

Eine modulare Erweiterung des User Experience Questionnaire

Hinweise zur Anwendung in praktischen Projekten

Martin Schrepp
SAP SE
Walldorf, Germany
martin.schrepp@sap.com

Jörg Thomaschewski
Hochschule Emden/Leer
Emden/Germany
joerg.thomaschewski@hs-emden-leer.de

ABSTRACT

Wir stellen einen modularen Ansatz vor, mit dem man sich aus einem Katalog von 16 UX Aspekten einen perfekt passenden UX Fragebogen bauen kann. Grundlage sind die 6 Skalen des UEQ, die um 10 Skalen erweitert wurden. Wir beschreiben die Konstruktion und erste Evaluationsergebnisse zu diesen neuen Skalen. Es wird weiterhin beschrieben, wie UX Professionals sich aus diesem Katalog einfach einen Fragebogen zusammenstellen können, wie die Daten ausgewertet werden, was man bei diesem Ansatz in der praktischen Anwendung beachten sollte, aber auch wo die Limitationen dieses Vorgehens sind, d.h. in welchen Fällen man besser einen vorhandenen Standardfragebogen verwendet.

KEYWORDS

User Experience Questionnaire, UX Messung, UX Fragebögen, Modularer Fragebogen

1 Einleitung

Es kann in vielen Fällen sehr nützlich sein, die User Experience (UX) eines Produkts quantitativ zu messen [1]. Solche objektiven Messwerte können z.B. benutzt werden, um notwendige Aufwände zur Verbesserung der UX Qualität gegenüber dem Management zu rechtfertigen. Auch wenn man die Entwicklung der UX eines Produkts über einen längeren Zeitraum kontinuierlich erfassen oder das eigene Produkt mit Konkurrenzprodukten vergleichen will, ist eine objektive Messung sehr wichtig [2].

Standardisierte UX Fragebögen, z.B. UEQ [3], AttrakDiff2 [4], VISAWI [5], meCUE [6], SUS [7], ISOMETRICS [8] oder SUMI [9] bieten eine einfache Möglichkeit, eine solche Messung vorzunehmen. Die Fragebögen sind in der Regel frei verfügbar. Oft stehen auch Tools bereit, die dem Anwender bei der Auswertung und Interpretation der Daten helfen. Viele Fragebögen bieten auch einen Benchmark, mit dem man das

eigene Ergebnis mit den Ergebnissen anderer Produkte vergleichen kann, was die Interpretation der Ergebnisse erleichtert.

Die UX ist ein sehr vielschichtiges Konstrukt. Um eine gute UX sicherzustellen, muss ein Produkt verschiedene positive UX Aspekte bzw. Qualitäten aufweisen, z.B. effizient bedienbar sein, leicht zu erlernen sein, ein schön gestaltetes User Interface haben, etc. [10]. Natürlich kann kein einzelner Fragebogen alle UX Qualitäten abdecken, was die Vielzahl der verfügbaren UX Fragebögen erklärt [2]. Die meisten Standardfragebögen enthalten mehrere Skalen, die einzelnen UX Aspekten entsprechen. Jeder UX Fragebogen misst daher immer nur eine Teilmenge der möglichen UX Aspekte, d.h. erfasst nur einen Ausschnitt der User Experience.

Weiterhin sind nicht alle UX Aspekte für alle Produkte gleich wichtig [11, 12]. Für eine betriebswirtschaftliche Anwendung, die ein Nutzer mehrere Stunden am Tag bedient, ist offensichtlich Effizienz der Bedienung ein entscheidender Qualitätsaspekt. Ein unnötiger Klick, den man am Tag 50-mal wiederholen muss, macht dem Nutzer in der Regel wenig Freude. Für ein Spiel oder einen Self-Service, z.B. Ändern der Kontoverbindung für die Auszahlung des Lohns, spielt Effizienz eher keine Rolle. Hier sind andere UX Aspekte, z.B. Immersion für ein Spiel oder intuitive Bedienung für einen Self-Service, deutlich wichtiger.

Als UX Professional, der das zu evaluierende Produkt gut kennt, hat man natürlich eine Vorstellung davon, welche UX Aspekte wichtig sind und daher evaluiert werden sollten. Allerdings kann man nicht immer davon ausgehen, dass es auch einen UX Fragebogen gibt, der genau diese UX Aspekte in seinen Skalen abbildet. Einen eigenen Fragebogen zu entwickeln ist in solchen Fällen meist zu aufwändig. Eine Kombination vorhandener Fragebögen zu benutzen, ist oft auch problematisch, da die Fragenformate der verschiedenen Fragebögen oft sehr verschieden sind und die Zusammenführung und gemeinsame Interpretation der Ergebnisse damit schwierig wird.

In diesem Artikel stellen wir einen modularen Ansatz vor, mit dem man sich aus einem Katalog von 16 UX Aspekten einen passenden Fragebogen bauen kann. Grundlage sind die 6 Skalen des UEQ (*Attraktivität, Effizienz, Durchschaubarkeit, Steuerbarkeit, Stimulation, Originalität*). Es wurden in der Vergangenheit bereits Erweiterungsskalen für *Vertrauen* [13] sowie *Haptik* und *Akustik* [14] konstruiert. Wir beschreiben hier die Konstruktion und erste Evaluation der zusätzlichen Skalen *Ästhetik, Anpassbarkeit,*

Nützlichkeit, Intuitive Bedienung, Wertigkeit, Inhalts-Seriosität und Inhalts-Qualität. Damit stehen jetzt 16 Skalen zur Verfügung, die zusammen ein breites Spektrum verschiedener UX Aspekte abbilden.

Wir zeigen auch, wie man sich aus diesem Katalog der 16 Skalen einen Fragebogen zusammenstellen kann, wie man die Daten auswertet, was man dabei beachten sollte, aber auch wo die Limitationen dieses Vorgehens sind, d.h. in welchen Fällen man besser einen Standardfragebogen wie den UEQ verwendet.

2 Vorarbeiten und Item-Format

Der User Experience Questionnaire (UEQ) ist ein weit verbreiteter UX Fragebogen, der 6 Skalen (*Attraktivität, Effizienz, Durchschaubarkeit, Steuerbarkeit, Stimulation und Originalität*) zur Messung verschiedener UX Aspekte enthält, die für viele Produkte relevant sind [3, 15]. Die Skalen sind jeweils durch 4 Items repräsentiert (mit Ausnahme der Skala *Attraktivität*, die 6 Items enthält).

Die Items bestehen jeweils aus zwei Gegensatzpaaren (semantisches Differential) zwischen denen der Teilnehmer auf einer 7-stufigen Lickert-Skala entscheiden kann, z.B.:

schwer zu lernen o o o o o o leicht zu lernen
 langweilig o o o o o o spannend
 langsam o o o o o o schnell

Die sechs Skalen des UEQ erfassen bereits eine Reihe von UX Qualitäten, die für sehr viele Produkte eine wichtige Rolle spielen, decken aber natürlich nicht das gesamte Spektrum von UX ab.

Für den UEQ wurden in der Vergangenheit daher bereits einige Erweiterungsskalen konstruiert. Für Anwendungen, bei denen sensible und kritische Daten verwaltet werden, z.B. Online-Banking, spielt das Vertrauen des Anwenders in die Sicherheit seiner Daten eine wichtige Rolle. Eine Skala zur Messung von *Vertrauen* wurde in [13] beschrieben. Für Haushaltsgeräte spielen neben den klassischen Usability Kriterien auch das Gefühl beim Anfassen des Geräts (*Haptik*) und die Wahrnehmung der beim Betrieb entstehenden Geräusche (*Akustik*) eine wichtige Rolle. Entsprechende Skalen wurden von [14] konstruiert.

Diese vorhandenen Erweiterungsskalen wurden in den UEQ+ übernommen, so dass hier schon 9 Skalen zur Verfügung stehen. Wir beschreiben in diesem Beitrag die Konstruktion weiterer 7 Skalen.

Der UEQ+ ist ein Katalog von Skalen mit identischem Item-Format, die man für eine konkrete Untersuchung frei kombinieren kann, d.h. man kann eine beliebige Menge solcher Skalen in einer beliebigen Reihenfolge auswählen. Das macht einige kleinere Änderungen am bisherigen Format der UEQ Skalen notwendig.

Im UEQ ist die Reihenfolge der Items zufällig gewählt. Das gilt auch für die Polarisierung der Items. In der Hälfte der UEQ Items steht der positive Begriff links (*schnell / langsam*), in der anderen Hälfte rechts (*langweilig / spannend*). Einige bisher nicht publizierte Untersuchungen haben gezeigt, dass die Polarisierung der Items auf die Skalenmittelwerte keinen Einfluss hat. Ähnliche Ergebnisse liegen auch für andere Fragebögen vor [16]. Daher wurde entschieden im UEQ+ für alle Items jeweils mit dem

negativen Begriff links und dem positiven Begriff rechts zu verwenden.

UX Skalen sind eigentlich nie völlig unabhängig, da die zugrundeliegenden Qualitätsaspekte natürlich semantische Überschneidungen haben. Betrachten wir dazu ein Beispiel. Ein Item (*überladen/aufgeräumt*) hat einen Bezug auf die Effizienz, mit der der Nutzer eine Anwendung bedienen kann. In einem aufgeräumten und übersichtlichen User Interface findet man sich schneller zurecht, was auch die Geschwindigkeit der Aufgabebearbeitung positiv beeinflusst. Aber ein solches Item hat auch einen Bezug zur visuellen Ästhetik [17], da es auch das Konzept der klassischen Ästhetik [18] repräsentiert.

In einigen Skalen des UEQ+ finden sich daher ähnliche oder teilweise identische Items. Da die Skalen beliebig kombiniert werden können, ist es notwendig die Items einer Skala zu gruppieren und einen klaren Kontext zu setzen, in dem die Items interpretiert werden. Dies wird durch einen kurzen einleitenden Satz sichergestellt, der den Kontext für die gemeinsame Interpretation aller Items einer Skala setzt. Zusätzlich wird für die Berechnung einer KPI (Key Performance Indikator) noch die Wichtigkeit der Skala für das betreffende Produkt abgefragt.

Das folgende Beispiel der Skala *Effizienz* zeigt das neue Format einer UEQ+ Skala:

Für das Erreichen meiner Ziele empfinde ich das Produkt als								
langsam	o	o	o	o	o	o	o	schnell
ineffizient	o	o	o	o	o	o	o	effizient
unpragmatisch	o	o	o	o	o	o	o	pragmatisch
überladen	o	o	o	o	o	o	o	aufgeräumt
Die durch diese Begriffe beschriebene Produkteigenschaft ist für mich								
Völlig unwichtig	o	o	o	o	o	o	o	Sehr wichtig

Die Ratings zu den Wichtigkeiten werden verwendet, um aus allen Skalen eine KPI zu berechnen, d.h. einen Kennwert, der als UX Gesamtzufriedenheit interpretiert werden kann. Die Grundidee dahinter ist es, dass der pro Person gemessene Skalenwert mit den Ratings der Wichtigkeit dieses Skalenwertes gewichtet wird. Die Berechnung ist völlig identisch zur entsprechenden Berechnung für den UEQ und in [19,20] ausführlich beschrieben.

3 Konstruktion der neuen Skalen

Wir beschreiben jetzt die Konstruktion der Erweiterungsskalen für *Visuelle Ästhetik, Anpassbarkeit, Nützlichkeit, Intuitive Bedienung, Wertigkeit* und *Inhaltsqualität*.

Diese Skalen wurden basierend auf einer Liste von 16 UX Aspekten aus [11,12] gewählt. Für jede zu konstruierende Skala wurde eine Menge von semantisch passenden Items im gewünschten Format durch zwei Experten konstruiert.

In einer empirischen Studie beurteilten 192 Teilnehmer (Studenten der Hochschule Emden/Leer, die auf freiwilliger Basis teilnahmen) mehrere gängige Produkte mit den neuen Skalen. Das mittlere Alter der Teilnehmer betrug 30,42 Jahre. 119 Teilnehmer waren männlich, 73 weiblich. Die Daten wurden über eine Hauptkomponentenanalyse ausgewertet (mit dem R Paket *psych*

[21]). Es wurde zuerst untersucht, ob wie erwartet ein einzelner Faktor die Daten ausreichend beschreibt. In diesem Fall wurden die 4 Items gewählt, die auf dem Faktor die höchste Ladung zeigten.

Wir beschreiben im Folgenden die Analyse der Daten für die Skalen *Visuelle Ästhetik* und *Inhaltsqualität*. Die Analyse aller Skalen ist in [22] ausführlich beschrieben. Die folgende Abbildung 1 zeigt die Ergebnisse der Hauptkomponentenanalyse für die Skala *Visuelle Ästhetik*.

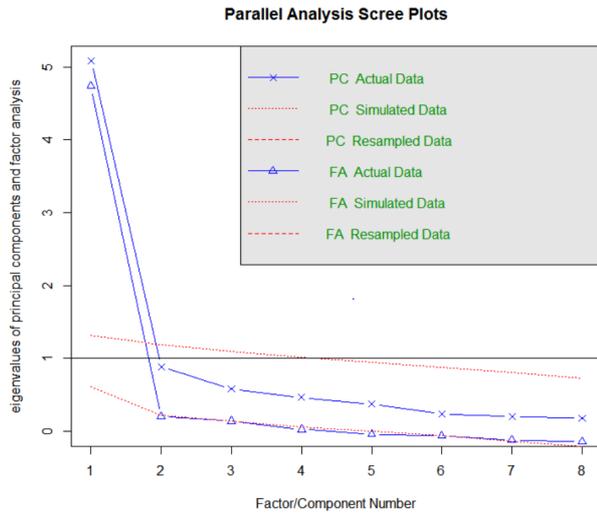


Abbildung 1: Scree Plot der Hauptkomponentenanalyse für Visuelle Ästhetik.

Die Ergebnisse zeigen, dass ein einzelner Faktor die Daten sehr gut beschreibt. Der Anteil der erklärten Varianz war 0,64 und der Fit basierend auf den Werten in der Diagonale (fit based on diagonal values) betrug 0,99 (Werte > 0,95 sind ein Indikator für eine gute Anpassung).

Die Ladung der einzelnen Items auf dem Faktor ist in Tabelle 1 dargestellt. Für die Skala *Visuelle Ästhetik* wurden anhand der Ladungen die Items *hässlich / schön*, *stillos / stilvoll*, *nicht ansprechend / ansprechend* und *unästhetisch / ästhetisch* gewählt.

Item	Ladung
<i>hässlich / schön</i>	0.89
<i>stillos / stilvoll</i>	0.86
<i>nicht ansprechend / ansprechend</i>	0.88
<i>farblich unschön / farblich schön</i>	0.79
<i>unharmonisch / harmonisch</i>	0.84
<i>unästhetisch / ästhetisch</i>	0.88
<i>nicht kunstvoll / kunstvoll</i>	0.63
<i>unüberlegt / durchdacht</i>	0.51

Tabelle 1: Ladung der Items auf dem Faktor Visuelle Ästhetik.

Für die zu konstruierenden Skalen *Anpassbarkeit*, *Nützlichkeit*, *Intuitive Bedienung*, und *Wertigkeit* ergab sich ebenfalls eine gute Anpassung mit einem Faktor. Details sind in [22] beschrieben.

Für die Items zur Inhaltsqualität ergab sich keine eindimensionale Lösung. Hier zeigte ein Modell mit zwei Faktoren eine bessere Passung. Da die beiden Faktoren auch gut zu interpretieren waren, wurden somit zwei Erweiterungsskalen konstruiert. Die folgende Abbildung 2 zeigt die Ergebnisse der Hauptkomponentenanalyse für die Items zur Inhaltsqualität.

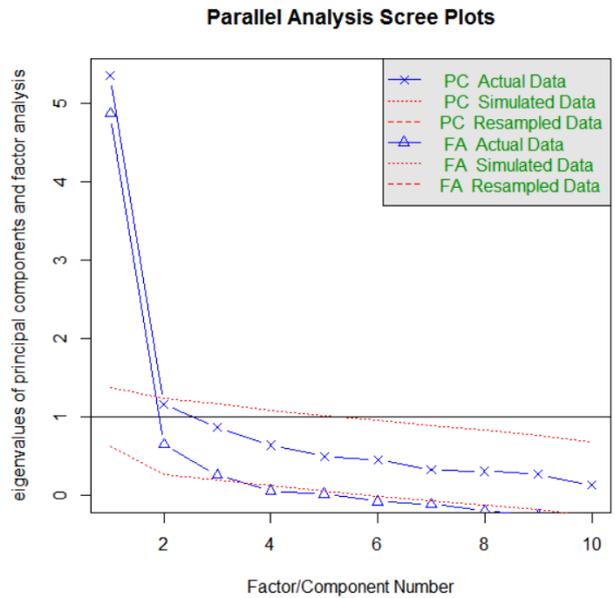


Abbildung 2: Scree Plot der Hauptkomponentenanalyse für Inhaltsqualität.

Die folgende Tabelle 2 zeigt die Ladungen der Items auf den beiden Faktoren.

Item	Faktor 1	Faktor 2
veraltet / aktuell	0.32	0.64
nicht informativ / informativ	0.55	0.56
uninteressant / interessant	0.21	0.68
schlecht aufbereitet / gut aufbereitet	0.30	0.77
unverständlich / verständlich	0.10	0.78
minderwertig / hochwertig	0.58	0.58
nutzlos / nützlich	0.68	0.33
unglaubwürdig / glaubwürdig	0.90	0.18
unseriös / seriös	0.89	0.20
ungenau / genau	0.77	0.28

Tabelle 2: Ladung der Items auf den beiden Faktoren der Inhaltsqualität.

Die Items mit hoher Ladung auf dem ersten Faktor beschreiben das Vertrauen in den Inhalt, den die Anwendung dem Nutzer bereitstellt. Die Items mit hoher Ladung auf dem zweiten Faktor stehen für die wahrgenommene Qualität des Inhalts. Entsprechend wurden die beiden Skalen mit *Inhalts-Seriosität* und *Inhalts-Qualität* benannt.

4 Übersicht der Skalen im UEQ+

Wir geben nun einen Überblick zu den vorhandenen Skalen im UEQ+ Framework. Pro Skala wird deren Bedeutung, der zusammenfassende Satz und die vier Items angegeben.

Die Skalen *Attraktivität*, *Effizienz*, *Durchschaubarkeit*, *Steuerbarkeit*, *Stimulation* und *Originalität* enthalten die Items der entsprechenden UEQ-Skalen. Hier wurde nur der Einleitungssatz ergänzt und die Polung der Items geändert. Bei der Skala *Attraktivität* (diese Skala enthält 6 Items im UEQ, hat aber mit 4 Items immer noch eine sehr hohe Reliabilität) wurden zwei Items entfernt, um auf eine einheitliche Zahl von vier Items pro Skala zu kommen. Die Skalen *Haptik* und *Akustik* sind in den Items identisch zu [14], die Items der Skala *Vertrauen* stammen aus [13]. Auch hier wurde lediglich die Polung der Items geändert und der Einleitungssatz passen ergänzt. Alle anderen Skalen wurden im Rahmen dieser Arbeit konstruiert.

4.1 Attraktivität

Gesamteindruck des Produkts. Mögen die Benutzer das Produkt oder nicht?

Einleitungssatz: Insgesamt empfinde ich das Produkt als:

Items: unerfreulich / erfreulich, schlecht / gut, unangenehm / angenehm, unsympathisch / sympathisch

4.2 Effizienz

Benutzer haben den Eindruck, dass sie ihre Aufgaben ohne unnötigen Aufwand lösen können.

Einleitungssatz: Für das Erreichen meiner Ziele empfinde ich das Produkt als:

Items: langsam / schnell, ineffizient / effizient, unpragmatisch / pragmatisch, überladen / aufgeräumt

4.3 Durchschaubarkeit

Subjektiver Eindruck des Benutzers, dass es leicht ist, sich mit dem Produkt vertraut zu machen. Es ist leicht zu lernen, wie man das Produkt benutzt.

Einleitungssatz: Die Bedienung des Produkts empfinde ich als:

Items: unverständlich / verständlich, schwer zu lernen / leicht zu lernen, kompliziert / einfach, verwirrend / übersichtlich

4.4 Steuerbarkeit

Der Benutzer hat den Eindruck, dass er die Interaktion steuert und stets die volle Kontrolle hat.

Einleitungssatz: Die Reaktion des Produkts auf meine Eingaben und Befehle empfinde ich als:

Items: unberechenbar / vorhersagbar, behindernd / unterstützend, unsicher / sicher, nicht erwartungskonform / erwartungskonform

4.5 Stimulation

Das Gefühl, dass es aufregend und motivierend ist, das Produkt einzusetzen.

Einleitungssatz: Die Beschäftigung bzw. das Arbeiten mit dem Produkt empfinde ich als:

Items: uninteressant / interessant, langweilig / spannend, minderwertig / wertvoll, einschläfernd / aktivierend

4.6 Originalität

Gefühl, dass das Produkt innovativ und kreativ ist und das Interesse des Anwenders weckt.

Einleitungssatz: Die Produktidee bzw. die Gestaltung des Produkts finde ich:

Items: phantasielos / kreativ, konventionell / originell, herkömmlich / neuartig, konservativ / innovativ

4.7 Vertrauen

Subjektiver Eindruck, dass die in das Produkt eingegebenen Daten in sicheren Händen sind und nicht dazu verwendet werden, den Benutzer zu schädigen.

Einleitungssatz: In Bezug auf die Verwendung meiner persönlichen Informationen und Daten ist das Produkt:

Items: unsicher / sicher, unseriös / seriös, unzuverlässig / zuverlässig, intransparent / transparent

4.8 Visuelle Ästhetik

Eindruck, dass das Produkt schön und ansprechend aussieht.

Einleitungssatz: Die visuelle Gestaltung des Produkts empfinde ich als:

Items: hässlich / schön, stilllos / stilvoll, nicht ansprechend / ansprechend, unästhetisch / ästhetisch

4.9 Anpassbarkeit

Subjektiver Eindruck des Benutzers, dass das Produkt leicht an persönliche Vorlieben oder persönliche Arbeitsstile angepasst werden kann.

Einleitungssatz: In Bezug auf meine persönlichen Anforderungen und Vorlieben ist das Produkt:

Items: nicht anpassbar / anpassbar, nicht veränderbar / veränderbar, starr / flexibel, nicht erweiterbar / erweiterbar

4.10 Nützlichkeit

Subjektiver Eindruck, dass die Verwendung des Produkts Vorteile bringt, Zeit spart oder die persönliche Produktivität verbessert.

Einleitungssatz: Die Möglichkeit das Produkt zu nutzen empfinde ich als:

Items: nutzlos / nützlich, nicht hilfreich / hilfreich, nicht vorteilhaft / vorteilhaft, nicht lohnend / lohnend

4.11 Intuitive Bedienung

Subjektiver Eindruck, dass das Produkt ohne Schulung, Anweisungen oder Hilfe von anderen Personen sofort verwendet werden kann.

Einleitungssatz: Die Bedienung des Produkts wirkt auf mich:

Items: mühevoll / mühelos, unlogisch / logisch, nicht einleuchtend / einleuchtend, nicht schlüssig / schlüssig

4.12 Wertigkeit

Subjektiver Eindruck, dass das Produkt von hoher Qualität und professionell gestaltet ist.

Einleitungssatz: Die Gestaltung des Produkts wirkt auf mich insgesamt:

Items: minderwertig / wertvoll, nicht vorzeigbar / vorzeigbar, nicht geschmackvoll / geschmackvoll, nicht elegant / elegant

4.13 Inhalts-Seriosität

Subjektiver Eindruck, dass die vom Produkt bereitgestellten Informationen zuverlässig und genau sind.

Einleitungssatz: Die Informationen und Daten, die mir das Produkt bereitstellt sind:

Items: nutzlos / nützlich, unglaubwürdig / glaubwürdig, unseriös / seriös, ungenau / genau

4.14 Inhalts-Qualität

Subjektiver Eindruck, dass die Informationen, die das Produkt dem Nutzer bereitstellt, aktuell, gut aufbereitet und interessant sind.

Einleitungssatz: Die Informationen und Daten, die mir das Produkt bereitstellt sind:

Items: veraltet / aktuell, uninteressant / interessant, schlecht aufbereitet / gut aufbereitet, unverständlich / verständlich

4.15 Haptik

Subjektive Gefühle, die sich aus der Berührung des Produkts ergeben.

Einleitungssatz: Die Oberfläche des Produkts empfinde ich als:

Items: instabil / stabil, unangenehm anzufassen / angenehm anzufassen, rau / glatt, rutschig / rutschfest

4.16 Akustik

Subjektiver Eindruck bezüglich der Betriebsgeräusche des Produkts.

Einleitungssatz: Die beim Betrieb des Produkts entstehenden Geräusche sind:

Items: lärmend / leise, missklingend / wohlklingend, dröhnend / gedämpft, schrill / sanft

4.17 Abhängigkeiten zwischen den Skalen

Im Prinzip können die Skalen des UEQ+ frei kombiniert werden, um einen passenden Fragebogen zu erzeugen. Allerdings machen wegen inhaltlicher Überschneidungen in der Praxis nicht alle Kombinationen Sinn. Betrachten wir zwei Beispiele.

Durchschaubarkeit beschreibt, ob die Nutzer das Erlernen einer Anwendung als einfach und mühelos wahrnehmen. Wenn die Anwendung aber schon als intuitiv bedienbar wahrgenommen wird, ist das automatisch gegeben. D.h. es wird in der Regel keinen Sinn machen in einem Fragebogen *Intuitive Bedienung* und *Durchschaubarkeit* zusammen zu verwenden. Evaluiert man eine Web-Seite oder einen Self-Service, wird man an intuitiver Bedienung interessiert sein. Untersucht man eine komplexe betriebswirtschaftliche Anwendung, wird man vermutlich nur *Durchschaubarkeit* verwenden.

Ähnliches gilt für *Inhalts-Seriosität* und *Vertrauen*. *Inhalts-Seriosität* ist in gewisser Weise eine Spezialisierung von *Vertrauen* auf eine bestimmte Art von Anwendung. Wird eine Web-Seite untersucht, z.B. der Online-Auftritt einer Zeitschrift, wird man *Inhalts-Seriosität* verwenden. Untersucht man eine Online-Banking Anwendung, wird dagegen *Vertrauen* die Skala der Wahl sein.

Weitere Infos zu Abhängigkeiten finden sich im UEQ+ Handbuch, dass auf der UEQ+ Web-Seite (ueqplus.ueq-research.org) zur Verfügung steht.

5 Erste Validierungen

In diesem Abschnitt werden nun erste Validierungsstudien zu den neu konstruierten Skalen beschrieben. Hierzu wurden für die drei Produktkategorien *Web-Shops*, *Video Plattformen* und *Programmierungsumgebungen* jeweils zwei häufig genutzt Produkte ausgewählt (Web-Shops: *Otto.de*, *Zalando.de*; Video Plattformen: *Netflix*, *Amazon Prime*; Programmier-Umgebungen: *Eclipse*, *Visual Studio*).

Pro Produktkategorie wurde ein spezieller Fragebogen aus den UEQ+ Skalen erstellt.

Für Web-Shops wurden die Skalen *Attraktivität*, *Steuerbarkeit*, *Intuitive Bedienung*, *Visuelle Ästhetik*, *Inhalts-Seriosität*, *Inhalts-Qualität*, *Vertrauen* und *Wertigkeit* gewählt.

Für Video-Plattformen wurden die Skalen *Attraktivität*, *Durchschaubarkeit*, *Intuitive Bedienung*, *Visuelle Ästhetik*, *Inhalts-Qualität*, *Inhalts-Seriosität* und *Vertrauen* verwendet.

Für Programmierumgebungen bestand der Fragebogen aus den Skalen *Attraktivität, Steuerbarkeit, Durchschaubarkeit, Effizienz, Nützlichkeit, Anpassbarkeit* und *Wertigkeit*.

Die Auswahl der für die Produkte relevanten Kategorien erfolgte weitgehend anhand des in [25] beschriebenen Prozesses.

Zusätzlich wurden Alter und Geschlecht und die Gesamtzufriedenheit mit der evaluierten Anwendung abgefragt, d.h. am Ende des Fragebogens erschien als zusätzliches Item:

*Insgesamt bin ich mit der Nutzerfreundlichkeit von <Produkt>
Sehr unzufrieden o o o o o o o Sehr zufrieden*

Zielsetzung der Untersuchung war die Validierung der Skalen. Deshalb wurden pro Fragebogen mehr Skalen verwendet, als man es in einer praktischen Anwendung tun würde und es wurden auch Skalen kombiniert, die man typischerweise nicht zusammen in einem Fragebogen verwenden wird.

Die Teilnehmer wurden über E-Mails und Links auf Webseiten rekrutiert. Sie konnten auswählen, welche Anwendung er oder sie bewerten wollte. Deshalb ergaben sich unterschiedliche Teilnehmerzahlen für die verschiedenen Anwendungen. Tabelle 3 zeigt Details zu den Teilnehmern aller Studien.

Produkt	N	Alter	Geschlecht	Zeit [ms]	Klicks
otto.de	42	34	16 m, 25 f, 1 NA	202.899	54
zalando.de	46	31	20 m, 24 f, 2 NA	187.803	53
Netflix	73	31	42 m, 27 f, 4 NA	211.112	48
Amazon Prime	57	32	36 m, 21 f	259.491	47
Eclipse	14	36	7 m, 4 f, 3 NA	368.552	42
Visual Studio	29	32	25 m, 1 f, 3 NA	225.006	50

Tabelle 3: Übersicht zu Teilnehmern aller 6 Fragebögen.

Das Ausfüllen der Fragebögen erforderte etwa 4 Minuten (entspricht 240.000 ms), d.h. es ging recht schnell. Die Zahl der Klicks deutet darauf hin, dass die Teilnehmer gewählte Antworten nicht sehr häufig gewechselt haben (bei 8 Skalen braucht man schon 40 Klicks für die Items und zusätzlich noch 3 weitere für die Angaben zu Alter, Geschlecht und Gesamtzufriedenheit).

Die folgenden Tabellen 4, 5 und 6 zeigen für jedes Produkt den Skalenmittelwert und dessen Standardabweichung, sowie den Cronbach-Alpha Koeffizienten.

	otto.de [N = 42]			zalando.de [N=46]		
	M	STD	α	M	STD	α
Attraktivität	1.30	1.19	0.93	1.68	1.09	0.92
Durchschaubarkeit	1.58	1.08	0.82	2.02	0.89	0.85
Intuitive Bedienung	1.57	1.09	0.94	2.13	0.87	0.90
Visuelle Ästhetik	0.89	1.41	0.95	1.47	1.29	0.95
Inhalts-Qualität	1.35	1.13	0.89	1.91	0.96	0.78
Inhalts-Seriosität	1.33	1.15	0.86	1.73	1.02	0.81
Vertrauen	1.28	1.20	0.90	1.26	1.19	0.93
Wertigkeit	0.93	1.24	0.93	1.58	1.16	0.88

Tabelle 4: Ergebnisse für Web-Shops.

Die Mittelwerte aller Skalen (mit Ausnahme von Vertrauen) sind für *otto.de* höher als für *zalando.de*. Beides sind etablierte Shops, d.h. in Bezug auf Vertrauen in die Sicherheit der eigenen Daten ist

hier auch kein Unterschied zu erwarten. Die Werte der anderen Skalen unterscheiden sich jedoch zwischen beiden Produkten.

Skala	Netflix [N=73]			Amazon Prime [N=57]		
	M	STD	α	M	STD	α
Attraktivität	2.13	1.06	0.95	1.61	1.14	0.90
Durchschaubarkeit	2.04	1.19	0.80	1.62	1.41	0.91
Intuitive Bedienung	1.86	1.16	0.90	1.36	1.38	0.94
Visuelle Ästhetik	1.58	1.17	0.89	1.01	1.28	0.94
Inhalts-Qualität	1.83	1.23	0.84	1.49	1.27	0.82
Inhalts-Seriosität	1.48	1.12	0.87	1.46	1.21	0.87
Vertrauen	1.03	1.40	0.90	0.71	1.73	0.96

Tabelle 5: Ergebnisse für Video-Plattformen.

Außer für Inhalts-Seriosität sind die Skalenmittelwerte aller anderen Skalen für Netflix höher als für Amazon-Prime. Beides sind Plattformen, die vergleichbare (Kino-)Filmangebote bereitstellen, d.h. es gibt hier eigentlich keinen erwartbaren Unterschied bzgl. der Seriosität der Inhalte.

Skala	Eclipse [N=14]			Visual Studio [N=29]		
	M	STD	α	M	STD	α
Attraktivität	0.48	1.71	0.93	1.67	0.83	0.76
Steuerbarkeit	0.84	1.80	0.97	1.77	0.82	0.83
Durchschaubarkeit	0.11	1.68	0.93	0.93	1.16	0.86
Effizienz	0.71	1.53	0.90	1.44	1.02	0.80
Nützlichkeit	1.21	1.74	0.98	2.00	0.96	0.82
Anpassbarkeit	1.25	1.56	0.98	1.78	0.91	0.80
Wertigkeit	0.32	1.64	0.93	1.66	1.08	0.79

Tabelle 6: Ergebnisse für Programmierumgebungen.

Visual Studio wird in allen Skalen deutlich höher bewertet als Eclipse. Hier ist aber zu berücksichtigen, dass die Produkte dieser Kategorie nur von wenigen Teilnehmern bewertet wurden. Die Cronbach-Alpha Werte aller Skalen sind für alle bewerteten Produkte sehr hoch. Typischerweise werden Werte > 0,7 als Indikator für gute Skalen-Qualität interpretiert.

Die Werte für die Wichtigkeit der Skalen zeigten in allen Fällen, dass die gewählten Skalen für die untersuchten Produkte auch von den Teilnehmern als relevant angesehen wurden. Diese Werte sind in [22] detailliert beschrieben.

Aus den Ratings zur Wichtigkeit der Skalen und den Ratings der Items kann man eine KPI berechnen, die den Gesamteindruck zum Produkt repräsentieren soll. Die folgende Tabelle zeigt diese KPI, die Bewertung der Gesamtzufriedenheit der Teilnehmer und die Korrelation beider Werte.

Produkt	Zufriedenheit		KPI		Korr.
	M	STD	M	STD	
otto.de	5.48	1.24	1.27	0.90	0.71
zalando.de	5.65	0.91	1.70	0.69	0.66
Netflix	6.06	0.99	1.73	0.74	0.77
Amazon Prime	5.30	1.08	1.35	0.87	0.78
Eclipse	4.21	1.74	0.40	1.37	0.83
Visual Studio	5.55	0.97	1.59	0.57	0.71

Tabelle 7: Gesamtzufriedenheit und KPI.

Die Korrelationen zwischen der berechneten KPI und der erfragten Gesamtzufriedenheit sind sehr hoch. Das stützt unsere

Interpretation der KPI. Zusätzlich besteht dieser hohe Zusammenhang offenbar auch für unterschiedliche Auswahl von Skalen und über sehr verschiedene Produktkategorien.

6 Vorgehen bei einer Anwendung mit dem UEQ+

Wie geht man vor, wenn man den UEQ+ praktisch anwenden will?

Im ersten Schritt muss man aus den vorhandenen UX Skalen die Skalen auswählen, die im konkreten Fragebogen verwendet werden sollen. Hierzu kann man auf einige empirisch ermittelte Empfehlungen [2, 11, 12] zurückgreifen, die abhängig von der Art des Produkts angeben, welche Skalen von Nutzern typischerweise als relevant angesehen werden. Allerdings sollte man diese Empfehlungen immer kritisch hinterfragen, da sich für spezielle Produkte Abweichungen ergeben können. Es ist mit wenig Aufwand möglich, in einer kleinen Vorstudie die Wichtigkeiten der Skalen von einer Gruppe von Nutzern zu erfragen. Oft haben aber auch Produktverantwortliche oder die Designer der Anwendung schon eine klare Vorstellung, welche UX Aspekte untersucht werden sollten, so dass Unternehmensstrategien in die Auswahl der Skalen mit einfließen können. Beispielsweise kann ein Unternehmen die Skala *Wertigkeit* als unternehmensrelevant im Vergleich zum Wettbewerb ansehen, auch wenn diese von den Nutzern nicht als vorrangige Skala ausgewählt wird.

Wir empfehlen nicht mehr als 6 Skalen in einer Untersuchung zu verwenden damit der Aufwand zum Ausfüllen überschaubar bleibt und genau zu überlegen, welche Skalen ausgewählt werden. Das Handbuch zum UEQ+ gibt zusätzliche Hinweise zu Abhängigkeiten zwischen den Skalen, d.h. Paaren von Skalen, die man eher nicht zusammen in einem Fragebogen verwenden sollte.

Im zweiten Schritt muss der konkrete Fragebogen erstellt werden. Hier stehen auf der Web-Seite des UEQ+ (ueqplus.ueq-research.org) Vorlagen (im Moment in Deutsch und Englisch) zur Verfügung.

Der finale Schritt ist die Durchführung und die Auswertung der Daten. Hier steht, wie beim UEQ, ein Excel-Sheet zur Datenanalyse auf der Web-Seite bereit. Da die Skalenauswahl beim UEQ+ nun dynamisch ist, bietet dieses weniger Analysen und Komfort, als das entsprechende Tool des Original-UEQ. Berechnet werden pro Skala Mittelwert, Standard-Abweichung und Konfidenzintervall, sowie über alle Skalen die KPI.

7 Abgrenzung zum UEQ

Der UEQ+ erlaubt es mit vertretbarem Aufwand einen UX Fragebogen zu erstellen, der für ein bestimmtes Produkt optimal passt. Ein solcher Fragebogen enthält genau die Skalen, die für die Nutzer des Produkts die höchste Relevanz haben und bildet damit den Eindruck der Nutzer zur UX des Produkts gut ab.

Aber diesen Vorteilen stehen einige Nachteile gegenüber. Das Aufsetzen des Fragebogens und die Auswertung der Daten ist etwas aufwändiger, als wenn man direkt den UEQ oder einen anderen Standardfragebogen verwendet.

Jede Anwendung des UEQ+ setzt voraus, dass die Skalen für die Untersuchung zuvor festgelegt wurden. Es gibt empirische Ergebnisse, die je nach Typ der zu evaluierenden Anwendung Hinweise auf die wichtigsten Skalen geben [11, 12], aber es ist auf jeden Fall ein gewisser Aufwand in die Auswahl der Skalen zu investieren.

Für die Auswertung der Daten eines UEQ reicht es völlig aus, die erhobenen Daten in das Data Analysis Excel-Sheet (frei verfügbar unter ueq-online.org) einzutragen. Alle relevanten Analysen sind hier bereits implementiert und können direkt übernommen werden.

Da die Skalen einer UEQ+ Umfrage variieren, kann das ebenfalls vorhandene UEQ+ Auswertungs-Excel keinen völlig gleichwertigen Komfort bieten, d.h. man muss unter Umständen einige relevante Fragen durch eigene Datenanalysen vornehmen.

Auch die Interpretation der Ergebnisse ist schwieriger. Die berechneten Skalenwerte sind erst einmal reine Zahlen, d.h. eigentlich nur in Beziehung zu anderen Ergebnissen gut interpretierbar. Ist ein beobachteter Wert von 1,3 für die Skala Effizienz jetzt gut oder nicht? Der UEQ bietet einen Benchmark [23,24], der hilft die Daten zu interpretieren. Dieser Benchmark basiert auf den Ergebnissen einer großen Zahl von UEQ Evaluationen verschiedener Produkte. Ein Vergleich des eigenen Ergebnisses mit den Ergebnissen der Produkte im Benchmark-Datensatz gibt also einen guten Einblick, wo man in Bezug auf UX im Vergleich mit typischen anderen Produkten steht. Für die neuen Skalen im UEQ+ steht ein solcher Benchmark im Moment noch nicht zur Verfügung. Da einige dieser Skalen auch nur für spezielle Produkte relevant sind, wird es auch eine längere Zeit dauern, bis für alle neuen Skalen ein guter und verlässlicher Benchmark zur Verfügung steht. Anstelle des Benchmarks lässt sich jedoch beim UEQ+ anhand der Werte für die Wichtigkeiten der einzelnen Skalen ermitteln, wie wichtig eine Skala aus Sicht der Nutzer ist. Wenn man diese Wichtigkeiten mit aller Vorsicht als „Soll-Werte“ interpretiert, kann ein Handlungsbedarf abgeleitet werden.

Anhand dieser Vor- und Nachteile des UEQ+ gegenüber dem klassischen UEQ kann man Hinweise geben, wann man welchen Ansatz verwenden sollte.

Wenn nur ein einzelnes Produkt evaluiert werden soll und man eher an einer ersten Einschätzung zur UX des Produkts interessiert ist, sollte man den Original-UEQ verwenden. Selbst wenn einige Skalen nicht perfekt zum evaluierten Produkt passen oder einige Skalen fehlen, die man als wichtig erachtet, überwiegen hier die Vorteile der leichten Anwendbarkeit und die Verfügbarkeit des UEQ Benchmarks.

Soll das gleiche Produkt über einen längeren Zeitraum mehrfach evaluiert werden (z.B. um die Verbesserung oder Verschlechterung der UX zwischen verschiedenen Versionen des Produkts zu messen) und passen die UEQ Skalen nicht gut zum Produkt, spricht dies eher für die Konstruktion eines speziellen Fragebogens mit Hilfe des UEQ+. In diesem Szenario spielt der noch nicht vorhandene Benchmark keine Rolle, da man ja verschiedene Versionen des Produkts miteinander vergleicht, d.h. man hat hier automatisch einen guten Bezugspunkt (die erste Messung) der hilft die Ergebnisse zu deuten.

Wenn man eine UX Messung für eine größere Menge von ähnlichen Produkten etablieren will und die UEQ Skalen nicht gut passen, sollte man ebenfalls über eine Verwendung des UEQ+ nachdenken. In diesem Fall ist der zusätzliche Aufwand für das Aufsetzen des Fragebogens vernachlässigbar, da dieser nur einmal anfällt und das Ergebnis dann für alle Produkte einer Produktgruppe wiederverwendet werden kann. In einem solchen Szenario wird man auch in der Lage sein, selbst schnell eine größere Menge von Messergebnissen zu generieren, d.h. kann leicht seinen eigenen Benchmark aufbauen.

8 Zusammenfassung und Ausblick

Wir haben eine modulare Erweiterung des UEQ vorgestellt, den UEQ+. Dieser UEQ+ erlaubt einen eigenen Fragebogen aus einer Menge von 16 vorhandenen Skalen zu erstellen. Ein solcher Fragebogen kann damit genau an die Anforderungen des zu evaluierenden Produkts anpasst werden. Allerdings erfordert die Anwendung des UEQ+ etwas mehr Aufwand als eine direkte Anwendung des UEQ und es ist im Moment auch etwas weniger komfortabel die Daten zu interpretieren. Vor einer Anwendung des UEQ+ sollte man also die Vor- und Nachteile beider Methoden gegeneinander abwägen.

Natürlich decken auch die Skalen des UEQ+ nicht das gesamte Spektrum von UX ab und das wird auch nicht erreichbar sein, da neue Produkte immer neue Interaktionsparadigmen und damit neue UX Aspekte mit sich bringen. Wir werden daher versuchen, den UEQ+ in Zukunft um weitere Skalen zu erweitern.

Eine weitere Aktivität ist die Übersetzung des UEQ+. Der Original-UEQ liegt im Moment in mehr als 20 Sprachen von (siehe ueq-online.org). Für den UEQ+ sind im Moment nur eine deutsche und englische Version verfügbar, was sich aber hoffentlich bald ändern wird.

Weiterhin ist geplant einen Benchmark für die KPI zu erstellen.

REFERENCES

- [1] Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In International Conference of Design, User Experience, and Usability, S. 383-392. Springer, Cham.
- [2] Schrepp, M. (2018). User Experience mit Fragebögen messen. Amazon Kindle Direct Publishing, ISBN: 9781986843768.
- [3] Laugwitz, B.; Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. A.M. Heinecke & H. Paul (Eds.): Mensch & Computer 2006 - Mensch und Computer im Strukturwandel. Oldenbourg Verlag, S. 125 - 134.
- [4] Hassenzahl, M.; Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität In: Ziegler, J. & Szwillus, G. (Hrsg.), Mensch & Computer 2003. Interaktion in Bewegung, S. 187-196. Stuttgart, Leipzig: B.G. Teubner.
- [5] Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. International Journal of Human-Computer Studies, 68, S. 689-709.
- [6] Minge, M. & Riedel, L. (2013). meCUE - Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In: S. Boll, S. Maaß & R. Malaka (Hrsg.): Mensch und Computer 2013: Interaktive Vielfalt, S. 89-98. München, Oldenbourg Verlag.
- [7] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability Evaluation in Industry, 189(194), S. 4-7.
- [8] Gediga, G., Hamborg, K. C., & Dütsch, I. (1999). The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. Behaviour & Information Technology, 18(3), S. 151-164.
- [9] Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. British journal of educational technology, 24(3), S. 210-212.
- [10] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). Human-Computer Interaction. Workinghal, Addison-Wesley.
- [11] Winter, D., Schrepp, M., & Thomaschewski, J. (2015). Faktoren der User Experience-Systematische Übersicht über produktrelevante UX-Qualitätsaspekte. Mensch und Computer 2015 - Usability Professionals.
- [12] Winter, D., Hinderks, A., Schrepp, M. & Thomaschewski, J., (2017). Welche UX Faktoren sind für mein Produkt wichtig? In: Hess, S. & Fischer, H. (Hrsg.), Mensch und Computer 2017 - Usability Professionals. Regensburg: Gesellschaft für Informatik e.V., S. 191 - 200.
- [13] Hinderks, A. (2016). Modifikation des User Experience Questionnaire (UEQ) zur Verbesserung der Reliabilität und Validität. Unveröffentlichte Masterarbeit, University of Applied Sciences Emden/Leer.
- [14] Boos, B. & Brau, H., (2017). Erweiterung des UEQ um die Dimensionen Akustik und Haptik. In: Hess, S. & Fischer, H. (Hrsg.), Mensch und Computer 2017 - Usability Professionals, Regensburg: Gesellschaft für Informatik e.V., S. 321 - 327.
- [15] Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In Holzinger, A. (Hrsg.): HCI and Usability for Education and Work, LNCS 5298, Berlin, Heidelberg: Springer, S. 63-76.
- [16] Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive?. In CHI Proceedings, S. 2215-2224.
- [17] Schrepp, M. & Müller, K (2015). Übersichtlichkeit als Mediator zwischen Ästhetik und Usability? In: Diefenbach, S.; Henze, N. & Pielot, M. (Eds.), Mensch & Computer 2015 - Tagungsband, S. 73-82, DE GRUYTER 2015. DOI: 10.1515/9783110443929-009.
- [18] Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer-Studies 60, S. 269-298.
- [19] Schrepp, M.; Hinderks, A. & Thomaschewski, J. (2017). Die UX KPI. Wunsch und Wirklichkeit. In: Hess, S. & Fischer, H. (Hrsg.), Mensch und Computer 2017 - Usability Professionals, Regensburg: Gesellschaft für Informatik e.V., S. 117-125.
- [20] Hinderks, A., Schrepp, M., Domínguez Mayo, F.J., Escalona, M.J., Thomaschewski, J. (2019). Developing a UX KPI based on the User Experience Questionnaire. Computer Standards & Interfaces.
- [21] Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 1.8.12
- [22] Schrepp, M. & Thomaschewski, J. (2019). Construction and first Validation of Extension Scales for the User Experience Questionnaire (UEQ). Research Report available on Research Gate. University of Applied Sciences Emden/Leer. DOI: 10.13140/RG.2.2.19260.08325
- [23] Schrepp, M.; Olschner, S. & Schubert, U. (2013). User Experience Questionnaire Benchmark - Praxiserfahrungen zum Einsatz im Business-Umfeld. In: Brau, H.; Lehmann, A.; Petrovic, K.; Schroeder, M. (Eds.); *Usability Professionals 2013* S. 348 - 353, 2013.
- [24] Schrepp, M.; Hinderks, A. & Thomaschewski, J. (2017). Construction of a benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No. 4, S. 40-44.
- [25] Thomaschewski, J., Hinderks, A., Schrepp, M. (2018): Welcher UX-Fragebogen passt zu meinem Produkt? In: Stefan Hess und Holger Fischer (Hg.): Mensch und Computer - Usability Professionals. Bonn, S. 437-445.

AUTOREN



Schrepp, Martin

Dr. Martin Schrepp studierte Mathematik und Psychologie an der Universität Heidelberg. 1990 Abschluss als Diplom-Mathematiker. 1990 – 1993 Promotion in Psychologie. Seit 1994 bei der SAP AG tätig. Bisherige Tätigkeitsfelder waren hier die Konzeption technischer Dokumentation, Software-Entwicklung, User Interface Design und Barrierefreiheit. Hauptinteressen sind die Anwendung kognitionswissenschaftlicher Erkenntnisse auf das Design interaktiver Anwendungen, Barrierefreiheit und die Entwicklung von Methoden zur Evaluation und Datenanalyse.



Thomaschewski, Jörg

Jörg Thomaschewski ist Professor an der Hochschule Emden/Leer mit den Lehr- und Forschungsschwerpunkten Usability und User Experience, Human Computer Interaction, Requirement Engineering, Agile Software Development, Internet-Programming, und E-Learning. Er ist Autor verschiedener Online-Module, u.a. „Mensch-Computer-Kommunikation“, das im Rahmen der Virtuellen Hochschule (VFH) an sechs Hochschul-Standorten eingesetzt wird. Er verfügt über umfangreiche Erfahrungen in Usability-Schulungen, agilen Methoden, IT-Analysen und Beratungen.