

User Experience mit Fragebögen evaluieren

Tipps und Tricks für Datenerhebung, Auswertung und Präsentation
der Ergebnisse

Martin Schrepp¹, Andreas Hinderks², Jörg Thomaschewski³

SAP SE¹

RMTSoft GmbH & Co. KG²

Hochschule Emden/Leer³

Zusammenfassung

Fragebögen erlauben es, die User Experience eines Produkts sehr effizient zu messen. Speziell bei Online-Fragebögen ist der notwendige Zeitaufwand für die Datenerhebung minimal, so dass sehr viele Nutzer befragt werden können. Bei der Anwendung eines Fragebogens gibt es aber auch einige Fallstricke, die den Erfolg einer Evaluation gefährden können. Wir zeigen, was man bei Datenerhebung, Auswertung und Kommunikation der Ergebnisse beachten sollte.

1 Einleitung

Fragebögen erlauben es, die User Experience eines Produkts sehr effizient quantitativ zu messen. Speziell bei Online-Fragebögen ist der notwendige Zeitaufwand für die Datenerhebung minimal, so dass sehr viele Nutzer befragt werden können.

Die Anwendung eines Fragebogens ist in der Regel nicht schwierig, da die meisten UX Fragebögen mit umfangreichem Material ausgestattet sind, z.B. einem Handbuch oder einem Tool zur Vereinfachung der Datenanalyse. Dennoch gibt es eine Reihe von Punkten, die der UX Professional beachten sollte, um Fehler bei Durchführung, Auswertung oder Präsentation der Ergebnisse zu vermeiden.

2 Ist ein Fragebogen überhaupt die richtige Wahl?

Ein Fragebogen liefert quantitative Daten zur User Experience, d.h. in der Regel eine Reihe von Skalenwerten zu bestimmten UX Dimensionen. Der UEQ (Laugwitz et al. 2006) hat z.B. die Skalen *Attraktivität*, *Effizienz*, *Durchschaubarkeit*, *Steuerbarkeit*, *Stimulation* und *Originalität*. Da numerische Skalenwerte vorliegen, kann man die Ergebnisse zu einem Produkt gut mit anderen Produkten bzw. Ergebnissen früherer Befragungen vergleichen. Folgende Fragestellungen lassen sich mit einem Fragebogen sehr gut beantworten:

- Ist die User Experience des Produkts ausreichend?
- Ist die neue Version des Produkts in Bezug auf UX besser als die alte?
- Ist das Produkt in Bezug auf UX besser als die Konkurrenzprodukte?
- Wo liegen die generellen Stärken und Schwächen des Produkts?

Ist man dagegen ausschließlich an konkreten Verbesserungsvorschlägen interessiert, d.h. am Aufspüren von Usability Problemen, sollte man einen klassischen Usability Test oder andere Methoden vorziehen.

3 Den richtigen Fragebogen auswählen

Es gibt eine Reihe von UX Fragebögen, z.B. AttrakDiff2 (Hassenzahl et al. 2003), UEQ (Laugwitz et al. 2006), VISAWI (Mooshagen & Thielsch 2010), meCUE (Minge & Riedel 2013), SUS (Brooke 1996), SUMI (Kirakowski 1996), etc. Diese messen jeweils unterschiedliche Aspekte von User Experience. Die erste Schwierigkeit für den Anwender liegt also in der Auswahl des für seine Fragestellung am besten geeigneten Fragebogens. Wir stellen die im deutschsprachigen Raum verbreitetsten Fragebögen und die von ihnen jeweils gemessenen UX-Aspekte kurz vor (siehe Tabelle 1).

In Winter et al. (2015) sind weitere, mögliche UX-Dimensionen aufgeführt und beschrieben.

Unter Umständen kann es vorkommen, dass keiner der verfügbaren Fragebögen, alle für die Evaluation eines Produkts interessanten Eigenschaften abdeckt. In diesem Fall kann man mehrere Fragebögen anwenden. Man sollte vermeiden, vorhandene Standardfragebögen zu modifizieren, z.B. in dem man einzelne Items weglässt. Das verkürzt zwar den Fragebogen, aber man zerstört die Messung der eigentlich gewünschten Dimensionen und man verliert wichtige Informationen, z.B. ist ein Vergleich mit den Benchmark-Daten dann nicht mehr sinnvoll möglich.

Dimension	UEQ	AttrakDiff2	VisAWI	meCUE	SUS	SUMI
Gesamtbeurteilung	--	--	--	--	Vollständig	Vollständig
Attraktivität	Vollständig	Vollständig	--	--	--	--
Effizienz	Vollständig	Teilweise	--	--	Teilweise	Vollständig
Durchschaubarkeit	Vollständig	Teilweise	Teilweise	Vollständig	Teilweise	Vollständig
Steuerbarkeit	Vollständig	Teilweise	--	--	--	Vollständig
Stimulation	Vollständig	Vollständig	--	--	--	--
Originalität	Vollständig	--	Vollständig	--	--	--
Identität	--	Vollständig	--	Vollständig	--	--
Schönheit	--	--	Vollständig	Vollständig	--	--
Wertigkeit	--	--	Vollständig	--	--	--
Nützlichkeit	--	--	--	Vollständig	--	Vollständig
Verbundenheit	--	--	--	Vollständig	--	--
Immersion	--	--	--	Vollständig	--	--
Emotionen	--	--	--	Vollständig	--	Vollständig

Tabelle 1: Übersicht über die Fragebögen und die darin verwendeten Dimensionen

4 Durchführung einer Fragebogen-Untersuchung

Die richtige Zielgruppe wählen

Eine 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 aber nicht immer möglich. Da die Einschätzung eines Produkts zwischen verschiedenen Personengruppen stark variieren kann, sollte man in diesem Fall die Ergebnisse nur mit Vorsicht interpretieren. Wichtig ist es, zusätzlich zum eigentlichen Fragebogen, sozio-demografische Daten zu erheben (Alter, Geschlecht etc.). Dies hilft, die Zusammensetzung der befragten Personen zumindest nachträglich auswerten zu können.

Verstehen die Teilnehmer die Items?

Grundvoraussetzung für die Datenerhebung ist es, dass die Teilnehmer die Fragen richtig verstehen. Kritisch ist, wenn einige der Teilnehmer keine Muttersprachler sind und 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). Daher sollte ein Fragebogen möglichst in der Muttersprache des Teilnehmers zur Verfügung stehen (für den UEQ existieren z.B. aktuell 13 Übersetzungen, www.ueq-online.org). Wenn man sozio-demografische Daten erhebt, dann bietet sich hier auch die Abfrage der Muttersprache an.

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.

Was ist der richtige Befragungszeitpunkt

Wird ein Fragebogen z.B. als Teil eines Usability Tests eingesetzt, sollte er direkt nach dem der Nutzer die letzte Aufgabe bearbeitet hat, ausgefüllt werden und nicht erst nachdem man mit dem Benutzer länger über den Test und vom Nutzer gefundene Probleme diskutiert hat.

Beim Einsatz mit Endbenutzern einer betriebswirtschaftlichen Software sollte man die Befragung zur User Experience erst durchzuführen, nachdem diese 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). Auch hier bietet es sich an, auch die Nutzungsdauer zu erheben.

5 Auswertung der Daten

Fehlerhafte Daten finden und eliminieren

Vor der Auswertung muss sichergestellt sein, 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. Dadurch ist ein Konsistenzcheck der Antworten möglich, der im Auswertungs-Excel auch zur Verfügung steht.

Wurden Items fehlinterpretiert?

Weichen die Mittelwerte der Items einer Dimension stark voneinander ab, kann das auf eine Fehlinterpretation einzelner Items hinweisen. Der UEQ für Jugendliche enthält z.B. ein Item *zeitsparend – zeitraubend*. Dieses soll die Effizienz der Nutzerinteraktion messen. In einer Untersuchung eines sozialen Netzwerks wurde dieses Item aber offenbar von vielen Nutzern als „das soziale Netzwerk raubt mir meine Zeit“ interpretiert. Solche Fehlinterpretationen kann man über die Item-Mittelwerte und Standardabweichungen entdecken. Ein weiterer Indikator für Fehlinterpretationen ist der Cronbach-Alpha Koeffizient. Dieser misst, wie konsistent die Items einer Skala beantwortet wurden. D.h. niedrige Werte dieses Koeffizienten deuten ebenfalls auf abweichende Interpretationen einiger Items durch eine größere Menge von Teilnehmern hin. Solche Fehlinterpretationen lassen sich nicht immer vermeiden, sind jedoch unkritisch, da man dies in den Daten gut erkennen kann und die Werte der betroffenen Dimension dann nur vorsichtig oder gar nicht interpretiert.

Überinterpretation der Ergebnisse vermeiden

Hat man nur eine kleine Stichprobe befragt oder gibt es bei den befragten Personen eine große Uneinigkeit bzgl. der Einschätzung des Produkts, so sind die Konfidenzintervalle der

Dimensionen in der Regel groß. Das Konfidenzintervall ist der Bereich, der den wahren Wert mit 95% Wahrscheinlichkeit enthält. D.h. in einem solchen Fall ist der gemessene Wert offenbar sehr ungenau und sollte daher vorsichtig interpretiert werden.

Kann man Dimensionswerte zu einem Gesamt-Score verrechnen?

Oft besteht der Wunsch aus dem Management eine einzelne Kennzahl für die User Experience bereit zu stellen. Einige Fragebögen liefern auch einen solchen Gesamtscore, andere nur Werte pro Dimension. 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 wird aber in konkreten Projekten in der Regel nicht zutreffen. 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.

Verwenden des Benchmarks

Hat man ein Produkt erstmals mit einem Fragebogen evaluiert, hat man in der Regel gewisse Probleme die Skalenmittelwerte zu interpretieren. Ist eine 1,5 auf der Skala Durchschaubarkeit ein gutes oder schlechtes Ergebnis? Diese Frage kann man nur dann sinnvoll beantworten, wenn man das Ergebnis mit den Ergebnissen anderer Produkte vergleicht. Das ist das Ziel eines Benchmarks. Die meisten standardisierten Fragebögen, z.B. UEQ, SUS, SUMI, oder VISAWI, stellen einen solchen Benchmark zur Verfügung. Hier kann man feststellen, wie gut das evaluierte Produkt im Vergleich mit anderen Produkten abgeschnitten hat.

Sind Unterschiede wirklich bedeutsam?

Will man zwei Produkte vergleichen (z.B. eine neue Version mit der alten oder ein Produkt mit einem Konkurrenzprodukt) so stellt sich die Frage nach der Bedeutsamkeit (Signifikanz) von gefundenen Unterschieden in den Werten der Dimensionen. Nehmen wir an, wir haben für eine neue Version des Produkts den Wert 1,8 auf der UEQ Skala Effizienz gemessen. Für die alte Version wurde 1,5 ermittelt. Natürlich sind nicht alle Personen bzgl. der Effizienz des Produkts einer Meinung. D.h. solch ein Unterschied kann entweder durch eine wirkliche Verbesserung des Produkts oder durch etwas Glück bei der zufälligen Auswahl der jeweils befragten Personen. Um diese Frage zu klären, ist es daher notwendig beide Ergebnisse über einen Signifikanztest zu vergleichen. Dies kann z.B. leicht über eine Statistiksoftware (auch Excel bietet solche Funktionen zur Datenanalyse) geschehen oder beim UEQ über ein bereits vorbereitetes Datenanalysetool (in Excel), in dem man nur noch die Werte der beiden Befragungen eingeben muss.

6 Darstellung der Ergebnisse

Hat man die Datenerhebung und die Auswertung der Daten abgeschlossen, so fehlt als letzter Schritt noch die geeignete Präsentation der Ergebnisse. Typische Adressaten einer solchen Präsentation sind Produkt-Owner, Manager oder Mitglieder des Entwicklungsteams. Diese Personen haben oft wenig oder keine Erfahrung mit dem Einsatz und der Interpretation von

Fragebögen. Wichtig ist es hier erst mal Vertrauen in die Methode zu schaffen. Man kann z.B. zu Beginn der Präsentation erwähnen, dass es sich um einen etablierten wissenschaftlich fundierten Fragebogen handelt und evtl. die entsprechende Publikation zitieren.

Ein zentraler Teil der Ergebnispräsentation sind die gemessenen Mittelwerte der Dimensionen. Hier ist es wichtig, dass die Teilnehmer die Bedeutung der Dimensionen verstehen (die reinen Dimensionsnamen werden oft missverstanden). D.h. für jede Skala sollte kurz erklärt werden, was damit inhaltlich gemeint ist. Falls man den Fragebogen nach einem Usability Test verwendet, bietet es sich auch an gefundene Probleme mit den Werten der Dimensionen in Beziehung zu setzen (z.B. viele Probleme mit der Terminologie und ein geringer Wert der UEQ Dimension Durchschaubarkeit). Wichtig ist auch die Werte der Dimensionen nicht alleine stehen zu lassen. Hat man keinen Vergleichswert (z.B. das Ergebnis aus dem Test der Vorversion), so sollte man immer einen Vergleich mit dem Benchmark heranziehen.

Kontaktinformationen

Martin Schrepp: martin.schrepp@sap.com, Andreas Hinderks: andreas@hinderks.org, Jörg Thomaschewski: joerg.thomaschewski@hs-enden-leer.de

Literaturverzeichnis

- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189, S. 4-7.
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Szwillus, G. & Ziegler, J. (Hrsg.): *Mensch & Computer 2003*. S. 187-196. Vieweg+ Teubner Verlag.
- 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.; Lehmann, A.; Petrovic, K.; Schroeder, M. (Eds.); *Usability Professionals 2012* S. 78 – 83, 2012.
- Kirakowski, J. (1996). The software usability measurement inventory: background and usage. *Usability evaluation in industry*, 169-178.
- Laugwitz, B.; Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: A.M. Heinecke & H. Paul (Eds.): *Mensch & Computer 2006 – Mensch und Computer im Strukturwandel*. Oldenbourg Verlag, S. 125 – 134.
- 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.
- Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68 (10), S. 689-709.
- 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..
- Winter, D.; Schrepp, M. & Thomaschewski, J. (2015). Faktoren der User Experience - Systematische Übersicht über produktrelevante UX-Qualitätsaspekte. In: Endmann, A.; Fischer, H. & Krökel, M. (Eds.), *Mensch und Computer 2015 – Usability Professionals*, S. 33-41, DE GRUYTER 2015.

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.

**Hinderks, Andreas**

Andreas Hinderks studierte Informatik an der Hochschule Emden/Leer. Danach war er geschäftsführender Gesellschafter der ebis GmbH, die ein branchenspezifisches Warenwirtschaftssystem für den Kücheneinzelhandel entwickelt hat. Seit 2010 ist er als Produkt Manager bei der RMT Soft GmbH & Co. KG beschäftigt. Dort ist er verantwortlich für die benutzerzentrierte und innovative Gestaltung von Geschäftsanwendungen mit Schwerpunkt Prozesssteuerung. Berufsbegleitend studiert er im Masterstudiengang Medieninformatik an der Hochschule Emden/Leer und engagiert sich in der Forschungsgruppe HS Emden/Leer im Bereich der Usability und User Experience.

**Thomaschewski, Jörg**

Dr. 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, Markup Languages, 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.