

Wohin mit der Toolbar?

Eine unmoderierte Online-Studie

Bernard Rummel, Christopher Lafleur

Global Design Enablement, SAP SE

Zusammenfassung

Betriebswirtschaftliche Software muss heute auf verschiedenen Geräteklassen laufen; Layouts werden mit Responsive Design an verschiedene Bildschirmgrößen angepasst. Dabei erwarten Benutzer eine gewisse Konsistenz zwischen Desktop und Mobilgerät. Die bei Mobilgeräten gängige Position der Anwendungstoolbar am Fuß des Bildschirms erweist sich dabei auf großen Monitoren als problematisch. In einer unmoderierten Online-Studie untersuchten wir drei verschiedene Designvarianten einer Anwendungstoolbar im Hinblick auf Fehlerhäufigkeit, zeitliche Effizienz, Benutzerpräferenz und – Zufriedenheit auf Desktop-Bildschirmen. Der Beitrag stellt das Studiendesign, ausgewählte Ergebnisse sowie Erfahrungen mit der Testplattform Userzoom vor.

1 Einleitung und Problemstellung

Für eine einheitliche Positionierung der Anwendungstoolbar in einer Suite von betriebswirtschaftlichen Anwendungen gibt es nur endlich viele Optionen. Die Toolbarposition soll aufgabengerecht, aber auch zwischen verschiedenen Anwendungen konsistent platziert sein. Soll die Anwendung auch auf Mobilgeräten laufen, liegt eine Position im Fußbereich des Bildschirms nahe, unter anderem weil der Bildschirmkopf oft bereits mit anderen Funktionen belegt ist (Identifikation des Geschäftsobjekts, Navigation, generische Funktionen). Auf großen Monitoren ist diese Position jedoch wenig erwartungskonform. In Tests hatte sich herausgestellt, dass eine Toolbar-Positionierung am Bildschirmfuß auf großen Monitoren problematisch ist. Testteilnehmer verwendeten teils erhebliche Zeit auf das Finden von Buttons, die auf Mobilgeräten und Tablets mühelos gefunden wurden.

Als Lösung wurde eine animierte Toolbar vorgeschlagen, die erscheint, sobald die Anwendung in einen Zustand kommt, in dem die Toolbar sinnvoll wird. Durch eine Animation, die auch im peripheren Gesichtsfeld wahrgenommen wird, ist das Erscheinen der Toolbar gut zu erkennen. Zu klären waren die Position und der genaue Inhalt der Toolbar.

Der vorliegende Beitrag beschreibt eine unmoderierte Online-Studie, die wir im Rahmen der Untersuchungen zu dieser Problemstellung durchführten. Das vorrangige Ziel des Beitrags ist die Beschreibung des Ansatzes sowie unserer Erfahrungen mit der Testplattform Userzoom. Ergebnisse, die aus unserer Sicht von allgemeinem Interesse und auf andere Designsysteme übertragbar sind, werden ausgewählt dargestellt.

2 Vorgehensweise

Als Testscenario wählten wir eine fiktive Intranet-Bestellanwendung, die browsergestützt über ein Online-Userpanel verteilt werden konnte. Ein bereits angelegter Einkaufskorb sollte bearbeitet und zum Einkauf abgeschickt werden. Die Toolbar war als eine floating toolbar (d.h. scroll-unabhängig) ausgelegt und erschien jeweils, sobald der Einkaufskorb in Bearbeitungsmodus geschaltet war. Drei Positionsvarianten wurden untersucht (Abb. 1):

- 1) Bildschirmfuß, Abschicken-Button (*Submit*) außerhalb der Toolbar am Bildschirmkopf.
- 2) Bildschirmfuß, Abschicken-Button erscheint innerhalb der Toolbar an der Stelle des Sichern-Buttons (*Save*), nachdem dieser geklickt wurde.
- 3) Toolbar wie 2) am unteren Ende des Bildschirmkopfes etwa im oberen Viertel des Bildschirms (der Bildschirmkopf selbst war mit identifizierenden Informationen zum Einkaufskorb sowie generischen Objektfunktionen wie „Bearbeiten“ belegt).

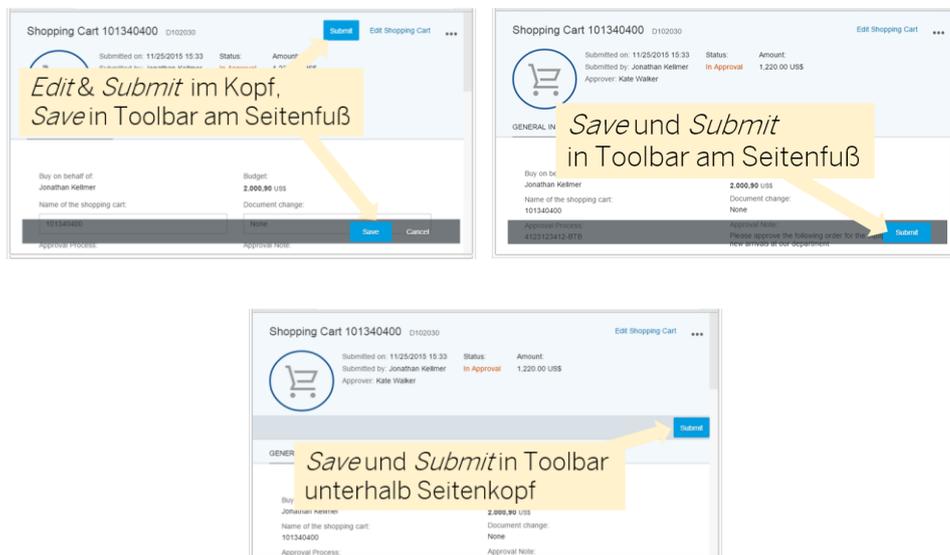


Abbildung 1: Positionsvarianten der animierten Toolbar (Bildschirmfenster verkleinert)

Für jede Variante bestanden Hypothesen hinsichtlich zeitlicher Effizienz, Fehlerquote sowie Benutzerpräferenz und -Zufriedenstellung.

Da sich durch die verschiedenen Button-Positionen unterschiedliche Cursorwege ergeben (Fitts' Law, Fitts 1954), entwickelten wir drei Aufgaben, die verschiedene Bereiche der Benutzungsoberfläche involvierten, und zwar im oberen, unteren sichtbaren und unteren nicht sichtbaren Bereich (Scrollen erforderlich). Die Aufgaben umfassten das Löschen einer Bestellposition im unteren sichtbaren Screenbereich, das Löschen eines Anhangdokuments im unteren nicht sichtbaren Screenbereich sowie das Ändern eines Feldinhaltes im oberen Screenbereich. Damit ergab sich ein Versuchsplan mit 3 Designvarianten x 3 Aufgaben = 9 Versuchsdurchgängen, wobei Reihenfolgeneffekte durch Randomisierung auszugleichen waren. Um die Testteilnehmer nicht mit ständig wechselnden Designvarianten zu verwirren, wurden die drei Aufgaben in jeder Designvariante nacheinander randomisiert dargeboten und die Reihenfolge der Designvarianten blockweise randomisiert.

Da die zu erwartenden Effekte bezüglich zeitlicher Unterschiede und Fehlerquoten eher gering waren, brauchten wir eine relativ hohe Anzahl an Testteilnehmern. Wir entschieden uns daher an Stelle einer Laborstudie für eine unmoderierte Online-Studie, mit der in relativ kurzer Zeit Daten in einer Größenordnung von etwa 100 Teilnehmern erhoben werden konnten. Mit der Testplattform Userzoom (<http://www.userzoom.com/>) war unser Studiendesign dergestalt umsetzbar, dass jeder Teilnehmer alle Versuchsbedingungen mit jeweils vollständiger Randomisierung von Designvarianten und Aufgaben durcharbeiten konnte. Die Plattform erlaubte auch das Tracken der uns interessierenden Interaktionsereignisse sowie Videomitschnitte der Aufgaben.

Zur Erfassung der Benutzerpräferenz zwischen den drei Designvarianten entschieden wir uns nach einigen Iterationen in Pilottests für ein forced ranking-Verfahren. Alternativen wie direktes Rating oder Paarvergleich erschienen uns aufgrund der bekanntermaßen eher geringen Motivation der Teilnehmer von Online-Studien vergleichsweise riskant. Userzoom bietet ein entsprechendes Frageformat an. Eine Schwierigkeit ergab sich bei der Identifikation der Designvarianten für die Testteilnehmer. Da die Varianten sich visuell sehr ähnelten – die Hauptunterschiede lagen ja im dynamischen Verhalten – entschieden wir uns nach einigen Iterationen dafür, die Designvarianten durch kurze Videos zu identifizieren und jeweils eindeutigen Screenshots zuzuweisen (etwa wie Abb. 1), die die Teilnehmer letztendlich in eine Reihenfolge bringen konnten.

Zuletzt wurden die Teilnehmer zu jeder Variante einzeln abschließend befragt, und zwar mit der Single Ease Question (SEQ, Sauro & Dumas 2009, Sauro 2012) sowie einem Freitextfeld für Kommentare.

3 Erfahrungen mit der Plattform Userzoom

Das Aufsetzen des Tests mit Userzoom war nicht zuletzt durch die Hilfe eines uns zugewiesenen Projektcoaches unproblematisch. Die Studien-Designoberfläche ist klar strukturiert. Die Anforderungen an den Studiendesigner sind eher fachlicher Art: da die

gesamte Studie selbsterklärend sein muss, bestehen hohe Anforderungen an Erklärungs- und Aufgabentexte. Weiterhin muss man darauf achten, auch scheinbar exotische Fehlerfälle abzufangen, wie etwa das unvollständige oder zu langsame Laden des Prototypen sowie die verschiedenen Formen des Scheiterns an einer Aufgabe. Insgesamt verwendeten wir 6 Pilotläufe bis zur endgültigen Version der Studie und anschließend einen „soft launch“ an einen zunächst begrenzten Teilnehmerkreis.

Die verwendeten Prototypen wurden passwortgeschützt auf einem Server gehalten; die Authentifizierung der Testteilnehmer erfolgte über das Userzoom-Framework. Teilnehmer hatten im Rekrutierungsprozess ein Non-Disclosure Agreement sowie die notwendigen Einverständniserklärungen unterzeichnet.

Die Rekrutierung erfolgte über Kooperationspartner von Userzoom in den USA. Nachdem wir die erwartete Sitzungsdauer von 20 Minuten auf (realistische) 15 Minuten heruntersetzt hatten, erreichten wir in drei Einladungswellen die gewünschte Größenordnung von 100 Teilnehmern innerhalb einer Woche. Insgesamt wurden 2048 Kandidaten angeschrieben, von denen 735 Interesse an der Studie zeigten. 550 wurden über den Eingangsfragebogen aussortiert, da sie z.B. keine Geschäftseinkäufe im Intranet tätigten oder in der IT-Branche arbeiteten. 127 Personen begannen die Studie (wegen weiterer Ausfallgründe sind die Zahlen nicht additiv).

Die von diesen Personen gesammelten Daten wurden eingehend inspiziert und bereinigt. „Professionelle Tester“ waren daran zu erkennen, dass sie parallel zu unserer Studie URLs anderer Testframeworks ansurften. Auffällig schnelle Teilnehmer waren mithilfe von Probability Plots (Rummel, 2014; Abbildung 2) der Studiendauer leicht zu identifizieren. Die Verifikation dieser Ausreißer erfolgte über Inspektion von Bildschirm-Videomitschnitten im Verdachtsfall. Insgesamt wurden 34 Personen mit dieser Vorgehensweise ausgeschlossen.

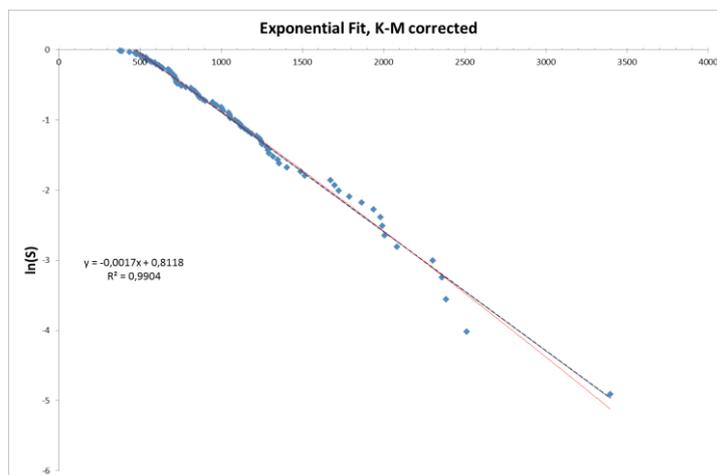


Abbildung 2: Probability Plot (Rummel, 2014) der Studiendauer. Die Studiendauern der Teilnehmer folgen einer Exponentialverteilung. Auffällig schnelle Teilnehmer sind daran erkennbar, dass sie im oberen Teil deutlich von der Regressionsgeraden abweichen.

Die Daten von 89 Personen wurden schließlich weiter ausgewertet, wobei es in einzelnen Versuchsbedingungen zu weiteren Verlusten durch unvollständig getrackte Sessions kam. Userzoom bietet verschiedene Möglichkeiten an, Benutzeraktionen zu tracken. In der von uns verwendeten Variante wurde auf bestimmte definierte Benutzeraktionen (Clicks) hin per Javascript die URL geändert und diese Änderung vom Framework getrackt. Trotz Einschränkung auf Google Chrome fanden wir einige Fälle, in denen Benutzeraktionen auf dem Videomitschnitt zu sehen waren, die aber nicht getrackt wurden. Das Problem wurde vom Userzoom Support aufgenommen, konnte aber im Studienzeitraum nicht behoben werden. Letztlich verblieben je nach Versuchsbedingung zwischen 73 und 68 auswertbare Datensätze.

Die von Userzoom angebotenen Standardauswertungen, die durchaus umfangreich sind, erwiesen sich als für uns nicht ausreichend. Die Standard-Zeitmessung erfolgt vom Aufgabenstart zum (vom Teilnehmer) selbsterklärten Erfolg oder Abbruch. Uns interessierte das Zeitintervall vom Einschalten des Bearbeitungsmodus (*Edit*) bis zum Abschicken des Einkaufskorbs (*Submit*). Die Validierung des Aufgabenerfolgs kann über das Erreichen einer bestimmten Ziel-URL sowie Kontrollfragen erfolgen. Wir benötigten dagegen eine Validierung gegen einen definierten Klickpfad, da Teilnehmer die gefragten Aufgabenaktionen auch durchaus überspringen konnten. Erfolgsquoten, Fehlertypen und Bearbeitungszeiten mussten also direkt aus dem Clickstream ermittelt werden. Das Herunterladen und Bearbeiten des Clickstreams in Excel war unproblematisch. Mittels Formatierungsregeln wurden die relevanten Clickstrecken farblich hervorgehoben und konnten innerhalb zweier Personentage ausgewertet werden.

4 Detailauswertung und ausgewählte Ergebnisse

Abbildung 3 zeigt die Erfolgsraten der Teilnehmer in den drei Aufgaben und den drei Varianten. Während die prozentualen Erfolgsquoten sich nicht signifikant unterscheiden und um 80% im erwarteten Bereich liegen, fällt ein erheblicher Abfall der Anzahl klassifizierbarer Teilnehmer in den Aufgaben 2 und 3 auf. In diesen Aufgaben mussten die Teilnehmer scrollen oder Eingaben machen, was bis zur Hälfte der Teilnehmer unterließen. Da ohne Detailkontrolle der Videomitschnitte nicht entscheidbar war, ob hier ein Trackingfehler oder Faulheit zugrunde lag, wurden die entsprechenden Teilnehmer für die jeweilige Versuchsbedingung ausgeschlossen.

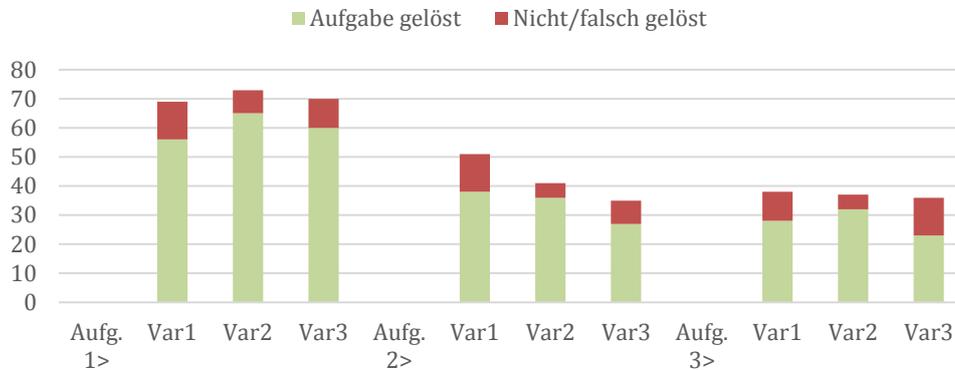


Abbildung 3: Anzahl klassifizierbarer Teilnehmer, die die jeweilige Aufgabe lösten bzw. nicht lösten.

Userzoom bietet eine Aufschlüsselung der verschiedenen Klickpfade an, die Testteilnehmer verfolgten. Daraus lässt sich ermitteln, an welchen Stellen im Benutzungsprozess welcher Anteil an Benutzern „verlorengeht“. Abbildung 4 zeigt das Ergebnis dieser Aufschlüsselung beispielhaft für Aufgabe 1 (graphische Aufbereitung von uns). Da unser Prototyp etwa 10s zum Laden benötigte, verließen 7-9% der Teilnehmer in der Zwischenzeit die Studie zugunsten für sie interessanterer Sites. Weitere fanden bzw. clickten den *Bearbeiten*-Button (Edit) nicht, unterließen die gefragte Aufgaben-Aktion (Task) – hier, das Löschen einer Bestellposition, etc.

Auffällig ist, dass 15% in Variante 3 nach dem Löschen der Bestellposition nicht auf den *Sichern*-Button clickten, vermutlich weil sie annahmen, dass das Sichern im Löschvorgang bereits enthalten war. In den beiden anderen Varianten trat das Problem seltener auf, da hier der *Sichern*-Button in unmittelbarer Nachbarschaft der Positionsliste erschien.

Eine weitere Auffälligkeit sind die hohen Verluste zwischen Sichern und Abschicken des Einkaufskorbes. In allen Designvarianten gab es Teilnehmer, die nach dem Sichern den Einkaufskorb nicht abschickten. Die größten Verluste gab es in Variante 1, wo der *Sichern*-Button in der Toolbar am Bildschirmfuß, der *Abschicken*-Button jedoch am Bildschirmkopf erschien.

Letztlich gab es noch einzelne Teilnehmer, die zwar den Einkaufskorb erfolgreich abschickten, jedoch vergaßen, diesen Erfolg im userzoom-Framework als solchen zu deklarieren. Insgesamt hat Variante 2 die geringsten Ausfallquoten.

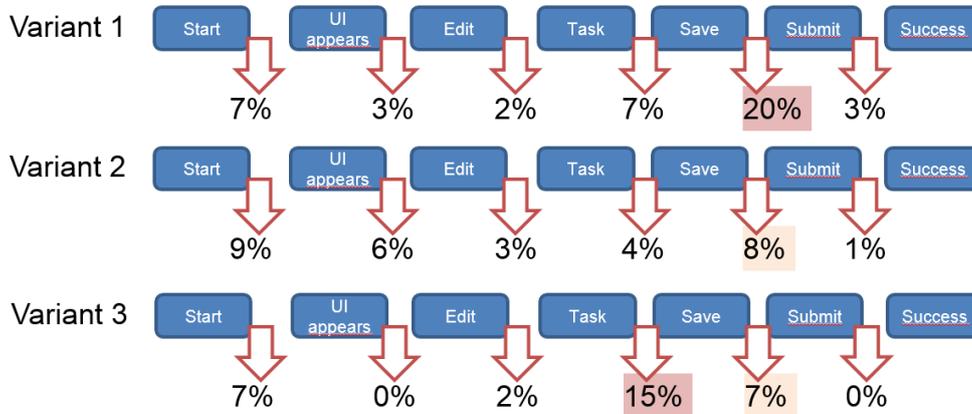


Abbildung 4: Aufschlüsselung der verschiedenen Klickpfade, die Testteilnehmer verfolgten, am Beispiel von Aufgabe 1. Erläuterungen im Text.

Abbildung 5 zeigt Probability Plots (Rummel, 2014) der Bearbeitungszeiten. Die horizontale Achse stellt jeweils die Bearbeitungszeit dar. Die vertikale Achse zeigt in logarithmischer Skalierung die Survival-Funktion, d.h. den prozentualen Anteil Teilnehmer, die zu der gegebenen Zeit noch an der Aufgabe arbeiten. Alle Bearbeitungszeiten verhalten sich näherungsweise exponentialverteilt, was an der geradlinigen Anordnung der Datenpunkte zu sehen ist. Die Regressionsgeraden schneiden die Zeitachse zu näherungsweise gleichen Zeitpunkten, d.h. alle Bearbeitungszeiten sind um einen etwa gleichen, konstanten Betrag verschoben. Diese Konstante entspricht der Zeit, die in etwa zum Durchklicken der Aufgabe erforderlich ist. Dass diese Zeiten sich praktisch nicht unterscheiden, deutet darauf hin, dass die unterschiedlichen Cursorwege in den hier untersuchten Designvarianten keine Rolle spielten. Deutlich unterschiedlich sind jedoch die Steigungen der Regressionsgeraden in den interaktionsintensiveren Aufgaben 2 und 3. Insbesondere Variante 3 hat eine deutlich höhere „Halbwertszeit“ des exponentialverteilten Prozessanteils, d.h. in dieser Variante verbringen die Teilnehmer mehr Zeit mit dem Suchen der Toolbar-Buttons.

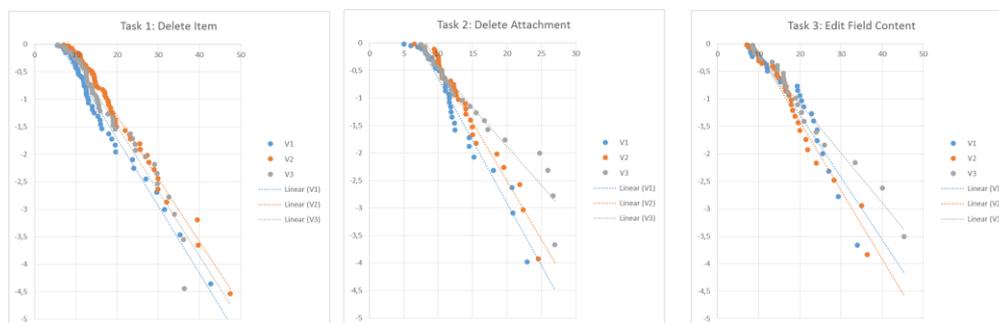


Abbildung 5: Probability Plots (Rummel, 2014) der Bearbeitungszeiten in den drei Aufgaben. Erläuterungen im Text

Abbildung 6 zeigt das Ergebnis der forced-ranking-Einstufung der drei Designvarianten. Variante 2 wurde am häufigsten (35 mal) an Position 1 und am seltensten (19 mal) an Position 3 eingestuft. Dies entspricht dem Ergebnis der direkten Einstufung mittels des SEQ-Fragebogens, in der Variante 2 mit 5,9 den besten, Variante 1 mit 5,2 den schlechtesten Score erzielte ($p < 0.01$, andere Vergleiche n.s.).



Abbildung 6: Ergebnis des forced ranking-Vergleichs der drei Designvarianten.

5 Diskussion

Unmoderierte Online-Studien werden als zeitgemäße, zeitsparende und ressourcenschonende Form des Usability Testing dargestellt (Albert, Tullis & Tedesco, 2010). Die vorliegende Studie wäre tatsächlich als klassische Laborstudie nicht oder nur zu hohen Kosten darstellbar gewesen – nicht nur wegen der geforderten hohen Teilnehmerzahl aus einem betriebswirtschaftlichen Profil, sondern auch wegen der insgesamt kurzen Sitzungsdauer, für die eine Anreise in ein Labor nicht sinnvoll wäre.

Der Aufwand zur Erstellung, Durchführung und Auswertung der Studie ist jedoch mit einer klassischen Laborstudie absolut vergleichbar, wenn nicht höher. Allein für Pilotstudien zur Klärung und Absicherung des Studienablaufs verwendeten wir etwa so viele Teilnehmer wie für ein gängiges formatives Testformat. Eine weitere Herausforderung besteht in der erwartungsgemäß geringen Motivation der Teilnehmer, die sich in erheblichem Schwund von der Ansprache bis zum auswertbaren Datensatz niederschlägt. Aufbereitung und Qualitätssicherung der Daten sind entsprechend aufwändig. Der Funktionsumfang der Plattform Userzoom war für diesen Aufwand notwendig, aber nicht hinreichend; zusätzlich fielen anspruchsvolle und aufwändige ad-hoc-Auswertungsaufgaben an.

Dem gegenüber steht eine Ergebnisqualität, die vor allem für quantitative Fragestellungen beeindruckend ist. Durch das weitgehende Fehlen von Beobachtungsmöglichkeiten (ca. 900 Videomitschnitte kann man sich nicht ansehen), dafür aber reichhaltige Verhaltens- und Befragungsdaten ergeben sich interessante Untersuchungsansätze, die fortgeschrittene statistische Methoden erfordern und teils überhaupt erst interessant machen.

Als Usability Professional bzw. Team ist man damit auf mehreren Ebenen gefordert: neben dem klassischen Anwendungs – und Testdesign ist eine selbsterklärende Testdarbietung, entsprechende Studien- und Rekrutierungskonzeption, Auswertungsplanung und letztlich - Durchführung zu leisten. Effizienzgewinne ergeben sich erst bei iterativem Vorgehen, z.B. Retest weiterentwickelter Varianten anhand desselben Testplans.

Hinsichtlich der Studienergebnisse ist überraschend, dass der Beitrag unterschiedlicher Cursorwege zur Gebrauchstauglichkeit einer Toolbar offenbar vernachlässigbar ist. Entscheidend ist, dass sie schnell gefunden wird und die Buttons enthält, die vom Benutzer erwartet werden.

Danksagung

Wir danken Roman Rommel (SAP SE) für die Erstellung der verwendeten Prototypen und Frau Kathrin Holliger von Userzoom für ihr geduldiges und effektives Coaching während des Projekts.

Literaturverzeichnis

- Albert, B., Tullis, T., Tedesco, D. (2010): *Beyond the Usability Lab*. Burlington, MA: Elsevier
- Fitts, Paul M. (1954): The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. In *Journal of Experimental Psychology*, 47 (0) pp. 381-391.
- Rummel, B. (2014): Probability Plotting: A Tool for Analyzing Task Completion Times. *Journal of Usability Studies*, August 2014, pp. 152-172. Zugriff Juni 2016 auf <http://uxpajournal.org/probability-plotting-a-tool-for-analyzing-task-completion-times-2/>
- Userzoom (2015): <http://www.userzoom.com/>. Letzter Zugriff Mai 2016
- Sauro, J. & Dumas, J.S. (2009): Comparison of Three One-Question, Post-Task Usability Questionnaires. In *CHI 2009*, April 4–9, 2009, Boston, Massachusetts, USA.
- Sauro, J. (2012): *10 Things To Know About The Single Ease Question (SEQ)*. <http://www.measuringu.com/blog/seq10.php>; Zugriff Mai 2016

Autoren



Rummel, Bernard

Bernard Rummel arbeitet seit mehr als 20 Jahren im Arbeitsfeld Ergonomie, Gebrauchstauglichkeit und Interaktionsdesign. Seit 2011 beschäftigt er sich bei SAP mit der Quantifizierung von Gebrauchstauglichkeit, UX Benchmarking und Web Analytics.



Lafleur, Christopher

Chris Lafleur has more than 6 years' experience working in the areas of usability and cognitive engineering. He has worked since 2010 at SAP as a User Researcher, responsible for both formative and summative research topics, along with the development of new research methods, which bring together both applied and experimental research methodologies. Furthermore, he has written on and been published on the topic of visual attention and its applications to applied usability.