

User Experience mit Fragebögen messen

Durchführung und Auswertung am Beispiel des UEQ



Maria Rauschenberger

MSP Medien Systempartner GmbH & Co. KG
Peterstraße 28–34
26121 Oldenburg
Maria.Rauschenberger@medien-systempartner.de

Jörg Thomaschewski

Hochschule Emden/Leer
Constantiaplatz 4
26723 Emden
jt@imut.de

Martin Schrepp

SAP AG – User Experience
Dietmar-Hopp-Allee 16
69190 Walldorf
Martin.schrepp@sap.com

Abstract

Über Fragebögen kann die subjektive User Experience zu einem Produkt effizient gemessen werden. Jedoch gibt es beim Einsatz und insbesondere bei der anschließenden Auswertung einige „Fallstricke“, auf die dieser Artikel hinweist. Hierdurch können in konkreten Projekten Fehleinschätzungen bzgl. der gemessenen User Experience vermieden werden. Am Beispiel des User Experience Questionnaire (UEQ) und dem UEQ-Excel-Tool zur Auswertung wird auf die Besonderheiten beim Einsatz eines Fragebogens eingegangen. Die Notwendigkeit dieser zusammenfassenden Darstellung ergibt sich aus der Tatsache, dass gerade erprobte Fragebögen oftmals „zu unbedacht“ eingesetzt werden, wie eine Reihe von Anfragen zum Einsatz des UEQ über die Webseite www.ueq-online.org zeigte.

Keywords:

/// User Experience
/// Evaluation
/// Fragebogen
/// UEQ

1.

Einleitung

Die User Experience (dt. Benutzererlebnis) ist laut der DIN EN ISO 9241–210 definiert als die „Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produktes, eines Systems oder einer Dienstleistung resultieren“ (ISO 9241–210:2010). Dabei handelt es sich um Empfindungen und jegliche Reaktion eines Benutzers vor, während und nach der Benutzung eines Produktes. Dies betrifft auch das Markenbild, die persönlichen Fähigkeiten des Benutzers oder die Erfahrung (ISO 9241–210:2010).

Es gibt inzwischen zahlreiche Methoden für die Messung der User Experience, z.B. durch die Erhebung von Gesichts-Emotionen während der Nutzung, mit Hilfe von Befragungstechniken (vgl. Kim et al. 2012, Mahlke 2006, Burmester et al. 2010) oder mit Hilfe von Fragebögen (Hassenzahl et al. 2008, Laugwitz et al. 2006). Die Verwendung von Fragebögen bietet den Vorteil, dass umfangreiche Benutzergruppen schnell befragt werden können.

Eine Längsschnittuntersuchung ist dann bedeutend, wenn die zeitliche Veränderung des Benutzererlebnisses erfasst werden soll (vgl. Wilamowitz-Moellendorff et al. 2007) und sich in agilen Entwicklungsprozessen auch das interaktive System kontinuierlich weiterentwickelt. Bei Längsschnittstudien ist insbesondere darauf zu achten, dass die Methode zur Messung der User Experience möglichst identisch zu den verschiedenen Zeitpunkten eingesetzt werden kann. Hierfür sind Fragebögen wegen des geringen Aufwands und der Standardisierung der Datenerhebung besonders gut geeignet.

Insgesamt hat der Einsatz von Fragebögen zur Ermittlung der User Experience folgende Vorteile

- einfach einsetzbar (online als auch offline)
- umfangreiche Benutzergruppen können befragt werden
- Einsatz in Längsschnittuntersuchungen möglich
- ermittelt das Benutzererlebnis nach der Benutzung

Im deutschsprachigen Raum werden zwei Fragebögen zur Messung der User Experience oft eingesetzt: Der „AttrakDiff2“ von Hassenzahl et al. (2008) und der „User Experience Questionnaire: UEQ“ von Laugwitz et al. (2006). Beide Fragebögen verwenden semantische Differentiale um die Qualitäten unterschiedliche Testobjekte ermitteln zu können.

2.

User Experience Questionnaire

Das Ziel des UEQ ist die effiziente Messung der User Experience eines Produktes. Die User Experience wird dabei als subjektiv empfundener Eindruck verstanden, den der Benutzer in Bezug auf das Produkt entwickelt hat.

Die Items und Dimensionen des UEQ sind über eine datenanalytische Vorgehensweise konstruiert. Im ersten Schritt ist eine Liste mit 229 potentiellen Items in mehreren Brainstorming-Sitzungen mit Usability-Experten erstellt worden. Diese Liste ist in einem zweiten Schritt von den Experten auf eine Rohversion des Fragebogens mit 80

Items reduziert. Diese Rohversion wurde dann in mehreren Studien zu interaktiven Produkten (z.B. Statistik-Software, Online Collaboration Software, betriebswirtschaftliche Produkte, Mobile Phone) von insgesamt 153 Teilnehmern ausgefüllt. Die Dimensionen des UEQ und die Items pro Dimension sind aus diesem Datensatz über eine Faktorenanalyse extrahiert.

Dieser Prozess resultierte in folgende 6 Dimensionen (jeweils mit den Items):

- **Attraktivität** (unerfreulich / erfreulich, gut / schlecht, abstoßend / anziehend, unangenehm / angenehm, attraktiv / unattraktiv, sympathisch / unsympathisch)
- **Effizienz** (schnell / langsam, ineffizient / effizient, unpragmatisch / pragmatisch, aufgeräumt / überladen)
- **Durchschaubarkeit** (unverständlich / verständlich, leicht zu lernen / schwer zu lernen, kompliziert/einfach, übersichtlich / verwirrend)
- **Steuerbarkeit** (unberechenbar/voraus-sagbar, behindernd / unterstützend, sicher / unsicher, erwartungskonform / nicht erwartungskonform)
- **Stimulation** (wertvoll / minderwertig, langweilig / spannend, uninteressant / interessant, aktivierend / einschläfernd)
- **Originalität** (kreativ / phantasielos, originell / konventionell, herkömmlich / neuartig, konservativ / innovativ)

Der UEQ besteht also aus 26 bipolaren Items, die sich auf sechs Dimensionen verteilen. Die Items sind als siebenstufige Lickert-Skalen realisiert, z.B.:

Attraktiv ○ ○ ○ ○ ○ ○ ○ **Unattraktiv**

Details zur Konstruktion und zur Validierung des Fragebogens sind in Laugwitz, Schrepp & Held (2006) und Laugwitz, Held & Schrepp (2008) beschrieben.

Zur effizienten Auswertung steht ein UEQ-Excel-Tool zur Verfügung (siehe Abbildung 1). Es genügt hier pro Teilnehmer die Bewertungen der 26 Items einzugeben. Die Dimensionsmittelwerte, deren Konfidenzintervalle und weitere statistische Kennwerte werden dann automatisch berechnet.

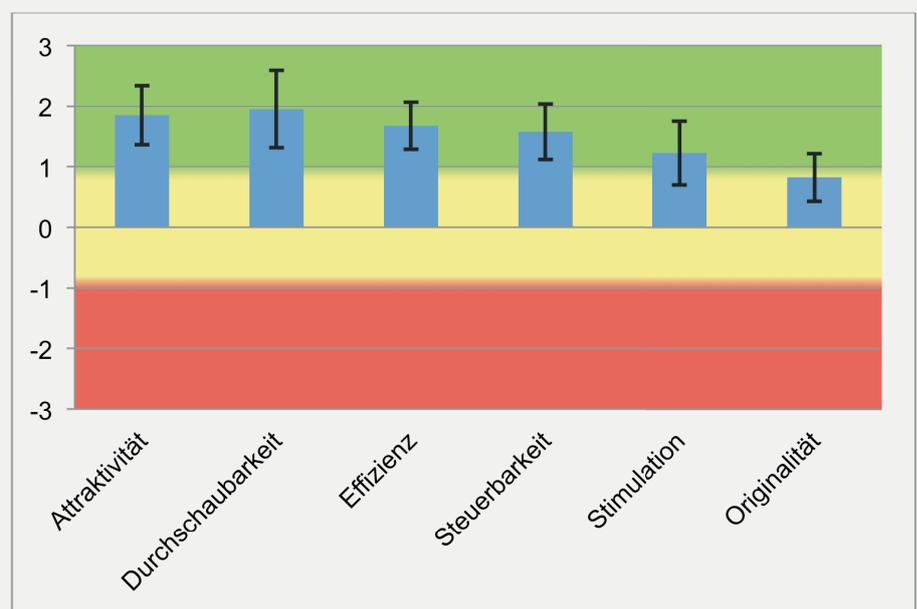


Abb. 1. Ergebnisse für ein hypothetisches Produkt (aus dem UEQ-Excel-Tool), d.h. Dimensionsmittelwerte und deren Konfidenzintervalle.

Der UEQ liefert als Ergebnis einer Befragung also die Werte der sechs Dimensionen (jeweils Mittelwert der Items pro Dimension), wie in Abbildung 1 dargestellt. **[Abb. 1]**

Pro Dimension wurden in der Konstruktion vier Items gewählt, um einerseits die Messgenauigkeit (Reliabilität), d.h. die Robustheit gegenüber zufälligen Antwortfehlern, zu erhöhen und andererseits den Fragebogen nicht unnötig lang werden zu lassen. Die Items einer Dimension sollen identische Qualitätsaspekte messen, d.h. hoch korrelieren.

Die Konsistenz einer Dimension kann durch Berechnung des Alpha-Coeffizienten $\alpha = n \cdot r / 1 + (n - 1) \cdot r$, wobei r die durchschnittliche Korrelation der Dimensionen und n die Anzahl der Dimensionen beschreibt) auch numerisch ausgedrückt werden. Es gibt keine klar definierten Regeln, wie groß Alpha sein sollte. Einige verbreitete Faustregeln fassen Werte $\alpha > 0,6$ oder $> 0,7$ als Indikator für eine zufriedenstellende Konsistenz der

Dimension auf. Ist der Alpha-Wert einer Dimension zu klein, kann das darauf hindeuten, dass einige der Items in der konkreten Untersuchung nicht wie vorgesehen interpretiert wurden. Die Alpha-Werte werden vom UEQ-Excel-Tool je Dimension berechnet, so dass eine einfache Überprüfung dieser Werte erfolgen kann.

Semantische Differentiale wie der UEQ können nicht einfach durch eine Übersetzung der Gegensatzpaare in andere Sprachen übertragen werden. Bei der Erstellung einer Sprachversion muss immer sichergestellt werden, dass durch die Übersetzung keine abweichenden Bedeutungen entstehen, die das Verhalten des Fragebogens beeinflussen. D.h. es ist auch notwendig, Daten mit der übersetzten Version zu erheben, um sicherzustellen, dass die Dimensionen in beiden Sprachen ähnliche Qualitäten messen.

Für den UEQ stehen im Moment eine Reihe von geprüften Sprachversionen zur Verfügung, z.B. Englisch, Spanisch (s. Rauschenberger et al., 2012) und in



Kürze Portugiesisch. Zusätzlich sind Übersetzungen in Französisch und Italienisch vorhanden, bei denen aber bzgl. der Qualität der Dimensionen noch keine ausreichenden Erkenntnisse vorliegen. Die vorhandenen Sprachversionen können unter www.ueq-online.org heruntergeladen werden.

Der UEQ eignet sich aufgrund seiner einfachen Struktur für den Einsatz in verschiedenen Szenarien, z.B. nach Usability Tests, als Online-Evaluation von Web-Seiten oder auch zur kontinuierlichen Prüfung größerer Anwendungspakete (siehe Laugwitz et al. 2009).

3. Typische Probleme bei der Anwendung von Fragebögen

Wie bei jedem Instrument zur Messung bestimmter Eigenschaften, sind auch beim Einsatz von Fragebögen gewisse Fallstricke vorhanden.

Eine sehr wichtige Rolle spielt die Auswahl der richtigen Zielgruppe für die Befragung. Idealerweise sollte hier eine möglichst repräsentative Gruppe von Endanwendern ausgewählt werden. In der Praxis ist dies je nach Projekt nicht immer möglich. Steht beispielsweise nur eine Gruppe von Beta-Testern zur Verfügung kann es sinnvoll sein, deren Eindruck zur User Experience per Fragebogen zu erfassen und die Bewertung später ggf. bewusst von den wirklichen Endbenutzern zu trennen.

Je nach Art des Fragebogens ist zu beachten, dass die adressierten Nutzer die Fragen richtig verstehen. Ein Beispiel für die Probleme dieser Art sind Teilnehmer, die keine Muttersprachler sind, d.h. evtl. Probleme haben sich Items wie **usual –leading edge** zu erschließen. Hier wurde bewusst ein Beispiel des englischen UEQ verwendet, um das Problem zu verdeutlichen.

Ein weiteres Problem kann auftreten, wenn es sich zwar um Muttersprachler handelt, aber aufgrund anderer Umstände ein eingeschränktes Sprachverständnis vorhanden ist.

Beim Versuch ein Web-Portal für jugendliche Benutzer mit dem UEQ zu evaluieren, zeigten sich z.B. massive Verständnisprobleme bei einigen Items des UEQ („Was bedeutet **konventionell?**“). Aufgrund dieser Rückmeldung wurde eine spezielle Sprachversion für jugendliche Benutzer erstellt (Hinderks et al., 2012). Die Verständlichkeit der Items für die intendierte Zielgruppe sollte deshalb vor dem Einsatz eines Fragebogens geprüft werden.

Enthält ein Fragebogen vollständig ausformulierte Sätze anstelle semantischer Differentiale, dann sind Probleme mit dem Sprachverständnis weniger wahrscheinlich. Bei semantischen Differentialen wie dem UEQ muss oft ein höheres Maß an Sprachverständnis voraussetzen werden, da der Benutzer hier nur zwei Gegensatzpaare zur Verfügung hat, um sich die Bedeutung der Fragestellung zu erschließen.

Einzelne Items könnten je nach Testobjekt unterschiedlich interpretiert werden oder es könnte sich die Wertigkeit im Extremfall sogar umkehren. Ein Beispiel hierzu ist das Item-Paar „sicher – unsicher“ zur Dimension „Steuerbarkeit“. Im Kontext von sozialen Netzwerken kann dieses Item-Paar anders interpretiert werden und der technischen Sicherheit von Daten zugeordnet werden. Somit ist je nach Testobjekt das Item unterschiedlich zu bewerten. Abhilfe kann ein Beta-Test schaffen, bei dem die Probanden im Nachgang zu den Items und den zugehörigen Assoziationen befragt werden.

In bestimmten Szenarien stellt sich die Frage, zu welchem Zeitpunkt der Fragebogen eingesetzt werden soll. Gemäß Donald A. Norman „Memory is more important than actuality“ (Norman 2009), ist es relevant das Benutzererlebnis nach der Benutzung zu ermitteln. Ist es Ziel des UEQ den subjektiven Eindruck des Benutzers zu messen, wird der UEQ am Besten im Rahmen eines Usability Tests eingesetzt. Diskussionen mit dem Benutzer über z.B. Interaktionsprobleme beeinflussen die Meinung des Benutzers und verfälschen damit das Fragebogen-Ergebnis und sollten erst

nach dem Ausfüllen des Fragebogens z.B. mit einem Interview ermittelt werden.

Beim Einsatz mit Endbenutzern einer betriebswirtschaftlichen Software sollte man versuchen, die Befragung zur User Experience erst durchzuführen, nachdem die Benutzer eine gewisse Zeitspanne mit dem Produkt gearbeitet haben. Ansonsten besteht die Gefahr einer Verfälschung der Ergebnisse durch mangelnde Erfahrung der Benutzer oder Probleme durch die Umstellung von einem alten Produkt auf eine neue Lösung (Rauschenberger et al., 2011).

Zu vermeiden ist eine „Belohnungsstruktur“ für das Ausfüllen des Fragebogens, z.B. eine Teilnahme an einem Gewinnspiel mit attraktiven Gewinnen. Es zeigt sich, dass hierdurch zwar schnell eine hohe Teilnehmerzahl erreicht werden kann, aber die Motivation der Teilnahme bei vielen Teilnehmern zu sehr auf das Gewinnspiel gerichtet ist. Somit werden die Ergebnisse gegebenenfalls weniger valide ausfallen.

Auch bei einem Abhängigkeitsverhältnis zu den Teilnehmern kann die Validität der Ergebnisse leiden, beispielsweise bei einer Verschickung des Fragebogens durch die Geschäftsleitung oder bei Fragebögen von Professoren an Studierende.

4. Typische Fehler bei der Interpretation der Ergebnisse

Nach der erfolgreichen Erhebung von Daten müssen diese entsprechend analysiert und interpretiert werden. Bietet der Fragebogen ein Tool zur Berechnung an, müssen nur die Ergebnisse interpretiert werden. Der UEQ bietet genau diese Möglichkeit und es müssen zunächst nur die erhobenen Daten in ein UEQ-Excel-Tool eingetragen werden. Leider besteht durch die einfache Auswertung auch die Gefahr, dass Ergebnisse unreflektiert verwendet werden.

Vor der Auswertung muss sichergestellt sein, so dass nur „ernsthaft ausgefüllte“ Fragebögen ausgewertet werden. Im UEQ

wurden die bipolaren Items bezüglich der Polung und der Reihenfolge randomisiert, sodass die Zugehörigkeit zu den Dimensionen von den Probanden nicht erkannt werden kann. Wenn im gesamten Fragebogen alle Werte „ganz links“ (entspricht dem Wert „1“) oder „ganz rechts“ (entspricht dem Wert „7“) vom Probanden angekreuzt wurden, sollte der Fragebogen von der weiteren Auswertung als „nicht ernsthaft ausgefüllt“ ausgeschlossen werden. Dagegen sollten Items ohne Antwort von einem Probanden als eine Meinung gezählt werden (Höpflinger, 2010).

Zu den typischen Fehlern gehört das „überinterpretieren“ der Ergebnisse, wenn beispielsweise Mittelwerte der Dimensionen verwendet werden, obwohl aufgrund von stark unterschiedlichen Einschätzungen der Teilnehmer die Konfidenzintervalle sehr groß sind. Große Konfidenzintervalle lassen entweder auf eine zu geringe Stichprobengröße schließen oder auf sehr unterschiedliche Antworten der Teilnehmer. Hierfür müssen die einzelnen Fragebögen gegenübergestellt und ggf. eine Abgleichung mit der Zielgruppe vorgenommen werden.

Beim Vergleich zweier Dimensionen sowie beim Vergleich zweier Produkte ist auf die Signifikanz von Unterschieden in den Mittelwerten zu achten. Das UEQ-Excel-Tool gibt hierzu Fehlerbalken an, welche eine mögliche Toleranz der Ergebnisse darstellt und Konfidenzintervalle von 95 Prozent angeben.

Vor einer Betrachtung der Mittelwerte zu einer Dimension sind die Ergebnisse der einzelnen Items kritisch zu prüfen. Wenn einzelne Items einer Dimension stark voneinander abweichen, dann könnte dies auf eine Fehlinterpretation einzelner Items durch die Probanden hinweisen. In Abbildung 2 ist ein Ausschnitt aus Verteilung der Mittelwerte der einzelnen Items dargestellt. Die Farben zeigen die Zugehörigkeit der Items zu einer Dimension an, beispielsweise steht die Farbe Gelb für die Dimension „Stimulation“. In der Abbildung kann erkannt werden, dass sich der Mittelwert für das Item-Paar „uninteressant“ – „interessant“

stark von den anderen Mittelwerten derselben Farben (und damit derselben Dimensionszugehörigkeit) unterscheidet. Somit sollten die Ergebnisse dieses Items und der zugehörigen Dimension kritisch hinterfragt werden. [Abb. 2]

Beim UEQ ist zu berücksichtigen, dass die Items für die Berechnung des Mittelwerts der Dimensionen benötigt werden und keine Items aus dem Fragebogen entfernt werden dürfen. Je mehr Items eine Skale hat, desto höher ist ihre Messgenauigkeit, d.h. Reliabilität. Die Reliabilität der Dimension ist beim Entfernen von Items nicht mehr gegeben und zusätzlich ist eine Vergleichbarkeit mit anderen Untersuchungen nicht mehr möglich.

Oft besteht der Wunsch aus dem Management eine Kennzahl für die Softwareüberprüfung bereit zu stellen. Aufgrund der

Konstruktion des Fragebogens ist dies nicht möglich. Ein Verrechnen der einzelnen Dimensionsmittelwerte zu einer einzigen KPI wäre nur sinnvoll, wenn man davon ausgeht, dass alle Dimensionen den gleichen Einfluss auf das Gesamturteil der befragten Personen haben. Diese Annahme kann aber in konkreten Projekten in der Regel nicht treffen. Erfahrungen aus bisherigen Anwendungen zum UEQ zeigen auch, dass der Einfluss der Dimensionen auf das Gesamturteil abhängig von der Art des evaluierten Produkts durchaus stark variieren kann.

5. Zusammenfassung

Fragebögen bieten sich zur Messung der User Experience in vielfältigen Situationen an. Empfohlen wird ein Pre-Test des Fragebogens mit der Zielgruppe und dem Testobjekt bei dem die Probanden

Bewertung pro Item

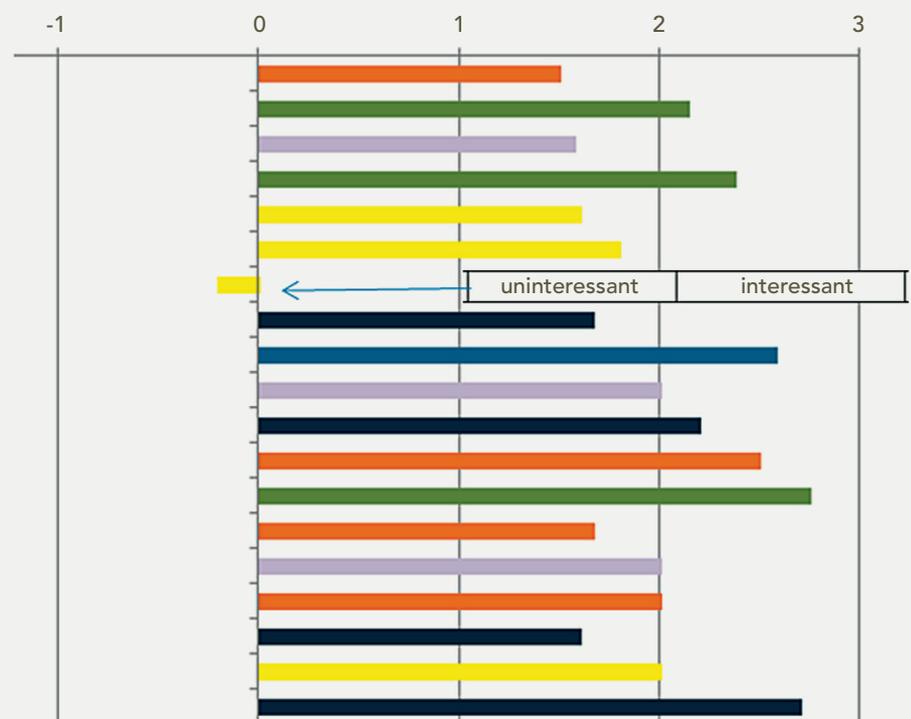


Abb. 2. Beispiel: Mittelwert für das Item-Paar „uninteressant“ – „interessant“ unterscheidet sich stark von den anderen Mittelwerten.



zur Interpretation der Items im Anschluss befragt werden. Dadurch kann eine korrekte Interpretation der Items durch die Zielgruppe und bezogen auf das Testobjekt erhöht werden.

Vor der Interpretation der Ergebnisse der Umfrage mittels Fragebogen ist eine Datenanalyse notwendig. Zur Berechnung liefert beim UEQ das UEQ-Excel-Tool alle hierfür notwendigen Algorithmen und Grundlagen (Cronbach-Alpha-Werte der einzelnen Dimensionen, Konfidenzintervalle je Item und Dimension, Varianz und Standardabweichung). Anhand der Bewertung der einzelnen Items können oftmals Interpretationsschwierigkeiten durch die Probanden vermutet oder erkannt werden.

Sofern die Datenanalyse keine negativen Auffälligkeiten ergeben hat, können sowohl die Mittelwerte der einzelnen Dimensionen als auch der einzelnen Items eine gute Aussage über den Handlungsbedarf am Produkt oder über die Verbesserung nach einem Relaunch liefern. Die Ergebnisse sollten allerdings immer kritisch reflektiert und Auffälligkeiten hinterfragt werden z.B. bei großen Abweichungen einzelner Items vom Dimensionsmittelwert.

Literatur

1. Burmester, M., Jäger, K., Mast, M., Peissner, M. & Sproll, S. (2010). Design verstehen – Formative Evaluation der User Experience. In: Brau, H. (Hrsg.). Usability Professionals 2010, 206–211. Stuttgart. Fraunhofer Verlag.
2. Hassenzahl, M., Burmester, M. & Koller, F. (2008). Der User Experience (UX) auf der Spur: Zum Einsatz von www.attraktdiff.de. In: Brau, H. (Hrsg.). Usability Professionals 2008, 78–82. Stuttgart: German Chapter der Usability Professionals Assoc.
3. Hinderks, A., Schrepp, M., Rauschenberger, M., Olschner, S. & Thomaschewski, J. (2012). Konstruktion eines Fragebogens für jugendliche Personen zur Messung der User Experience. In: Brau, H. (Hrsg.). Usability Professionals 2012, 78 – 83. Stuttgart. German UPA e.V.
4. Höpflinger, F. (2010). Praktische Regeln zur Formulierung von Fragen für Fragebögen. online verfügbar unter <http://arbeitsblaetter.stangl-taller.at/FORSCHUNGSMETHODEN/FrageformulierungDetail.shtml>, zuletzt geprüft am 14.06.2013.
5. ISO 9241–210:2010 (2011). Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.
6. Kim, K., Kolbe, K. & Eisele, S. (2012). Es steht Dir ins Gesicht geschrieben! In: I-Com (1), 63–67. online verfügbar unter <http://dx.doi.org/10.1524/icom.2012.0016>. zuletzt geprüft am 29.06.2013.
7. Laugwitz, B., Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: Heinecke, A.M (Hrsg.). Mensch & Computer 2006, 125–134. München. Oldenbourg Verlag S.
8. Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In: Holzinger, A. (Hrsg.). Lecture Notes in Computer Science. Berlin Heidelberg. Springer Berlin Heidelberg.
9. Laugwitz, B., Schubert, U., Ilmberger, W., Tamm, N., Held, T. & Schrepp, M. (2009). Subjektive Benutzerzufriedenheit quantitativ erfassen: Erfahrungen mit dem User Experience Questionnaire UEQ. In: Brau, H (Hrsg.). Usability Professionals 2009, 220 – 225. Stuttgart: Fraunhofer Verlag.
10. Mahlke, S. (2006). Emotionen als Aspekt des Nutzungserlebens. In: Bosenick, T. (Hrsg.). Usability Professionals 2006, 140–145. Stuttgart. German Chapter der Usability Professionals Assoc.
11. Rauschenberger, M., Hinderks, A. & Thomaschewski, T. (2011). Benutzererlebnis bei Unternehmenssoftware: Ein Praxisbericht über die Umsetzung attraktiver Unternehmenssoftware In: Brau, H. (Hrsg.). Usability Professionals 2011, 158 – 163. Stuttgart. German UPA e.V.
12. Norman, Donald A. (2009). THE WAY I SEE IT: Memory is more important than actuality. In: interactions (2), S. 24.
13. 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. (Hrsg.). Sistemas y Tecnologías de Información 2 (2013), 39–45. Madrid.
14. Wilamowitz-Moellendorff, M. von; Hassenzahl, M. & Platz, A. (2007). Veränderung in der Wahrnehmung und Bewertung interaktiver Produkte. In: Gross (Hrsg.). Mensch & Computer. 7, 49–58. München. Oldenbourg Verlag

