

Instant Usability Report: Nutzertests mit Ergebnisdoku in 8 Stunden

Ron Reckin¹, Vera Burckhardt²

Kognitionspsychologie und Kognitive Ergonomie, Technische Universität Berlin¹
Pidoco GmbH²

Zusammenfassung

Herkömmliche Methoden für Durchführung und Auswertung von Usabilitytests mit Nutzern erfüllen in vielen Aspekten die Anforderungen (agiler) Entwicklungspraxis von Software-KMU nicht. Basierend auf Praxiserfahrungen eigener Usability-Projekte und in Studien berichteter Anforderungen wurde deshalb nutzerzentriert die praxisgerechte Methode „Instant Usability Report“ (IUR) zur Testung der Usability von Workflows entwickelt und evaluiert. Innerhalb eines Arbeitstages erfolgen Vorbereitung, Durchführung und Aufbereitung der Ergebnisse. Praxistauglichkeit und Akzeptanz der Methode wurde in einer aufwendigen Studie mit 8 Unternehmen evaluiert. Die Durchführung war für die teilnehmenden Unternehmensvertreter sehr zufriedenstellend, zudem bestätigten sie einen hohen Nutzen sowie eine gute Durchführbarkeit in ihrer Unternehmenspraxis. Die vorliegende Methodenbeschreibung soll jeden Produktverantwortlichen befähigen, die Durchführung zu organisieren bzw. selbst zu leiten. (Ausführliche Methodenbeschreibung ist bereitgestellt auf <http://instant-usability-report.blogspot.de>)

1 Anforderungen an agile Usabilitymethoden

Die Integration von Usabilitymaßnahmen in den Entwicklungsprozess wird häufig als zusätzlicher Kostenfaktor empfunden, der die Entwicklung verzögert (Trappe et al., 2015; Trauzettel et al., 2015; Woywode et al., 2012). In vielen Unternehmen wird nur wenig oder gar kein Budget für Usability bereitgestellt (Weber, 2015; Trauzettel et al., 2015; Rößler & Kauer, 2015) und Usabilityaufgaben werden von Entwicklern oder Designern statt von Usabilityexperten übernommen (Woywode et al., 2012). Usabilitymethoden werden häufig unsystematisch oder sogar falsch angewendet und es wird nicht genügend Zeit eingeplant, um die Erkenntnisse aus Testungen einzuarbeiten (Weber, 2015). Besonders für agile Arbeitsprozesse sind schlanke Usabilitymethoden gefragt, die einfach umsetzbar sind, so dass sie

in den modernen Entwicklungsprozess integriert werden können ohne diesen zu verlangsamen oder den Kostenrahmen zu sprengen. Das ist vor allem für kleine und mittelständige Unternehmen (KMU) von großer Bedeutung, da diesen in der Regel nicht ausreichend Ressourcen für Usabilitymaßnahmen zur Verfügung stehen (Woywode et al., 2012; Reckin & Brandenburg, 2013). Einige Studien haben gezeigt, dass die Bedeutung von Usability zwar in Unternehmen anerkannt, die tatsächliche Integration aber insbesondere in KMU eher schlecht ist (Woywode et al., 2012; Stickel et al., 2015). Besonders bei agilen Entwicklungsprozessen wie Scrum oder Kanban, mit schneller Konzeption und Implementierung einzelner Programmteile in kurzen Sprints von teilweise nur ein bis zwei Wochen, ist die Integration von Usabilitymethoden nur schwer möglich. Um also Usabilitymethoden in agile Entwicklungsprozesse zu integrieren, müssen sie schnell genug durchführbar sein, um die Lieferung von funktionaler Software bis zum Ende des Sprints gewährleisten zu können. Die Umsetzung von Ansätzen, die strukturelle Anpassungen des Entwicklungsprozesses erfordern (z.B. da Silva et al., 2012; Singh, 2008), scheitert bisher an ungenügender Anerkennung der Bedeutsamkeit von Usability im Management (Woywode et al., 2012).

In vielen Unternehmen werden Usabilitymaßnahmen von Entwicklern oder Designern übernommen, denen die Expertise fehlt, um für eine bestimmte Fragestellung die richtige Methode auszuwählen und die Ergebnisse korrekt zu interpretieren (Stickel et al., 2015; Rößler & Kauer, 2015). Vorhandene Methodenbeschreibungen fokussieren vor allem auf deren Durchführung und beinhalten nur selten die Testvorbereitungen wie beispielsweise die Zielstellung, Aufgabenauswahl oder geeignete Testumgebungen. Besonders die Ergebnisberichte werden als zeitaufwändig empfunden. Eine Erstelldauer von bis zu einer Woche (Rösler & Thölke, 2015) bedeutet eine Verzögerung des Entwicklungsprozesses oder provoziert dessen Fortführung ohne Berücksichtigung dieser Ergebnisse. Zusätzlich sind die Entwicklerteams häufig von der großen Anzahl der berichteten Usabilityprobleme und Ergebnisse überfordert (Weber, 2015). Um Zeit zu sparen, können je nach Testsituation auch sogenannte Quickfindings-Berichte ohne formelle Beschreibungen des Testablaufs und aufwendige Ergebnispräsentationen erstellt werden (Nielsen, 2005). Ergebnisse werden hier einfach direkt im Anschluss an die Testung diskutiert und nur kurz in einem Bericht zusammengefasst. Sie eignen sich vor allem, wenn mehrere Tests sukzessive durchgeführt werden oder die Ergebnisse nicht länger aufbewahrt werden sollen.

Rösler & Thölke (2015) haben eine Testmethode für Testungen innerhalb agiler Entwicklungsprozesse entwickelt. Mit ihrer sogenannten agile testing Methode können Testergebnisse innerhalb von 1-2 Tagen zu Verfügung gestellt werden. Dafür werden die Testszenarien von den Entwicklern vorbereitet, die auch während der Testung anwesend sein müssen. Die Auswertung findet direkt im Anschluss an die Testung statt, so dass der Ergebnisreport auf das Wesentliche reduziert werden kann und nur als Gedächtnisstütze dient. Daneben wurden einige Programme entwickelt, die die Erstellung von Usability Reports vereinfachen. Sie bieten eine strukturierte Vorgehensweise in dem sie beispielsweise ein Klassifizierungsschema und eine Schweregradbewertung anbieten oder die Verwaltung der gefundenen Probleme unterstützen (z.B. Andre et al., 2001; Hvannberg et al., 2007).

Die berichteten Probleme bei der Umsetzung von Usabilitymethoden in agiler Entwicklung erfordern also eine Methode, die sich ohne die Notwendigkeit von strukturellen Anpassungen innerhalb der kurzen Sprints durchführen lässt. Eine Testung darf keine lange Vorbereitung benötigen und Ergebnisse müssen unmittelbar zur Verfügung stehen, so dass auch mehrere Tests innerhalb eines Iterationszyklus möglich sind. Insbesondere die Ergebnisreports sollen auf die relevanten Aspekte des aktuellen Entwicklungszyklus fokussiert sein, statt sämtliche Probleme aufzulisten (Molich & Dumas, 2008) und die Dokumentation auf die notwendigen Informationen beschränkt sein (Chamberlain et al., 2006; Sy, 2007). Die Methode muss auch für Teammitglieder mit geringen Vorkenntnissen schnell erlernbar sein (Woywode et al., 2012) und darf nur wenig Ressourcen verbrauchen (Reckin & Brandenburg, 2013). Diesen Anforderungen folgend, wurde eine neue Vorgehensweise entwickelt. Die „Instant Usability Report“ Methode (IUR) ermöglicht innerhalb eines Tages die Durchführung von remote User Tests mit schneller Auswertung und unmittelbarer Verfügbarkeit der aufbereiteten und auf die relevanten Aspekte fokussierten Ergebnisse.

2 Methodenbeschreibung Instant Usability Report

Das hier knapp beschriebene methodische Vorgehen für die Durchführung von Nutzertests in agilen Entwicklungsprozessen (ausführlich auf <http://instant-usability-report.blogspot.de>) ist schnell und kann flexibel den Unternehmensbedingungen angepasst werden: mit Testern Inhouse oder via

Konferenzsystem (remote), durch interne Usabilityverantwortliche oder externe Experten. Zwingend ist nur die Anwesenheit eines Produktverantwortlichen (PV) des Unternehmens bei Testdurchführung und -auswertung. Ziel ist die Reduktion des Aufwands durch Ermöglichen von remote Tests.

Testobjekt und Testaufgaben

Ziel der IUR Methode ist das Aufdecken von Usabilityproblemen in web- oder desktopbasierter Software, insbesondere für vorher definierte Workflows. Die Testaufgaben beinhalten realistische Nutzerszenarien, die Testzeit pro Testperson (inkl. Fragebögen und Interviews) soll 45 Minuten nicht überschreiten, bei max. 30-minütiger Interaktionszeit mit den Workflows. Nicht geeignet ist die Methode für die umfangreiche Analyse komplexer Software oder das Gewinnen eines ausführlichen qualitativen Nutzerfeedbacks.

Vorbereitung und Testaufbau

Testziele, Workflows und Nutzungsszenarien werden in einem Vorgespräch mit den Produktverantwortlichen abgestimmt und sollten kurzfristige Entwicklungsziele unterstützen. Hierauf basierend werden Testaufgaben erstellt, deren Bearbeitung idealerweise aktuell relevante Probleme aufzeigen oder direkt implementierbare Verbesserungsansätze generieren könnte. Die Testleitung bearbeitet bei der Erstellung des Testleitfadens alle Aufgaben testweise selbst und formuliert einen verständlichen Instruktionstext. Hierbei werden auch die relevanten Screens des Workflows jeweils zweifach ausgedruckt und in der Reihenfolge der Aufgabenbearbeitung sortiert. Auf einem Ausdruck wird bei der Testung das Protokoll erstellt, auf dem Zweiten später der Ergebnisreport. Zusätzlich werden Klebezettel (z.B. Post-its) in vier verschiedenen Farben bereitgelegt, auf denen die Erkenntnisse während der Testung festgehalten werden. Für die remote Testdurchführung wird eine Konferenzsoftware verwendet, die das Teilen von Bildschirm, Maus und Tastatur ermöglicht. Testperson und Testleitung müssen kommunizieren können, eine Videoaufzeichnung der Testperson ist nicht notwendig. Protokollant und PV können der Konferenz zugeschaltet werden oder mit der Testleitung im gleichen Raum sein. Bei geübter Testleitung und etablierter Kommunikation mit den PV ist eine Testvorbereitung innerhalb weniger Stunden möglich und sollte idealerweise am Vortag durchgeführt werden (kein Zeitdruck vor Test).

Ablauf der Nutzertests

Die Nutzertests werden hintereinander mit insgesamt vier Testnutzern in einem remote Testsetup durchgeführt. Dabei werden sie von der Testleitung

instruiert und alle Interaktionen von einem Protokollanten und einem PV beobachtet. Ein geübter PV mit Usabilityexpertise kann hier ggf. den separaten Protokollanten ersetzen. Eine Testung dauert maximal 45 Minuten, mit max. 30 Minuten für die Aufgabenbearbeitung, und wird direkt im Anschluss in einer ca. zehnminütigen Zwischenbesprechung von Testleitung, Protokollant und PV vorläufig ausgewertet. Eine kurze Erholungspause nach zwei Tests wird empfohlen. Nach der letzten Testperson folgt eine ausführliche Auswertung, bei der ein Gesamtbericht der relevanten Erkenntnisse erstellt wird.

Protokollierung

Während der Testung notiert der Protokollant jede Erkenntnis einzeln auf einem farbigen Klebezettel und klebt diese unmittelbar an die entsprechende Stelle auf dem gedruckten Screen. Die Kategorie der Erkenntnis bestimmt die gewählte Farbe. Zur Erleichterung und Präzisierung der Beschreibung steht eine Anleitung zur Verfügung. In dieser ist für jede Farbe eine Auswahl von Problemarten oder zu notierender Beobachtungen vorgegeben, detailliert mit möglichen Ursachen um das Erkennen zu erleichtern. Im Folgenden die Kategorien und Beispiele für die Hilfestellung:

- **Rot: Usabilityproblem** (verursacht Probleme und Verzögerungen in der Aufgabenbearbeitung) – „Verständnisproblem“: Wording nicht aussagekräftig, Inhalte nicht konsistent benannt, keine Erklärungen/ Infos vorhanden – oder „Nicht gesehen“; „Unsicherheit bei der Nutzung“; „Fehlerhafte Nutzung“
- **Gelb: Hinweis** (Kommentar, hat keine Auswirkung auf den Erfolg der Aufgabenbearbeitung) – „Erwartung“: Darstellung von Inhalten/ Symbolen, Interaktionsmöglichkeiten, weitere Schritte – oder „Persönliche Präferenz“; „Wunsch/ Vorschlag“; „Hinweis“
- **Grün: Positive Erkenntnisse** (positive Äußerungen oder Verhalten in der beabsichtigten Weise) – „Explizit“: Äußerung; „Implizit“: Handlung
- **Blau: Bug** (Fehlfunktion der Software) – „Funktion“: z.B. falsche/ fehlende Verlinkung, falsche Interaktion; „Inhalt“: z.B. falsche/fehlende Inhalte, copy-paste Fehler, Tippfehler

Ergebnisse aus nachfolgenden Testungen werden auf dieselben Ausdrucke geklebt. Identische Erkenntnisse können auf demselben Klebezettel durch Angeben der Anzahl der betroffenen Testpersonen angegeben werden. Der Versuchsleiter achtet darauf, dass der Protokollant genügend Zeit zum Notieren hat und macht ggf. kurze Pausen. Die Klebezettel werden so strukturiert,

dass sie ausreichend Platz für Problemspezifizierung, Verbesserungsvorschlag, Schweregrad und Relevanzeinstufung bieten (siehe Abbildung 1). Dafür empfiehlt sich eine Größe von 76x76mm.

Problembeschreibung	
Hinweis/ Verbesserungs- ansatz	Schweregrad
	Relevanz
	Anzahl Nutzer

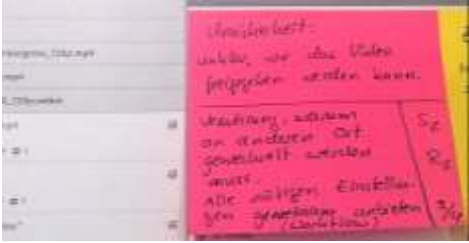


Abbildung 1: Klebezettel Layout und Beispielfoto Klebezettel auf Bildschirmdruck

Auswertung und Erstellung des Ergebnisreports

Zwischenauswertung – Direkt im Anschluss an jeden Test erfolgt eine kurze Zwischenauswertung von Protokollant, Versuchsleitung und PV. Dabei werden die Erkenntnisse der vorangegangenen Testung weiter spezifiziert, zusätzliche Eindrücke ergänzt, falls nötig neu kategorisiert oder mit vorherigen Ergebnissen zusammengefasst. Insbesondere die Zuweisung zu Usabilityproblemen (rote Klebezettel) wird diskutiert, da diese in der Abschlussauswertung besonders berücksichtigt werden. Bei der Zwischenauswertung soll nur sichergestellt werden, dass alle wichtigen Erkenntnisse richtig festgehalten worden sind, mögliche Ursachen der Probleme werden noch nicht diskutiert.

Endauswertung – In der Endauswertung werden alle als Usabilityproblem eingestuft Erkenntnisse einzeln analysiert. Die Problembeschreibung wird so konkretisiert, dass sie für Nicht-Anwesende Mitarbeiter eindeutig ist, Kommentare oder Verbesserungsvorschläge können hinzugefügt werden. Jedem Problem wird ein Schweregrad zugeordnet: (S1) „Verzögert Nutzer kurz.“; (S2) „Verzögert Nutzer wesentlich, erlaubt ihnen aber letztlich die Aufgabe zu erfüllen.“; (S3) „Hindert Nutzer an der Aufgabenerfüllung.“. Diese Einstufung der Probleme wird ergänzt durch die Einschätzung der Relevanz des Problems für den aktuellen Entwicklungsprozess durch den Produktverantwortlichen. Ist das Problem wenig relevant (R1), „Muss aktuell nicht bearbeitet werden.“, ist es relevant (R2), stark relevant (R3) oder sehr stark relevant (R4) und damit „Top Priorität: Unbedingt zu bearbeiten – möglichst als erstes und möglichst schnell.“ Ebenfalls notiert wird, bei wie vielen Testnutzern dieses Problem auftrat (siehe Abb. 1).

Ergebnisreport – Für den Ergebnisreport wird das zweite Set der ausgedruckten Screens verwendet. Jedes analysierte Problem wird einzeln auf dem zugehörigen Screen platziert und fotografiert, so dass ein Fotoreport mit den einzeln aufgeführten Usabilityproblemen entsteht. Zusätzlich wird eine kurze Executive Summary erstellt, auf denen die wichtigsten Eindrücke festgehalten und übergeordnete Probleme notiert werden. Alle weiteren Klebezettel (nicht rote) werden screenweise fotografiert und ebenfalls kurz zusammengefasst. Sie werden nicht individuell analysiert, sondern dienen nur als zusätzliche Information. Der Ergebnisreport dient als Gedächtnisstütze für eine zeitnahe Verarbeitung in den nächsten Iterationszyklen. Er bietet keine ausführliche Beschreibung der aufgetretenen oder zugrundeliegenden Probleme, weshalb zur korrekten Einordnung der Erkenntnisse die Anwesenheit eines Produktverantwortlichen während der Testung erforderlich ist.

3 Methodenevaluation

Ziele der Evaluationsstudie waren a) Überprüfung der Durchführbarkeit der Methode gemäß der definierten Zielstellungen und Parameter; b) die Bewertung der Methode auf Praxistauglichkeit in Unternehmen; c) die spezifische Bewertung des unmittelbar erstellten Ergebnisreports. Die Validität der Bewertung wurde durch Auswahl relevanter Unternehmensvertreter sichergestellt. Diese waren sowohl mit Usability im Allgemeinen hinreichend vertraut (sie können Nutzen bewerten und kennen Usabilitymethoden) und hatten zudem auch Erfahrung mit Nutzertests im Speziellen. So konnten sie die Anwendbarkeit dieser spezifischen Methode bewerten und einschätzen, ob diese für einen Einsatz im Alltag der Softwareentwicklung besser geeignet ist, als andere Nutzertestmethoden, die sie bisher kennengelernt oder eingesetzt haben.

3.1 Studiendesign und Datenerhebung

Mit 8 Unternehmen wurde jeweils ein produktspezifischer Usabilitytest ausgearbeitet und mit der beschriebenen Methode (IUR) durchgeführt. Mindestens ein PV war bei Vorbesprechung und Zieldefinition, Durchführung und Auswertung des Tests beteiligt. Vor Studienbeginn wurden mit einem Fragebogen Daten zum Unternehmen, dessen Softwareentwicklungsprozess und Usability-Aktivitäten erfasst; sowie für

jeden beteiligten PV mit separatem Fragebogen dessen persönliche Erfahrung mit Nutzertests, seine Rolle im Unternehmen und der Umfang seiner Produktverantwortung. Unmittelbar nach Durchführung des Tests folgte die zweite Fragebogenerhebung mit jedem PV zur Beurteilung der Methode sowie ein ca. 3-minütiges Abschlussinterview zum Nutzen und Verwendbarkeit des erstellten Instant Usability Reports.

3.2 Stichprobe

Zur Sicherstellung der definierten Voraussetzungen, wurde die Stichprobe aus Unternehmen ausgewählt, die im Rahmen des UseTree Forschungsprojekts (vgl. Brandenburg et al., 2015) an einem Usability-Projekt teilgenommen haben. Die beteiligten Personen hatten in diesem Projekt grundlegende Kenntnisse über Usability-Prinzipien und Methoden erworben, inkl. Erfahrung mit Vorbereitung, Durchführung und Ergebnissen eines klassischen Usability-Tests mit Nutzern und lautem Denken. Die 8 ausgewählten Unternehmen (N=8) repräsentierten die Vielschichtigkeit der Softwareindustrie. Die Unternehmensgröße variierte zwischen sehr klein (1-5 MA), klein und mittel (bis 50 MA) bis größer (bis 100 MA), die Hälfte der Unternehmen produziert nur ein einziges Softwareprodukt, die jeweils getestete Software befand sich in unterschiedlichen Phasen der Weiterentwicklung: "Live: schnelle Veränderung, häufige Feature-Erweiterung in kurzen Abständen" (6 Produkte) oder "Live, graduelle Verbesserung, gelegentliche Ergänzung von Feature" (2 Produkte). Ihren Softwareentwicklungsprozess bezeichneten 5 Unternehmen als agil, 3 als Mischform.

Alle Unternehmen gaben an, dass aktuell Usability-Aktivitäten durchgeführt werden. Nur bei einem Unternehmen waren dies jedoch Nutzertests. Die anderen Unternehmen nannten Aktivitäten wie Prototyping, Anwenderbefragungen oder Auswerten von Kunden- und Nutzerfeedback für die Produktverbesserung. Innerhalb des letzten Jahres kamen Nutzertests nur in 2 von 8 Unternehmen regelmäßig zum Einsatz ("Wenige Male eingesetzt", "Häufiger eingesetzt" oder "Immer, ist regelmäßiger Bestandteil unseres Entwicklungsprozesses"). Die Nutzung anderer Methoden lässt jedoch darauf schließen, dass die definierten Voraussetzungen für die Studienteilnahme und Kompetenzen für Nutzertests gegeben waren. 6 der 8 Unternehmen hatten innerhalb des letzten Jahres mind. 4 Usabilitymethoden angewendet. Ein Unternehmen hatte sich auf Prototyping (Low Cost sowie High Cost) konzentriert, aber aktuell mit Nutzertests begonnen. Ein Unternehmen begann erst mit der Durchführung von Usability-Aktivitäten. Der hierfür verantwortliche Mitar-

beiter qualifizierte sich jedoch als Teilnehmer dieser Untersuchung, weil er vorher für ein anderes Unternehmen beim UseTree Usability-Projekt teilgenommen und hier die erforderliche Usability-Erfahrung gewonnen hatte.

Insgesamt nahmen 10 Unternehmensvertreter ($N=10$) am Versuch teil. Von jedem der 8 Unternehmen war mindestens eine Person dabei, zwei Unternehmen stellten einen zweiten Teilnehmer. Die Altersverteilung war heterogen ($<30J = 4TN$); ($30-40J = 4TN$); ($40-50J = 2TN$), nur eine Person weiblich. Alle Teilnehmer hatten vorher an einem UseTree Usability-Projekt teilgenommen und konnten mindestens produktbezogene Entscheidungen für das getestete Produkt treffen oder waren für dessen Usability verantwortlich. 3 Personen trugen die Gesamtverantwortung für die Entwicklung und Verbesserung des Produkts. Ihre Erfahrung mit Nutzertests schätzten 9 von 10 Teilnehmer als eher hoch ein, ein Teilnehmer als gering.¹ Nur zwei Teilnehmer hatten im letzten halben Jahr gar keinen Nutzertest durchgeführt. Alle Personen hatten eine positive Einstellung zum Thema Usability: $M = 6,88$; $SD = 0,33$ „Es ist notwendig, das Thema Usability dauerhaft zu unterstützen.“ (1=Stimme gar nicht zu - 7=Stimme voll zu). Der Nutzen von Nutzertests wurde ebenfalls von allen als hoch eingeschätzt mit a) Nutzeneinschätzung: $M = 6,1$ / $SD = 0,7$ (1=sehr gering - 7=sehr hoch) und b) Beitrag zur Wettbewerbsfähigkeit und zum Produkterfolg: $M = 6,1$ / $SD = 0,94$ (1=sehr gering - 7=sehr hoch). Von einigen Teilnehmern wurde jedoch der Aufwand für die Durchführung von Nutzertests als relativ hoch bewertet: $M = 3,8$ / $SD = 1,17$ (1=sehr hoch - 7=sehr gering)*².

4 Ergebnisse

Mit der beschriebenen Vorbereitung im Vorfeld konnten alle 8 Usabilitytests an einem Arbeitstag in 6 bis 8 Stunden ($M = 7,61h$; $SD = 0,4$) durchgeführt werden. Die durchschnittliche Dauer der Abschlussanalyse mit Erstellung des Instant Usability Report war hierbei 116 Minuten ($SD = 24,2$). Die Durch-

¹ Mittelwerte ($N=9$) ohne Ausreißer auf der Antwortskala (1=sehr gering - 7=sehr hoch): a) $M = 5,0$; $SD = 0,94$ „Mein derzeitiges Wissen über Nutzertests ist ...“ und b) $M = 5,22$; $SD = 0,63$ „Meine praktische Erfahrung bei der Durchführung von Nutzertests ist ...“. Eine Person gab auf dem Fragebogen eine sehr geringe Erfahrung mit Nutzertests an und wurde bei der Mittelwertberechnung als Ausreißer behandelt. Auf einen Ausschluß wurde jedoch verzichtet, weil diese Person angab im letzten halben Jahr Nutzertests durchgeführt zu haben.

² Mit * gekennzeichnete Skalen wurden für die Ergebnisdarstellung im Vergleich zum Fragebogen invertiert.

führbarkeit mit den Parametern der Methodenbeschreibung wurde bestätigt, die Praxistauglichkeit in Unternehmen quantitativ bewertet, der Nutzen des Ergebnisreports zusätzlich auch qualitativ.

4.1 Praxistauglichkeit der Nutzertestmethode

Die Bewertung von Ablauf und Durchführung der 8 Nutzertests durch die jeweiligen Unternehmensvertreter war sehr gut. Neben einer hohen Zufriedenheit gaben sie an, über Nutzertests dazugelernt zu haben und dass sie Nutzen im Arbeitsalltag erwarten (siehe Abb. 2).

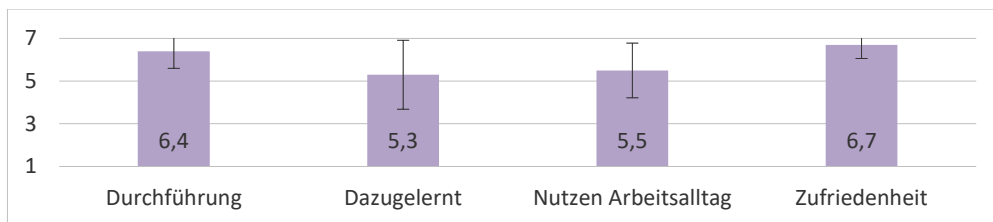


Abbildung 2: Mittelwerte (N=10) für die Frage "Wie beurteilen Sie Ablauf und Durchführung dieses Nutzertests?" mit der Antwortskala (1-7) und den Items "Die Durchführung dieses Nutzertests war ..." (Sehr schlecht - Sehr gut), "Über Nutzertests habe ich dazugelernt ..." (Sehr wenig - Sehr viel), "Das Gelernte über Ablauf und Durchführung wird mir in meinem Arbeitsalltag nützen ..." (Sehr wenig - Sehr viel), "Ich bin mit dem Nutzertest zufrieden." (Stimme gar nicht zu - Stimme voll zu); Fehlerbalken = SD.

Nach Erleben der Testmethode bewerteten sie den Aufwand für die Durchführung von Nutzertests als eher mittel, tendenziell etwas besser als vorher (MW +0,6; n. sign. $t(9) = -2.25$; $p = .051$), und den Nutzen für eine Produktverbesserung unverändert (+0,1) als sehr hoch. Persönlich würden sie gern sehr häufig Nutzertests durchführen, für eine Anwendung in ihrem Unternehmen reduziert sich der Wert jedoch signifikant ($Z = -2$, $p = .03$) (siehe Abb. 3).

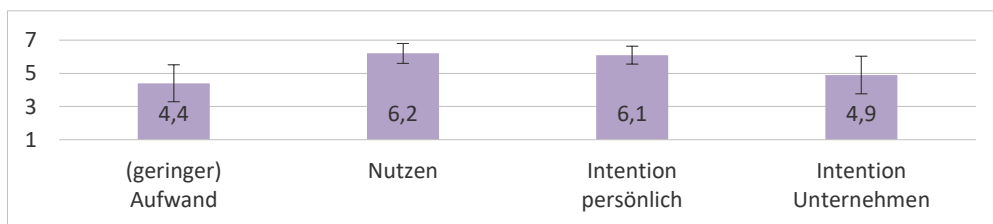


Abbildung 3: Mittelwerte (N=10) für die Frage "Wie würden Sie Nutzertests nach der Durchführung einschätzen?" mit der Antwortskala (1-7) und den Items "Der Aufwand für die Durchführung von Nutzertests ist ..." (Sehr hoch - Sehr gering)*, "Der Nutzen von Nutzertests für die Verbesserung des Softwareprodukts ist ..." (Sehr gering - Sehr hoch), "Wenn ich könnte wie ich wollte (unabhängig von der Umsetzbarkeit im Unternehmen), würde ich künftig Nutzertests durchführen ..." (Gar nicht - Sehr oft), "In unserem Unternehmen möchte ich im nächsten halben Jahr Nutzertests durchführen ..." (Gar nicht - Sehr oft); Fehlerbalken = SD.

Die Unternehmensvertreter wollten die kennengelernte Methode anwenden: Sie gaben an („Bei meinen Arbeitstätigkeiten im nächsten halben Jahr werde ich die vermittelten Inhalte ...“), „einmal ausprobieren“ (N=1), „einige Male durchführen“ (N=5) oder sogar „regelmäßig anwenden“ (N=3) zu wollen. Nur eine Person plante keine Berücksichtigung von Nutzertests im Unternehmen, begründete dies jedoch mit fehlender Passung zu den Zielen der aktuellen Produktentwicklungsphase und deren Ressourcenplanung.

Die Bewertung der Methode war insgesamt gut (siehe Abb. 4). Durchführbarkeit und Erlernbarkeit wurden als eher einfach eingeschätzt, die benötigte Lernzeit eher gering und die benötigte Kompetenz mittelhoch. Der Durchführungsaufwand wurde mittelhoch eingeschätzt, der technische Aufwand gering. Die Befragten sahen in ihrer Unternehmenspraxis eine gute Durchführbarkeit (bezogen auf Aufwand) und sehr gute Durchführbarkeit (bezogen auf technischen Aufwand). Den Nutzen des erstellten Reports für die Produktverbesserung schätzten sie hoch ein.

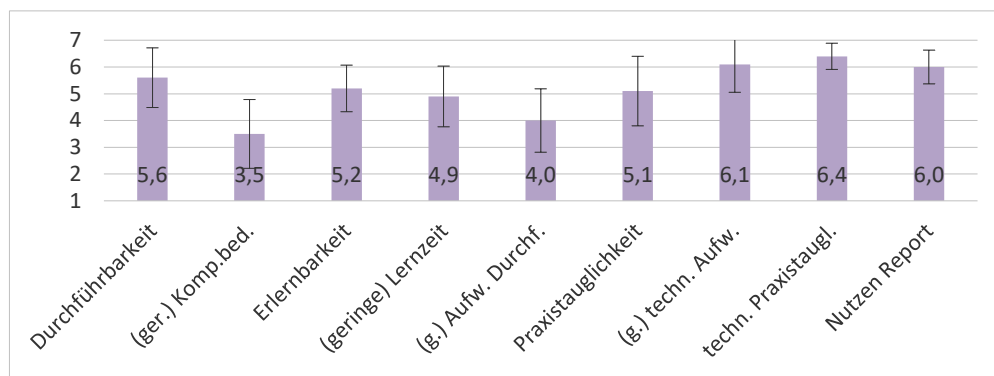


Abbildung 4: Mittelwerte (N=10) für die Frage "Stellen Sie sich vor, Sie sollen künftig Nutzertests nach der gezeigten Methode im Unternehmen durchführen. Ich bewerte diese Methode als ...?" mit der Antwortskala (1-7) und den Items "Die Durchführung von Nutzertests mit dieser Methode ist ..." (Sehr schwierig - Sehr einfach)*, "Die benötigte Kompetenz für Vorbereitung und Durchführung der Methode ist ..." (Sehr hoch - Sehr gering)*, "Das Erlernen der Methode ist ..." (Sehr schwierig - Sehr einfach)*, "Die benötigte Zeit für das Erlernen dieser Methode ist ..." (Sehr lang - Sehr kurz)*, "Der Aufwand für die Durchführung ist ..." (Sehr hoch - Sehr gering)*, "Mit diesem Aufwand ist die Methode in unserer Unternehmenspraxis ..." (Gar nicht gut durchführbar - Sehr gut durchführbar), "Der technische Aufwand ist ..." (Sehr hoch - Sehr gering)*, "Mit diesem technischen Aufwand ist die Methode in unserer Unternehmenspraxis ..." (Gar nicht gut durchführbar - Sehr gut durchführbar), "Der Nutzen des erstellten Reports für die Verbesserung des Produkts ist ..." (Sehr gering - Sehr hoch); Fehlerbalken = SD.

4.2 Nutzen der Reporterstellung

Das im Rahmen der allgemeinen Methodenbewertung gute Rating des Reports (vgl. Abb. 4: hoher Nutzen für die Produktverbesserung; MW = 6,0, SD

= 1,18), zeigt sich auch, wenn nach dem Einfluss in einem Produktverbesserungsprozess gefragt wird („Ich denke dieser Report ist in diesem Produktverbesserungsprozess ...“): Der Report wird als gut anwendbar ($M = 6,0$; $SD = 1,34$; 1=Sehr schlecht anwendbar – 7=Sehr gut anwendbar) und hilfreich eingeschätzt ($M = 6,0$; $SD = 1,18$; 1=Gar nicht hilfreich – 7=Sehr hilfreich), die hierin dokumentierten Erkenntnisse als sehr relevant ($M = 6,3$; $SD = 0,64$; 1=Gar nicht relevant – 7=Sehr relevant).

Eine unmittelbare Verwendung dieser Erkenntnisse für eine schnelle Produktverbesserung ("Im Regelfall würden wir bei uns im Unternehmen den erstellten Report für die Produktverbesserung verwenden innerhalb von ...") war jedoch bei der Stichprobe weniger relevant. Nur 2 von 8 Unternehmen wählten die Antwortoptionen „1 Tag (sofort)“ oder „2-4 Tage“. Die Mehrzahl plante dies erst innerhalb mehrerer Wochen bis einem Monat oder im Rahmen der nächsten geplanten Produktverbesserungsaktivitäten.

Den über eine reine Testbeobachtung mit eigenen Notizen hinausgehenden Mehrwert des erstellten Reports für die Produktentwicklung werteten 8 von 10 Personen im Rahmen des Abschlussinterviews als ziemlich hoch, eine Person als sehr hoch. (siehe Abb. 6) Die Nennung "wenig Mehrwert" wurde mit hoher eigener Usability-Expertise und folglich guter Qualität der Notizen begründet. Die Person ergänzte aber, dass Personen mit weniger Usability-Erfahrung die Reporterstellung auf jeden Fall bräuchten.



Abbildung 6: Antworthäufigkeiten ($N=10$) für die Frage "Verglichen mit einer reinen Beobachtung eines Usability-Tests und meinen resultierenden Notizen: Hat der erstellte Report für die Produktentwicklung einen Mehrwert? Der Report hat ..."; Skala (1-5).

Die inhaltsanalytische Auswertung der Interviewantworten auf die Frage "Wäre der Report verzichtbar, würde Ihnen also die Teilnahme an den Tests und ihre Beobachtungen und Notizen ausreichen?" ergab, dass keine der teilnehmenden Personen die gemeinsame Erstellung des Reports im Anschluss an die Testung für verzichtbar hielt ("Er hat definitiv einen Mehrwert und ist nicht ohne weiteres verzichtbar."; "Der ist sehr schön als Ausgangsbasis und

Zusammenfassung für die erkannten Probleme."). Den Report direkt im Anschluss an die Tests gemeinsam mit Experten zu erarbeiten, wurde als wichtig empfunden. Als wichtiger Aspekt bewertet wurde die strukturierte und inhaltliche Leitung durch die Usability-Experten. Die gute Kommunizierbarkeit der Ergebnisse³ sowie Einfachheit der Weiterverarbeitung der gefundenen Probleme wurden als Mehrwerte des fertiggestellten Reports genannt. Dies sind auch für die geplante Verwendung des Reports die entscheidenden Eigenschaften. Alle Teilnehmer gaben an, die Ergebnisse für die Bearbeitung in ein internes System (z.B. Ticketsystem, Product Backlog) zu überführen und/oder sie mit den Verantwortlichen zu diskutieren. Auch die Unterstützung bei einer Argumentation für die Notwendigkeit von Produktverbesserungen wurde genannt. Als wichtigste Eigenschaften des Reports ("Was am erarbeiteten Report bringt den größten Nutzen für ihr Unternehmen und die Produktentwicklung?") benannt wurden a) die Priorisierung der Usabilityprobleme für die Bearbeitung, b) die anschauliche, visuelle Darstellung der Probleme und c) deren konkrete Benennung mit Lösungsansätzen sowie d) die unmittelbare Management Summary.

5 Diskussion und Ausblick

Die durchgeführte Evaluationsstudie konnte zeigen, dass mit der entwickelten Methode "Instant Usability Report" (IUR) die in (agiler) Entwicklungspraxis und Literatur identifizierten Hindernisse für die Durchführung von Nutzertests erfolgreich adressiert werden konnten. Die unmittelbare Verwertbarkeit der im Ergebnisreport spezifisch aufbereiteten Erkenntnisse für die zielgerichtete Produktverbesserung am Folgetag und deren Priorisierung bedeutete für die Vertreter der 8 Unternehmen einen hohen Mehrwert. Sie bewerteten die Methode als sehr gut durchführbar in ihrer Unternehmenspraxis, mit großem Nutzen und angemessenem Aufwand, und konnten sich gut vorstellen diese künftig einzusetzen. Durch Umfang der Studie und Auswahl repräsentativer Unternehmen konnte eine hohe Ergebnisvalidität gesichert werden. Die Unternehmensvertreter verfügten über die notwendige Erfahrung

³ Eine Einschränkung der Verwendbarkeit als Kommunikationsmittel nannte ein Unternehmen: "Was ich skeptisch sehe bei dem Report: In dieser Form können wir den wahrscheinlich nicht einsetzen in Präsentationen... Man müsste Nacharbeiten vornehmen und das in eine Powerpoint überführen, weil ich mir vorstelle wenn wir zu den Entscheidern gehen, bräuchten wir schon ein aufgeräumtes Dokument. Es ist gut als Grundlage, aber ist nicht das was ich denjenigen präsentieren würde."

mit Usabilitymethoden und Nutzertests, um die durchgeführte Methode spezifisch bewerten zu können. Alle hatten schon vor der Studie eine positive Einstellung zu Usability und führten in ihren Unternehmen Usabilityaktivitäten durch. Zudem war die Durchführung kostenlos. Diesen positive Bias gilt es bei der Interpretation der Studie zu berücksichtigen.

Große Stärken der Methode sind schnelle Durchführung und effiziente, zielgerichtete Ergebniserhebung. Der geringe Zeiteinsatz je Testperson sowie insbesondere der durch das „remote“ Testen mögliche Verzicht auf eine Anreise, erleichtert das Gewinnen von Testpersonen und die Einbindung von Kunden und Key Usern direkt im Nutzungskontext (z.B. am Arbeitsplatz). Die Gesamtkosten für eine Testsession werden erheblich verringert, der direkt wahrnehmbare Nutzen durch die nun mögliche unmittelbare und priorisierte Produktverbesserung erhöht. Durchführung und Ergebnisverwendung lassen sich nahtlos in den agilen Entwicklungsprozess integrieren. Durch die Einbindung von Produktverantwortlichen, aber idealerweise auch weiterer Personen, wie Entwicklern oder Führungskräften, kann das Team für Usabilityprobleme und die Notwendigkeit von Usabilityaktivitäten sensibilisiert werden. Interaktionsdesigner und Usabilityverantwortliche können ihre Annahmen schnell testen und Unsicherheiten ausräumen. Aufgrund der geringen Stichprobengröße sind die Ergebnisse allerdings nicht wissenschaftlich abgesichert. Im Vergleich mit herkömmlichen Methoden und umfangreichen Tests ist deren absolute Aussagekraft sicher eingeschränkt. Das Ausmaß dieser Einschränkung kann nur durch Vergleichsstudien ermittelt werden. Die Vorteile der Methode für ihren Einsatzzweck (schrittweise Entwicklung und Verbesserung definierter Workflows und Funktionalitäten, nicht jedoch tiefgehende Analysen komplexer Systeme und Exploration von Nutzerbedürfnissen) wiegen das Risiko, einige Usabilityprobleme zu übersehen, deutlich auf.

Literaturverzeichnis

- Andre, T. S., Rex Hartson, H., Belz, S. M., & McCreary, F. A. (2001). The user action framework: a reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54(1), 107–136.
- Brandenburg, S., Sachse, K., Jäger, A., Reckin, R., & Becker, M. (2015). Usability für kleine und mittelständische Unternehmen im Bereich mobiler und internetbasierter Anwendungen: Abschlussbericht von UseTree, dem Berliner Kompetenzzentrum für Usability-Maßnahmen. Berlin: UseTree - Berliner Kompetenzzentrum für Usability-Maßnahmen.

- Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a framework for integrating agile development and user-centred design. In *Extreme programming and agile processes in software engineering* (pp. 143–153). Springer.
- Da Silva, T. S., Silveira, M. S., Maurer, F., Hellmann, T., et al. (2012). User experience design and agile development: From theory to practice. *Journal of Software Engineering and Applications*, 5(10), 743.
- Hvannberg, E. T., Law, E. L.-C., & Lárusdóttir, M. K. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, 19(2), 225–240. doi: 10.1016/j.intcom.2006.10.001
- Molich, R., & Dumas, J. S. (2008). Comparative usability evaluation (CUE-4). *Behaviour & Information Technology*, 27(3), 263–281.
- Nielsen, J. (2005). Formal usability reports vs. quick findings. Abgerufen am 05.07.2017, von <https://www.nngroup.com/articles/formal-vs-quick-usability-reports/>
- Reckin, R. & Brandenburg, S. (2013). Discount Usability-Maßnahmen als erste Schritte von KMU auf dem Weg zum Agile Usability Engineering. In E. Brandenburg, L. Doria, A. Gross, T. Günzler & H. Smie (Hrsg.), *Grundlagen und Anwendungen der Mensch-Maschine-Interaktion: Proceedings der 10. Berliner Werkstatt Mensch-Maschine-Systeme 10.-12. Oktober 2013* (S. 238-244). Berlin: Universitätsverlag der TU Berlin.
- Rösler, A., & Thölke, M. (2015). Usability-Testing in agilen Entwicklungsprojekten. *Mensch und Computer 2015–Usability Professionals*.
- Rößler, T., & Kauer, M. (2015). Analyse des aktuellen Stands der Berücksichtigung von Usability bei der Softwareentwicklung in deutschen kleinen und mittleren Unternehmen. *Mensch und Computer 2015–Usability Professionals*.
- Singh, M. (2008). U-scrum: An agile methodology for promoting usability. In *Agile, 2008. agile'08. conference* (pp. 555–560).
- Stickel, O., Ogonowski, C., Jakobi, T., Stevens, G., Pipek, V., & Wulf, V. (2015). Praktiken der Nutzereintegration im Entwicklungsprozess von KMU. *Mensch und Computer 2015-Proceedings*.
- Sy, D. (2007). Adapting usability investigations for agile user-centered design. *Journal of Usability Studies*, 2(3), 112–132.
- Trappe, C., Heuten, W., Boll, S., Timmermanns, S., Rahner, S., Wolff, D., & Gräfe, B. (2015). Zentrale Faktoren bei der Umsetzung von Usability-Engineering bei einem mittelständischen Softwarehersteller in der Pflegebranche. *Mensch und Computer 2015–Workshopband*.
- Trauzettel, F., Minge, M., & Thüring, M. (2015). Usability in medizintechnischen Softwarefirmen: Eine Interviewstudie zur aktuellen Lage. *Mensch und Computer 2015– Workshopband*, 85.
- Weber, M. (2015). Usability Testing als Kunst des Möglichen-Wie man Usability Testing erfolgreich in Projekte einbringt und umsetzt. *Mensch und Computer 2015–Usability Professionals*.
- Woywode, M., Maedche, A., Wallach, D., & Plach, M. (2012). Gebrauchstauglichkeit von Anwendungssoftware als Wettbewerbsfaktor für kleine und mittlere Unternehmen (KMU): Abschlussbericht.

Ausführliche Methodenbeschreibung und Materialien: <http://instant-usability-report.blogspot.de>

Kontakt: ronny.reckin@tu-berlin.de; vera.burckhardt@pidoco.com