

Usability Engineering

Vorlesung 4: Usability Testing (Teil 1)

VU 183.123

Christoph Wimmer

Übersicht

- Usability Testing
 - Arten von Usability Tests
 - Vorgehensweise
 - Vorbereitung und Planung
 - Durchführung
 - Auswertung und Analyse der Ergebnisse
 - Kommunikation der Ergebnisse & Reporting

Usability Testing

Usability Testing

- Bei einem **Usability Test** verwenden BenutzerInnen ein Softwaresystem in einer kontrollierten Umgebung und werden bei der Bearbeitung vordefinierter Aufgaben beobachtet



Quelle: Danny Hope

Usability Testing

If you ask ten usability experts how to run a valid usability test, you'll get about twenty different answers.

- Lukas Mathis

Formative vs. Summative Evaluierung

- **Formative** Evaluierung
 - Während Design und Entwicklung
 - Wie werden die Bedürfnisse der BenutzerInnen erfüllt?
 - Evaluieren von Alternativen
 - Probleme voraussehen
 - Ständige Verbesserung
 - Meist Fokus auf qualitative Beobachtungen und Daten

- **Summative** Evaluierung
 - Nachdem Design feststeht
 - Überprüfung von Hypothesen
 - Prüfen auf Standards
 - Benchmarking
 - Fine-Tuning & Polishing
 - Meist Fokus auf quantitative Messungen und Daten

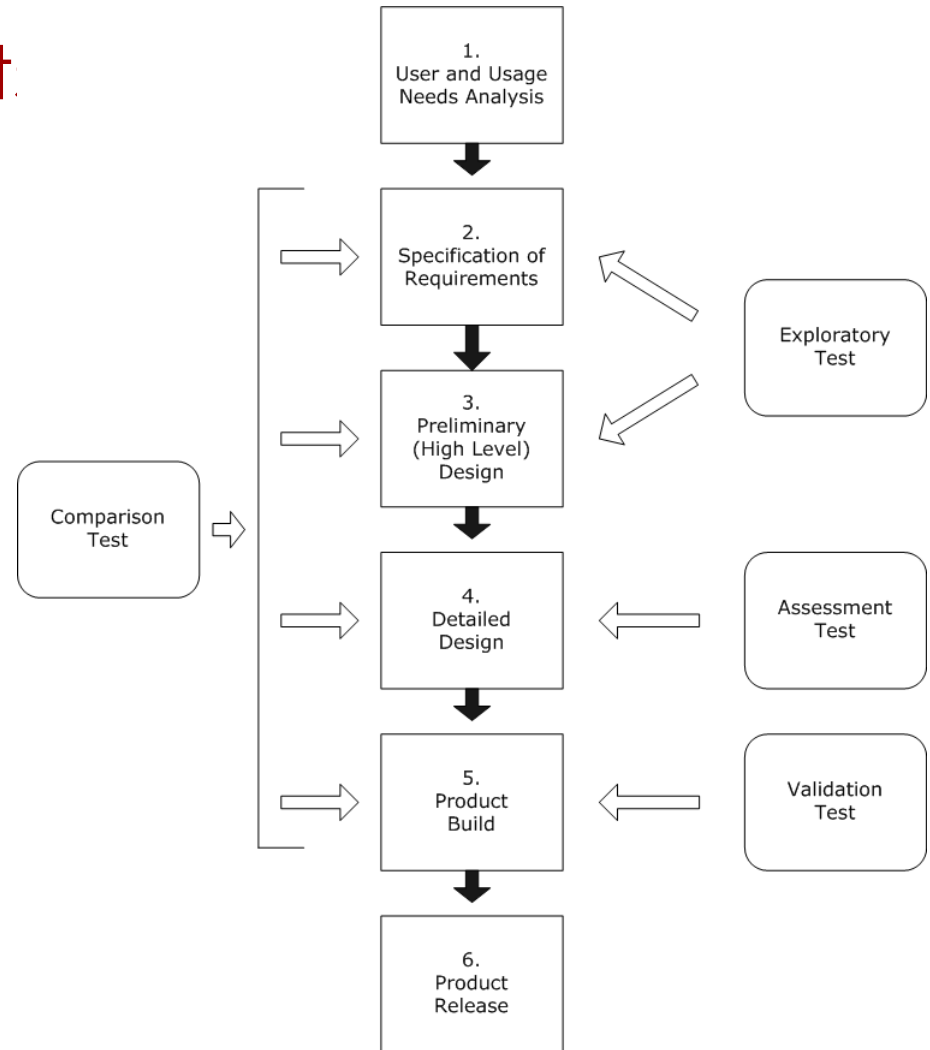
Formell vs. Informell

- **Formell** (typischerweise für summativ notwendig):
 - Viele Testpersonen
 - Test-Labor
 - Großes Team
 - Umfassende Protokollierung & Aufzeichnung (Audio, Video)
 - Formelle Dokumentation der Ergebnisse
- **Informell** (typischerweise eher formativ möglich):
 - Eine oder wenige Testpersonen
 - Vor Ort, z.B. direkt am Arbeitsplatz
 - Einzelne Test-LeiterIn oder kleine Teams
 - Wenig oder gar keine Protokollierung/Aufzeichnung
 - EntwicklerInnen beobachten direkt vor Ort

Arten von Usability Test:

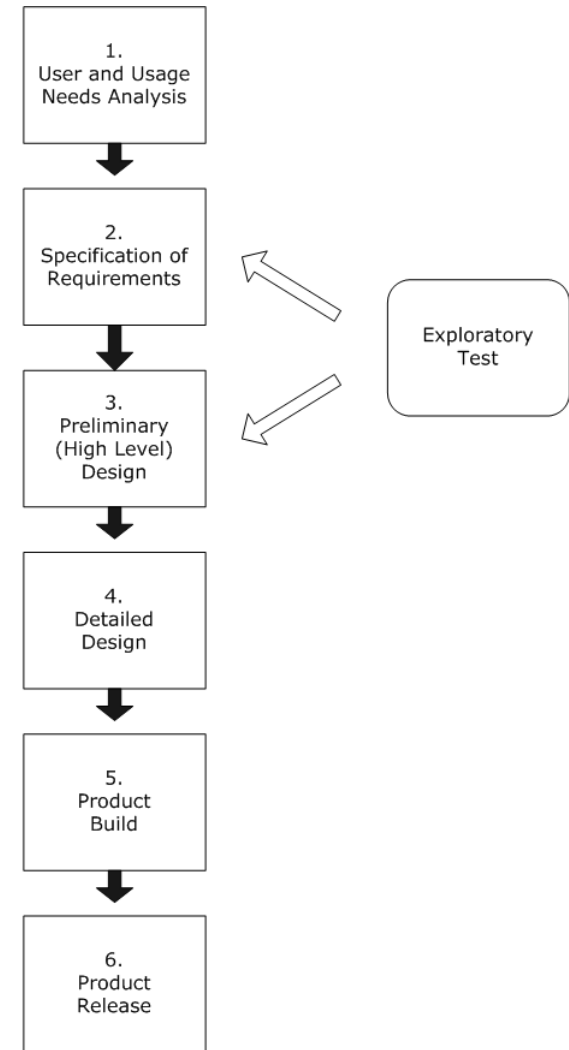
Rubin (1994) beschreibt vier Arten von Usability Tests:

- Exploratory Testing
- Assessment Testing
- Validation Testing
- Comparison Testing



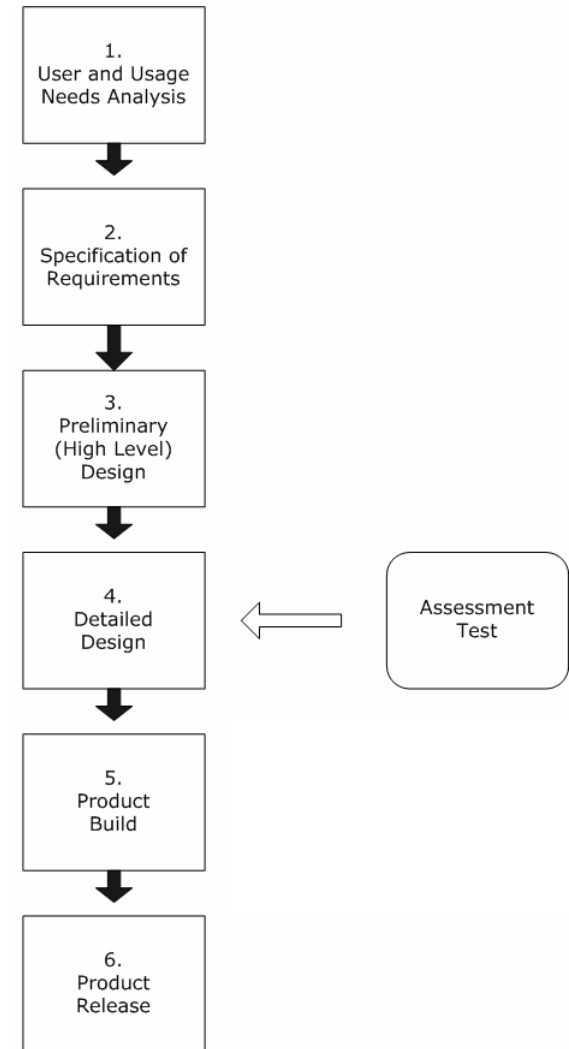
Exploratory Test

- Überwiegend formativ
- Früh im Design-Prozess
- Noch keine Funktionalität implementiert
- High level, grundlegende Designfragen
- Ziel: Qualitative Aussagen zur Gestaltung der Benutzerschnittstellen
- Viel Interaktion zwischen Testperson und TestleiterIn erforderlich



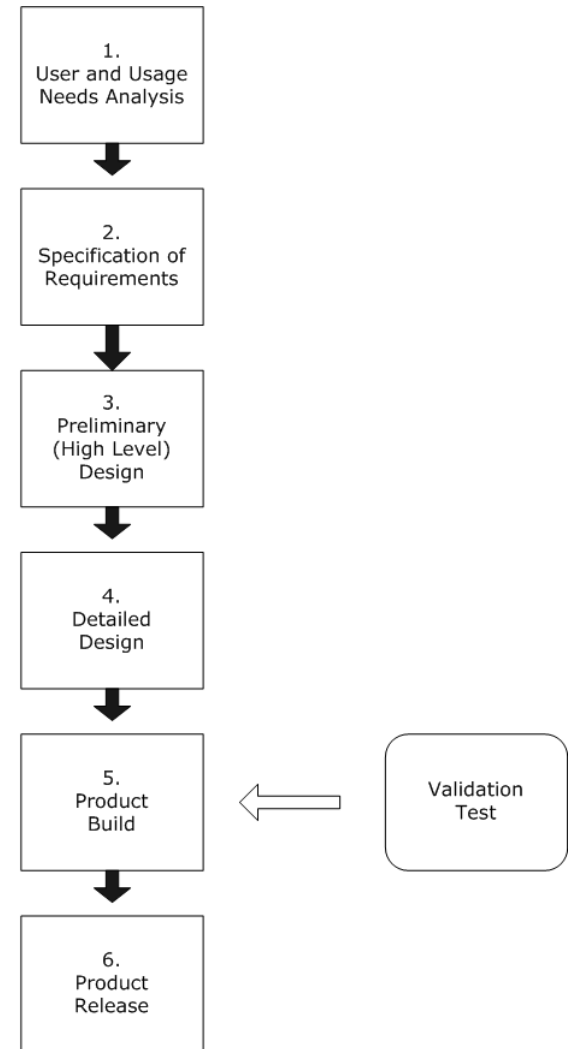
Assessment Test

- Formativ und summativ
- Mittlere Phase im Entwicklungsprozess
- Konzepte zum User Interface sind klar definiert
- Erfassung erster quantitativer Daten möglich (User performance)
- Erfüllung von vordefinierten Tasks
- Geringe Interaktion zwischen Testperson und TestleiterIn (eigentlich mehr BeobachterIn)



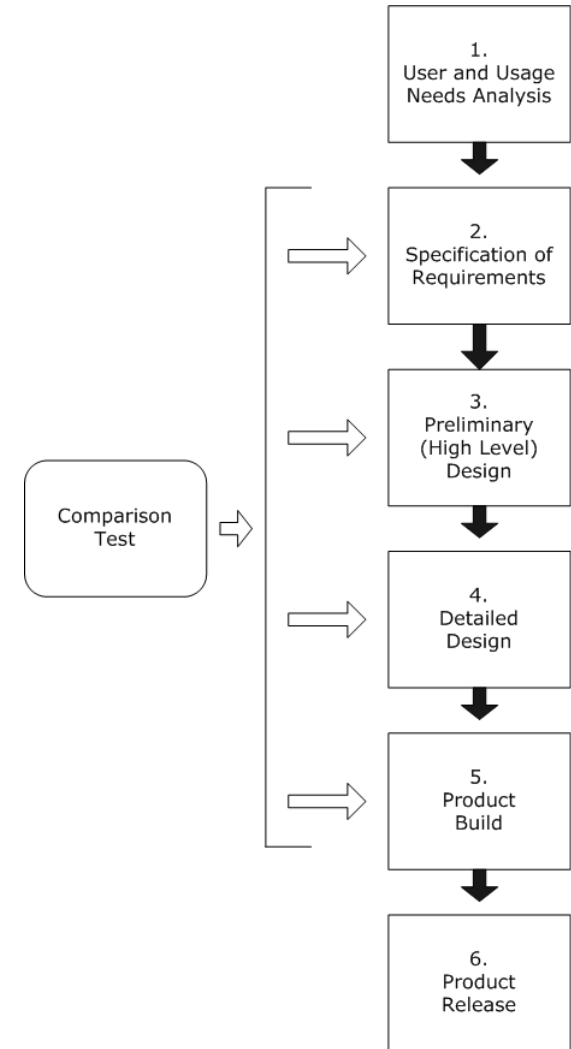
Validation Test

- Überwiegend summativ
- Fertiges Produkt
- Test gegen bestehende Standards (intern oder Konkurrenzprodukte)
- Benchmarking von Performance-Kriterien (Zeit, Fehlerrate, ...)
- (Fast) keine Interaktion zwischen Testperson und TestleiterIn



Comparison Test

- In jeder Phase des Design-Prozesses
- Zum Vergleich verschiedener Design-Konzepte oder mit Konkurrenzprodukten
- Wird in Verbindung mit den anderen drei Testtypen verwendet
- Methodik orientiert sich am Haupttesttyp



Vorgehensweise

Usability Testing

- Testpersonen repräsentieren reale User.
- Testpersonen führen realistische, praxisnahe Aufgaben durch.
- Es wird beobachtet und aufgezeichnet, was die Testpersonen machen und sagen.
- Die aufgezeichneten Daten werden analysiert, um die Ursachen der beobachteten Usability Probleme zu finden. Es werden Vorschläge zu deren Lösung gemacht.

Usability Testing: Flow of Information



NNGROUP.COM **NN/g**

Ablauf eines Usability Tests

- Grobe Unterteilung in 4 Phasen:
 1. Vorbereitung und Planung
 - Was muss alles für einen Test bedacht und im Vorfeld organisiert werden?
 2. Durchführung
 - Abwicklung der eigentlichen Test-Sessions mit den Testpersonen
 3. Auswertung
 - Sichtung und Analyse der gesammelten Daten
 - Erarbeitung von Verbesserungsvorschlägen
 4. Kommunikation der Ergebnisse / Reporting
 - Kunde/EntwicklerInnen werden mit den Ergebnissen vertraut gemacht

Aufbau Testbericht

Management Summary

Problemstellung (Produktbeschreibung, Testziele)

Methode (Testpersonen, Testszenarien, Testmaterial und Einrichtung, Design, Ablauf, generelle bzw. taskspezifische Anweisungen an die Testpersonen)

Ergebnisse (nach Testszenario gruppiert)

Analyse und Verbesserungsvorschläge

Conclusio

Anhänge (Interviewleitfaden, Fragebögen, Testmaterial, etc.)

Aufbau Testbericht: Vor dem Test

Management Summary

Problemstellung (Produktbeschreibung, Testziele)

Methode (Testpersonen, Testszenarien, Testmaterial und Einrichtung, Design, Ablauf, generelle bzw. taskspezifische Anweisungen an die Testpersonen)

Ergebnisse (nach Testszenario gruppiert)

Analyse und Verbesserungsvorschläge

Conclusio

Anhänge (Interviewleitfaden, Fragebögen, Testmaterial, etc.)

Vorbereitung und Planung

Everything needs a good plan

- „Good planning is absolutely necessary for a smooth and useful test. If you skimp on the planning, you will regret it during and after the test.“ (Dumas, Redish 1999)
- Was nicht funktioniert:
 - Wir holen ein paar Leute und schauen uns an, wie sie arbeiten.
- Warum? Es ist unklar,
 - ob „ein paar Leute“ der Zielgruppe entsprechen
 - welche Aufgaben die User lösen wollen / sollen
 - welche Informationen durch die Beobachtung gesammelt werden sollen
 - wie man die gewonnen Informationen analysiert und was man daraus lernen will

Vorbereitung und Planung

- Umfang festlegen
- Testziel festlegen
- Metriken und gesammelte Daten festlegen
- Zeitplan erstellen
- Taskliste und Szenarien ausarbeiten
- Benutzerprofil definieren und Testpersonen rekrutieren
- Testraum und Equipment vorbereiten
- Testablauf planen und Test-Team einweisen

Umfang

- Welches Produkt will man testen?
- Auf welcher Plattform / in welcher Betriebsumgebung will man das Produkt testen?
- Welche Teile des Produkts will man testen?
- Welche Teile des Produkts will man nicht testen?
- Welche Benutzergruppen nutzen diesen Teil des Produkts?
- Welche Aufgaben kann man mit diesem Teil des Produkts erledigen?

Ziel des Tests

- Abstrakte Fragestellungen als Ausgangspunkt, aber oft zu unpräzise
 - „Ist das System gut benutzbar?“
 - „Ist das System effizient zu verwenden?“
- Besser: Als konkrete Fragestellung formuliert (Beispiele)
 - Können BenutzerInnen das System selbständig und ohne Hilfestellung in Betrieb nehmen?
 - Wissen BenutzerInnen, was sie mit dem System machen können?
 - Können BenutzerInnen mit dem System [Aufgabe X] selbständig erfüllen?
 - Ist die Antwortzeit des Systems ausreichend schnell?
 - Sind die Icons verständlich?
 - Sind die vorhandenen Konfigurationsmöglichkeiten verständlich?
 - Sind die vorhandenen Konfigurationsmöglichkeiten ausreichend?
 - Lässt sich das System ausschließlich mit einem Touchscreen bedienen?
 - Verwenden die BenutzerInnen Tastatur-Shortcuts?
 - ...

Metriken und gesammelte Daten

- Qualitative Daten:
 - Beobachtungen (des Teams)
 - Aussagen, Kommentare, Anekdoten, Beschwerden, ... (der TeilnehmerInnen)
- Quantitative Daten:
 - Anzahl der erfolgreich abgeschlossenen Tasks / Erfolgsrate
 - Fehlerrate (kritische / nicht-kritische)
 - Anzahl der benötigten Hilfestellungen
 - Zeit welche für einen Task benötigt wird
 - Subjektive Eindrücke und Zufriedenheit (mittels Fragebogen)
 - ...

Zeitplan

- Wann ist das Produkt bereit für den Test?
- Wann werden die Ergebnisse benötigt?
- Wann und wo soll der Test stattfinden?
- Wie lange dauert eine einzelne Test-Session?
- Wie viele Test-Sessions pro Tag sollen stattfinden?
- Wann haben die TeilnehmerInnen Zeit für den Test?
- Wann und mit wem werden die Ergebnisse analysiert?
- Wann werden die Ergebnisse präsentiert?

Taskliste

- Tasks sind vordefinierte Aufgaben zur Bearbeitung durch die Testpersonen
 - Tasks beschreiben und lenken die Interaktion von Testpersonen mit dem System
- Basierend auf Anforderungen der BenutzerInnen (nicht technischen oder geschäftlichen)
 - Relevante Tasks in Abhängigkeit der Testziele auf Basis realistischer Szenarien identifizieren
- Tasks für kritische Aspekte, zur Beantwortung offener Fragen oder zur Erkundung bekannter Usability Probleme

- Daumenregel: ~10 Tasks, max. 60 min

Taskliste

- Herausforderung: Offen oder spezifisch?
 - Tasks sollten ausreichend offen sein, um eine autonome Lösungsfindung der Testpersonen zu ermöglichen
 - Tasks sollten ausreichend spezifisch sein, um den Fokus auf relevante und kritische Aspekte zu lenken
- Daumenregel:
 - Eher offene Tasks zur Erhebung qualitativer Daten
 - Eher spezifische Tasks zur Erhebung quantitativer Daten

Taskliste

- Formulierung der Tasks ausgehend von einem realen Szenario
 - Beispiel Szenario als Ausgangspunkt:
“Die Batterien in meiner Fernbedienung werden langsam leer. Ich will mir online Batterien bestellen, ich will möglichst wenig dafür bezahlen, und ich will den Versand gratis.”
- Herausforderung: Offene, realistische Szenarien haben manchmal keinen klaren Fokus, wenig Struktur und keine eindeutige Zielsetzung.
 - Beispiel: Eine Testperson bestellt nicht die billigsten, sondern die zweitbilligsten Batterien - ist das Ergebnis als Erfolg, Teilerfolg oder Misserfolg zu bewerten?

Taskliste

- Lösungsansatz: Allgemeine Szenarien in spezifische Tasks zerlegen
 - Vorteil: Besser vorhersehbar und überprüfbar
 - Nachteil: Autonomen Handlungsspielraum der Testpersonen wird eingeschränkt
- Ausgangspunkt: Zerlegung des Szenarios in Teilkomponenten
 1. Registrierung und/oder Login
 2. Passendes Produkt finden
 3. Mehrere passende Produkte vergleichen
 4. Produkt in den Warenkorb legen
 5. Versandkosten ermitteln
 6. Bestellung absenden
 7. ...

Taskliste

- Zusammenfassung der Teilkomponenten zu konkreten Tasks [die jeweiligen Teilkomponenten stehen in Klammer]:
 1. Registrierung eines / einer neuen BenutzerIn [1]
 2. Login eines / einer bestehenden BenutzerIn [1]
 3. Produkt bestellen [2, 4, 6]
 4. Mehrere Produkte miteinander vergleichen [2, 3]
 5. Versandkosten ermitteln [5]
- Bei der Definition von Tasks darauf achten, dass diese...
 - ... klar abgegrenzt sind und ein klar definiertes Ziel haben
 - ... ein realistisches Anwendungsszenario abbilden
 - ... keine schrittweisen Handlungsanweisungen enthalten

Taskliste: Beispiel-Task

Task Komponente	Beschreibung
Task	Produkt bestellen
Vorbedingung	Das Produkt (Batterien) ist im System vorhanden. Die Startseite des Online-Shops ist im Browser geöffnet. Der Benutzer ist eingeloggt und sämtliche relevanten Daten (Name, Adresse, Zahlungsinformation) sind bei dem verwendeten Account hinterlegt.
Endzustand	Die Bestellung wurde erfolgreich abgeschickt.
Benchmark	10 Minuten
Script / Anweisung an Testperson	Sie bemerken, dass die Batterien in Ihrer Fernbedienung langsam schwächer werden. Bestellen Sie 4 Stück neue AA-Batterien zu sich nach Hause.

Taskliste: Beispiel-Task

<ul style="list-style-type: none">■ Der Testperson wird nur das Script / Anweisung an die Testperson vorgelegt (!)	Produkt bestellen
<ul style="list-style-type: none">■ Alle anderen Informationen dienen nur der Planung und internen Verwendung durch das Test-Team■ Das Script sollte keine schrittweise Anleitung oder Hinweise darauf enthalten, wie der Tasks zu erledigen ist<ul style="list-style-type: none">■ Was zu tun ist■ Nicht wie man es macht	Online-Shops ist im Browser geöffnet. Der Benutzer ist eingeloggt und sämtliche relevanten Daten (Name, Adresse, Zahlungsinformation) sind bei dem
Endzustand	Die Bestellung wurde erfolgreich abgeschickt.
Benchmark	10 Minuten
Script / Anweisung an Testperson	Sie bemerken, dass die Batterien in Ihrer Fernbedienung langsam schwächer werden. Bestellen Sie 4 Stück neue AA-Batterien zu sich nach Hause.

Lerneffekte: Balancing der Taskreihenfolge

- Aus der Taskreihenfolge können sich Lerneffekte ergeben
- Vermeidung von Lerneffekten: Balancieren der Task-Reihenfolge mittels **Latin Square Balancing**

Testperson	1	2	3	4
<i>TP1</i>	A	B	C	D
<i>TP2</i>	B	D	A	C
<i>TP3</i>	D	C	B	A
<i>TP4</i>	C	A	D	B

Task

- Balancieren der Task-Reihenfolge ist nicht immer notwendig, aber manchmal sinnvoll (abhängig von Testdesign und Testziel)

Lerneffekte: Comparison Testing

- Within-Subject vs. Between-Subject Design
- Beispiel: Kann man mit Maus oder Touchscreen das Interface schneller bedienen?
 - Unabhängige Variable – Eingabegerät (Maus/Touchscreen)
 - Abhängige Variable – Zeit für die Erfüllung

- Within Subject Design:

- Jede TP in jeder Bedingung
- Vorteil: weniger Testpersonen
- Nachteil: Lerneffekte
(Lösung: Balancing – Latin Square)

	Maus	Touchscreen
TP1	12 Sek.	17 Sek.
TP2	19 Sek.	15 Sek.
TP3	13 Sek.	21 Sek.

- Between Subject Design:

- Jede TP nur in einer Bedingung
- Nachteil: Gruppen müssen vergleichbar sein

	Maus		Touchscreen
TP1	12 Sek.	TP2	17 Sek.
TP3	19 Sek.	TP4	15 Sek.
TP5	13 Sek.	TP6	21 Sek.

Auswahl der Testpersonen

- Repräsentative User der Zielgruppe → Benutzerprofil
 - Benutzerprofil im Testplan
 - Evt. Screening Fragebogen für die Auswahl der Testpersonen
 - Wichtig für aussagekräftige Ergebnisse (Validität!)
- In verschiedene Kategorien unterteilen
 - z.B. AnfängerIn, Experte/Expertin
 - Personas Konzept (Cooper, 1999)
- Rekrutieren
 - Freiwillige
 - Bezahlte (Motivation)

Benutzerprofil: Beispiel

Charakteristik	Kriterium	TeilnehmerInnen
Android-Erfahrung	Über ein Jahr	45% 1 - 3 Jahre 35% 3 - 5 Jahre 20% 5+ Jahre
Alter	18+ Jahre	Mittelwert: 26,7 Median: 24 Min: 18 Max: 47
Geschlecht	Möglichst gleich verteilt	45% Männlich, 55% Weiblich
Erfahrung mit dem Produkt	Nicht vorhanden	70% kannten das Produkt nicht 30% hatten davon gehört, aber es noch nie benutzt

Anzahl der Testpersonen

“The best results come from testing no more than 5 users and running as many small tests as you can afford.

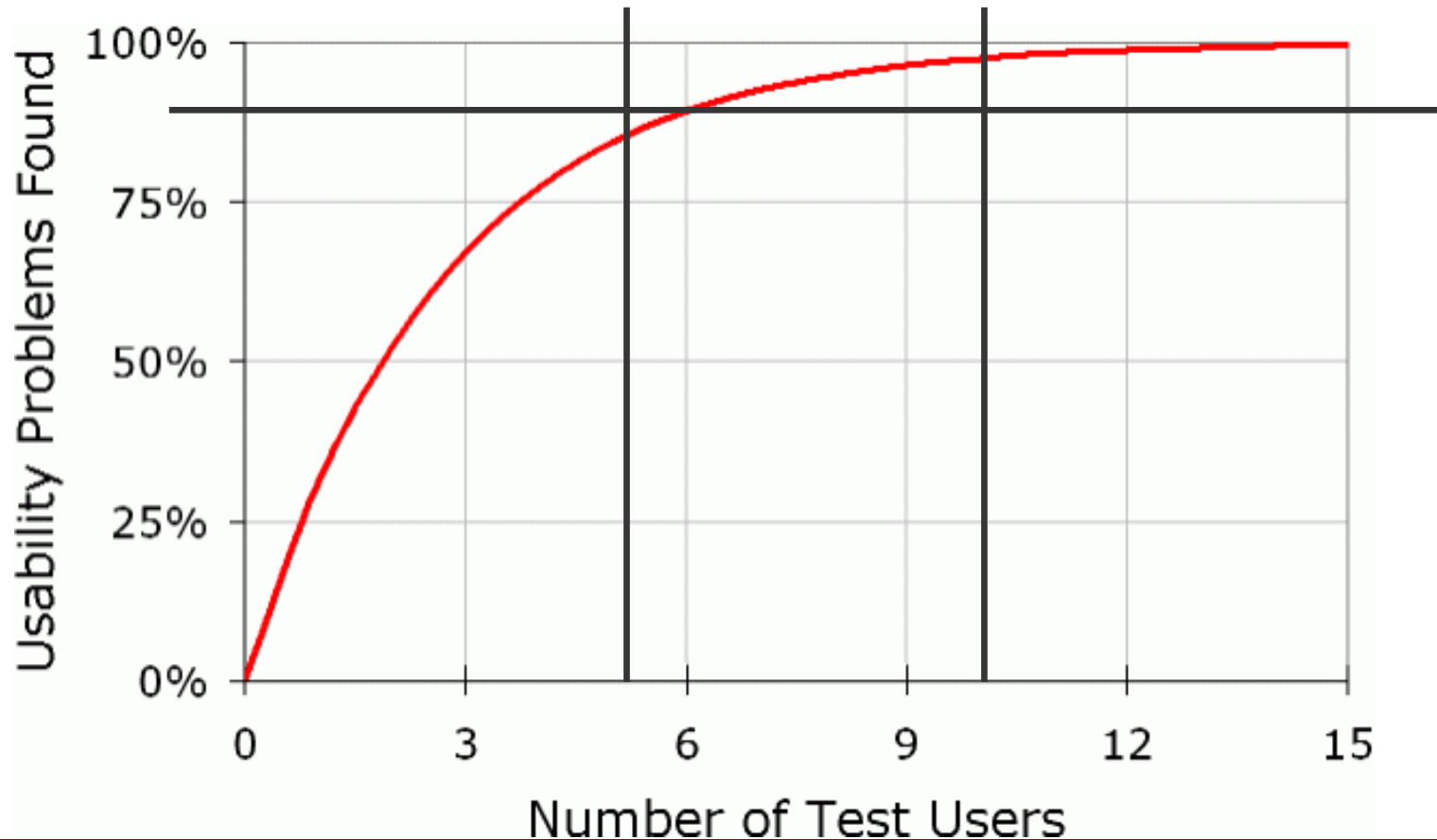
[T]he number of usability problems found in a usability test with n users is:

$$N(1-(1-L)^n)$$

where N is the total number of usability problems in the design and L is the proportion of usability problems discovered while testing a single user. The typical value of L is 31%, averaged across a large number of projects we studied.”

Anzahl der Testpersonen (Nielsen, 2000)

- Plotting the curve for $L=31\%$ gives the following result:



Anzahl der Testpersonen

- Für qualitative, formative Tests: 3-5 Testpersonen
- Bei mehr Ressourcen: Mehrere Tests ...
 - ... mit unterschiedlichen Testzielen oder
 - ... in mehreren Iterationen
- RITE: Rapid Iterative Testing & Evaluation
 - Testen bis man ein Problem gefunden & bestätigt hat (ggf. nach nur einer Testperson)
 - Direkte Behebung des Problems bevor man mit den Tests fortfährt
- Für Benchmarking, wissenschaftliche Experimente, ...: 20+

Testumgebung und Equipment

- Equipment zur Durchführung des Tests
- Equipment zur Datenerfassung
 - Abhängig von Art der erhobenen Daten

- Hardware
- Software
- Testumgebung



Quelle: <http://www.flickr.com/photos/nzdave/491411546/>

Testumgebung: Beispiel

“Die Testumgebung besteht aus einem abgetrennten Büro mit Schreibtisch und zwei Beobachter-Stühlen. Der Computer ist ein Laptop mit 13,3” Screen (1280x800 Auflösung), QWERTZ-Keyboard, Touchpad und einer Zwei-Button-Mouse mit Scrollrad.

Als Betriebssystem kommt Mac OS X 10.15 zum Einsatz. TeilnehmerInnen haben die Wahl zwischen der aktuellsten Version von Firefox, Chrome oder Safari je nach persönlicher Präferenz und Erfahrung. Der Browser ist zu Testbeginn mit der Startseite www.google.com geöffnet.

Audio und Video werden während des Tests mit der im Laptop integrierten Videokamera bzw. Mikrofon aufgezeichnet. Bildschirminhalte werden mit Quicktime aufgezeichnet. Zeitmessungen werden manuell von einem Beobachter durchgeführt.”

Ablauf der Test-Sessions: Beispielplanung (1)

Der Ablauf einzelner Test-Sessions gliedert sich in vier Phasen:

1. Begrüßung und Background-Fragebogen

Jede/r TeilnehmerIn wird persönlich vom Testleiter begrüßt und gebeten, den Background-Fragebogen (Anhang A) auszufüllen.

2. Orientierung

Jede/r TeilnehmerIn erhält eine verbale Einführung (Orientierungsskript, Anhang B), welche den Zweck und Ziel des Tests beschreibt. Anschließend wird den TeilnehmerInnen das Test-Equipment vorgestellt und sie werden darüber informiert, dass während des Tests Audio- und Video-Aufzeichnungen gemacht werden.

Ablauf der Test-Sessions: Beispielplanung (2)

3. Durchführung der Test-Session

Der Testlauf besteht aus einer Reihe von vordefinierten Tasks, welche die TeilnehmerInnen selbständig unter Beobachtung durchführen. Die Tasks werden folgendermaßen durchgeführt:

- Der / die TestleiterIn liest den Task vor und anschließend wird dem/der TeilnehmerIn die Aufgabe schriftlich (separat für jeden Task) ausgehändigt. Danach versucht der/die TeilnehmerIn, den Task erfolgreich durchzuführen.
- Nach der Erledigung aller Tasks füllt der/die TeilnehmerIn einen Post-Test-Fragebogen aus.
- Während des Testlaufs dokumentiert der/die TestleiterIn Fehler, Beobachtungen und verstrichene Zeit.

Ablauf der Test-Sessions: Beispielplanung (3)

4. TeilnehmerInnen Debriefing

Debriefing durch den/die TestleiterIn. Die Themen der Fragen umfassen:

- Gesamt-Zufriedenheit des Teilnehmers/ der Teilnehmerin mit dem User Interface
- Kommentare und Empfehlungen des Teilnehmers/ der Teilnehmerin betreffend User Interface
- Antworten auf Fragen des Testleiters/ der Testleiterin bezüglich spezifischer Fehler und Beobachtungen, die während des Tests aufgetreten sind

Rollen der Teammitglieder

- TestleiterIn (Verantwortlicher)
 - ModeratorIn (Kommunikation mit Testpersonen)
 - ProtokollantIn
 - ZeitnehmerIn
 - Video / Audio OperatorIn
 - Produktspezialist / technischer SpezialistIn
 - Zusätzliche Testrollen
 - Testbeobachter
-
- Es können natürlich auch mehrere Rollen von einer Person übernommen werden

Aufgaben der Teammitglieder: Beispiel

„Der/die **TestleiterIn** sitzt mit dem/der TeilnehmerIn **in einem Raum** während der Durchführung des Tests. Der/die TestleiterIn **startet die einzelnen Tasks** nach erfolgreicher Vorbereitung und **dokumentiert Fehler, Beobachtungen** und **Zeitmessungen**. Der/die TestleiterIn leistet **keine Hilfestellungen** außer der/die TeilnehmerIn weicht **gravierend vom Lösungsweg ab** und beantwortet nur Fragen, die die Testergebnisse nicht beeinflussen. Der/die TestleiterIn steht im Fall von **Equipment-Fehlern** und technischen Problemen zur Verfügung. TeilnehmerInnen erhalten die Aufgabenbeschreibung in **schriftlicher Form** und sind angehalten, die Aufgaben **selbständig und möglichst rasch** mit dem vorhandenen Test-Equipment durchzuführen.“

Vor Durchführung des Usability Tests

- Den Test selbst durchführen
 - Zeitplan (Session)
 - Testdesign
- Einen Pilottest durchführen
 - Tasks
 - Testplan
 - Material
- Das zu testende Produkt überprüfen
 - Bugs sollten nicht überraschend während des Tests auftreten
- Testzubehör und Testumgebung überprüfen



A large, bold, red question mark is centered on a white background. The question mark is composed of a thick, curved line that forms the main body, and a solid red dot below it. The background is white, with a solid red horizontal bar at the top and another solid red horizontal bar at the bottom, framing the central white area.



deco.inso.tuwien.ac.at

