



Usability & User Experience Engineering

Einheit 7: Evaluation

Univ. Ass. Alexander Meschtscherjakov



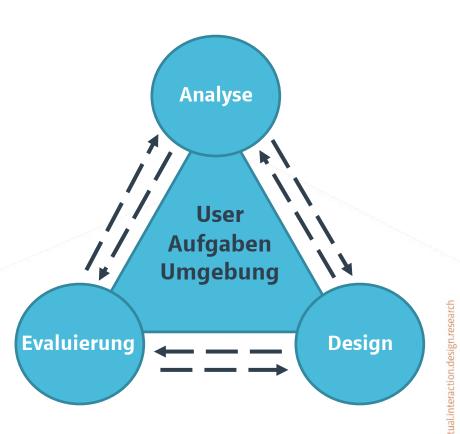
EVALUATION

Was ist User-Centered Design?



⇒ Entwicklungsansatz in Human-Computer Interaction

Ein interaktiver Designprozess für interaktive Produkte und Anwendungen, in dem User, ihre Aufgaben und ihre Umgebung eine zentrale Rolle spielen bei der Analyse der Anforderungen, der Entwicklung alternativer Design und bei der Evaluierung statischer oder interaktiver Prototypen.



Evaluierung (I)



Evaluierung ist eine Qualitätssicherung

- Konzept
- Design
- Prototypen

Implementierung

- Aufdeckung konkreter (Usability)
 Probleme
 - Strukturen
 - Abläufe
 - Navigation
 - Layout
 - Textierung
 - Grafiken ...
- Erfassen der Ursachen für die Probleme
- Identifizierung der optimalen
 Verbesserungsmöglichkeiten unter den gegebenen Rahmenbedingungen

Evaluation



Schnelles, erstes und objektives Feedback auf Conceptual Model

Design Alternativen

Erste objektive Evaluation sichert langfristig Zeit und Geld

Änderungen sind früh günstig und einfach durchzuführen

Evaluation



Testziel entscheiden
Basiert auf User Profile und Usability Goals

Entscheidung hinsichtlich User, Aufgabenfokus

Typ und Bandbreite der Benutzer

Zentrale, häufige, wichtige Aufgaben

Design von Testaufgaben, Entscheidung von Fokusierung Bezugnahme auf Analyse Ergebnisse

Evaluierung (II)



- Unterscheidung Wer:
 - Experten (Usabilty)
 - Nutzer
- Unterscheidung **Wo**:
 - Labor
 - Feld
- Unterscheidung Was:
 - Qualitativ
 - Quantitativ
- Unterscheidung Wann:
 - Formativ: während der Entwicklung (iterativ)
 - Summativ: am Ende der Entwicklung

2 (Hauptarten) von Evaluation



Experten basierte Evaluation

Guidelines/Heuristics

Walkthroughs

Video Analyse

User basierte Evaluation

Fokus Gruppen

Fragebögen

User Tests / Experimente / Feld Studien

Observation

Auswahl Evaluierungsmethode



Auswahl der Methode nach unterschiedlichen Aspekten: Fokus auf Usability, User Experience, User Acceptance Zeitaufwand bzw. Detail der Ergebnisse Qualitative vs. Quantitative Daten Labor vs. Feld

...

Es gibt nicht DIE richtige Methode, sondern pro Fall muss der richtige "Methoden-Mix" definiert werden.

Jede Methode birgt "Fallen", welche zu falschen Ergebnissen führen können.

- z.B. zu viel Hilfestellungen beim Usability Test beeinflusst die Erfolgsrate bei den Aufgaben (Effektivität)
- z.B. Auswahl unpassender Testpersonen (falsches Rekrutierungsprofil)



Expertenevaluierung

Expertenevaluierung



"Discount Evaluation Technique" ("Vorhersage" von Usability Problemen) mit einer kleinen Anzahl an Experten

Usability Experten analysieren ein System anhand einer Liste von Regeln oder Guidelines.

Heuristische Evaluierung (Nielsen) http://www.useit.com/papers/heuristic Walkthrough (kognitiv, pluralistisch, gruppen-basiert)

Meist durchgeführt vor User Studien im Labor.

Für Probleme, die man selbst finden kann, braucht man keine Userstudie Nur nützlich und sinnvoll wenn Probleme noch geändert werden vor dem User Studien.

ABER nicht geeignet als Ersatz, wenn keine User verfügbar sind.

Ziel

Finden von Problemen (Verletzungen der Prinzipien/Heuristiken) Finden von Problemlösungsvorschläge/Verbesserungsvorschlägen

Experten Evaluation



Basiert auf Experteneinschätzung

Evaluatoren prüfen Usabilityaspekte eines UI

2 bis 4 Experten

Methoden für implementierte Uls als auch Ul

Prototypen anwendbar

Effektiv in frühen Designstadien, kein Ersatz für User Tests

Experten Evaluation



Verschiedene Methoden

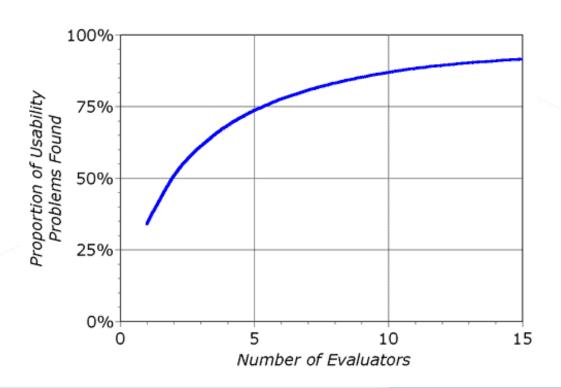
Unterschiede in der Durchführung Unterschiede hinsichtlich Kriterien

Wichtigsten Probleme können erkannt werden Experten Evaluationen können innerhalb 2 bis 4 Tagen durchgeführt werden

Expertenevaluierung



3-5 Experten finden 75% der Usability Probleme (Nielsen)



Experten Evaluation



Ziele

Probleme in einem existierenden User Interface Design finden

Empfehlungen für Verbesserungen

Problem/Lösungsrationalität

Reihung der Probleme

Experten Evaluation



Experten?

Usability

und / oder

Domaine / Fachlich

Zumindest Basiswissen in beiden Bereichen notwendig



Beispiel einer Experten-basierten Methode

HEURISTISCHE EVALUIERUNG

Heuristische Evaluierung



"Discount Evaluation Technique" ("Vorhersage" von Usability – predictive method)

Kleine Anzahl von Experten/Evaluatoren

Kennen das Interface noch nicht

Verständnis von User Perspektive

Interface wird ausprobiert und dabei mit gültigen Usability-Prinzipien (Heuristiken) verglichen

Ziel:

Finden von Problemen (Verletzungen der Prinzipien/Heuristiken) Problemlösungsvorschläge

Expertenevaluierung



Vorgehen (Dauer: ca. 1 – 1,5 Stunden)

Inputs definieren (Szenarien, Tasks, Heuristiken, Guidelines)

Briefing (Testleiter führt den Experten in die Methode ein)

System evaluieren mit Heuristiken und Guidelines (Experten arbeiten alleine)

Nielsen 10 Usability Heuristics

Norman 7 Design Principles

Shneiderman's 8 Golden Rules

Datensammlung (aller Probleme) und Analyse (Zusammenfassung)

Severity Ranking

Wie häufig tritt das Problem auf?,

Welche Auswirkungen hat es?

Wie hartnäckig ist das Problem?

Severity Rating

0 = I don't agree that this is a usability problem at all

1 = Cosmetic problem only: need not be fixed unless extra time is available on project

2 = Minor usability problem: fixing this should be given low priority

3 = Major usability problem: important to fix, so should be given high priority

4 = Usability catastrophe: imperative to fix this before product can be released

Heuristische Evaluierung



1: Briefing

Testleiter führt den Experten in die Methode ein (v.a. Prinizipien)

2: Evaluierung

Jeder Experte arbeitet alleine

Experte geht anhand von Liste von Usability Prinzipien das Interface mehrmals durch um Usability Probleme zu finden, zusätzlich fließt sein Expertenwissen mit ein

Resultate werden durch Experten oder durch externen Beobachter festgehalten

Unterstützung durch Verwendung von Szenarien und Aufgaben

Zwei Durchgänge:

1: Probleme im Ablauf (Interaktionsfluss)

2: Spezielle Interface Elemente

Heuristische Evaluierung



3: Auswertung der Ergebnisse

(Testleiter) → Problemliste

4: Severity Ranking

Wie häufig tritt das Problem auf? Welche Auswirkungen hat es? Wie hartnäckig ist das Problem?

Experten einzeln, dann Errechnung des Durchschnitts

5: Zusammenfassung & Diskussion Ergebnisse

(Testleiter + Experte)

untextual interaction design research

Severity Rating - Beispiel



- 0 = I don't agree that this is a usability problem at all
- 1 = Cosmetic problem only: need not be fixed unless extra time is available on project
- 2 = Minor usability problem: fixing this should be given low priority
- 3 = Major usability problem: important to fix, so should be given high priority
- 4 = Usability catastrophe: imperative to fix this before product can be released

Heuristische Evaluierung - Ergebnis



Liste mit Usability Problemen, inklusive Referenz zu verletzten Usability Problemen und Severity-Rating

Reihung der Probleme nach Problembereichen oder Schweregrad (Severity)

Unterstützung zb. durch Screenshots

Mögliche Lösungsvorschläge durch Experten integrieren

Nielsen's 10 Usability Heuristics

- 1. Visibility of the system status
- 2. Match between the system and the real world
- 3. User control and freedom
- 4. Consistency and standards
- 5. Flexibility and efficiency of use
- 6. Aesthetic and minimalist design
- 7. Error prevention
- 8. Recognition rather than recall
- 9. Help and documentation
- 10. Help users recognize, diagnose, and recover from errors



Online Quelle:

http://www.useit.com/papers/heuristic/heuristic_list.html
Jakob Nielsen: Heuristic Evaluation. In: Usability Inspection Methods. Ed. by Jakob Nielsen and Robert L. Mack. - New York u.a.: Wiley, 1994. - S. 47.

10 Usability Heuristics





Visibility

Show system status, tell what's happening



Mapping

Use familiar metaphors & language



Freedom

Provide good defaults & undo



Consistency

Use same interface and language throughout



Error Prevention

Help users avoid making mistakes



Recognition

Make information easy to discover



Flexibility

Make advanced tasks fluid and efficient



Minimalism

Provide only necessary information in an elegant way



Error Recovery

Help users recognize, diagonize and recover from errors



Help

Use proactive and in-place hints to guide users

Weitere Guidelines und Quellen



- 944 GUIDELINES FOR DESIGNING USER INTERFACE SOFTWARE (Sidney Smith & Jane Mosier, 1986)
- http://hcibib.org/sam/contents.html
- Online Tool
 http://www.usabilitest.com/ux6sUV9

Beispiele



Heuristic Evaluation

Issues and	Cuannes A	rane Enume	Heina Ha	wietie Eu	dution M	lathade
issues and	SUCCESS A	reas Found	i usina He	uristic Ev	aiution iv	etnoas

	Issue	Heuristics Violated	Severity
10	The user must initialize a grid before creating a grid. However, the system allows users to 'create grid isovist' before the user has clicked 'initialize grid.'	#3, #5, #9	4
31	When clicking on the "New" button or the same menu item, you will lose your current work without warning.	#5, #3	4
33	There are no coordinates displayed on the interface of the Syntax2D.	#1, #3	4
36	The system cannot recognize layers in a file. Users are required to specify layer names. The system does not provide feedback that the layer name the user specified was incorrect or undetected.	#5, #6, & #9	4
41	There is no measure shown by default after the path isovist is created.	#1	4
42	No "undo" function.	#9	4
46	There is no color spectrum on the left side of the screen to help the user understand the path isovist measures. This is inconsistent with the grid isovist presentation.	#1,#4,#6	4
47	Overall, errors dialog boxes do not provide helpful diagnostic information, for example when a user skips to the create path isovist without preparing the path or observation points, an error dialog box simply states "invalid path" rather than providing useful instructions.	#5, #9	4
50	System requires polylines but does not give the user feedback that he/she has uploaded a drawing without complete polylines. If there is a gap in between lines it floods the area with cells during grid initialization. The user must go back to AutoCAD, redo the drawing, then re-import to Syntax2D.	#4, #5 & #9	4
28	Cannot tell what measure you are looking at (default measure = area) after the grid isovist has been created	#1	3

Beispiele



Heuristic Evaluation Checklist

Heuristic	Rating		g	Comments	
Appearance/Aesthetics – First impression are im	portant	t-it	can	make the difference between users staying	
Primary goal/purpose is clear	*	1	×		
Clean, simple design	*	1	×		
Pleasing color scheme	*	1	×		
Appropriate use of white space	*	✓	×		
Consistent design	*	1	×		
Text and colors are consistent	*	✓	×		
Icons are universally understood	*	✓	×		
Images are meaningful and serve a purpose	*	1	×		

Major headings are easy to understand	*	1	×	
Easy to scan	*	1	×	
Minimal text/information presented	*	1	×	
Clear terminology, no jargon	*	1	×	
Links are clear and follow conventions	*	1	×	
Help is available on every page	*	✓	×	
Important content is above the fold	*	1	*	
Search box is easy to identify and easy to use	*	1	×	

Navigation - Makes getting around your site easy and takes out the guess work of a user's visit.

Consistent Navigation	*	1	*	
Easy to identify your location on the site (breadcrumbs, headers, colors)	*	1	*	
Consistent way to return Home	*	1	×	
Limited number of buttons & links	*	1	*	
Organization of information makes sense	*	1	×	

untextual interaction design research

Beispiele



Heuristics Evaluation of [enter product name]

By [Enter Your Name]

Date [Enter Date]

1. Visibility of system status

- Always keep users informed about what is going on.
- Provide appropriate feedback within reasonable time.

Evaluation

[Enter your observation and evaluation of the degree to which this Heuristic has been satisfied. Use as much space as you see fit.]

2. Match between system and the real world

- Speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms.
- · Follow real-world conventions, making information appear in a natural and logical order.

Evaluation

[Enter your observation and evaluation of the degree to which this Heuristic has been satisfied. Use as much space as you see fit.]

3. User control and freedom

- Users often choose system functions by mistake.
- Provide a clearly marked "out" to leave an unwanted state without having to go through an
 extended dialogue.
- Support undo and redo.

Evaluation

[Enter your observation and evaluation of the degree to which this Heuristic has been satisfied. Use as much space as you see fit.]

Vor- und Nachteile



Vorteile:

- Heuristiken helfen, sich auf bestimmte Themen zu konzentrieren
- Dadurch, dass keine NutzerInnen involviert sind, sind keine praktischen/ ethischen Probleme damit verbunden
- Einzelne Probleme können identifiziert werden, welche sich auf die gesamte UX auswirken können

Nachteile:

- Die richtige Wahl der Heuristiken ist wichtig, um keine Probleme zu übersehen
- Es können sehr zeit-intensiv werden, wenn die ExpertInnen erst trainiert werden können
- Vorerfahrungen, was "gute Usability" ist, können Verzerrungen hervorrufen
- Probleme werden "gemacht"

Heuristische Evaluierung



Führen Sie auf der Site http://www.volkswagen.at/ mit Car Configurator eine Heuristische Evaluierung durch

Zeit: 10 Minuten

Danach: besprechen wir die Usability Fehler und reihen diese.



Beispiel einer Nutzer-basierten Methode

USABILITY TESTING / TEST

Was ist Usability Testing?



Usability testing ist ein Verfahren um zu gewährleisten dass Nutzer eines Systems die beabsichtigten Aufgaben mit dem System effizient, effektiv und zufriedenstellend ausführen können.

Warum Usability Tests machen?



Ziel ist es, die Benutzerfreundlichkeit eines Produkts zu verbessern

"Echte" Nutzer als Teilnehmer, welche "echte" Aufgaben durchführen -> realistisches Nutzerverhalten beobachtbar

Warum Usability Tests machen?



Beobachtetes Nutzerverhalten zeigt Usability Probleme auf, die einen Entwickler möglicherweise verborgen bleiben

Um herauszufinden welche Interaktionsschritte für Nutzer einfach oder schwierig sind, und welche Vorstellungen Nutzer über die Interaktion mit einem System haben

ontextual interaction design researc

Wann im Designprozess?



Iterativ und in verschiedenen Phases der System-Entwicklung

Vor dem Start des Designprozesses
Teste ähnliche Systeme
Teste bereits existierende Systeme (wenn dieses ersetzt werden soll)

Zu Beginn des Designprozesses Evaluierung von low-fidelity (Papier)Prototypen

In der Mitte des Designprozesses Teste unterschiedliche Aspekte des Designs Teste Veränderungen im Design

Am Ende des Designprozesses Überprüfe ob die Usability Ziele erfüllt werden Vergleiche mit ähnlichen oder vorherigen Systemen

ontextual interaction design research

Wie wird's gemacht? - Zentrale Schritte



Check ob Voraussetzungen gegeben sind (kann ich das im System testen, was ich testen will?)

Erstellung des Test-Plans

Vorbereitung der Test-Materialien & der Testumgebung

Durchführung von Pilot-Tests

Rekrutierung der Teilnehmer

Durchführung der Tests

Analyse der gesammelten Daten

Berichterstattung

Analyse



Analyse des Systems / Benutzergruppen

Ziele der Benutzer

Was ist für den Erfolg relevant

Systemteile die getestet werden sollen

Welche Probleme können erwartet werden

Welche Designfragen sollten geklärt werden

Informationen führen zu

Aufgaben definition

Interviewfragen, Fragebögen

extual.interaction.design.research

Aufgaben



Was sind gute Aufgaben für einen Test?

Basiert auf Zielen der Benutzergruppen

Beantworten kritischer Erfolgsfragen

Nicht zu generell, nicht zu spezifisch

Bekannte Anzahl an möglichen Lösungen

Hat Endpunkt der vom Benutzer erkannt werden kann

Untersucht Aktionen, nicht nur Meinungen

Entweder kritisch oder oft ausgeführt

Aufgaben



"A good task to select is one that has the potential to uncover a usability problem"

A Practical Guide to Usability Testing (Dumas/Redish 1999)

ontextual interaction design research

Wer nimmt am Usability Test teil?



Eine Auswahl an Nutzern, die der/den beabsichtigten Nutzergruppe(n) entsprechen und das sinnvollste Feedback geben.

Beispiel: Usability Test einer Website für den Verkauf von hochwertigem Besteck -> die idealen Testteilnehmer sind Personen, die hochwertiges Besteck kaufen wollen

Beim Rekrutierungsprozess ist wesentlich ein klar definiertes "Nutzerprofil" zu verwenden

contextual interaction design research

Wieviele Teilnehmer?



Zumindest 5 Teilnehmer

Ziel des Test ist nicht eine statistische Auswertung!

Nach 6 Teilnehmern kommen eher die gleichen Usability Probleme zum Vorschein

Jeweils pro Zielgruppe!

Eine iterative Evaluierung ist wichtig um alle Usability Probleme zu finden

Wo findet ein Usability Test statt?



In der "realen" Nutzungsumgebung

Dort wo der Nutzer im Alltag ein

System nutzt

Bequem für den Test-Nutzer

Testbedingungen schwierig zu

kontrollieren

Mögliche Unterbrechungen des Tests

Videoaufnahme möglicherweise

unangenehm für Nutzer

. . .

Im Usability Labor

Künstliche Umgebung

Unbequem für den Test-Nutzer (muss

zum Lab kommen)

Testbedingungen sind gut zu

kontrollieren

Keine Unterbrechungen

Einfache Videoaufnahme

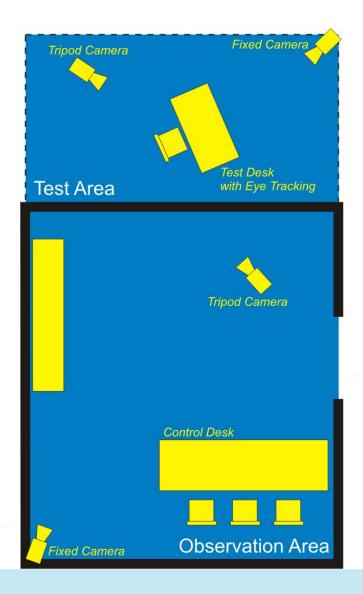
...

... Und noch viel mehr zu bedenken. Trend zu Usability Laboren, die reale Nutzungsumgebungen immer besser simulieren.

contextual.interaction.design.research

Usability Lab Umgebung

















contextual.interac

contextual.interaction.design.researc

Phasen eines Usability Tests



Briefing Phase

Pre-interview

Task Phase

Fragebögen

Qualitative Interviews

Personen



Testleiter

Führt durch den Usability Test Die einzige Person die in Kontakt mit der Testperson kommt Löst alle Probleme

Beobachter

Macht Notizen

Bedient Technik (Kamers, DVD-Recorder)

Unterbricht bei kritischen Problemen (z.B. wenn Mikrofon nicht funktioniert)

Verpflegung



TESTPLAN

ontextual interaction design research

Testplan



Ziele des Tests

Was sind die (quantitativen) Ziele für diese Kriterien?

Spezifiziere Methoden und Techniken

z.B. Thinking aloud

Benutzeprofile (Anzahl der Benutzer, Screening Kriterien)

Ablauf des Tests im Detail (Briefing, Task Scenarios, interview)

Zeitablauf planen

Rollen definieren: Testleiter, Assistent, Technischer Support

Definieren des Testortes, des Testzeitpunkts, der Testdauer, der

Aufwandsentschädigung

Rekrutierungsplan (Quellen, Datenbanken, etc.)



Kurze Testbeschreibung	XIII II	
Auftraggeber	ANET	
Name des Projekts	ETC mobility lab	
Inhalt (Was wird getestet)	etPlanner	

Personenanzahl	12 Personen:					
		weiblich	männlich			
	15 bis 24 Jahre	1 Person	1 Person			
	25 bis 34 Jahre	1 Person	1 Person			
	35 bis 49 Jahre	2 Personen	1 Person			
	50 bis 59 Jahre	1 Person	2 Personen			
	Über 60 Jahre	1 Person	1 Person			
Puffer-Personen Anzahl: 1-2 Testpersonen mehr rekrutieren, um Ausfallssicherheit zu haben. Diese Testpersonen sollten "uns" gewogen sein, sie werden rechtzeitig (24h vorher) darüber unterrichtet, dass der Test eventuell nicht stattfindet.	1x Mann über 50; Pil	occi nuizen für f	crationen			
Kriterien (Alter, Geschlecht, Nutzung, Sprachen etc.); z.B. 10 TPn 5 Kunden (3 m, 2w), 5 Nicht-Kunden (2m, 3w), Interesse an mobilen Services, Interesse an Musik	- 50% männlich/ 50% - eine gleichmäßige / Jahren besteht - eine häufige (täglich besteht - eine häufige (täglich eine zumindest selt Nutzung von MMS einemand aus Tirola alle potentielle Innsalle Englisch-Basis	Altersverteilung v he) Nutzung des he) Nutzung von ene (mind. 1x ver der mobilem Into abstammt abruck-Touristen	Mobiletelefons SMS besteht wendet) ernet besteht sind			
Benutzergruppen						
Besondere Anforderungen (z. B. Exectracking – keine Brillenträger)	keine					

Testleiter: Welche Mitarbeiter testen (wann, in welchem Ausmaß)	AW, AB, EB Es wird als 2er-Team getestet an allen Testtagen; Testleitung: EB; Technikassistenz: AW, AB
Ort der Testdurchführung (Labor, im Feld)	Lab
Labor: In welchem(n) Labor(s) wird getestet (Wird parallel getestet?)	ICT&S Usability Lab
Falls im Feld: Treffpunkt, evtl. Mobilnummer des Testleiters	

Testzeitraum (Datum, KW)	23. Oktober – 31. Oktober
(, , , , , , , , , , , , , , , , , , ,	2 Stunden Takt:
Testdauer (Stunden) und Abstand	
	 90 Minuten pro Test
	 30 Minuten Pause zwischen den Tests
Zeiträume, die nicht für Tests zur Verfügung stehen	Do 26.10, Fr. 27.10. und 1.11. (Feiertag)
(neben Berücksichtigung der Kalender der Testleiter)	
Kalender: Sollen besondere Infos in den Kalender	
eingetragen werden? (z. B. Gruppe, Organisation)	
Aufwandsentschädigung: Betrag	€ 20 pro TP
Testmitschrift: Pfad zur xls-Testmitschrift (Eintragen	
der Testpersonen und ihrer Daten)	

Rekrutierung		
Aus welchen Qu	ellen wird rekrutiert:	
•	Kontaktdaten vom Aufraggeber	
•	Datenbank	Interne Kontaktdatenbank
•	Private Kontakte von ICT&S-	Bekannte, Freunde der ICT&S Mitarbeiter Mitarbeiter.
	Mitarbeitern	
•	Sonstige	E-Mail-Verteiler (RB Studenten)
•	Ausschließungsgründe (Mitarbeiter	
	des Auftraggebers, Testpersonen des	
	vorherigen Tests, nicht aus der	
	Testdatenbank etc.)	

Testablauf					
Maximale Aufgabendauer	Aufgabe 1: Aufgabe 2: Aufgabe 3: Aufgabe 4: Aufgabe 5: Aufgabe 6:				

Report				
Report	Report auf Deutsch			

contextual.interaction.design.researd

Testplan Beispiel



http://www.usability.gov/how-to-andtools/resources/templates/usability-test-plantemplate.html

contextual interaction design research

Entwicklung des Test-Plans



Test-Ablauf

Einführung in den Testablauf, Zweck des Tests

Datenverwertungserlaubnis (Videoaufzeichnung)

Vorinterview Fragen (und/oder Vor-Fragebogen)

Test-Aufgaben und Szenarios

Abschluss-Fragebögen

Abschlussinterview Fragen

Entwicklung des Test-Plans



Definition der Testteilnehmer

Männlich / weiblich?

Altersgruppen?

Internet-Erfahrung: Häufigkeit der Nutzung? Fähigkeiten?

Interessen, Ziele,?

Einschränkungen?

ntextual interaction design research

Entwicklung des Test-Plans



1.	Funktionen auswählen	2. Aufgaben definieren	3. Vor- und Abschluss- interview Fragen		
•	Oft verwendet? Neu? Als problematisch eingeschätzt? Von Nutzern als wichtig eingeschätzt?	 Typische Handlungen von Nutzern? Auf ein End-Ziel des Nutzers bezogen? Spezifisch genug? Möglich? In einer realistischen Abfolge? Vernünftige Länge? Geschätzte Aufgabendauer? 	 Vor-Interview: Background and Routinen der Testperson? Ähnliche Erfahrungen? Abschluss-Interview: Gesamt-Einschätzung des Systems? Elemente die verändert werden sollen? Die besten/schlimmsten 3 Dinge 		

Entwicklung des Test-Plans



Vorinterview

	Testperson	Testleiter	Technik-Assistent
Vorinterview		Dann starten wir mit einem kurzen Vorinterview TL schreibt mit	Ass. schreibt am Observer mit

- Seit wie vielen Jahren verwenden Sie das Handy?
- 2. Welche Funktionen bzw. Dienste Ihres Handys haben und nutzen Sie? (beruflich und privat) (frei aufzählen lassen)

Aufgabe 4: Restaurant suchen und Bewertung lesen

Szenario:

Sie planen am Abend italienisch essen zu gehen.

Aufgabe:

Finden Sie mit dem Planer ein für Sie ansprechendes Restaurant. Nennen Sie mir die Durchschnitts-Bewertung des ausgesuchten Restaurants und kehren Sie dann wieder zur Startseite zurück.

contextual interaction design research

Entwicklung des Test-Plans



Auswahl der Usability / Performance Metriken

z.B.:

Die Geschwindigkeit mit der jemand eine Aufgabe erfüllt

Wie viele Fehler gemacht werden

Wie oft ein bestimmter Fehler auftritt

Wie oft Testpersonen einen Fehler wieder gut machen können

Wie viele Testpersonen eine Aufgabe erfolgreich erfüllen

Wie zufrieden/frustriert die Testpersonen mit der Erledigung der Aufgaben sind

http://www.measuringu.com/blog/essential-metrics.php

Fragebögen



Abschluss Usability Fragebögen

SUS - System Usability Scale [Brook 1986]

http://www.measuringusability.com/sus.php Ease of use, learnability, satisfaction, ...

SUMI – Software Usability Measurement Inventory [Kirakowski and Corbett 2003] http://sumi.ucc.ie/

Perceived quality of use of software

WAMMI - Website Analysis and Measurement Inventory http://www.wammi.com/

Satisfaction

CSUQ - Computer System Usability Questionnaire [Lewis 1995] http://hcibib.org/perlman/question.cgi

System usefulness, information quality, interface quality

AttrakDiff - Measuring the attractiveness of interactive products [Hassenzahl, M., Burmester, M., & Koller, F. 2003]

http://attrakdiff.de/

Hedonic (Identity and Stimulation), pragmatic Quality, Attractiveness

contextual interaction design research

Sytstem Usabilty Scale



Explorieren Sie die Seite http://www.volkswagen.at/ mit Car Configurator

Zeit: 10 Minuten

Danach: Füllen sie den SUS aus

Berechnen Sie Ihren eigenen SUS Score

Zusammenführung der Ergebnisse (Mittelwert, Standardabweichung) & Interpretation

SUS

Ich stimme dem zu					
nicht	weniger	mittel	ziemlich	sehr	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5 5	
	1 1 1 1 1	nicht weniger 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2	nicht weniger mittel 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3	nicht weniger mittel ziemlich 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4	

SUS



Berechnung SUS Score:

Umwandeln der Werte:

Ungerade Items: Position minus 1

Gerade Items: 5 minus Position

Summe bilden und mit 2,5 multiplizieren

untextual interaction design research

Auswertung

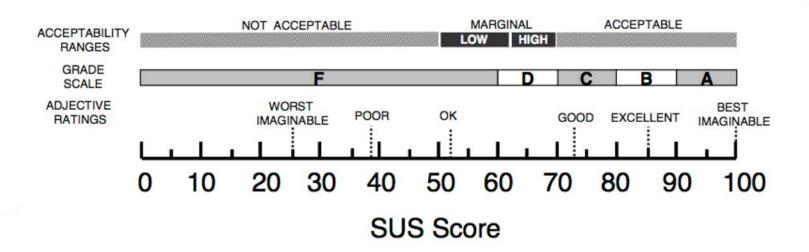


100% entsprechen einem perfekten System ohne Usability-Probleme.

Werte über 80% deuten auf eine gute bis exzellente Usability hin.

Werte zwischen 60% und 80% sind als *grenzwertig* bis *gut* zu interpretieren.

Werte unter 60% sind Hinweise auf erhebliche Usability-Probleme.







Fragebogen zum ersten Eindruck (AttrakDiff) BlackBerry Studie

TP Nr.:		
Salzburg,	Juni/Juli	2008

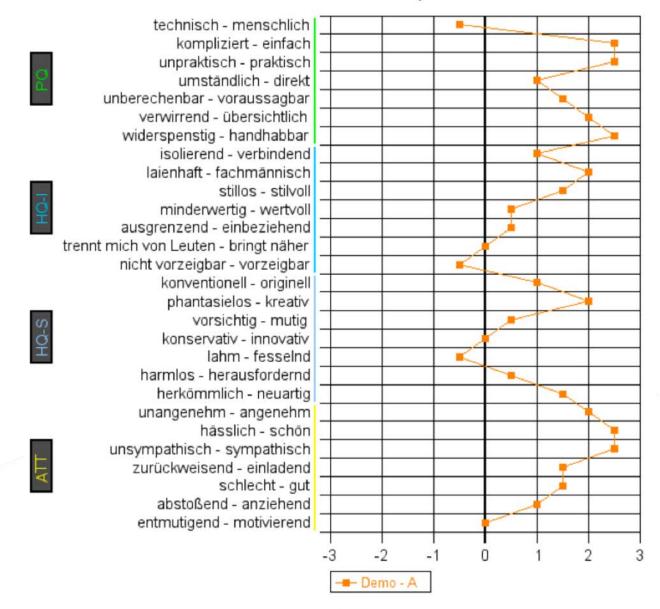


Bitte geben Sie mit Hilfe der folgenden Wortpaare Ihren Eindruck zum BlackBerry wieder. Kreuzen Sie bitte die entsprechenden Kästchen ohne lang Nachzudenken an.

Ich finde das BlackBerry ...

menschlich	Ш							technisch
isolierend								verbindend
angenehm								unangenehm
originell								konventionell
einfach								kompliziert
fachmännisch								laienhaft
hässlich								schön
praktisch								unpraktisch
sympathisch								unsympathisch
umständlich								direkt
stilvoll								stillos
voraussagbar								unberechenbar
minderwertig	Н		Н					wertvoll
ausgrenzend								einbeziehend
ŭ .								Trennt mich von
bringt mich Leuten		Ш	Ш	Ш	ш	Ш	Ш	
								Leuten
nicht vorzeigbar				Ц			Ц	vorzeigbar
								einladend
phantasielos	Ц					Н	片	kreativ
gut	Ш	Ш	Ш	Ш		Ш	Ш	schlecht
verwirrend								übersichtlich
abstoßend								anziehend
mutig								vorsichtig
innovativ								konservativ
lahm								fesselnd
harmlos								herausfordernd
motivierend								entmutigend
neuartig								herkömmlich
widerspenstig								handhabbar

Profil der Wortpaare





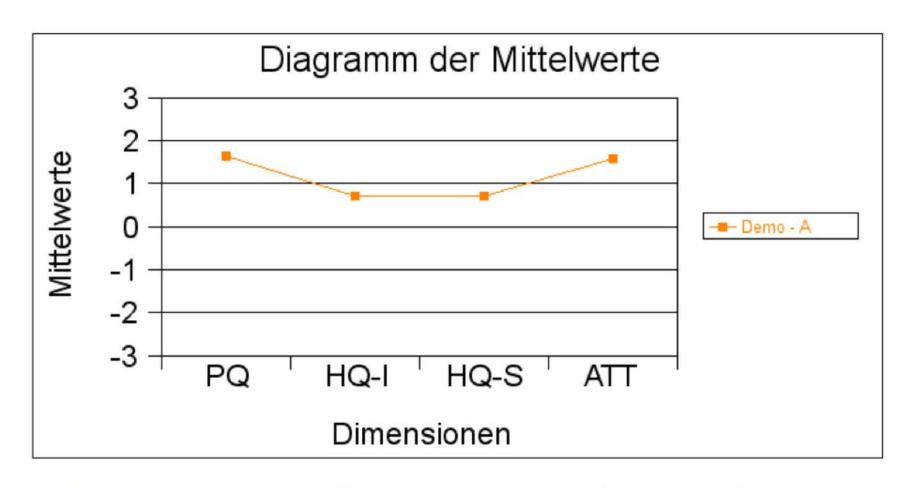


Abbildung 2: Mittlere Ausprägung der vier Dimensionen des AttrakDiff™ für das Produkt "Demo - A"



https://esurvey.uid.com

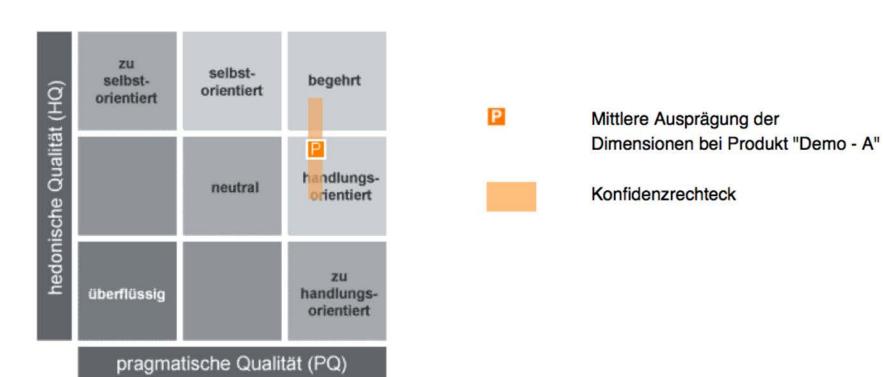


Abbildung 1: Portfolio mit der durchschnittlichen Ausprägung der Dimensionen PQ und HQ und dem Konfidenz-Rechteck des Produkts "Demo - A"



Online Auswertung:

https://esurvey.uid.com

contextual interaction design researc

Eye-Tracking: Allgemeines



- Eye-Tracking: Aufzeichnen der Blickbewegungen einer Person
- Various measures of usability are necessary to gather the whole picture [...]. Eye Tracking alone is not a complete usability engineering approach, but it can make a significant contribution to the assessment of usability. (Jacob & Karn, 2003)

ontextual interaction design research

Eye-Tracking: Allgemeines



Sakkaden

- schnelle ballistische Bewegungen, die das Auge auf ein Blickobjekt richten
- Auslösung spontan durch gezielte Verhaltenspläne oder unwillkürlich

Fixationen

- Zustand, in dem das Auge sich bezüglich eines Sehobjekts in "relativem" Stillstand befindet
- Aufnahme visueller Informationen

contextual interaction design research

Eye-Tracking: Allgemeines



- Klassifizierung nach Kahnemann (1973):
 - Spontaneous looking
 occurs when the subject views a scene without any specific
 task in mind
 - Task-relevant looking
 is performed when the observer views the scene with a
 particular question or task in mind
 - Orientation of thought looking
 occurs when the observer is not paying much attention to
 where he is looking, but is attending to some "inner thought"

















Task-relevant looking



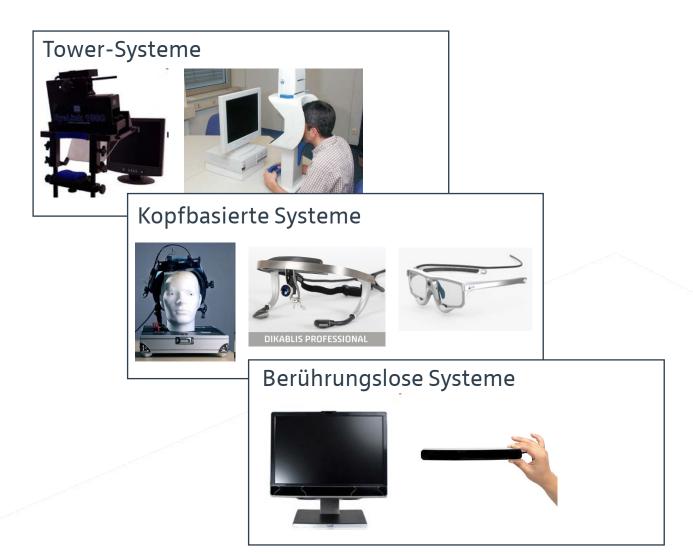
- 1) Free examination
- 2) Estimate the material circumstances of the family
- 3) Give the ages of the people
- 4) Surmise what the family had been doing before the arrival of the "unexpected visitor"
- Remember the clothes worn by the people
- 6) Remember the position and objects in the room
- 7) Estimate how long the "unexpected visitor" had been away from the family

(Yarbus, 1967)

contextual interaction design research

Eye-Tracking: Systeme





Eye-Tracking: Systeme



SMI Eye Tracking Glasses

Dikablis Professional



http://www.smivision.com

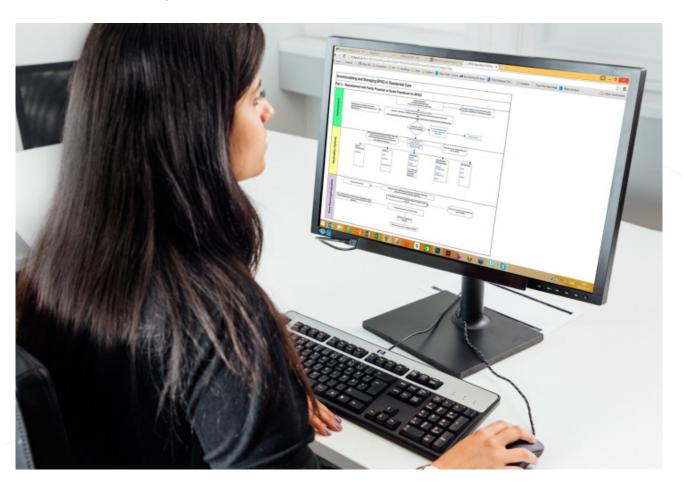


http://www.ergoneers.com/eye-tracking/

Eye-Tracking: Systeme



Tobii Remote System



http://www.topii.com/

Eye-Tracking: Systeme



The Eye Tribe



Eye-Tracking: Systeme



SmartEye Pro System with 4 cameras



www.smarteye.se

Eye-Tracking: Daten und Analyse



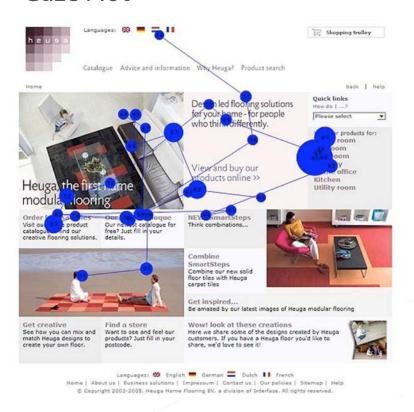
- Verwendet während der Aufgabendurchführung
- Zum Sammeln von zusätzlichen "harten" Daten/Fakten für Problembereiche, z.B.
 - Gesamtanzahl der Fixationen
 - Blicke (%) auf jede Area of Interest (AOI)
 - Mittlere Fixationsdauer gesamt
 - Anzahl von Fixationen auf jeder AOI
 - Mittlere Blickdauer auf jeder AOI
 - Blickpfad
 - Zeit bis zur ersten Fixation der Ziel AOI

•

Eye-Tracking: Daten und Analyse



Gaze Plot



Heat Map



Eye-Tracking: Herausforderungen



- KISS: Keep It Short and Simple
- Definition der Aufgaben, Area-of-Interests, relevanten Blickparameter
- Kalibrierung
- Synchronisation
- Wechselnder Bildschirminhalt / Dynamik
- Wechselnde Lichtverhältnisse
- Brillen / Schminke
- Auswertung
- •



LEITFADEN

Test Material: Test-Leitfaden



Einleitung

Danke für Ihre Teilnahme & Information über das System

Wir testen das Interface, nicht Sie!

Sie werden gebeten Aufgaben mit dem System durchzuführen. Dabei können Sie nichts falsch machen.

Wenn Sie Fragen haben, fragen Sie bitte und ich werde – wenn möglich – antworten (nicht immer, da ich wissen möchte wie sie mit dem System alleine zurecht kommen)

Einverständniserklärung

Nachdem der Teilnehmer diese Erklärung unterzeichnet hat, startet die Videoaufnahme

Thinking Aloud (Laut Denken Methode)

Test Material: Test-Leitfaden



Vor-Interview

Background und Gewohnheiten der Test-Teilnehmer (für das getestete System relevante Bereiche)?

Bisherige Erfahrungen?

Vor-Fragebogen (falls benötigt)

Erwartungen

Sozio-demografische Daten

Nutzungsverhalten

Test Material: Test-Leitfaden



Test Aufgaben & Szenarios

Startseite für die Aufgabe

Endseite für die Aufgabe

Schritte der Aufgabendurchführung

Dauer der Aufgabe (max)

Hilfestellungen/Hinweise pro Aufgabe

Einschätzung des Schwierigkeit der Aufgabenerfüllung durch Test-

Teilnehmer

Offene Frage: Wie kann man das System verbessern?

Beispiel: Aufgabenbeschreibung



Aufgabe 1: Display-Helligkeit einstellen

Szenario:

Stellen Sie sich vor, Sie sitzen daheim vor Ihrem Fernseher und probieren Ihre neu gekaufte Universal-Fernbedienung aus.

Aufgabe:

Stellen Sie bitte eine für Sie angenehme Display-Helligkeit ein. Und kehren Sie dann wieder zum ersten Menü in der Fernbedienung zurück. (Alles Verstanden?)

Startseite: Sprachauswahl Endseite: Hauptmenü

Aufgabenerfüllung: passende Helligkeit eingestellt

Maximale Dauer: 5 Minuten

Hinweis: nach 2 Minuten (falls DVD gesehen): Erinnern Sie sich an die Videoclips.

nach 3 Minuten: Versuchen Sie den Button zu rotieren.

Bei Nicht-Erfüllt: TP wird der Lösungsweg gezeigt.

ggf. Ausnahmehinweis (falscher Lösungsweg am Ende als richtig gesehen)

Bewertung:

Wenn sie nach einem Schulnotensystem – wo 1 "sehr einfach" und 5 "sehr schwierig" ist – die Aufgabe bewerten müssen, wie bewerten sie diese?

Sehr einfach		Neutral		Sehr schwierig	
1	2	3	4	5	

Frage nach der Aufgabe:

Was würden Sie am System verändern, damit die Erledigung dieser Aufgabe in Zukunft einfacher wird?

Beispiel Aufgabenbeschreibung



Aurgabe 1	Aufgabe 1: (min)	wizard:					
	Szenario:						
	Aufgabe:						
Laut denken!	[Testleiter: Sind Sie bereit, dann geht es						
	Aufgabenerfüllung: ohne Hilfe□	mit Hilfe 🗆	nicht erfüllt □				
	Meilensteine:						
	Hilfestellungen: □						
	[Testleiter: Danke, Sie haben die Aufgabe erfüllt!]						
	 Wie zufrieden waren Sie mit der Nutzung des Systems w\u00e4hrend dieser Aufgabe? Bitte bewerten Sie nach einem Schulnotensystem – wo 1 "sehr zufrieden" und 5 "gar nicht zufrieden" ist: 						
	Sehr zufrieden	Neutral Gar nicht zufrieden		en			
	1 2	3	4 5	_			
	Wie einfach oder schwierig fiel es Ihnen, diese Aufgabe durchzuführen? Bitte bewerten Sie nach einem Schulnotensystem – wo 1 "sehr einfach" und 5 "sehr schwierig" ist:						
	Sehr einfach Neutral Sehr schwierig			ig			
	1 2	3	4 5	_			
	3. Was hat Sie bei der Durchführung dies	er Aufgabe am meister	n gestört, wo gab es Probleme	9?			
	Haben Sie irgendwelche Verbesserungsvorschläge oder haben Sie irgendetwas vermisst?						

Pilot Tests



Zentrales Ziel:

Überprüfung der Aufgabenbeschreibung und -dauer Überprüfung des technischen Set-ups

Rekrutierung der Pilot-Tester

Testumgebung aufsetzen

Pilot test durchführen

Überarbeitung der Materialien und Verbesserung des Test Set-ups

Anpassung der gesetzten Dauer der Aufgaben

Diesen Schritt nicht überspringen!

Teilnehmer Rekrutierung



6-12 Test Teilnehmer

Gender ausgeglichen (Männer und Frauen)

1-2 Stunden pro Test, 30 Minuten zwischen den Tests

Aufwandsentschädigung definieren

(bei uns ca. 20-30 EUR)

Wichtig ist ein klar definierter Rekrutierungsplan; die richtigen Testpersonen zu finden ist harte Arbeit (bspw. ältere Nutzer, Manager,..)

Unterschiedlich Wege der Rekrutierung nutzen (Test Teilnehmer Datenbank, Netzwerke, E-mail,...)

Durchführung



Testleiter: führt den Teilnehmer, beobachtet, macht Notizen, führt die Interviews

Eine Session pro Teilnehmer

Videoaufzeichnung

System (screen, user interface)

Teilnehmer

Kommentare

Beobachtung des Tests







Rollen definieren

Testleiter

Auf "think aloud" achten, den Teilnehmer motivieren seine Gedanken laut auszusprechen

Den Teilnehmer nicht führen oder ihm helfen

Fragen stellen um Ansichten/Wahrnehmungen des Nutzers zu erforschen

Vermeiden, den Nutzer abzulenken während des Tests

Beobachterrolle

Notizen machen

Den Schritten im Leitfaden folgen

Durchführung



Abschluss-Fragebogen

Standard Usability Fragebogen (System Usability Scale - SUS) AttrakDiff Fragebogen (hedonische und pragmatische Faktoren) ... andere Fragebögen

Abschluss-Interview

Gesamt-Erfahrung Elemente, die (nicht) verändert werden sollen Die besten/schlimmsten 3 Dinge

contextual.interaction.design.resea

Beispiel Testbericht



http://www.usability.gov/how-to-andtools/resources/templates/report-template-usabilitytest.html

Analyse der Daten & Bericht



Zusammenfassung & Analyse der Daten

Gemessene Zeit pro Aufgabe, Anzahl der erfüllen Aufgaben, Einschätzungen, Verbesserungsvorschläge

Haben die Teilnehmer die Aufgaben erfolgreich erfüllt?

Wenn ja, wie schnell haben sie die Aufgaben erfüllt?

Schätzen die Teilnehmer die Aufgabenerfüllung als schnell genug ein?

Welche Wege haben die Teilnehmer zur Aufgabenerfüllung genommen?

Waren die Teilnehmer effizient bei der Erfüllung der Aufgaben?

Wo hatten die Teilnehmer Probleme?

Welche Vorschläge haben die Teilnehmer um die Aufgabenerfüllung zu verbessern?

Analyse der Daten & Bericht



Bericht & Empfehlungen

Beschreibung der Test Teilnehmer (Background) & der Test Umgebung

Auflistung an gefundenen Usability Problemen (ggf. mit Screenshots)

Anzahl an Teilnehmern, die eine Aufgabe erfolgreich erfüllt hat,

Vergleich der Dauer der Aufgabendurchführung

Vorschläge, Statements (Kommentare) der Teilnehmer

Verbesserungsvorschläge machen (Expertenkommentare)

Zusammenfassung der Interview-Ergebnisse

Fragebogen-Ergebnisse

This is not a typical user studie



Hopefully ;-)











http://www.bugbash.net/

References



- Bradley, M.M. and Lang, P.J. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49-59.
- Brooke, J. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189, 194.
- Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M. and Pidruzny, J.N. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing.
 Journal of Experimental Social Psychology, 45(4), 624-634.
- de Kort, Y.A.W., IJsselsteijn, W.A. and Poels, K. 2007. Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ). *Proceedings of PRESENCE*, 195-203.
- Gabrys-Barker D. 2010. Appraisal sysytems in L2 Vs. L3 Learning Experiences

References



- Hassenzahl, M., Platz, A., Burmester, M., and Lehner, K. 2000. Hedonic and ergonomic quality aspects determine a software's appeal. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 201-208). ACM.
- IJsselsteijn, W., Poels, K. and De Kort, Y.A.W. 2008. The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games. *TU Eindhoven, Eindhoven, The Netherlands*.
- Kirakowski, J. and Corbett, M. 1993. SUMI: The software usability measurement inventory. British journal of educational technology, 24(3), 210-212.
- Kim, J., and Moon, J.Y. 1998. Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with computers*, 10(1), 1-29.

References



- Lewis, J. 1995. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1), 57-78.
- Mehrabian, A. and Russell, J.A. 1974. An approach to environmental psychology. Cambridge, MA, MIT.
- Moser, C., Fuchsberger, V. and Tscheligi, M. 2012. Rapid assessment of game experiences in public settings. In *Proceedings of the 4th International Conference on Fun and Games* (FNG '12), ACM, 73-82.
- Sonderegger, A., and Sauer, J. 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, 41(3), 403-410.

Noch Fragen...







Kontakt

Dr. Alexander Meschtscherjakov

alexander.meschtscherjakov@sbg.ac.at