faktoren extrahiert (KMO-Kriterium: .82; Bartlett-Test: .00). Die erklärte Gesamtvarianz beträgt 66,3%. Auf den ersten Faktor laden alle Items der Skala EFF sowie jeweils zwei Items der Dimensionen VER („behindernd – unterstützend“, „erwartungskonform – nicht erwartungskonform“) und DUR („übersichtlich – verwirrend“, „kompliziert – einfach“). Es wäre daher denkbar, dass dieser Faktor entweder die Skala „Effizienz“ repräsentiert oder die Dimension „Pragmatische Qualität“. Auf den zweiten Faktor laden die vier Items der Skala STI, jedoch ist die Ladung des Items „wertvoll – minderwertig“ nicht eindeutig. Auf den Faktor 3 laden die Items „kreativ – phantasielos“, „originell – konventionell“ und „herkömmlich – neuartig“ der Skala ORI, wobei letzteres

keine eindeutige Ladung hat. Zudem lädt auf diesen Faktor das Item „leicht zu lernen - schwer zu lernen“ der Skala DUR. Auf den vierten Faktor laden die restlichen Items der Skala VER und auf den fünften Faktor das verbleibende Item der Skala ORI. Das Item „unverständlich – verständlich“ (DUR) kann keinem Faktor eindeutig zugeordnet werden. Bei der zweiten explorativen Faktorenanalyse (ATT-Items) wurde ein Faktor extrahiert (KMO-Kriterium= .89; Bartlett-Test= .00). Die erklärte Gesamtvarianz beträgt 67,9%.

Bei der Experimental- und Kontrollgruppe korreliert die Dimension ATT jeweils am stärksten mit den Zufriedenheits-Items. Die PQ der Expertengruppe hängt noch stärker mit der Zufriedenheit zusammen als die ATT. Die r-Werte der Experimentalgruppe sind insgesamt etwas höher, wobei der Zusammenhang zwischen HQ und den Zufriedenheits-Items stärker positiv ist als zwischen der PQ und diesen. Bei der Kontrollgruppe korreliert die Dimension ATT ebenfalls am stärksten mit den Zufriedenheits-Items. Insgesamt liegen bei der Korrelationsmatrize der Kontrollgruppe kaum signifikante Skalen-Werte vor. Bei allen Gruppen erzielt die Skala ORI nicht signifikante oder nur schwache Korrelationen. Es besteht bei den drei Gruppen ein sehr starker Zusammenhang zwischen der Dimension ATT und der PQ oder HQ. Außer bei der Kontrollgruppe korrelieren die Dimensionen PQ und HQ sehr stark signifikant miteinander (vgl. Tabellen 1-3).

	<b>Gesamteindruck</b>	<b>Zufriedenheit mit Problemlösung</b>	<b>Verwendung im täglichen Leben</b>
ATT-MW	.55**	.54**	.32
DUR_MW	.50**	.49**	.19
ORI_MW	.36*	.39*	.29
STI_MW	.54**	.33	.04
VER_MW	.45*	.58**	.06
EFF_MW	.64**	.55**	.31
PQ_MW	.61**	.60**	.22
HQ_MW	.53**	.41*	.20

\*\* Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Tabelle 1: Korrelationswerte der Skalen und Dimensionen der Expertengruppe (n=31) mit den Zufriedenheits-Items.

	<b>Gesamteindruck</b>	<b>Zufriedenheit mit Problemlösung</b>	<b>Verwendung im täglichen Leben</b>	<b>Weiterempfehlung</b>
ATT-MW	.80**	.64**	.80**	.78**
DUR_MW	.50**	.40*	.49**	.56**
ORI_MW	.54**	.43*	.49**	.53**
STI_MW	.67**	.43*	.60**	.74**
VER_MW	.38*	.28	.56**	.48**
EFF_MW	.45*	.54**	.59**	.53**
PQ_MW	.50**	.47*	.63**	.59**
HQ_MW	.68**	.48**	.62**	.72**

\*\* Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Tabelle 2: Korrelationswerte der Skalen und Dimensionen der Experimentalgruppe (n=28) mit den Zufriedenheits-Items.

	Gesamteindruck	Zufriedenheit mit Problemlösung	Verwendung im täglichen Leben	Weiterempfehlung
ATT_MW	.27	.56**	.65**	.74**
DUR_MW	.07	.14	.21	.26
ORI_MW	-.19	-.33	.17	-.01
STI_MW	.60**	.22	.61**	.55**
VER_MW	.11	.36	.44*	.47*
EFF_MW	.23	.32	.38	.54**
PQ_MW	.17	.33	.41*	.52**
HQ_MW	.30	-.04	.52**	.37

\*\* . Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

\* . Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Tabelle 3: Korrelationswerte der Skalen und Dimensionen der Kontrollgruppe (n=26) mit den Zufriedenheits-Items.

## 5 Diskussion

Die Ergebnisse lassen eine Validität des UEQ zur Anwendung auf Services als zweifelhaft erscheinen: Zwei der sechs Skalen des UEQ erwiesen sich als nicht reliabel (ORI, DUR). Eine mögliche Erklärung hierfür ist, dass die Items dieser Skalen von den Nutzern – bedingt durch den Service-Kontext – fehlinterpretiert wurden oder für die Bewertung von Servicedienstleistungen ungeeignet sind (Rauschenberger et al., 2013). Des Weiteren erlaubt die durchgeführte explorative Faktorenanalyse nur teilweise eine Abbildung der Modellstruktur, die dem UEQ zugrunde liegt. Auf Dimensionsebene des UEQ wurden zwar befriedigende Korrelationswerte festgestellt, nicht aber auf Skalenebene. Es erscheint insbesondere ratsam, die Skalen ORI und DUR nicht für die Bewertung der UX von Dienstleistungen zu verwenden, um so die Reliabilität und Validität zu erhöhen. Ferner sollten neue Items hinzugefügt werden, welche spezifisch auf die Erfassung der UX von Services zugeschnitten sind. Es gilt anzumerken, dass der untersuchte Service nur eine sehr spezifische Art von Dienstleistungen darstellt und lediglich als Online-Prototyp untersucht wurde. Zudem konnten die Teilnehmer mit dem Prototyp die Konversation nicht frei lenken, wobei auch die Interaktions-Grade jeweils variieren. Dies könnte ihnen das Gefühl eines künstlich geschaffenen Erlebnisses gegeben haben. Folgeuntersuchungen sind daher notwendig, um die Ergebnisse der vorliegenden Studie zu untermauern und auf andere Services auszuweiten.

## 6 Fazit

In der vorliegenden Studie wurde der UEQ erstmals zur Anwendung auf Servicedienstleistungen überprüft. Wie aufgezeigt wurde, eignet sich der ursprüngliche Fragebogen nur begrenzt zur Messung der UX von Services und widerlegt somit teilweise die anfänglich beschriebene Aussage der UEQ-Autoren. Dies fußt in der Annahme, dass bei Servicedienstleistungen andere Aspekte die UX beeinflussen als bei materiellen Produkte.

Durch diese Validierung des UEQ kann dem kooperierenden Unternehmen aufgezeigt werden, in welchem Umfang der UEQ zur Messung der UX von Servicedienstleistungen einsetzbar ist.

## Literaturverzeichnis

- Bevan, N. (2008). Classifying and selecting UX and usability measures. <http://www.nigelbevan.com/papers/Classifying%20and%20selecting%20UX%20and%20usability%20measures.pdf> (abgerufen am 10.06.2016).
- Bühl, A. (2014): SPSS 22: Einführung in die moderne Datenanalyse. Hallbergmoos: Pearson.
- Bühner, M. & Ziegler, M. (2009). Statistik für Psychologen und Sozialwissenschaftler. München: Pearson Studium.
- Cota, M. P.; Thomaschewski, J.; Schrepp, M. & Gonçalves, R. (2014): Efficient Measurement of the User Experience. A Portuguese Version. In: Cota, M. P.; Barroso, J.; Ferreira, S. B. L.; Fonseca, B.; Mikropoulos, T. & Paredes, H. (Hrsg.): *Procedia Computer Science*. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 27. (2014), S. 491-498.
- George, D. & Mallery, P. (2003). SPSS for Windows step by step. A simple guide and reference. 11.0 Update. 4. Aufl. Boston, MA: Allyn & Bacon.
- Hassenzahl, M. (2008). User Experience (UX). Towards an experiential perspective on product quality. In Brangier, E.; Michel, G.; Bastien, J. M. C. & Carbonell, N. (Hrsg.): *Proceedings of the 20th international conference on Association Francophone d'Interaction Homme-Machine*. Metz: Vieweg & Teubner. S. 11-15.
- Hassenzahl, M.; Burmester, M. & Koller, F. (2008). Der User Experience (UX) auf der Spur. Zum Einsatz von [www.attrakdiff.de](http://www.attrakdiff.de). In Brau, H.; Diefenbach, S.; Hassenzahl, M.; Koller, F., Peissner, M. & Röse, K. (Hrsg.): *Usability Professionals 2008*. Stuttgart: German UPA e.V. S. 78-82.
- Hussy, W.; Schreier, M. & Echterhoff, G. (2010): *Forschungsmethoden in Psychologie und Sozialwissenschaften*. Berlin: Springer.
- Laugwitz, B.; Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In Heinecke, A. M. & Paul, H. (Hrsg.): *Mensch & Computer 2006*. München: De Gruyter. S. 125-134.
- Rauschenberger, M.; Schrepp, M.; Cota, M. P.; Olschner, S. & Thomaschewski, J. (2013). Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version. In *International Journal of Interactive Multimedia and Artificial Intelligence*, 2. (2013), Nr. 1, S. 39-45.
- Schrepp, M. (2015). *User Experience Questionnaire Handbook*. All you need to know to apply the UEQ successfully in your projects. <http://www.ueq-online.org/> (abgerufen am 30.05.2016).

## Autoren



### **Alberola, Catherine**

Catherine Alberola studierte Wirtschaftspsychologie an der International School of Management in München. Sie hat diese Studie im Rahmen ihrer Masterarbeit zum Thema "Erweiterung des User Experience Questionnaire zur Anwendung auf Produkte und Services in der Hausgeräteindustrie" in Kooperation mit der UX-Abteilung eines Hausgeräteunternehmens durchgeführt. In diesem Zusammenhang konnte sie erste praktische Erfahrungen im Bereich der UX sammeln.



### **Brau, Henning**

Henning Brau studierte Psychologie an der TU Berlin. Danach folgte freiberufliche Tätigkeit als Marktforscher und Usability-Consultant. Er war ab Mai 2003 zunächst in der Forschung, später im zentralen IT Management der Daimler AG tätig. Dort war er bis 2010 verantwortlich für das Themenfeld User-Centered Technologies. Ab 2010 leitete er als 'Director of User Experience Design' den Münchener Standort der User Interface Design GmbH („UID“). Seit 2015 ist er als Manager Corporate UX verantwortlich für die Entwicklung des UX Reifegrads aller Geschäftsbereiche der BSH Hausgeräte GmbH in München sowie für die Einführung von Design Thinking. Von 2007 bis 2014 war er im Vorstand der German UPA tätig.