

# HCI & Usability

## Unterlagen zur ILV

Christina Hochleitner  
[christina.a.hochleitner@gmail.com](mailto:christina.a.hochleitner@gmail.com)

WS 2017/18

## Themenblock 5

# Usability Evaluation – Teil 2

Wie in Kapitel 4 beschrieben, unterscheidet man zwischen empirischen und formal-analytischen Evaluationsmethoden. Während die formal-analytischen Methoden auf Experten, die die NutzerInnen verkörpern bzw. auf Formeln und automatisch errechnete Kennzahlen zurückgreifen, werden im Rahmen von empirischen Methoden (z. B. Usability Studien oder Fragebögen) die benötigten Informationen über NutzerInnenbefragungen und Beobachtungen gewonnen. Analytische (vor allem kosteneffiziente) Methoden werden allgemein eher früher im Design-Prozess eingesetzt, empirische Methoden werden hingegen vermehrt gegen Ende der Entwicklungsphase verwendet.

Um eine erfolgreiche Evaluation durchführen zu können, müssen zuerst Ziele gesetzt werden. Laut Gediga und Hamborg [3] kann man Evaluationen nach folgenden Zielen kategorisieren:

- *Vergleich:* Es werden zwei oder mehrere Systeme oder Designvarianten anhand von bestimmten Kriterien verglichen.
- *Systemeigenschaften:* Hier werden bestimmte Systemeigenschaften und deren Ausprägung anhand bestimmter Kriterien evaluiert.
- *Schwachstellen:* Bei einer (meist formativen) Evaluation sollen Schwachstellen gefunden werden, die dann ein Ausgangspunkt für Verbesserungen bzw. die weitere Entwicklung sein sollen.

### 5.1 Usability Studien

Usability Studien sind die wohl bekannteste Methode der Usability Evaluation. Usability Studien involvieren meist eine Reihe von Aufgaben, die von NutzerInnen erfüllt werden sollen, während Probleme, mit denen die NutzerInnen in Kontakt kommen, festgehalten werden. Im Regelfall werden die NutzerInnen während der Ausführung der Aufgaben von Usability-Experten beobachtet und Daten (Zeit- oder Fehlermetriken, Meinungen, Verhalten) gesammelt. Die Ergebnisse werden verwendet, um Schlussfolgerungen über das System zu treffen, bzw. Verbesserungsmöglichkeiten vorzuschlagen.

Jakob Nielsen beschreibt Usability Studien in seinem Buch *Usability Engineering* [5] folgendermaßen:

User testing with real users is the most fundamental usability method and is

in some sense irreplaceable, since it provides direct information about how people use computers and what their exact problems are with the concrete interface being tested.

Usability Studien mit realen Nutzern sind die grundlegendste Usability Methode. Sie sind auf ihre Art unersetzbar, da sie direkte Informationen darüber liefern, wie Menschen Computer benutzen und was ihre exakten Probleme mit dem getesteten Interface sind.

In den folgenden Absätzen werden die wesentlichen Punkte in Zusammenhang mit der Planung und Durchführung einer Usability Studie erörtert.

**Art der Evaluation.** Grundsätzlich unterscheidet man zwischen experimentellen, deskriptiven und korrelativen wissenschaftlichen Studien [1]. Bei einem Experiment werden zwei oder mehrere Gruppen miteinander verglichen und abhängige und unabhängige Variablen bestimmt, im Falle von deskriptiven Studien werden Informationen beschrieben (z. B. die Anzahl der gefundenen Usability Probleme) und korrelative Studien untersuchen Zusammenhänge (z. B. zwischen Alter und gefundenen Usability Problemen). Die im Rahmen dieser Lehrveranstaltung durchgeführten Studien fallen—je nach der Art der Untersuchung—unter die letzteren beiden Punkte.

**Forschungsfrage.** Eine Forschungsfrage beschreibt das zu untersuchende Thema und ist der erste wichtige Schritt im Rahmen der Durchführung einer empirischen Studie [4]. Die Forschungsfrage muss im Rahmen der Studie bzw. deren Ergebnisse beantwortet werden können. Somit definiert die Forschungsfrage das Ziel der Studie. Kepplinger und Mair definieren folgende Richtlinien für die Formulierung von Forschungsfragen [4]:

- Eine Forschungsfrage muss möglichst präzise formuliert sein.
- Die Fragestellung sollte aus der Literatur abgeleitet werden.
- Forschungsfragen müssen empirisch überprüfbar sein.
- Die Beantwortung der Forschungsfrage leistet einen neuen, originären Beitrag zum Wissen mit akademischem oder praktischem Nutzen.
- Forschungsfragen müssen sowohl das inhaltliche Ziel des Projektes („was will man wissen“) als auch die ihm zugrunde gelegte Grundgesamtheit („von wem will man es wissen“) zum Ausdruck bringen.
- Forschungsfragen müssen geographisch und zeitlich eindeutig abgegrenzt sein.
- Forschungsfragen müssen als ein Satz klar und widerspruchsfrei formuliert werden.

Anschließend an die Forschungsfragen werden die Hypothesen formuliert bzw. aus der Forschungsfrage abgeleitet.

**Zielgruppe.** Die rekrutierten NutzerInnen sollten so gut als möglich der gewählten Zielgruppe (z. B. den Personas) des Systems entsprechen bzw. für diese Zielgruppe repräsentativ sein. Dadurch werden die erhaltenen Ergebnisse aussagekräftig und das System wird passend für die gewählte Zielgruppe entwickelt. Des Weiteren sollten die StudienteilnehmerInnen nicht aus dem Kreis der EntwicklerInnen bzw. des Projektteams

gewählt werden und das zu evaluierende System, wenn es dem Ziel der Evaluation entspricht, nicht kennen. Die Anzahl der, für die Evaluation benötigten NutzerInnen kann durch die Anzahl der zu evaluierenden Faktoren und der Signifikanz der Ergebnisse berechnet werden. Wichtig ist jedoch, dass die Ergebnisse ausreichend sind, um das System zur Zufriedenstellung der NutzerInnen zu gestalten und Fehler und Interaktionsprobleme zu eliminieren. Während die Probleme selbst bereits mit sehr wenigen NutzerInnen erkannt werden können, ist es erst durch das Feedback von vielen NutzerInnen möglich, die Auswirkungen der Probleme festzustellen. Von Dix et al. werden 10 TeilnehmerInnen als das absolute Minimum für Usability Studien angegeben [2].

**Evaluationsort.** Die Studien können in einem Labor (kontrollierbar), in einer realen, aber kontrollierbaren Umgebung (z. B. Büroräumlichkeiten) oder in einer schlecht bzw. nicht kontrollierbaren Umgebung („in the wild“) statt finden. Während Labors noch vor einigen Jahren mit einem Einwegspiegel ausgestattet waren, hinter dem sich die EvaluatorInnen befanden, sind die heutigen Labors viel flexibler und mobiler. Meist werden die NutzerInnen per Kamera aufgenommen, der/die EvaluatorIn sitzt jedoch in den häufigsten Fällen direkt neben dem/der NutzerIn und notiert das erhaltene Feedback. Zusätzlich gibt es noch die Möglichkeit, Studien *remote*, also über das Internet durchzuführen. Dazu gibt es spezielle Software, jedoch kann es auch genügen, wenn die EvaluatorInnen den Bildschirm der TeilnehmerInnen sehen und deren Stimme hören.

**Verwendeter Prototyp.** Abhängig von Art und Ziel der Studie können wenig bis weit entwickelte Prototypen verwendet werden. Während bei low-level Prototypen mit entsprechendem prototypischem (sketchy) Design eher die Funktionalität im Vordergrund steht, kann bei weiter fortgeschrittenen high-level Prototypen das Design in den Fokus der NutzerInnen rücken. Daher ist es bei high-level Prototypen besonders wichtig, die Interaktion der NutzerInnen mit dem System zu beobachten, um auch nicht geäußerte Probleme mit dem System feststellen zu können.

**Aufgaben.** Die Aufgaben müssen so gewählt werden, dass mit den gesammelten Daten die gesetzten Evaluationsziele erreicht werden und die zu testenden Funktionen abgedeckt sind. Die Anzahl der Aufgaben muss dem System (und dessen Entwicklungsstatus), als auch der Evaluationssituation (benötigte Zeit, verfügbare Ressourcen, Entschädigung der TeilnehmerInnen, etc.) entsprechen. Die Aufgaben sollen den NutzerInnen sinnvoll und stichhaltig erscheinen.

**Dokumentation.** Um die Daten analysieren zu können, müssen diese in angebrachter Form (Granularität, Umfang, etc.) gesammelt werden. Die einfachste Methode, um Kommentare von NutzerInnen festzuhalten ist mit Papier und Bleistift. Kommentare können entweder in Klartext oder in kodierter Form niedergeschrieben werden. Ein Nachteil dieser Methode ist, dass sehr selten alle Details einer NutzerInnenaussage festgehalten werden können. Um einfache Thinking Aloud Evaluationen festzuhalten, können Audioaufnahmen verwendet werden. Hier kann allerdings nicht mehr die gesamte Interaktion (z. B. die Darstellung am Display während der NutzerInnenaussage) vollständig nachvollzogen werden. Detailliertere Informationen können über Videoaufnahmen gesammelt werden. Videos können von den NutzerInnen selbst und/oder vom

Bildschirm mit der Interaktion gemacht werden. Gemeinsam mit Audioaufnahmen führt das zu einer fast lückenlosen Dokumentation der Studie. Log-Files ermöglichen es, Interaktionsabfolgen, aber auch Interaktionszeiten detailliert festzuhalten und auszuwerten [2].

**Metriken.** Um die gesetzten Ziele einer Evaluation zu erreichen und sinnvolle Daten zu erhalten, ist es notwendig bereits im Vorfeld Metriken festzulegen, die während der Evaluation festgehalten bzw. gemessen werden. Die folgende Liste enthält einen Überblick über oft verwendete Metriken [5, 6]:

- Effektivität.
- Effizienz.
- Zufriedenheit.
- Zeit um eine Aufgabe zu erledigen (Task completion time).
- Anzahl der Aufgaben, die in einer bestimmten Zeit bearbeitet/abgeschlossen werden.
- Anzahl der Fehler.
- Verhältnis zwischen erfolgreichen Interaktionen und Fehlern.
- Zeit die benötigt wird, um sich von Fehlern zu erholen.
- Anzahl der verwendeten/nicht verwendeten Funktionen.
- Funktionen, an die sich die NutzerInnen noch nach der Verwendung des Systems erinnern.
- Wie oft musste der/die EvaluatorIn helfen?
- Kritische NutzerInnenaussagen (positiv und negativ).
- Verhältnis zwischen positiven und negativen NutzerInnenaussagen.
- Dauer der inaktiven Zeit (z. B. Warten auf Eingaben durch die NutzerInnen).

### 5.1.1 Wissenschaftliches Experiment

Ein Experiment ist eine spezielle Form einer (sozialwissenschaftlichen) Studie, bei der zwei oder mehrere Gruppen miteinander verglichen werden. Dazu werden Hypothesen angenommen, abhängige und unabhängige Variablen bestimmt und ein entsprechendes Studiendesign gewählt. Ein Experiment erlaubt es, Schlussfolgerungen über Ursache und Wirkung zu formulieren [1].

**Hypothese.** Eine Hypothese ist die Vorhersage des Ergebnisses eines Experiments. Die initial angenommene Forschungsfrage wird durch die Hypothese weiter konkretisiert. Hypothesen beschreiben Zusammenhänge, Unterschiede, Veränderungen oder operationale Faktoren [1]. Die Hypothese wird durch abhängige und unabhängige Variablen bestimmt und besagt, dass die Änderung einer unabhängigen Variable eine Änderung der abhängigen Variable verursacht. Das Ziel einer Studie ist es, zu zeigen, dass diese Voraussage korrekt ist. Dies wird durch Ablehnung der Null-Hypothese, die besagt, dass sich nichts ändern wird, erreicht. Die Signifikanz eines Ergebnisses sagt aus, ob die erhaltenen Unterschiede zufällig zustande kommen konnten [2].

**Variablen.** Experimente manipulieren und messen Variablen unter kontrollierbaren Bedingungen, um eine Hypothese zu testen. Dazu gibt es 2 Typen von Variablen: abhängige und unabhängige Variablen. Unabhängige Variablen können verändert werden, abhängige Variablen werden gemessen. Unabhängige Variablen werden verwendet, um die unterschiedlichen Studienbedingungen herzustellen, z. B. Art des Systems, Interface Stil, etc. Komplexere Studien können über mehr als nur eine unabhängige Variable verfügen. So müsste eine Studie mit 2 unabhängigen Variablen (System A / System B, Design 1 / Design 2) 6 Bedingungen untersuchen, um alle Möglichkeiten evaluieren zu können. Abhängige Variablen werden hingegen durch die Studie gemessen. Daher müssen abhängige Variablen auch messbar sein (z. B. die Anzahl der Fehler) [2].

**Studiendesign.** Bei Usability Studien wird zwischen *between subjects* und *within subjects* Studien unterschieden. Bei *within subjects* Studien erledigen alle NutzerInnen alle vorgegebenen Aufgaben mit allen zur Verfügung stehenden Systemen. Werden z. B. System A und System B gegeneinander evaluiert, so werden alle TeilnehmerInnen System A und System B in gegengewichteter Reihenfolge (Counterbalancing) verwenden. Bei *between subjects* Studien werden die TeilnehmerInnen pro zu evaluierender Variante in möglichst ähnliche (homogene) Gruppen geteilt. Gruppe 1 würde z. B. nur System A evaluieren und Gruppe 2 nur System B. Während *between subjects* Studien Probleme hinsichtlich der Reihenfolge der Systemnutzung (z. B. System A zuerst, dann System B) eliminieren, reduzieren *within subjects* Designs mögliche Unterschiede in der Zusammensetzung der Gruppen.

### 5.1.2 Erhebungsmethoden

Verschiedene Methoden können in Usability Studien eingesetzt werden, um entsprechendes Feedback von den NutzerInnen zu erhalten. Diese Methoden helfen, direkt während oder kurz nach der Interaktion mit dem System Informationen zu erhalten. Zusätzlich gibt es Methoden der Befragung, um weitere Informationen zu Präferenzen, Abneigungen und Anforderungen der NutzerInnen zu erfahren. Zu diesen Methoden zählen Feldbeobachtung, Fokusgruppe, Interview, pro-aktive Feldstudie und Fragebögen.

Die folgende Liste gibt einen Überblick über häufig verwendete Erhebungsmethoden, die auch miteinander kombiniert werden können [6, 7]:

- *Thinking Aloud:* Die NutzerInnen werden gebeten, während der Interaktion mit dem System ihre Gedanken zu verbalisieren (i.e. laut mitzudenken).
- *Coaching:* Die Teilnehmer können einem Experten (Coach, meist der/die EvaluatorIn selbst) jegliche System-relevanten Fragen stellen.
- *Co-discovery Learning:* Zwei NutzerInnen lösen eine Aufgabe gemeinsam. Die TeilnehmerInnen dürfen sich gegenseitig helfen, als ob sie real zusammen am System arbeiten würden.
- *Performance Measurement:* Hier werden quantitative Daten aus der Interaktion der NutzerInnen mit dem System (i.e. dem Lösen der Aufgaben) gewonnen. Jegliche Interaktion mit dem/der EvaluatorIn (z. B. Fragen, Verbalisierung der Gedanken) wirkt sich hier auf die erhaltenen Daten aus. Zusätzlich können auch Log-Files analysiert werden.

- *Question-asking Protocol*: Die EvaluatorenInnen stellen den NutzerInnen Fragen zum System, um ihre Denkweise und mögliche Probleme zu verstehen.
- *Remote Testing*: Die EvaluatorenInnen sind räumlich/zeitlich von den NutzerInnen getrennt. Meist wird ein Video der Interaktion aufgenommen bzw. der Bildschirm übertragen, um zur selben Zeit die Interaktion zu steuern oder die Daten im Nachhinein zu analysieren.
- *Retrospective Testing*: Die EvaluatorenInnen sammeln weitere Informationen, indem sie nachträglich eine Aufzeichnung (Video) der Interaktion gemeinsam mit den NutzerInnen ansehen und ihnen Fragen zu ihrem Verhalten während der Evaluation stellen.
- *Shadowing*: Ein/e ExpertIn (Domänen-ExpertIn) sitzt zusätzlich neben dem/der EvaluatorIn, um ihm/ihr die Aktionen der NutzerInnen zu erklären. Diese Methode ist hilfreich, wenn die Thinking-aloud Methode nicht angewandt werden kann.
- *Teaching*: Die NutzerInnen interagieren mit dem System, um sich Wissen im Umgang mit dem System anzueignen. Danach werden sie gebeten, unerfahrenen NutzerInnen das System zu erklären bzw. eine Reihe von vorgegebenen Aufgaben zu zeigen.
- *Eyetracking*: Mittels Eyetracking kann der Blickverlauf der NutzerInnen während der Interaktion analysiert werden. Probleme mit der Erkennung von wichtigen Interaktionselementen können somit festgestellt werden.

### 5.1.3 Evaluationsplan

Um aussagekräftige und valide Ergebnisse zu erhalten, muss die Studie sorgfältig geplant und ein Evaluationsplan erstellt werden. Der Evaluationsplan stellt – ähnlich einem Interviewleitfaden – sicher, dass alle nötigen Überlegungen und Fakten für die Evaluation gesammelt wurden und dass – vor allem bei mehreren EvaluatorenInnen – die Evaluation immer nach dem gleichen Schema und mit den gleichen Informationen abläuft. Folgende Informationen müssen in einem Evaluationsplan enthalten sein:

- Ziel der Studie.
- Forschungsfragen und Hypothese.
- Methode:
  - Zielgruppe (inkl. Rekrutierung).
  - Evaluationsort und -Setup.
  - Evaluationszeitraum.
  - EvaluatorenInnen.
  - Studiendesign (within vs. between subjects Evaluation).
  - Erhebungsmethode.
  - Metriken.
  - Verwendeter Prototyp.
  - Dokumentationsart.
  - Aufgaben.

- Ablauf der Studie (inkl. Begrüßung/Verabschiedung und möglicher Entschädigung für die TeilnehmerInnen).

#### 5.1.4 Fehlerquellen

Fehler in der Planung und Durchführung einer Usability Studie können schwerwiegende Auswirkungen auf die Ergebnisse haben und diese auch nicht verwendbar machen. Daher ist es wichtig, mögliche Fehlerquellen zu vermeiden:

- Pilot Test: Vor jeder Studie sollte probeweise eine Evaluation außer Konkurrenz statt finden, um mögliche Probleme mit Ablauf, Zeit und Aufgabenstellungen zu vermeiden.
- Ausgereifte Prototypen: Die zu evaluierenden Funktionen des Prototypen müssen in ausreichendem Umfang umgesetzt werden. NutzerInnen halten sich nicht an geplante Interaktionsabläufe. Daher müssen auch alternative Interaktionsmöglichkeiten gegeben und funktionstüchtig sein.
- Pause zwischen Evaluationen: Wenn mehrere Evaluationen nacheinander statt finden, ist es wichtig, dass genügend Zeit zwischen diesen Evaluationen eingeplant ist. Dadurch können eventuelle Fragen der NutzerInnen nach der Evaluation in Ruhe beantwortet werden.
- Gutes Aufgabendesign: Die zu lösenden Aufgaben müssen durchdacht und für die NutzerInnen sinnvoll sein.
- Informationspreisgabe: Es sollen nur solche Informationen preis gegeben werden, die für die Interaktion selbst benötigt werden. Weitere Informationen können die NutzerInnen beeinflussen bzw. unabsichtliche Agenden der Evaluation aufdecken.
- Frustration: Frustration durch den Testleiter oder mangelndes Interesse (z. B. nach der zehnten Evaluation und dem Wiederholen einiger Probleme) können gravierende Auswirkungen auf die NutzerInnen haben.
- Beeinflussende Fragestellungen: Durch leitende Fragen, können die Antworten und auch die Auffassung der NutzerInnen beeinflusst werden. Das sollte tunlichst vermieden werden.



# Quellenverzeichnis

- [1] Jürgen Bortz und Nicola Döring. *Forschungsmethoden und Evaluation - für Human- und Sozialwissenschaftler ; mit ... 70 Tabellen*. 3. Aufl. Berlin, Heidelberg: Springer, 2002 (siehe S. 5-2, 5-4).
- [2] Alan Dix u. a. *Human-Computer Interaction*. 3. Aufl. Harlow, UK: Prentice Hall, 2004 (siehe S. 5-3, 5-4, 5-5).
- [3] Günther Gediga und Kai-Christoph Hamborg. „Evaluation in der Software-Ergonomie: Methoden und Modelle im Software-Entwicklungsprozess“. *Zeitschrift für Psychologie / Journal of Psychology* 210 (2002), S. 40–57 (siehe S. 5-1).
- [4] Dietmar Kepplinger und Michael Mair. „Forschungsfragen und Hypothesenmodell: So machen Sie (sich) Ihre Projektziele klar!“ *Tourismus Wissen - quarterly* 5-6 (2016) (siehe S. 5-2).
- [5] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1994 (siehe S. 5-1, 5-4).
- [6] Florian Sarodnick und Henning Brau. *Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung*. Bern, CH: Hans Huber, 2006 (siehe S. 5-4, 5-5).
- [7] Wikipedia. *Usability Testing*. 2017. URL: [https://en.wikipedia.org/wiki/Usability\\_testing](https://en.wikipedia.org/wiki/Usability_testing) (besucht am 13.12.2017) (siehe S. 5-5).