

Beurteilung der UX Qualität durch Experten

Dominik Mayer¹, Martin Schrepp¹, Theo Held¹

SAP SE¹

dominik.mayer@sap.com, martin.schrepp@sap.com, theo.held@sap.com

Zusammenfassung

Experten-Reviews sind eine sehr schnell durchführbare und kostengünstige Methode, um die User Experience einer Anwendung zu beurteilen. Allerdings sind die bekannten Review-Verfahren stark auf das Auffinden von konkreten Nutzungsproblemen fokussiert. Heute hat sich aber eine deutlich weitere Sicht auf UX etabliert, d.h. es spielen auch nicht-aufgabenbezogen Aspekte, z.B. Joy of Use, originelle Gestaltung oder Schönheit der Benutzeroberfläche, eine zentrale Rolle für die Bewertung der Qualität einer Anwendung. Die bekannten Review-Verfahren sind nicht darauf angelegt, diese hedonischen Aspekte angemessen zu berücksichtigen. Wir stellen eine Variante des Experten-Reviews vor, die versucht, hedonische und pragmatische Qualitätsaspekte gleichermaßen zu berücksichtigen.

1 Experten-Reviews

Es gibt mehrere Varianten von UX Experten-Reviews. Sehr bekannt ist die Heuristische Evaluation (Nielsen & Molich, 1990), bei der Reviewer das User Interface eines Systems anhand von abstrakt formulierten Heuristiken (eine Übersicht verwendeter UX Heuristiken findet sich in Bader et al., 2017) testen. Eine andere bekannte Methode ist der Cognitive Walkthrough (Wharton et al., 1994). Hier spielen Experten typische Anwendungsszenarien eines Systems aus Sicht von Nutzern durch und versuchen so, bestehende Probleme aufzudecken. Eine weitere verbreitete Art des Experten-Reviews sind einfache Prüfungen, ob eine Anwendung einer vorgegebenen UI Guideline entspricht.

Das Ziel eines Experten-Reviews ist in allen Fällen, die potenziellen Probleme eines User Interfaces zu ermitteln, zu klassifizieren, zu priorisieren und ggfs. auch Vorschläge zur Vermeidung dieser Probleme zu erstellen. Damit erhält man als Ergebnis des Reviews eine Liste von konkreten Problemen und kann über den Schweregrad der Probleme entscheiden und mit welcher Priorisierung man diese beheben sollte, um die Qualität des Produkts zu verbessern.

Experten-Reviews sind eine sehr kostengünstige Möglichkeit, die User Experience eines Produkts zu beurteilen. Es genügt eine kleine Zahl von Reviewern, die die Nutzungsszenarien der zu evaluierenden Anwendung kennen sollten. Insbesondere müssen keine Nutzer rekrutiert werden. Experten können erfahrene UX Designer sein, aber mit entsprechender Anleitung auch Berater oder sogar Endbenutzer (siehe Mack & Nielsen, 1994).

Ein Problem von Experten-Reviews ist, dass sie stark auf klassische Usability-Aspekte fokussieren, z.B. Probleme mit Effizienz oder Erlernbarkeit. Heute hat sich aber eine deutlich weitere Sicht auf UX etabliert. Für die Bewertung der Gesamt-Qualität eines Produkts spielen auch Aspekte, wie z.B. Joy of Use, originelle Gestaltung oder Schönheit, eine zentrale Rolle. Die bekannten Review-Verfahren sind nicht darauf angelegt, solche hedonischen Aspekte angemessen zu berücksichtigen. Ein weiteres Problem ist, dass Experten-Reviews stark auf bestehende Schwächen eines Produkts fokussieren und Stärken oft nicht adäquat berücksichtigt werden. Das kann durchaus problematisch sein, da man unter Umständen bei der Beseitigung vorhandener Probleme auch unwissentlich vorhandene Stärken eines Produkts beschädigt.

Wir stellen ein Verfahren zum Experten-Review vor, das explizit auch hedonische UX Faktoren berücksichtigen soll und gleichermaßen geeignet ist, Stärken und Schwächen eines Produkts zu beurteilen.

2 Beschreibung des neuen Review-Verfahrens

Kern des Verfahrens ist eine Liste von allgemein gehaltenen Beschreibungen von UX Qualitäten. Diese werden jeweils in einem kurzen Text erklärt. Das erlaubt es auch hedonische Aspekte gleichwertig einzubinden. Betrachten wir zwei Beispiele:

- *Steuerbarkeit*: Das Produkt reagiert vorhersehbar und konsistent auf die Eingaben und Befehle des Nutzers. Er hat die Kontrolle über die Interaktion.
- *Stimulation*: Das Produkt ist anregend und spannend für den Nutzer. Er hat Spaß daran, sich mit dem Produkt zu beschäftigen und damit zu arbeiten.

Die UX Qualitäten wurden aus einer Untersuchung von Winter et al. (2017) abgeleitet, wobei die Beschreibungen der UX Qualitäten an einigen Stellen angepasst wurden, um deren Verständlichkeit zu erhöhen. Eine detailliertere Erläuterung zu diesen UX Qualitäten findet sich in Schrepp (2018).

Es sind folgende UX Qualitäten beschrieben: *Inhaltsqualität* (IN), *Anpassbarkeit* (AN), *Durchschaubarkeit* (DU), *Effizienz* (EF), *Intuitive Bedienung* (IB), *Nützlichkeit* (NÜ), *Originalität* (OR), *Schönheit* (SC), *Steuerbarkeit* (ST), *Stimulation* (SL), *Übersichtlichkeit* (ÜB), *Vertrauen* (VT), *Wertigkeit* (WE), *Immersion* (IM), *Identität* (ID) und *Verbundenheit* (VB).

2.1 Relevante UX Kriterien bestimmen

Es wird nicht angenommen, dass alle UX Aspekte für alle Anwendungen relevant sind. Zum Beispiel wird man für die Evaluation einer betriebswirtschaftlichen Anwendung sicher andere UX Aspekte wählen als für ein Spiel (siehe z.B. die Untersuchung von Winter et al., 2017). D.h. pro Anwendung wird in einem ersten vorbereitenden Schritt eine Auswahl an UX Aspekten getroffen, die von den Reviewern dann berücksichtigt werden sollen.

Hier spielen zwei Betrachtungsweisen eine Rolle. Erstens sollen die ausgewählten UX Aspekte natürlich aus Sicht der Nutzer des Produkts wichtig sein. Aber hier spielt auch die Sicht des Anbieters eine Rolle. Auch UX Kriterien, die Nutzer als relativ unwichtig ansehen, können aus Sicht des Anbieters für den Produkterfolg zentral sein. Nehmen wir z.B. eine Programmierumgebung. Für solche Anwendungen ist eine originelle Gestaltung des UI für Nutzer irrelevant (siehe z.B. die Daten in Winter et al., 2017). Wenn eine neue Programmierumgebung aber auf Veranstaltungen über Produkt-Demonstrationen platziert werden soll, sieht die Sache anders aus. Hier will der Anbieter sich von der Konkurrenz abheben und Aufmerksamkeit auf die Anwendung lenken und dafür ist eine originelle Gestaltung hilfreich.

Für die Auswahl der relevanten UX Kriterien kann man auf bestehende Daten zurückgreifen. Zum Beispiel wird in Winter et al. (2017) in einer empirischen Untersuchung die Wichtigkeit von UX Aspekten für verschiedene Produktkategorien aus Sicht von Nutzern untersucht. Man kann solche Daten aber auch leicht selbst durch Anwendung von Ranking-Verfahren erheben.

2.2 Durchführung des Reviews

Im eigentlichen Review gehen die Experten durch die Anwendung und beurteilen diese. Sie können dabei die Anwendung entweder ohne Vorgaben frei explorieren oder es wird ein aufgabenbasiertes Szenario vorgegeben, an dem sie sich orientieren. Die Entscheidung, welche dieser Varianten vorzuziehen ist, hängt von der Art der Anwendung und der Expertise der Reviewer ab.

Wichtig ist, dass vor der Beschäftigung mit der Anwendung die UX Aspekte vorgegeben werden und ausreichend Zeit vorhanden ist, sich mit diesen vertraut zu machen und bei Bedarf Fragen zu stellen. Damit wird ein einheitliches Verständnis der UX Aspekte gefördert.

Weiterhin sollten die Reviewer instruiert werden, sowohl positive als auch negative Kommentare zu erfassen. Diese Kommentare werden den Aspekten zugeordnet, z.B. indem Sie direkt unter der entsprechenden Beschreibung des UX Aspekts erfasst werden (am besten druckt man pro UX Aspekt ein Blatt, siehe Abbildung 1). Nachdem ein Reviewer alle Kommentare erfasst hat, gibt er noch auf einer 7-stufigen Likert-Skala eine Einschätzung ab, inwieweit die untersuchte Anwendung die durch die UX Aspekte beschriebenen Qualitäten besitzt.

Als Beispiel zeigen wir in Abbildung 1 die entsprechenden Seiten des Fragebogens für die UX Aspekte *Effizienz* und *Schönheit*, wie sie auch in dieser Form in den beiden in dieser Arbeit beschriebenen Studien verwendet wurden.

<p><u>Bitte beurteilen Sie die Schönheit des Produkts</u> Schönheit bedeutet für uns: Das Produkt ist schön und ansprechend gestaltet.</p> <hr/> <p>Gibt es unschöne Elemente im Produkt? Welche Elemente sind besonders schön gestaltet?</p> <div style="border: 1px solid black; height: 60px; width: 100%;"></div> <hr/> <p>Wie würden Sie als Experte die Schönheit des Produkts beurteilen? Sehr niedrig <input type="radio"/> Sehr hoch</p>	<p><u>Bitte beurteilen Sie die Effizienz des Produkts</u> Effizienz bedeutet für uns: Der Nutzer kann seine Ziele mit minimalem Aufwand erreichen. Zur Erledigung seiner Aufgaben sind keine unnötigen Eingaben oder <u>Navigationen</u> erforderlich. Das Produkt reagiert schnell auf die Eingaben und Befehle des Nutzers.</p> <hr/> <p>Haben Sie im Produkt Beispiele für Effizienzprobleme gefunden? Oder positive Beispiele, die ein effizientes Arbeiten massiv unterstützen?</p> <div style="border: 1px solid black; height: 60px; width: 100%;"></div> <hr/> <p>Wie würden Sie als Experte die Effizienz des Produkts beurteilen? Sehr niedrig <input type="radio"/> Sehr hoch</p>
---	--

Abbildung 1: Materialien zu Schönheit (links) und Effizienz (rechts).

2.3 Resultierende Daten

Am Ende des Reviews hat man einerseits eine Menge von konkreten Kommentaren zu den einzelnen UX Aspekten und andererseits eine quantitative Bewertung dieser Aspekte. Die Kommentare kann man direkt für mögliche Verbesserungen verwenden, während die quantitative Bewertung eher eine allgemeine Einschätzung der Experten zur UX Qualität zeigt.

3 Studie 1

In dieser Studie wurden die Anwendungen *Skype for Business*, *Amazon.de* und *Eclipse* untersucht. Als Reviewer agierten Studierende aus Studiengängen mit Bezug zu UX, d.h. wir haben hier Reviewer mit eher geringer UX Expertise.

3.1 Auswahl der relevanten UX Aspekte

Als Vorbereitung des eigentlichen Reviews wurde ein Workshop mit acht UX Experten durchgeführt. Diese wurden gebeten, für die entsprechenden Produktkategorien zu ermitteln, wie wichtig die vorhandenen UX Aspekte für diese Art von Produkten sind. Die Experten sollten dabei die Wichtigkeit aus Sicht der Nutzer und aus Sicht des Anbieters beurteilen. Für die Beurteilung die Anbieter-Sicht sollten sich die Experten vorstellen, dass das entsprechende

Produkt vor potenziellen Kunden präsentiert wird und festlegen, welche UX Aspekte hier für den Eindruck des Publikums relevant sind.

Die UX Aspekte *Verbundenheit*, *Immersion* und *Identität* wurden aus theoretischen Überlegungen als irrelevant für die untersuchten Anwendungen eingeschätzt und im Rahmen der Voruntersuchung nicht vorgegeben.

Um es den Teilnehmern zu erleichtern, alle Dimensionen in eine Reihenfolge zu bringen, wird ihnen das *Ranking by Elimination* (Wickelmaier et al., 2012) Verfahren vorgeschlagen. Hierbei „eliminieren“ die Teilnehmer nach und nach die einzelnen Dimensionen. Sie legen also zuerst fest, welche Dimension aus allen verfügbaren die unwichtigste ist und schreiben diese auf die letzte Position, die Dimension ist damit „eliminiert“. Für die übrigen Dimensionen gehen sie genauso vor und schreiben die am wenigsten wichtige auf die vorletzte Position, usw. Dies wird solange wiederholt, bis die gesamte Rangliste gefüllt ist.

Dieses Bewertungsverfahren wurde gewählt, um die Teilnehmer zu einer Priorisierung zu zwingen. Bei anderen Verfahren (z.B. der Verwendung von Likert-Skalen) besteht die Gefahr, dass Teilnehmer die Wichtigkeit aller Dimensionen als sehr hoch einstufen, da sie alle positive „Soll-Eigenschaften“ eines Produkts beschreiben. Durch die Bildung der Rangliste werden Unterschiede zwischen den Dimensionen deutlicher.

Pro Anwendungstyp wurde dieses Verfahren zuerst aus Sicht der Nutzer und direkt danach aus Sicht des Anbieters durchgeführt, so dass die Teilnehmer insgesamt sechs Rangreihen erzeugten. Die Rangreihen wurden dann verwendet, um über eine BTL-Skalierung (Bradley & Terry, 1952) Skalenwerte zu erzeugen, die die Wichtigkeiten der UX Aspekte relativ zueinander ausdrücken.

Abbildung 2 zeigt die Ergebnisse für die Produktkategorie *Entwicklungsumgebungen*. Man sieht direkt, dass die Experten die Wichtigkeit einiger UX Aspekte aus Nutzer- und Demosicht stark unterschiedlich einschätzen. Insgesamt waren die Experten in der Lage, die Rangreihen ohne Probleme zu erstellen.

Die Skalenwerte wurden nun benutzt, um die für das Review relevanten Aspekte auszuwählen. Hierzu wurden die UX Aspekte gewählt, die in einem der beiden Kontexte eine gewisse Wichtigkeit aufwiesen. Man kann hier keine wirkliche Grenze für den BTL-Skalenwert angeben, ab dem man einen UX Aspekt ausschließt. Wie viele UX Aspekte man wählt, ist hier immer eine subjektive Entscheidung, die auch davon abhängt, wie viel Zeit die Reviewer aufbringen können, d.h. wie viele unterschiedliche Aspekte man diesen zumuten will.

Für die Evaluation von Eclipse wurden die Aspekte *Effizienz* (EF), *Übersichtlichkeit* (ÜB), *Anpassbarkeit* (AN), *Durchschaubarkeit* (DU), *Intuitive Bedienung* (IB), *Nützlichkeit* (NÜ), *Steuerbarkeit* (ST), *Originalität* (OR), *Wertigkeit* (WE) und *Schönheit* (SC) gewählt.

Für die beiden anderen Produktkategorien wurde analog vorgegangen, um die Liste der für ein Review relevanten UX Aspekte zu ermitteln.

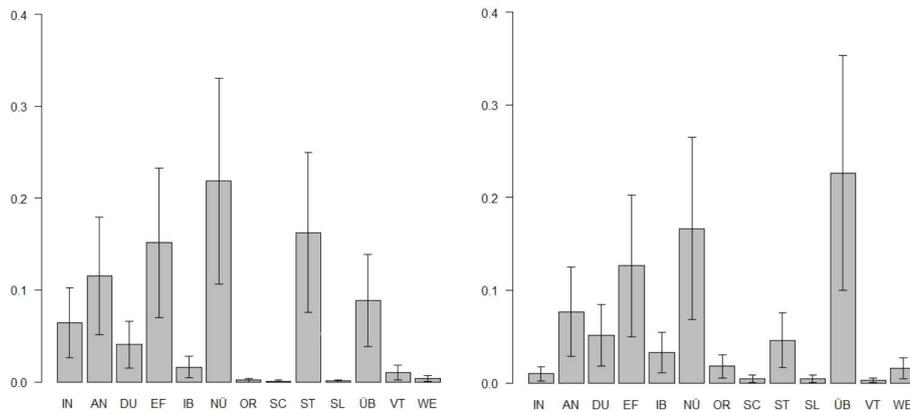


Abbildung 2: Wichtigkeiten der UX Aspekte Nutzersicht (links) und aus Anbietersicht (rechts).

3.2 Ablauf des Reviews

Alle Reviewer waren mit den untersuchten Anwendungen bereits vertraut. Alle drei Anwendungen wurden in einer Sitzung nacheinander evaluiert.

Zu Beginn wurde den Studierenden die Vorgehensweise erklärt. Hier wurde auch explizit erwähnt, dass sie sowohl auf negative als auch auf positive Punkte zur Anwendung achten sollten. Der Fragebogen mit der Beschreibung der UX Aspekte wurde ausgeteilt (siehe Abbildung 1) und die Reviewer erhielten ausreichend Zeit, sich damit vertraut zu machen und ggfs. Fragen zu den UX Aspekten und zum Vorgehen zu stellen. Danach öffneten die Reviewer die entsprechende Anwendung auf ihrem Laptop und hatten 30 Minuten Zeit, diese frei zu explorieren und Kommentare zur Anwendung zu notieren. Danach wurde die Bewertung der einzelnen UX Aspekte vorgenommen.

3.3 Auswertung

Betrachten wir zunächst einen Vergleich der quantitativen (Likert-Skala) und qualitativen (Kommentare) Befunde. Die von den Reviewern notierten Kommentare wurden in positive, negative und neutrale Punkte gruppiert. Die folgenden Tabellen zeigen die Differenz aus positiven und negativen Bemerkungen (d.h. eine negative Zahl zeigt, dass mehr negative als positive Bemerkungen notiert wurden) und die mittlere Bewertung eines UX Aspekts über alle Reviewer.

UX Aspekt	EF	OR	AN	DU	IB	NÜ	SC	ST	ÜB	WE
Diff. Positiv/Negativ	-5	-4	5	-5	-5	3	-13	-1	-4	-3
Bewertung	4,25	3	5,75	4,75	2,75	5,25	2,5	5	3,5	4,5

Tabelle 1: Positive und negative Kommentare und quantitative Bewertung für Eclipse.

UX Aspekt	EF	IN	DU	IB	NÜ	SC	ST	ÜB	VT	WE
Diff. Positiv/Negativ	12	10	0	10	16	-5	4	-8	6	-1
Bewertung	5,85	5,71	5,42	6	6,28	4,28	6,42	4,57	5	5,85

Tabella 2: Positive und negative Kommentare und quantitative Bewertung für Amazon.de.

UX Aspekt	EF	IB	NÜ	ST	VT	ID
Diff. Positiv/Negativ	4	1	11	-3	1	7
Bewertung	5,71	5,28	5,71	5,14	6	4,28

Tabella 3: Positive und negative Kommentare und quantitative Bewertung für Skype for Business.

Es gibt einen starken Zusammenhang zwischen der Differenz aus positiven und negativen Kommentaren und der quantitativen Bewertung. Über alle drei Anwendungen hinweg beträgt die Korrelation dieser Werte 0,725. Die positiven und negativen Befunde haben offenbar einen starken Einfluss auf die Bewertung der UX Qualität. Weiter zeigte sich, dass die Reviewer offenbar wie gewünscht sowohl Schwächen als auch Stärken berücksichtigt haben.

Parallel zum Review wurden in einer Online-Studie Daten aus einer studentischen Zielgruppe zu den drei Anwendungen erhoben. Die Teilnehmer erhielten die gleichen UX Aspekte wie die Reviewer vorgegeben. Hier wurde aber nur die Bewertung über die Likert-Skala vorgenommen, d.h. die Teilnehmer sollten die Anwendungen nicht genauer untersuchen, sondern nur ihren Eindruck wiedergeben. Die Beschreibungen der UX Aspekte war aber mit der Beschreibung für die Reviewer identisch. Die folgende Abbildung zeigt die Unterschiede zwischen den Bewertungen der sieben Reviewer und den Bewertungen der befragten Nutzer. Dabei wurde Eclipse von 22, Amazon.de von 24 und Skype for Business von 25 Personen bewertet.

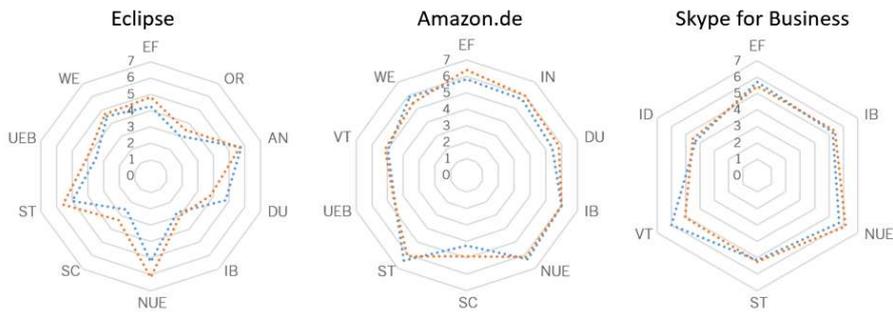


Abbildung 3: Die blaue Linie zeigt die Einschätzung der Reviewer, die rote Linie die der befragten Nutzer.

Die Einschätzungen der Reviewer decken sich sehr gut mit der Einschätzung der befragten studentischen Zielgruppe. Allerdings muss man hier einschränken, dass eine unrealistisch hohe Zahl von Reviewern mitwirkten. In typischen Projekten wird man selten mehr als zwei bis drei Reviewer finden und daher werden die Übereinstimmungen unter Umständen geringer ausfallen.

4 Studie 2

In der zweiten Studie waren die Reviewer durchweg UX Experten mit langjähriger Berufserfahrung. Hier wurde ein komplexes Softwareprodukt zur Administration von Cloud Anwendungen untersucht. Für dieses Produkt liegen auch Ergebnisse eines Usability Tests und Ergebnisse von mehreren UX Fragebögen vor. Durch einen Vergleich mit diesen Ergebnissen kann auch hier die Qualität der Ergebnisse des Experten-Reviews überprüft werden.

4.1 Auswahl der relevanten UX Aspekte

Die untersuchte Anwendung ist eine Entwicklungsumgebung und von den Nutzungsszenarien ähnlich zu *Eclipse*. Daher können die für *Eclipse* verwendeten UX Aspekte auch hier verwendet werden. Da es sich um ein Tool für erfahrene Entwickler handelt und die beiden UX Aspekte *Durchschaubarkeit* und *Intuitive Bedienung* recht ähnlich sind, wurde entschieden in dieser Untersuchung nur *Durchschaubarkeit* zu verwenden und auf die Evaluation von *Intuitiver Bedienung* zu verzichten, so dass 9 UX Aspekte verblieben.

4.2 Ablauf des Reviews

Die vier Reviewer evaluierten die Anwendung unabhängig voneinander in zweistündigen Sitzungen. Zu Beginn wurden dem Reviewer die Methode erklärt und der Fragebogen ausgeteilt. Dann hatte der Reviewer Zeit, sich mit den UX Aspekten vertraut zu machen und Fragen zu diesen UX Aspekten und zur generellen Vorgehensweise zu stellen. Danach bekamen die Reviewer einen Link auf die zu evaluierende Anwendung. Zusätzlich wurden die Aufgaben aus einem Usability Test zu dieser Anwendung ausgeteilt. Die Reviewer sollten die Anwendung anhand der dort vorgegebenen Nutzungsszenarien beurteilen. Die Reviewer versuchten zunächst in der Rolle des Nutzers die Aufgaben selbst zu lösen. Falls Sie dabei auf Schwierigkeiten stießen, stand ihnen eine Lösung auf einem separaten Lösungsblatt zur Verfügung.

Diese Vorgehensweise wurde gewählt, da die evaluierte Anwendung sehr komplex ist und viele verschiedene Nutzungsszenarien unterstützt. Bei einer rein freien Exploration hätte hier die Gefahr bestanden, dass die Reviewer in völlig verschiedene Bereiche der Anwendung abtauchen. Zusätzlich erlaubt diese Vorgehensweise einen Vergleich der Review-Ergebnisse mit den Daten des Usability Tests, die zu diesem Zeitpunkt schon vorlagen.

Die Reviewer notierten während der Bearbeitung der Aufgaben relevante Kommentare zu den UX Aspekten. Die Bewertung der UX Aspekte wurde dann nach Bearbeitung aller Aufgaben

vorgenommen. Am Ende des Reviews wurde noch ein kurzes Interview mit den UX Experten geführt, um zu ermitteln, wie gut diese mit dem Verfahren zurechtkamen.

4.3 Auswertung

4.3.1 Vergleich mit Ergebnissen des Usability Tests

Es lagen Daten aus einem Usability Test der Anwendung mit 10 Teilnehmern vor. Die im Usability Test gefundenen Punkte wurden den neun UX Aspekten zugeteilt. Falls ein Punkt nicht klar zuzuordnen war, wurde er für die folgende Auswertung ignoriert.

Wie schon in Studie 1 beschrieben, wurden auch die Kommentare der Reviewer in positive, negative und neutrale Kommentare eingeteilt. Abbildung 4 zeigt die Anzahl der positiven und negativen Kommentare pro UX Aspekt, die von den vier Reviewern erfasst wurden. Abbildung 5 stellt die entsprechenden positiven und negativen Befunde aus dem Usability Test dar.

Abbildung 4 zeigt, dass fast gleichmäßig viele positive und negative Kommentare erfasst wurden, d.h. Stärken und Schwächen der Anwendung werden berücksichtigt. Für die pragmatischen UX Aspekte werden zwar geringfügig mehr Kommentare erfasst, aber die hedonischen UX Aspekte sind ebenfalls ausreichend vertreten.

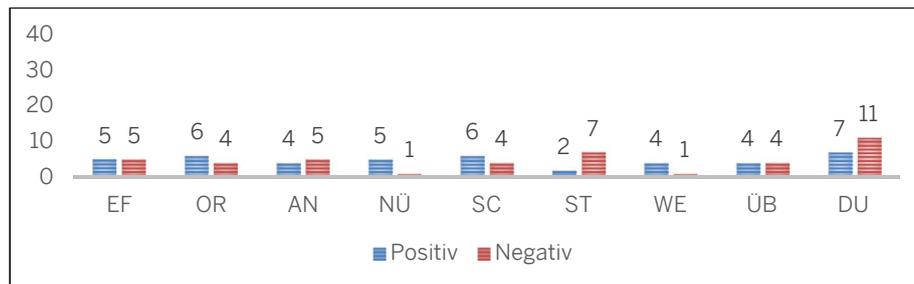


Abbildung 4: Positive und negative Kommentare pro UX Aspekt aus dem Review.

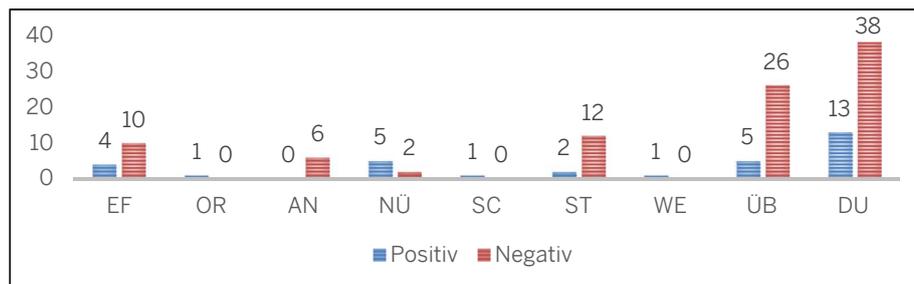


Abbildung 5: Positive und negative Befunde pro UX Aspekt aus dem Usability Test.

Die unterschiedlichen Verteilungen der Befunde auf die Dimensionen zeigen, dass eine Beurteilung mit dem UX-Fragebogen eine andere Qualität besitzt als Usability-Tests und dass beide Verfahren unterschiedliche Ergebnisse hervorbringen: Der UX-Fragebogen liefert eine umfassende Bewertung der UX eines Produkts mit positiven und negativen Eigenschaften, während Usability-Tests eher Aufschluss über konkrete Probleme mit pragmatischen Qualitäten eines Produkts aufdecken. Beide Methoden lassen sich also sehr gut kombinieren, um unterschiedliche Aspekte der User Experience eines Produkts zu untersuchen.

4.3.2 Vergleich mit Fragebögen

Für das evaluierte Produkt liegen auch Befragungen von Nutzern mit dem VISAWI (Moshagen & Thielsch, 2010) und dem UEQ (Laugwitz et al., 2006) vor. Die folgende Abbildung 6 zeigt die Bewertungen der Reviewer und die dazu passenden Skalenwerte der Fragebögen (EF, DU, ST, OR aus dem UEQ und SC aus dem VISAWI). Auch hier zeigt sich bis auf Originalität (wo die befragten Nutzer eine deutlich positivere Einschätzung abgaben als die Experten) eine überraschend gute Übereinstimmung.

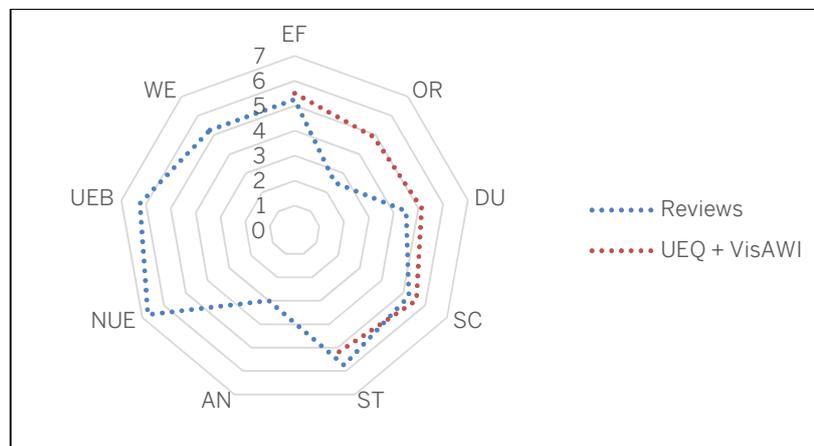


Abbildung 6: Einschätzung der Reviewer zur UX Qualität (blaue Linie) und Ergebnisse von Fragebögen (rote Linie), wobei nicht für alle UX Aspekte vergleichbare Skalen in den Fragebögen vorhanden sind.

4.3.3 Ergebnisse der Interviews

Alle Teilnehmer merken an, dass die verfügbare Zeit für das Review sehr knapp war. Sie hatten den Eindruck, dass mit mehr Zeit noch erheblich mehr Befunde anfallen würden. Nahezu alle Teilnehmer gaben an, dass die Vorgabe der Dimensionen ihnen hilft, das Produkt unter unterschiedlichen Gesichtspunkten zu bewerten und dass auch viele positive Eigenschaften des Produkts gefunden werden konnten.

Zwei Teilnehmer merken an, dass sie bei manchen Befunden nicht sicher waren, welcher Dimension diese zugeordnet werden sollen und sie dann der „passendsten“ zuordneten. Dies ist

auch zu erwarten, da die UX Aspekte nicht völlig unabhängig sind, d.h. hier durchaus semantische Überschneidungen bestehen.

Die Teilnehmer sind vom verwendeten UX-Fragebogen als Evaluationsinstrument für UX größtenteils überzeugt, vor allem in frühen Entwicklungsphasen eines Produkts sei es eine gute Methode, um ein Produkt zu verbessern. Insbesondere dann, wenn ein Usability Test wegen zu hoher Kosten oder zu großem Aufwand nicht möglich ist.

5 Zusammenfassung

Wir haben ein Verfahren zum Experten-Review vorgestellt, das sowohl pragmatische als auch hedonische UX Aspekte berücksichtigen soll. Erste Studien mit diesem Verfahren verliefen positiv. Die Reviewer waren in der Lage, auch zu hedonischen UX Aspekten sinnvolle Kommentare und Bewertungen abzugeben. Zusätzlich haben die Reviewer offenbar sowohl Schwächen als auch Stärken der Anwendungen in ihren Kommentaren erfasst, d.h. es ergibt sich eine ausgewogene Einschätzung zur UX Qualität. Auch das erste qualitative Feedback der Reviewer der zweiten Studie zum Verfahren ist sehr positiv.

Natürlich waren diese ersten Studien auf die Untersuchung des Verfahrens selbst ausgelegt. Es ist geplant dieses Verfahren in Zukunft in weiteren Projekten einzusetzen. Hier werden hoffentlich zusätzliche Daten anfallen, die helfen, die Vor- und Nachteile des Verfahrens besser einzuschätzen und die beschriebene Vorgehensweise weiter zu verbessern.

Literaturverzeichnis

- Bader, F.; Schön, E.M. & Thomaschewski, J. (2017). Heuristics considering UX and Quality Criteria for Heuristics. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol 4(6), S. 48-53.
- Bradley, R. A.; Terry, M. E. (1952): Rank Analysis of Incomplete Block Designs: I. The method of Paired Comparisons. *Biometrika*, Vol. 39, S. 324-345.
- Laugwitz, B.; Schrepp, M.; Held, T. (2006): Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: A. M. Heinecke und H. Paul (Hg.): *Mensch und Computer. Mensch und Computer im Strukturwandel*, S. 125-134.
- Moshagen, M.; Thielsch, M. T. (2010): Facets of visual aesthetics. In: *International Journal of Human-Computer Studies* 68 (10), S. 689-709.
- Nielsen, J. & Molich, R. (1990). Heuristic Evaluation of User Interfaces. *Proc. ACM CHI90 Conference*, Seattle WA, S. 249-256.
- Nielsen, J. & Mack, R.L. (1994). *Usability Inspection Methods*. New York: John Wiley & Sons.
- Schrepp, M. (2018). *User Experience mit Fragebögen messen*. Amazon CreateSpace, ISBN: 9781986843768.

- Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994). The cognitive walkthrough method. A practitioner's guide. In: Jakob Nielsen, Robert L. Mack (Hrsg.): Usability Inspection Methods. John Wiley & Sons, New York NY, S. 105–140.
- Wickelmaier, F.; Umbach, N.; Sering, K.; Choisel, S. (2012): Scaling sound quality using models for paired-comparison and ranking data. In: H. Hanselka (Hg.): Fortschritte der Akustik: DAGA 2012. Darmstadt.
- 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).

Autoren



Mayer, Dominik

Dominik Mayer studiert seit 2015 Wirtschaftsinformatik an der Dualen Hochschule Karlsruhe. Durch verschiedene Praxisphasen während des Studiums im Bereich User Research konnte er einen tiefgreifenden Einblick in die Forschung zu User Experience und Usability erlangen. Basierend darauf schrieb er seine Bachelorarbeit zum Thema „Entwicklung einer neuen Methode zur Evaluierung von Software bezüglich ihrer User Experience mit Expertenreviews“.



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. Er ist einer der Autoren des User Experience Questionnaire (UEQ) und hat Erfahrungen in der praktischen Anwendung zahlreicher anderer UX Fragebögen. Er ist auch Autor zahlreicher Beiträge zu methodischen Fragen im UX Bereich.



Held, Theo

Dr. Theo Held studierte Psychologie an der Universität Regensburg. Nach der Promotion (Universität Heidelberg) war er in Forschung und Lehre an den Universitäten Heidelberg, Graz und Halle/Saale tätig. Seine Forschungsinteressen liegen in den Bereichen Wahrnehmung und Wissensrepräsentation, sowie der Evaluation von Softwareprodukten. Seit 2001 gehört er dem User Experience Team der SAP an. Bis Ende 2010 war er für eine Reihe zentraler Designkonzepte der SAP CRMLösung verantwortlich. Seit 2011 ist er als User Experience Researcher tätig.