



Fachhochschule Nordwestschweiz
Hochschule für Angewandte Psychologie

**Empirische Untersuchung zweier Varianten des Usability-Tests:
Klassischer Usability-Test und Usability-Walkthrough**

Bachelor Thesis

2014

Autorin

Stefanie Friedli

Angewandte Psychologie Fachhochschule Nordwestschweiz

Begleiter

Prof. Dr. Fred Van den Anker

Praxispartner

Marcel Uhr

soultank AG

Zusicherung Selbständigkeit

Ich versichere hiermit, dass ich die vorliegende Bachelor Thesis selbständig und ohne Mithilfe Dritter verfasst habe. Alle Stellen, die wörtlich oder sinngemäss aus Schriften entnommen sind, habe ich in jedem einzelnen Fall durch Angabe der Quelle, auch der verwendeten Sekundärliteratur, als Entlehnung kenntlich gemacht.

Schaffhausen, 4.06.2014

Abstract

Die vorliegende Arbeit vergleicht zwei Varianten des Usability-Tests. Es wird untersucht, wie sich der klassische Usability-Test und der Usability-Walkthrough hinsichtlich des Erlebens der Testpersonen und der Effektivität und Effizienz der Testergebnisse unterscheiden. Mit einer Stichprobe von 40 Testpersonen werden die abhängigen Variablen Testerleben, Anzahl gefundener Usability-Probleme und Anzahl gelöster Testaufgaben erhoben und ausgewertet. Die Regressionsanalysen zeigen, dass der Usability-Walkthrough von den Testpersonen signifikant positiver als der klassische Usability-Test erlebt wird. Die qualitative Analyse zeigt, dass der Usability-Walkthrough eher als Gespräch und der klassische Usability-Test eher als Test wahrgenommen werden. Keinen signifikanten Zusammenhang lässt sich zwischen der Testvariante und der Anzahl oder Art der Usability-Probleme feststellen. Während 30 Minuten können beim Usability Walkthrough signifikant mehr Testaufgaben gelöst werden als beim klassischen Usability-Test. Die Arbeit umfasst 124'860 Zeichen inkl. Leerzeichen ohne Anhang.

Schlüsselwörter: Klassischer Usability-Test, Usability-Walkthrough, Testleitereffekt, Testerleben, Effektivität, Effizienz

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Ausgangslage.....	2
1.2	Fragestellung.....	3
2	Einführung in Usability.....	4
2.1	Definition Usability.....	4
2.2	Relevanz von Usability.....	4
2.3	Usability Normen, Kriterien und Guidelines.....	5
3	Theoretischer Hintergrund des Usability-Tests.....	6
3.1	Definition Usability-Test.....	6
3.2	Einordnung des Usability-Tests in die Usability-Methoden.....	6
3.3	Beurteilung des Usability-Tests anhand methodischer Gütekriterien.....	7
3.4	Vorbereitung.....	9
3.4.1	Testpersonen rekrutieren.....	9
3.5	Durchführung.....	10
3.5.1	Die goldenen Moderationsregeln.....	10
3.5.2	Hilfestellung anbieten.....	10
3.5.3	Testablauf.....	12
3.6	Usability-Tests sind überall möglich.....	12
3.6.1	Usability-Test im stationären Labor.....	13
3.6.2	Usability-Test mit dem mobilen Labor.....	13
3.7	Stärken des Usability-Tests.....	14
3.8	Schwächen des Usability-Tests.....	14
3.8.1	Testleitereffekt.....	15
3.8.2	Testangst.....	16
3.9	Fazit zum Vergleich der Testvarianten.....	16
4	Methodisches Vorgehen.....	18
4.1	Forschungsdesign.....	18
4.1.1	Abhängige Variablen.....	20
4.1.2	Unabhängige Variablen.....	22
4.2	Rekrutierung der Testpersonen.....	22
4.3	Beschreibung der Stichprobe.....	23

4.4	Materialien und Setting	25
4.5	Durchführung	27
4.5.1	Durchführung des klassischen Usability-Tests	27
4.5.2	Unterschiede in der Durchführung vom klassischen Usability-Test zum Usability-Walkthrough	28
4.6	Auswertung der Daten	29
4.6.1	Quantitative Auswertung	29
4.6.2	Qualitative Auswertung	30
5	Ergebnisse	31
5.1	Testerleben der zwei Varianten des Usability-Tests	31
5.1.1	Quantitative Ergebnisse zum Gesamtwert FTE	31
5.1.2	Qualitative Ergebnisse zum Testerleben	33
5.2	Effektivität der zwei Varianten des Usability-Tests	34
5.2.1	Quantitative Ergebnisse zur Anzahl der Usability-Probleme	34
5.2.2	Qualitative Ergebnisse zur Art der Usability-Probleme	36
5.3	Zeitliche Effizienz der zwei Varianten des Usability-Tests	42
5.3.1	Quantitative Ergebnisse zur Anzahl gelöster Aufgaben	42
6	Diskussion der Ergebnisse	44
6.1	Zusammenfassung der Ergebnisse	44
6.2	Ergebnisdiskussion	45
6.3	Stärken und Schwächen der Untersuchung	47
6.4	Schlussfolgerungen und Ausblick	48
6.4.1	Praxistransfer: Was bedeuten die Ergebnisse für die Praxis?	48
6.4.2	Ideen für zukünftige Forschungen	48
7	Literaturverzeichnis	50
8	Verzeichnisse	52
8.1	Abbildungsverzeichnis	52
8.2	Tabellenverzeichnis	53
8.3	Abkürzungsverzeichnis	54
9	Anhang	55

1 Einleitung

Nahezu jeder hat schon Erfahrungen mit unverständlichen Programmen, langen Einarbeitungszeiten oder technischen Pannen gemacht. Mit zunehmender Fachspezialisierung, den sich ständig verändernden Arbeitsbedingungen und der hohen Aufgabenkomplexität vergrößerte sich der Abstand zwischen den Entwicklern und den Endnutzern rapide. Viele Mitarbeitende müssen mit Systemen arbeiten, die nicht oder nur teilweise brauchbar sind (z.B. Abbildung 1).



Abbildung 1: Unbrauchbares Produkt (Weikel, 2014)

Die Überprüfung der Benutzbarkeit der eingesetzten Systeme ist folglich für eine effiziente, effektive und zufriedenstellende Nutzung unumgänglich (Sarodnick & Brau, 2010). Es gibt verschiedene Möglichkeiten und Usability-Methoden, um die Schwachstellen eines Systems aufzudecken und zu verbessern. Eine davon ist der Usability-Test. Das Interesse der vorliegenden Arbeit liegt in der Untersuchung der Unterschiede zwischen zwei Varianten des Usability-Tests. Effektivität und Effizienz spielen in der Arbeitswelt eine bedeutende Rolle. Da Usability-Tests nur mit Hilfe von Testpersonen möglich sind, liegt ein Fokus der Untersuchung darin, wie die beiden Testvarianten von den Testpersonen erlebt werden. Zusammenfassend steht die Frage nach den Unterschieden der zwei Testvarianten hinsichtlich Testerleben, Effektivität und Effizienz der Tests im Zentrum. Die Fragestellung wird mit einem Versuch geprüft und anhand der Datenauswertung interpretiert. Auf andere Usability-Evaluationsmethoden, die ebenfalls mit Testpersonen arbeiten, wie bspw. Interviews, Card Sorting, Fokusgruppen oder Fragebogen, wird nicht eingegangen.

Die vorliegende Arbeit gliedert sich wie folgt: Im Kapitel 1 werden Problemstellung, Zielsetzung und Fragestellungen beschrieben. Das Kapitel 2 gibt eine Einführung in die Usability. Im Kapitel 3 wird detailliert auf die Theorie zum Usability-Test eingegangen. Das Kapitel 4 informiert über das methodische Vorgehen der Untersuchung. Die Ergebnisse werden im Kapitel 5 vorgestellt. Im letzten Kapitel werden die Ergebnisse diskutiert, Stärken und Schwächen der Untersuchung reflektiert, Schlussfolgerungen für die Anwendung der Tests in der Praxis gezogen und Ideen für zukünftige Forschungen vorgestellt.

1.1 Ausgangslage

Betriebliche Rahmenbedingungen

Das Unternehmen soultank AG wurde im Jahr 2000 von Marcel Brunswiler, Michael Fischer, Thomas Fux und Marcel Uhr gegründet. Im Jahr 2008 übernahm Marcel Uhr sämtliche Aktienanteile und wurde Inhaber von der soultank AG. Der Hauptsitz des Unternehmens befindet sich in Zug. Eine kleinere Zweigstelle befindet sich in St. Gallen. Zurzeit beschäftigt die soultank AG elf Mitarbeitende und einen Lehrling. Das interdisziplinäre Team setzt sich aus Experten aus den Bereichen Software-Ergonomie, Psychologie, Design, Journalistik und Informatik zusammen.

Zu den Kernkompetenzen des Unternehmens gehören die Bereiche User Experience, Usability und Ergonomie. Die soultank AG analysiert, konzipiert, visualisiert und überprüft interaktive Systeme in enger Zusammenarbeit mit den Kunden mit dem Ziel, interaktive Systeme zu verbessern, wie bspw. Webseiten, Intranets, Softwareprodukte, Automaten, Touchscreensysteme, Haushaltsgeräte, Einkaufserlebnisse on- und offline etc.

Problemstellung

Am Standort in Zug betreibt die soultank AG ein Usability-Labor, in dem Nutzer direkt bei der Interaktion mit einem System beobachtet werden können. Der Usability-Test kann in verschiedenen Varianten durchgeführt werden. Beim **klassischen Usability-Test** befindet sich die Testleitung im Nebenraum und beobachtet, bspw. durch einen Einwegspiegel, die Testperson, die verschiedene Aufgaben mit dem Testsystem löst. Wenn der Prototyp nicht sehr ausgereift ist, empfiehlt die soultank AG den **Usability-Walkthrough** durchzuführen. Bei dieser Variante des Usability-Tests sitzt die Testleitung neben der Testperson und moderiert den Ablauf.

In früheren Jahren setzte die soultank AG vor allem die klassische Variante des Usability-Tests ein. Heute wird vermehrt die Variante des Usability-Walkthroughs durchgeführt, sowohl für wenig ausgereifte Prototypen als auch für bestehende Produkte. Die Gründe hierfür liegen einerseits darin, dass der Usability-Walkthrough einfacher eingesetzt werden kann als der klassische Usability-Test, da kein Einwegspiegel oder vergleichbare technische Einrichtung zur Verfügung stehen muss. Zudem wird die Situation im Usability-Walkthrough, nach Angaben von der soultank AG, von den Testpersonen häufig als natürlicher erlebt als der klassische Usability-Test. Durch den persönlichen Kontakt beim Usability-Walkthrough ist es für die Testleitung einfacher, individuell auf die Testpersonen einzugehen, bspw. bei Testangst, Unwohlsein oder Nervosität. Wenn sich die Testleitung ausserhalb des Sichtfeldes der Testperson befindet, ist dies schwieriger. Neben den beschriebenen Vorteilen des Usability-Walkthroughs, müssen auch mögliche Gefahren beachtet werden. Der Einfluss der Testleitung auf die Testpersonen kann sich negativ auswirken und die Testergebnisse verfälschen. In der Literatur ist dieser Einfluss als Testleitereffekt bekannt. Da es hierfür kaum empirische Untersuchungen im Bereich Usability gibt, ist es für die soultank AG wichtig, die möglichen Einflüsse und Auswirkungen des klassischen Usability-Tests und des Usability-Walkthroughs zu untersuchen.

Zielsetzung

Aus der beschriebenen Problemstellung wurden folgende Ziele für die vorliegende Arbeit abgeleitet:

- Empirischer Methodenvergleich
- Untersuchung des Einflusses der Anwesenheit der Testleitung auf die Ergebnisse des Usability-Tests und auf das Erleben der Testpersonen

Aus der Darlegung der Problemstellung und Zielsetzung werden im folgenden Kapitel die Hauptfragestellung und die Unterfragestellungen für die Untersuchung abgeleitet.

1.2 Fragestellung

Die folgende Fragestellung steht im Zentrum der vorliegenden Arbeit:

Wie unterscheiden sich der klassische Usability-Test und der Usability-Walkthrough?

Um die Hauptfragestellung detaillierter beantworten zu können, wurden zusätzlich drei Unterfragestellungen ausgearbeitet. Diese dienen als Orientierung und Fokussierung auf die zu untersuchende Thematik während der gesamten Datenerhebung und –auswertung.

1. Wie werden die zwei Varianten des Usability-Tests von den Testpersonen erlebt?
2. Wie unterscheiden sich die zwei Varianten des Usability-Tests hinsichtlich ihrer Effektivität?
3. Wie unterscheiden sich die zwei Varianten des Usability-Tests hinsichtlich der zeitlichen Effizienz?

Die beiden untersuchten Testvarianten unterscheiden sich hauptsächlich durch die Anwesenheit bzw. Abwesenheit der Testleitung und deren Auswirkung auf die Testpersonen. Folglich wurde untersucht, wie die Testpersonen die beiden Testsettings erleben. Das zentrale Ziel des Usability-Tests ist gemäss Sarodnick und Brau (2010) die Aufdeckung von Schwachstellen eines Systems und die Gewinnung von Gestaltungs- und Verbesserungsmöglichkeiten. Deshalb war die Frage nach den gefundenen Usability-Problemen und nach deren Art in beiden Testvarianten relevant. Für den praktischen Einsatz des Usability-Tests könnte der Faktor Zeit sowohl für die Testleitung als auch für die Auftraggeber eine Rolle spielen. Wenn bei einer Testvariante mehr Aufgaben gelöst werden können, könnte dies ein Kriterium für die Wahl des Tests sein.

Praktische Relevanz der Fragestellung

Für die soultank AG ist die Fragestellung von Bedeutung, um fundierte Kenntnisse über den Usability-Walkthrough zu erhalten. Aus den gewonnenen Ergebnissen der vorliegenden Arbeit können relevante Erkenntnisse für den Einsatz der Methode in der Praxis abgeleitet werden. Die Resultate dieser Untersuchung können auf Vorgehensweise, Durchführung und Auswertung zukünftiger Usability-Tests übertragen werden. Mithilfe der Erkenntnisse können Usability-Tests hinsichtlich der Effektivität, der Effizienz und dem Erleben der Testpersonen optimiert sowie Verhaltensregeln für die Testleitung abgeleitet werden.

2 Einführung in Usability

„In zehn Minuten beginnt die Konferenz, und der Rechner hat gerade die fertige Präsentation gefressen. Blöde Kiste! (...) Genauso bringen langsame und abstürzende Computer ihre Nutzer immer öfters in Rage“ (Oberhuber, 2004). Nach Oberhuber (2004) haben fast alle Büroangestellte den Computer schon angeschrien und verflucht, oder sie wurden sogar handgreiflich. Gemäss Sarodnick und Brau (2010) wird bei der Entwicklung und beim Design von Systemen zu wenig an die Endnutzer gedacht. Dieses Kapitel soll in die Thematik Usability einführen und einen Überblick verschaffen. Nachdem zentrale Begriffe im Bereich Usability erläutert werden, wird die Relevanz des Einsatzes von Usability behandelt. Die Darstellung der Usability Kriterien und Normen runden das Kapitel ab.

2.1 Definition Usability

Nach Richter und Flückiger (2010) lässt sich der Begriff Usability im deutschen Sprachraum am treffendsten mit der Benutzbarkeit eines Systems übersetzen. Inzwischen hat sich aber der englische Begriff Usability auch im deutschen Sprachgebrauch etabliert (Richter & Flückiger, 2010). Gemäss Heinsen (2003) gibt es mittlerweile eine allgemein akzeptierte Definition von Usability. Diese ist in der ISO-Norm 9241-11 festgelegt und definiert Gebrauchstauglichkeit als „das Ausmass, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ (DIN EN ISO 924-11, 1998; zitiert nach Bahner & Timpe, 2005, S.213). Effektivität meint nach Heinsen (2003) „ob der Benutzer eines Produktes, einer Software oder Website seine Ziele erreichen kann“ (S.3). Das Kriterium Effizienz wird beschrieben als „Aufwand, der zur Erreichung des Ziels nötig ist“ (Heinsen, 2003, S.3). Zufriedenstellend meint „(...) ob die Benutzer einer Software, Website oder ein Produkt mögen und sich damit wohl fühlen“ (Heinsen, 2003, S.3). Zusammenfassend beurteilt Usability, wie gut die Nutzer ein Werkzeug in ihrem Umfeld zur Bewältigung von Aufgaben nutzen können (Richter & Flückiger, 2010). Weitere Begriffe im Zusammenhang mit Usability werden im Glossar erläutert (Anhang A, S.56).

2.2 Relevanz von Usability

Nach Zwingmann (2004, zitiert nach Oberhuber, 2004) stellen Technikstress und Computerfrust grosse Probleme dar und richten Schaden in Milliardenhöhe an. Die Universität von Maryland schätzt, dass rund ein Drittel der Arbeitszeit der Mitarbeitenden am Rechner durch frustrierende Erlebnisse verloren geht (Oberhuber, 2004). Brinks (2005) konnte in ihrer Studie einen Zusammenhang zwischen Computerpannen und Frustration der Nutzer feststellen. 70% der Computernutzer verhalten sich aggressiv gegenüber dem Computer. Diese Computerfrustration wirkt sich negativ auf die Arbeitsmotivation und Zufriedenheit der Mitarbeitenden aus und kann schlimmstenfalls zu Depressionen führen (Brinks, 2005). Gemäss dem Bundesverband der Betriebskrankenkassen (BKK) nimmt die Zahl psychischer Erkrankungen bei Bildschirmarbeiten enorm zu (Oberhuber, 2004). Allerdings ist ungewiss, welchen Anteil die Frustration durch Soft- und Hardwareprobleme ausmacht. Einige weitere Vorteile von Usability sind nach Heinsen (2003) Einsparung von Kosten, Gewinn neuer Kunden, Erhöhung des Markenwerts und des Spassfaktors. Nach Sarodnick und Brau (2006) ist eine gelungene Gestaltung der technischen Systeme deshalb aus verschiedenen Gründen von grosser Bedeutung.

2.3 Usability Normen, Kriterien und Guidelines

„Usability Guidelines beinhalten Regeln für die Gestaltung von Benutzeroberflächen“ (Richter & Flückiger, 2010, S.54). Normen haben zum Ziel, durch Gestaltungsvorgaben die Benutzung von Technologien zu standardisieren und zu vereinfachen. Eine bekannte Norm ist die internationale ISO-Norm 9241. Die sieben Kriterien in der Tabelle 1 sind die zentralen Kriterien, die ein benutzerfreundliches System auszeichnen.

Tabelle 1: Grundsätze der Dialoggestaltung (Richter & Flückiger, 2010, S.55)

Kriterium	Beschreibung
Aufgabenangemessenheit	Das System unterstützt die Erledigung der Aufgaben und den Arbeitsablauf der Benutzer.
Selbstbeschreibungsfähigkeit	Das System enthält Erläuterungen und ist ausreichend verständlich.
Steuerbarkeit	Der Benutzer kann den Dialogablauf beeinflussen.
Erwartungskonformität	Erwartungen, Eigenschaften und Gewohnheiten der Benutzer werden unterstützt.
Fehlertoleranz	Fehler erfordern keinen oder nur geringen Korrekturaufwand.
Individualisierbarkeit	Das System kann an die individuellen Bedürfnisse angepasst werden.
Lernförderlichkeit	Das System erfordert einen geringen Lernaufwand und unterstützt das Erlernen neuer Funktionen.

3 Theoretischer Hintergrund des Usability-Tests

Dieses Kapitel dient der Einführung in die Usability-Evaluation und stellt den Usability-Test als Hauptthematik vor. Nach Nielsen (1993) ist der Usability-Test in gewisser Hinsicht unersetzlich, da direkte Informationen darüber geliefert werden, wie Menschen ein Produkt benutzen und was dabei die konkreten Probleme sind. Zunächst setzt sich das folgende Kapitel mit der Definition des Usability-Tests auseinander, anschliessend wird der Usability-Test in Zusammenhang mit weiteren Usability-Methoden gesetzt. Im Kapitel 3.3 wird der Usability-Test anhand von Gütekriterien beurteilt. Die Beurteilung des Usability-Tests durch Gütekriterien wird im letzten Kapitel der vorliegenden Arbeit nochmals aufgegriffen. Weiter werden Testvorbereitung, Durchführung und Unterschiede des stationären und mobilen Labors aufgezeigt. Zum Schluss des Kapitels werden die Stärken und Schwächen des Usability-Tests reflektiert sowie ein Fazit zu den Unterschieden der beiden Testvarianten gezogen.

3.1 Definition Usability-Test

Rubin und Chisnell (2008) gebrauchen den Begriff Usability im Zusammenhang mit einem Prozess, in dem Testpersonen, die zum potentiellen Zielpublikum des Produktes gehören, beauftragt werden zu evaluieren, inwieweit das Produkt die Usability Kriterien einhält. Barnum (2011) fasst den Begriff Usability-Test zusammen als „aktiver Prozess, der auf die Beobachtung der Benutzer fokussiert ist, die Aufgaben lösen, die realistisch und bedeutsam für sie sind“ (S.13, eigene Übersetzung). Usability-Tests sind eine Kombination aus verhaltens- und befragungsbasierten Methoden, die zur Ermittlung von Schwachstellen eines Produktes eingesetzt werden (Rubin, 1994; Dumas & Redish, 1999; zitiert nach Hamborg & Gediga, 2006). Barnum (2011) unterscheidet zwischen formativen und summativen Tests. Formative Tests werden während der Entwicklung des Produkts eingesetzt und haben zum Ziel, Probleme zu finden. Summative Tests werden eingesetzt als Qualitätskontrolle, nachdem ein Produkt entwickelt wurde (Barnum, 2011).

3.2 Einordnung des Usability-Tests in die Usability-Methoden

Dieses Kapitel soll den Zusammenhang des Usability-Tests zu den anderen Usability-Methoden herstellen. Grundsätzlich können Usability-Evaluationen in empirische und analytische Methoden eingeteilt werden (Sarodnick & Brau, 2010). Empirische Methoden gewinnen Informationen durch Befragung und Beobachtung der Nutzer. Bei den analytischen Methoden hingegen wird die Beurteilung von Usability-Experten übernommen, die sich in die Nutzer hineinversetzen. Analytische Verfahren können vor allem gut in der frühen Entwicklungsphase eingesetzt werden. Die daraus resultierenden Erkenntnisse können in einer späteren Phase mittels empirischen Methoden überprüft werden. Der Usability-Test sowie das Fragebogenverfahren sind typische empirische Methoden. Typische analytische Methoden sind nach Sarodnick und Brau (2010) die heuristische Evaluation und der Cognitive Walkthrough. Diese Verfahren können bereits in der frühen Entwicklungsphase eines Systems eingesetzt werden, da sie mit wesentlich weniger weit entwickelten Entwürfen durchgeführt werden können als der Usability-Test, der typischerweise in der späteren Entwicklungsphase eines Systems eingesetzt wird.

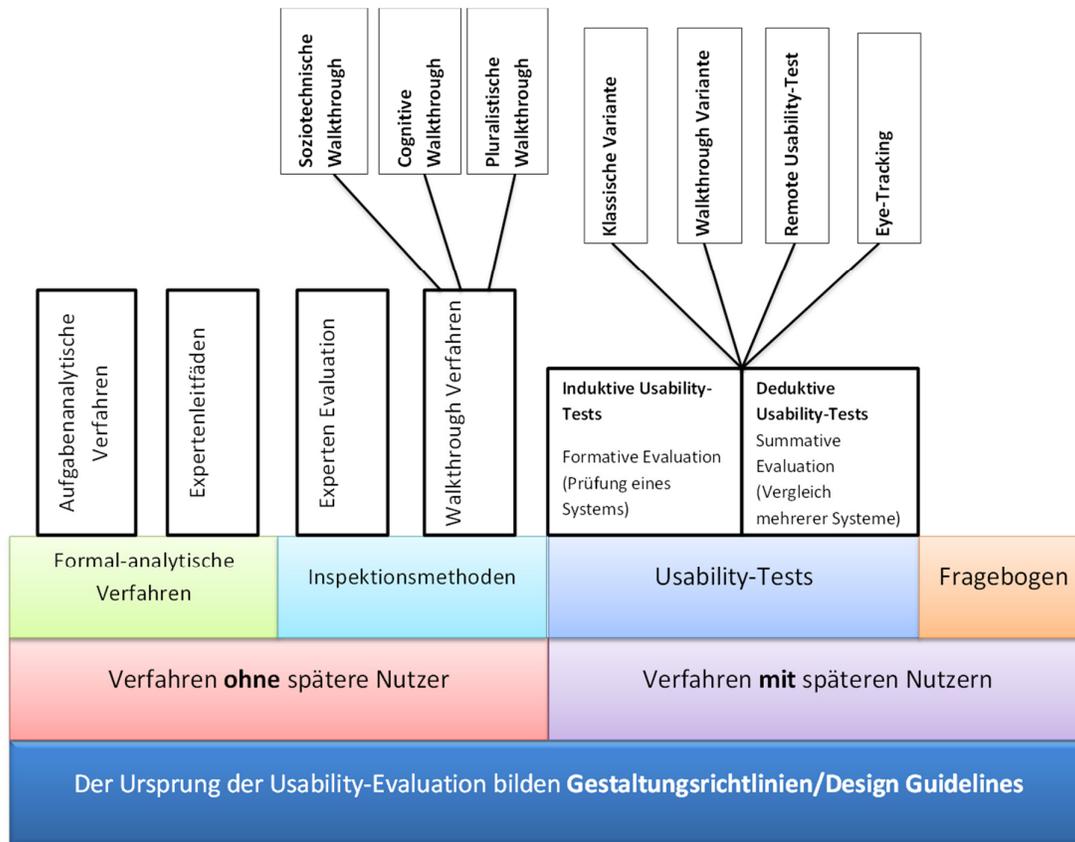


Abbildung 2: Überblick über Usability-Methoden (Eigene Darstellung, basierend auf Sarodnick & Brau, 2010)

Wie in Abbildung 2 ersichtlich ist, bilden die Gestaltungsrichtlinien den Ursprung der Usability-Methoden. Es gibt Usability-Methoden, die mit oder ohne Einbezug der späteren Nutzer als Testpersonen arbeiten. Bei Usability-Tests und Fragebogen werden mit potentiellen Nutzern zusammengearbeitet. Die formal-analytischen Verfahren und die Inspektionsmethoden arbeiten ohne direkten Einbezug von späteren Nutzern. Zu der formal-analytischen Methode gehören aufgabenanalytische Verfahren und Expertenleitfäden. Inspektionsmethoden unterscheiden zwischen Expertenevaluation und Walkthrough-Verfahren. Diese können in soziotechnische, cognitive und pluralistische Walkthrough unterschieden werden. Nach Sarodnick und Brau (2010) können induktive Usability-Tests (Verbesserung eines Systems) und deduktive Usability-Tests (Vergleich mehrerer Alternativen) unterschieden werden. Usability-Tests können als klassischer Test, Walkthrough, Remote-Test oder Test mit Eyetracking durchgeführt werden.

3.3 Beurteilung des Usability-Tests anhand methodischer Gütekriterien

Das Ziel einer Messung besteht nach Raithel (2006) darin, möglichst exakte und fehlerfreie Daten zu erheben. Allerdings können nach Raithel (2006) Messungen in der Praxis diesem Ziel kaum vollständig gerecht werden, da der wahre Messwert meist von Messfehlern beeinflusst wird. „Messungen sollen möglichst *objektiv*, *zuverlässig* und *gültig* (...) sein“ (Linert, 1969; Linert & Raatz, 1998; zitiert nach Raithel, 2006, S.42). Die Tabelle 2 (S.8) gibt einen Überblick über die wichtigsten Gütekriterien

und ihre Hauptfragestellung. Die Gütekriterien werden in Zusammenhang mit den durchgeführten Usability-Tests später in dieser Arbeit näher erläutert.

Tabelle 2: Übersicht über die wichtigsten Gütekriterien (Schmidt-Atzert & Amelang, 2012, S.131)

Gütekriterium	Leitfrage
Objektivität	Wie stark hängt das Ergebnis davon ab, wer die Testdurchführung leitet, den Test auswertet und interpretiert?
Reliabilität	Wie genau oder zuverlässig ist das Messergebnis? Wie stark verändert sich das Ergebnis beispielsweise bei einer Testwiederholung?
Validität	Wie gut gelingt es, genau das Merkmal zu messen, das mit dem Test gemessen werden soll (und nicht ein anderes)?

Objektivität

Eine vollständige Objektivität eines Usability-Tests würde dann vorliegen, wenn zwei Testleitungen den gleichen Test durchführen und exakt die gleichen Ergebnisse erhalten würden (Raithel, 2006). Die Auswertungsobjektivität ist bei quantitativen Erhebungsmethoden am höchsten (Raithel, 2006). Nach Schmidt-Atzert und Amelang (2012) ist ein standardisiertes Vorgehen wesentlich für die Qualität der Messungen. Bedingungen für die Durchführung müssen exakt festgelegt werden. Ein präzises Drehbuch für die Durchführung des Tests kann die Durchführungsobjektivität erhöhen. Jedoch ist in der Praxis eine vollständige Standardisierung meist nicht möglich. Viele Variablen wie Lichtverhältnisse, Geräusche, Raumtemperatur, Luftqualität, Sitzkomfort etc. sind nach Schmidt-Atzert und Amelang (2012) nicht bei jedem Test identisch. Auch die Testperson selbst kann die Objektivität negativ beeinflussen. So variieren Motivation, momentaner Gefühlszustand und geistige Verfassung. Das kann die Ergebnisse beeinflussen. Weitere Einflüsse werden durch die Testleitung verursacht, wie bspw. Geschlecht, Alter, Aussehen, Tonfall, Mimik und Gestik oder Erwartungen ans Testergebnis (vgl. Kapitel 3.8.1). Jedoch kann, wie bereits erwähnt, die Durchführungsobjektivität als gegeben angesehen werden, wenn alle Bedingungen, die das Testverhalten beeinflussen, festgehalten werden (Schmidt-Atzert & Amelang, 2012).

Reliabilität

Nach Nielsen (1993) ist die Reliabilität von Usability-Tests wegen der grossen individuellen Unterschiede zwischen den Testpersonen problematisch. Es ist nicht unüblich, dass die schnellste Testperson zehnmal so schnell ist wie die langsamste Testperson. Deshalb kann kein genereller Schluss daraus gezogen werden, dass bspw. Testperson A eine bestimmte Aufgabe mit der Schnittstelle X 40% schneller lösen kann als die Testperson B. mit der Schnittstelle Y. Wenn der Test mit Testperson C und D wiederholt wird, könnten die Resultate umgekehrt sein. So müssen Usability Experten oft Entscheide aufgrund von unzuverlässigen Daten fällen. Nielsen (1993) konnte in einer Untersuchungen zur Genauigkeit des Usability-Tests in Zusammenhang mit der Stichprobengrösse zeigen, dass mit einer Stichprobe von fünf Testpersonen auf einem 90%igen Konfidenzlevel der wahre Mittelwert um plus/minus (ca.) 34% abgebildet wird. Gemäss Nielsen (1993) ist dieser Genauigkeitsgrad für viele Projekte in der Praxis bereits genügend.

Validität

Die Validität ist das wichtigste Güte Merkmal eines Tests und vergleichbar mit dem Wirkungsnachweis eines Medikaments (Schmidt-Atzert & Amelang, 2012). Eine hohe Validität erlaubt nach Schmidt-Atzert und Amelang (2012), Schlüsse aus den Ergebnissen eines Tests auf ein Verhalten ausserhalb der Testsituation zu übertragen. Eine hohe Validität eines Usability-Tests bedingt nach Nielsen (1993) grosses methodisches Wissen und die Eliminierung von ungewollten Einflüssen. Typische Probleme der Validität können falsche Testpersonen, ungeeignete Testaufgaben, das Nichtbeachten des Zeitlimits oder sozialen Einflüssen sein (Nielsen, 1993). Eine Generalisierbarkeit der Ergebnisse kann verbessert werden durch eine höhere Anzahl Testpersonen oder durch den Einsatz weiterer Evaluationsmethoden (Richter & Flückiger, 2010).

3.4 Vorbereitung

Dieses Kapitel zeigt, was bei der Vorbereitung für den klassischen Usability-Test und den Usability-Walkthrough zu beachten ist. Vor jedem Usability-Test muss zwingend das Ziel des Tests mit den Auftraggebern geklärt werden, da dies grossen Einfluss auf die Art des Tests hat (Nielsen, 1993). Anschliessend wird der Testplan erstellt (Anhang B, S.57). Sehr wichtig in der Vorbereitungsphase ist die Ausarbeitung der Testaufgaben, die von den Testpersonen mit dem zu prüfenden System gelöst werden sollen. Die Qualität der Ergebnisse des Usability-Tests hängt im Wesentlichen von der Güte dieser Aufgaben ab. Sie sollen realistisch und relevant für die Nutzer sein (Nielsen, 1993). Zur Vorbereitung gehört zwingend ein Pretest, bei dem der Testablauf ausprobiert wird, Fehler im Experimentdesign aufgedeckt werden und die Zeiteinteilung überprüft wird. Der Pretest dient ausserdem dem Verfeinern des Testablaufs, dem Festlegen von Nutzerfehlern sowie dem Aufdecken von Inkonsistenzen und Schwächen im Testplan (Nielsen, 1993).

3.4.1 Testpersonen rekrutieren

Ein Usability-Test kann ohne Testperson nicht stattfinden, sie sind ein zentraler Bestandteil bei Usability-Tests. Nach Heinsen (2003) erlaubt der Einsatz von Testpersonen „wirklich Neues und sogar Überraschendes über die jeweilige Anwendung zu erfahren“ (S.188). Dieses Kapitel informiert über die Rekrutierung von Testpersonen. Das Vorgehen der Rekrutierung ist bei beiden Testvarianten dasselbe.

Da Testpersonen von grosser Bedeutung sind für Usability-Tests, ist es wichtig, die „richtigen“ Testpersonen zu rekrutieren (Heinsen, 2003). Richter und Flückiger (2010) empfehlen, Testpersonen einzuladen, die möglichst zur Benutzergruppe des zu prüfenden Systems gehören, d.h. tatsächlich zu den späteren Nutzern gehören könnten. Sie sollten nach Heinsen (2003) „das Persönlichkeitsprofil, die Soziodemografie und das Nutzungsverhalten abbilden“ (S.189), das der späteren Benutzergruppe entspricht. Mit den „falschen“ Testpersonen können Navigationsprobleme erkannt werden, nicht aber inhaltliche Probleme. So entsteht eine falsche Einschätzung des überprüften Produktes. Die richtigen Testpersonen zu rekrutieren ist nach Heinsen (2003) deshalb ein zentrales Qualitätsmerkmal guter Usability-Tests. Die Anzahl der benötigten Testpersonen hängt von den Zielen des Usability-Tests ab. Nach Richter und Flückiger (2010) genügen normalerweise fünf bis sieben Testpersonen, um die wichtigsten Anwendungsszenarien mittels eines Prototyps zu prüfen. Bei einer umfassenden

Qualitätskontrolle eines Systems empfehlen sie, mehr als zehn Testpersonen einzuladen. Die Testleitung trägt eine Verantwortung gegenüber den Testpersonen. Die Anleitung (Anhang B, S.58) soll als Hilfsmittel für einen ethisch korrekten Umgang mit Testpersonen dienen.

3.5 Durchführung

In diesem Kapitel wird detailliert auf die Durchführung von Usability-Tests eingegangen. Zunächst wird auf die Rolle der Testleitung Bezug genommen. Anschliessend wird ein weiteres wichtiges Qualitätsmerkmal der Testleitung (Hilfestellung anbieten) behandelt. Zum Schluss des Kapitels wird der Ablauf des Usability-Tests beschrieben.

3.5.1 Die goldenen Moderationsregeln

Die Testleitung muss nach Dumas und Loring (2008) beim klassischen Usability-Test und beim Usability-Walkthrough unvoreingenommen und neutral in Bezug aufs Produkt und offen und aufgeschlossen gegenüber den Testpersonen sein. Nach Dumas und Loring (2008) ist die Fähigkeit der Testleitung entscheidend für den Erfolg eines Usability-Tests, ebenso für die Validität der Daten. Die Mischung aus Neutralität und Zugänglichkeit stellt eine Herausforderung dar. Als Hilfestellung entwickelten Dumas und Loring die „goldenen Moderationsregeln“, die in der Tabelle 3 aufgelistet werden.

Tabelle 3: Die zehn goldenen Moderationsregeln (Dumas & Loring, 2008, S.3, eigene Übersetzung)

Nummer	Beschreibung der goldenen Moderationsregeln
1	Entscheide aufgrund des Testzieles, wie zu interagieren ist!
2	Respektiere das Recht der Testpersonen!
3	Denke an deine Verantwortung für zukünftige Nutzer!
4	Respektiere die Testpersonen als Experten, aber behalte die Verantwortung!
5	Sei professionell und authentisch!
6	Lass die Testpersonen sprechen!
7	Denke daran, dass deine Intuition dir schaden oder helfen kann!
8	Sei unvoreingenommen!
9	Gib keine Informationen unbeabsichtigt heraus!
10	Beobachte dich selbst!

3.5.2 Hilfestellung anbieten

Hilfestellung anzubieten kann eine Testperson ermutigen und den Prozess vorwärtsbringen. Allerdings könnten dadurch Informationen preisgegeben werden, über welche die Testpersonen ohne Hilfestellung nicht verfügt hätten. Die Intervention könnte nach Dumas und Loring (2008) verändern, wie die Testpersonen mit dem Produkt arbeiten und wie sie es wahrnehmen. Den Testpersonen weiterzuhelfen und dabei keine Informationen preiszugeben, die bei späteren Testaufgaben helfen könnten, gehört zu den wichtigsten Kompetenzen der Testleitung, sowohl beim klassischen Usability-Test als auch beim Usability-Walkthrough.

Ein Hinweis auf die Notwendigkeit des Anbieten von Hilfe ist ein Muster von wiederholtem unproduktivem Verhalten (Dumas & Loring, 2008). Wenn eine Testperson einige Alternativen ausprobiert hat und nach Hilfe fragt, wenn die Testperson auf dem Weg ist, das System zusammenbrechen zu lassen oder das Zeitlimit übersteigt, kann Hilfe angeboten werden. Eine grosse Herausforderung ist es nach Dumas und Loring (2008), einer Testperson bei einem Problem zu helfen, das bei anderen Aufgaben wieder vorkommt. Nicht alle Interventionen sind Hilfestellungen. Beispielsweise ist eine Klarstellung der Testaufgabe nach Dumas und Loring (2008) keine Hilfestellung. Die Entscheidung, ob Hilfe angeboten werden soll, hängt von der Einschätzung ab, ob die Testperson bereits lange genug probiert hat.

Stufen der Hilfestellung

Um das Anbieten von Hilfe zu vereinheitlichen, empfehlen Dumas und Loring (2008) ein vierstufiges Hilfsmodell. Dieses kann sowohl beim klassischen Usability-Test als auch beim Usability-Walkthrough angewandt werden.

Stufe 1: Eine wiederholende Sequenz unterbrechen

Wenn eine Testperson eine Sequenz mehrmals wiederholt und nicht weiterkommt, kann bereits eine Veränderung der Konzentration, wie bspw. die Frage „Was denken Sie, was hier vorgeht?“ weiterhelfen. Mit dem Hinweis „Lesen Sie bitte nochmals die Aufgabe!“ kann der Fokus aus Ziel zurückgebracht werden.

Stufe 2: Einen allgemeinen Tipp geben

Häufig kommen die Testpersonen dem Gesuchten nahe. Beispielsweise wenn das richtige Menu geöffnet ist, die Testpersonen aber nicht alles lesen oder zu schnell darüber fliegen und so die richtige Option verpassen. Die Testleitung kann eingreifen mit dem Hinweis „Erinnern Sie sich, wie Sie die Aufgabe begonnen haben? Sie waren nahe dran.“ Wenn die Testperson immer noch nicht das Gesuchte findet, kann die Testleitung den Tipp geben „Lehnen Sie sich zurück und betrachten Sie den ganzen Bildschirm.“

Stufe 3: Einen spezifischen Tipp geben

Wenn ein allgemeiner Tipp die Testpersonen nicht weiterbringt, muss der Tipp spezifischer sein. Die Testleitung kann bspw. sagen „Die Option befindet sich im Menu X.“ Diese Hilfestellung bringt die Testperson an die richtige Stelle im System, lässt aber der Testleitung die Möglichkeit offen, Informationen über bspw. die Begrifflichkeiten der Optionen zu sammeln.

Stufe 4: Den nächsten Schritt verraten

In bestimmten Situationen kann den Testpersonen gesagt werden, wie der nächste Schritt gemacht wird, bspw. „Klicken Sie auf den dritten Button.“

Wichtig ist, jede Hilfestellung und die abgebrochenen Aufgaben zu notieren. Übernehmen mehrere Personen die Testleitung, ist es für die Standardisierung wichtig, bereits im Vorfeld den Prozess der Hilfestellung festzulegen (Dumas & Loring, 2008).

3.5.3 Testablauf

Dieses Kapitel erläutert den Testablauf. Dabei wird auf die Methode des lauten Denkens eingegangen. Diese kann sowohl im klassischen Usability-Test als auch im Usability-Walkthrough angewandt werden.

Testdurchführung

Nielsen (1993) empfiehlt, vor Usability-Tests die nötigen Vorbereitungen zu treffen, wie bspw. alle Materialien bereitzustellen. Die Tests finden nach Richter und Flückiger (2010) in einem speziell eingerichteten Labor oder mithilfe eines mobilen Labors statt (vgl. Kapitel 3.6). Die Testperson bekommt die schriftliche Instruktion. Beim klassischen Usability-Test arbeitet die Testperson selbstständig im Testraum, während sich die Testleitung im Beobachtungsraum befindet. Die Testleitung soll nach Nielsen (1993) möglichst nicht mit der Testperson kommunizieren und keine persönliche Meinung vermitteln. Beim Usability-Walkthrough sitzt die Testleitung neben der Testperson im gleichen Raum, übernimmt die Moderation und begleitet die Testperson durch den Test (Richter & Flückiger, 2010). Hilfestellung wird nur wenn nötig angeboten (vgl. Kapitel 3.5.2). Nach dem Test wird die Testperson gebeten, einen Fragebogen zur Benutzbarkeit des getesteten Systems auszufüllen und wird nach ihrer Meinung und Verbesserungsvorschlägen zum System befragt (Nielsen, 1993).

Lautes Denken

Die Testpersonen werden bei beiden Testvarianten gebeten, laut zu denken. Indem nach Hamborg und Gediga (2006) die Testpersonen ihre Gedanken, Reaktionen, Freude und Missgunst während des Tests laut verbalisieren, wird auf die Problempunkte des untersuchten Produktes geschlossen. Die Methode des lauten Denkens hat nach Barnum (2011) den Vorteil, dass nachvollzogen werden kann, warum die Testperson so handelt und was sie über den Prozess denkt. Jedoch können sich nach Hamborg und Gediga (2006) das gleichzeitige Handeln und Verbalisieren gegenseitig beeinflussen und verändern so das Verhalten der Testpersonen sowie die Qualität und Vollständigkeit der Verbalisierungen. Dadurch wird gemäss Sarodnick und Brau (2010) die Bearbeitungsgeschwindigkeit gesenkt. Um diese Problematik zu umgehen, kann die Variante des postaktionalen lauten Denkens eingesetzt werden (Bowers & Snyder, 1990; Ohnemus & Biers, 1993; zitiert nach Hamborg & Gediga, 2006). Dabei wird zunächst das Verhalten der Testpersonen ohne Aufforderung zum lauten Denken auf Video aufgenommen. Erst bei der anschliessenden Betrachtung der Aufzeichnung werden die Testpersonen aufgefordert, laut zu denken. Der Vorteil dieser Variante ist nach Hamborg und Gediga (2006), dass sich eine mögliche Beeinflussung von gleichzeitigem Handeln und Verbalisieren vermeiden lässt. Der Nachteil liegt darin, dass es zu einer Beeinträchtigung der Ergebnisse durch fehlerhafte Erinnerungen kommen kann (Hamborg & Gediga, 2006).

3.6 Usability-Tests sind überall möglich

Der klassische Usability Test sowie der Usability-Walkthrough sind nach Nielsen (1993) überall möglich. Beide Testarten können unter kontrollierten Bedingungen im stationären Labor oder an einem beliebigen Ort mit dem mobilen Labor durchgeführt werden. In diesem Kapitel werden das stationäre und das mobile Labor beschrieben sowie deren Vor- und Nachteile aufgezeigt.

3.6.1 Usability-Test im stationären Labor

Die Abbildung 3 zeigt den Aufbau eines stationären Labors. Im Beobachtungsraum können sich nach Heinsen (2003) Auftraggeber und Entwickler aufhalten. Die Testleitung und eventuelle Beobachter verfolgen den Test in einem separaten Raum, der durch eine Einweg-Glasscheibe vom Testraum getrennt sein kann. Um mögliches Unbehagen der Testpersonen zu reduzieren, empfiehlt Nielsen (1993) den Testpersonen zu zeigen, wer sich hinter dem Spiegel befindet. Das zu prüfende System, die Testperson und die Situation als Ganzes werden auf Video aufgenommen. Die Testleitung und die Beobachter notieren Fehler und Schwierigkeiten der Testperson im Umgang mit dem System (Richter & Flückiger, 2010).

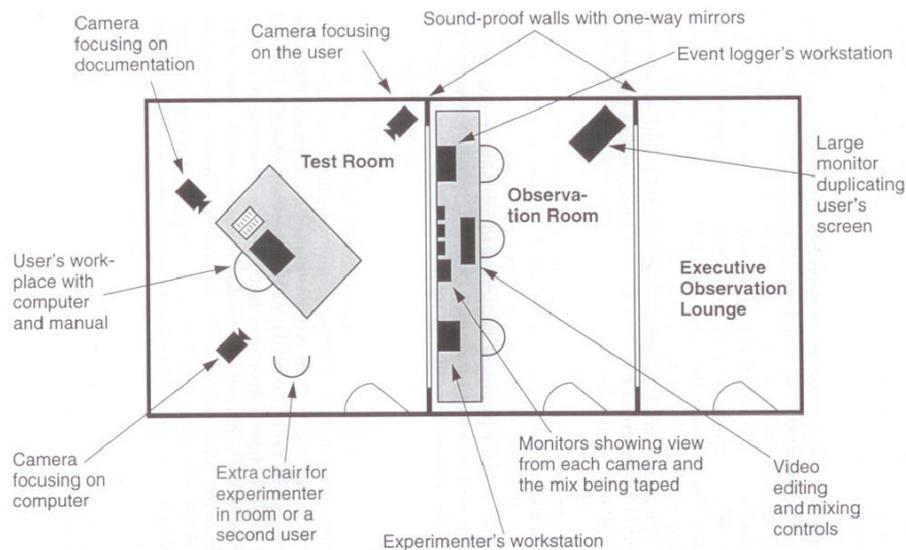


Abbildung 3: Beispielhafter Aufbau eines stationären Labors (Nielsen, 1993, S.201)

Die Vorteile des stationären Labors liegen darin, dass die Einrichtung erlaubt, die Testsituation und wichtige Erkenntnisse sofort miteinander im Besprechungsraum zu diskutieren, ohne die Testperson zu stören. So können bspw. die Entwickler eines Systems live miterleben, woran die Nutzer bei der Anwendung scheitern (Nielsen, 1993). Nach Nielsen (1993) ist der Aufbau eines stationären Labors häufig mit hohen Kosten verbunden, die in manchen Fällen nicht nötig sind, da mit sehr einfachen Mitteln überall getestet werden kann (z.B. nur mit einem Laptop). Heinsen (2003) gibt zu bedenken, dass die Anschaffung von teurer Ausrüstung dazu verleiten kann, diese zu nutzen, um die Investition zu rechtfertigen, auch wenn es nicht notwendig wäre.

3.6.2 Usability-Test mit dem mobilen Labor

Mittels Software, Kamera und Mikrofon können Usability-Tests überall aufgezeichnet werden (Barnum, 2011). Die Abbildung 4 (S.14) zeigt den Aufbau eines mobilen Labors. Heinsen (2003) empfiehlt, das mobile Labor so zu gestalten, dass es robust ist, leicht zu transportieren und schnell auf- und abzubauen ist. Dennoch soll es hochwertige Aufzeichnungen liefern und den Kunden eine Beobachtung des Tests an einem externen Gerät erlauben. Das mobile Labor soll nach Heinsen (2003) fähig sein, die Interaktion der Testperson mit dem System, das Testsetting und die Kommentare der Testperson aufzunehmen.



Abbildung 4: Aufbau des Settings eines mobilen Testlabors (soultank AG, 2014b)

Tests mit dem mobilen Labor haben laut Branum (2011) den Vorteil, dass der aktuelle Kontext, in dem das Produkt verwendet wird, beobachtet werden kann, wie bspw. Arbeitsplatz, Lichtverhältnisse, Störungen oder Ablenkungen. Die Schwierigkeit des Testens mit dem mobilen Labor liegt nach Barnum (2011) darin, die Umgebung zu kontrollieren, was sich negativ auf die Reliabilität der Testergebnisse auswirken kann. Möglicherweise kann sich die Testperson aufgrund der Störungen der Umgebung nicht so gut konzentrieren, wie es in einem Labor der Fall wäre. Zudem könnte die Testperson gehemmt sein, laut zu denken, wenn andere zuhören.

3.7 Stärken des Usability-Tests

Dieses Kapitel reflektiert die Stärken beider Testvarianten. Das folgende Kapitel setzt sich mit den Schwächen beider Testvarianten auseinander.

Einer der Stärken des **klassischen Usability-Tests** ist nach Richter und Flückiger (2010), dass die methodischen Gütekriterien (vgl. Kapitel 3.3) grösstenteils eingehalten werden können. Die Testleiterunabhängigkeit und Objektivität können beim klassischen Usability-Test als gegeben betrachtet werden. Da die Testleitung während des Tests weniger Einfluss auf die Testpersonen hat, wirkt der Testleitereffekt nicht so stark.

Wie bereits im Kapitel 1.1 erwähnt liegt die Stärke des **Usability-Walkthroughs** für die soultank AG einerseits im einfachen Einsatz der Methode, da kein Einwegspiegel notwendig ist oder vergleichbare technische Ausrüstung. Andererseits wirkt die Situation im Usability-Walkthrough für die Testperson häufig natürlicher. Durch den persönlichen Face-to-Face-Kontakt ist es gemäss Köppel (2007) einfacher, Vertrauen zwischen Testleitung und Testperson aufzubauen. So kann auf Unwohlsein, Nervosität oder Testangst der Testpersonen vermutlich besser eingegangen werden.

3.8 Schwächen des Usability-Tests

Zu den Schwächen des **klassischen Usability-Tests** gehört nach Richter und Flückiger (2010), dass er erst spät in der Entwicklungsphase eines Produktes eingesetzt werden kann, da die Testpersonen selbständig mit den Prototyp arbeiten müssen. Dies verursacht höhere Kosten als bei einer Evaluation in einer früheren Entwicklungsphase.

Die wahrscheinlich grösste Schwäche des **Usability-Walkthrou**s liegt in der grösseren Beeinflussung der Testpersonen durch die Testleitung. Die Testleiterunabhängigkeit ist nach Richter und Flückiger (2010) bei dieser Methode nicht gegeben. Mit dieser Herausforderung setzt sich das folgende Kapitel auseinander.

3.8.1 Testleitereffekt

Im Folgenden werden Faktoren beschrieben, welche die Interaktion zwischen der Testperson und der Testleitung beeinflussen und sich auf das Resultat des Tests auswirken können. Der nachfolgend beschriebene Testleitereffekt betrifft vor allem den Usability-Walkthrough, da sich Testleitung und Testperson im gleichen Raum befinden und die Beeinflussung dadurch erhöht ist. Anschliessend wird die Thematik Testangst, welche beide Testvarianten betrifft, beschrieben. Später in der vorliegenden Arbeit wird die Thematik Testangst nochmals aufgegriffen.

Bei Untersuchungen in den Sozialwissenschaften ist das Phänomen des Testleitereffekts bekannt (Schnepper, 2004). Dieser tritt nach Schnepper (2004) in Situationen auf, in denen die Anwesenheit der Testleitung (ggf. kombiniert mit Erwartungsausserungen) die Testpersonen zu Verhaltensänderungen anregt, die ohne die Anwesenheit der Versuchsleitung nicht stattgefunden hätten. Die Testleiterunabhängigkeit ist nach Jäger und Kubinger (2003) dann erfüllt, „wenn das Testverhalten der Tp und damit der für sie resultierende Testwert unabhängig von zufälligen oder systematischen Verhaltensvariation ist, wie sie durch die Person des TI bedingt sind“ (S.196). Gemäss Jäger und Kubinger (2003) bleibt stets ein „menschliches“ Risiko, dass die Testleitung bestimmte Erwartungen an die Testperson hat. Gemäss dem Halo-Effekt besteht die Möglichkeit, dass die Testleitung unwissentlich aufgrund der Erwartungshaltung das Verhalten der Testperson mittels nonverbaler Kommunikation konsequent verstärkt. Kubinger und Wurst (2000, zitiert nach Jäger & Kubinger, 2003) beobachteten Abweichungen bei Intelligenztests von bis zu 6-IQ-Punkten, wenn zwei verschiedene Personen die Testleitung übernahmen. Eine weitere Möglichkeit des Einflusses der Testleitung betrifft das Geschlecht. Hawell (1979, zitiert nach Jäger & Kubinger, 2003) entdeckte geschlechtsspezifische Testleitereffekte. Mädchen, die in einem Intelligenztest von Männern getestet wurden, schnitten signifikant besser ab, als Mädchen, die von Frauen getestet wurden. Aus diesen Erkenntnissen lässt sich schliessen, dass der Testleitereffekt auch bei völlig identischer Durchführung des gleichen Tests Einfluss auf die Resultate ausübt. Die Personencharakteristika der Testleitung (z.B. Geschlecht, Hautfarbe etc.) können bei einer Testperson unterschiedliches Verhalten hervorrufen und führen so zu unterschiedlichen Testresultaten (Jäger & Kubinger, 2003). Bis zum jetzigen Zeitpunkt konnten mittels Literaturrecherche keine Studien gefunden werden, die den Testleitereffekt in Usability-Tests untersuchen. Forschung zu Usability wird an vielen Hochschulen, wie bspw. an der Technischen Universität Chemnitz am Institut für Psychologie, welches den Human- und Sozialwissenschaften angegliedert ist, betrieben (Technische Universität Chemnitz, 2014). Demnach könnte angenommen werden, dass der in der Sozialwissenschaft bekannte Testleitereffekt auch in Usability-Tests auftreten könnte. Dies könnte durch zukünftige Forschung bestätigt werden.

3.8.2 Testangst

Weitere mögliche Einflüsse auf die Testresultate können durch Testangst der Testpersonen hervorgerufen werden. Eine Testsituation ist nach Margraf (2009) für viele Menschen, unabhängig vom Geschlecht oder Alter, eine besondere Herausforderung. Der Anspruch, dass eine Testsituation ganz ohne Lampenfieber gemeistert wird, ist nahezu unhaltbar. In solchen Situationen ist ein gewisser Pegel an Aufregung „ganz normal“. Jedoch kann es im schlimmsten Fall zu Leistungsängsten und Leistungsblockaden kommen, welche einen Teufelskreis aus Leistungsangst, Leistungsdefizit und Versagen auslösen (Margraf, 2009). Obwohl im Usability-Test den Testpersonen explizit mitgeteilt wird, dass nicht ihre Fähigkeiten getestet werden, sondern das System, empfinden viele Testpersonen die Situation dennoch als Test. In ihrer Studie konnten Hembree und College (1988) zeigen, dass Testangst eine schlechtere Leistung in Tests zur Folge hat. Die Testangst bezieht sich auf das Selbstwertgefühl und die Angst vor negativer Bewertung, Abwehr, und andere Formen von Angst. Die Literaturrecherche brachte keine Ergebnisse zu Studien, die Testangst in Usability-Tests untersuchten, zutage. Folglich ist es sinnvoll, diese Thematik zukünftig zu erforschen. Der Usability-Test weist einige Charaktereigenschaften eines Tests auf, bspw. dass Personen unter Beobachtung und in einer bestimmten Zeit Aufgaben lösen müssen und dabei gefilmt werden. Diese Gemeinsamkeiten könnten darauf hinweisen, dass Testangst auch bei Usability-Tests auftreten könnte. So kann angenommen werden, dass bisherige Forschungsergebnisse anderer Studienrichtungen teilweise für die Situation im Usability-Test übertragen werden können.

3.9 Fazit zum Vergleich der Testvarianten

Die Tabelle 4 (S.17) verdeutlicht die Gemeinsamkeiten und Unterschiede zwischen dem klassischen Usability-Test und dem Usability-Walkthrough. Die Gemeinsamkeiten der Testvarianten bestehen vor allem im Testaufbau, der Vorbereitung, der Rekrutierung von Testpersonen und teilweise in der Durchführung. Unterschiede können in der Art des getesteten Systems bestehen (ausgereifter oder noch nicht sehr ausgereifter Prototyp) und vor allem in den Konsequenzen durch die An- bzw. Abwesenheit der Testleitung. Weiter unterscheiden sich die Testvarianten in den Bereichen Testleitereffekt, Begleitung, Kommunikation und möglicherweise Testangst.

Tabelle 4: Fazit Vergleich zwischen klassischem Usability-Test und Usability-Walkthrough

Merkmal	Klassischer Usability-Test	Usability-Walkthrough
Art des Testsystems	Eher ausgereifte Prototypen	Ausgereifte und nicht sehr ausgereifte Prototypen
Testkosten	X	X
Testaufgaben	X	X
Pretest	X	X
Testpersonen rekrutieren	X	X
Anzahl Testpersonen	X	X
Goldene Moderationsregeln	X	X
Lautes Denken	X	X
Hilfestellung anbieten	X	X
Stationäres Labor	X	X
Mobiles Labor	X	X
Testangst	Eher mehr	Eher weniger
Testleitereffekt	Weniger	Mehr
Kommunikation	Kontrolliert über Gegensprechanlage (ohne Face-to-Face), so wenig wie möglich	Mehr Kommunikation (Face-to-Face)
Begleitung durch den Test	Die Testpersonen arbeiten selbstständig	Die Testleitung begleitet die Testperson durch den Test

4 Methodisches Vorgehen

Dieses Kapitel beschreibt das methodische Vorgehen, das Forschungsdesign, die Hypothesen, die Variablen, die Rekrutierung der Testpersonen, die Beschreibung der Stichprobe sowie die Durchführung der Tests, das eingesetzte Material und die Auswertung der Daten.

4.1 Forschungsdesign

Testsystem

Das zentrale Ziel dieser Forschungsarbeit war ein methodischer Vergleich zweier Varianten des Usability-Tests. Um die Vergleichbarkeit zweier Tests zu ermöglichen, musste bei beiden Tests das gleiche System mit den gleichen Aufgaben getestet werden. Beim klassischen Usability-Test lösten die Testpersonen eigenständig Aufgaben mit dem System, weshalb das Testsystem in einem eher ausgereiften Entwicklungszustand sein musste. Aus diesem Grund wurde als Testsystem eine Website gewählt, die bereits auf dem Internet vorhanden war. Die Wahl fiel auf die Website der Klubschule Migros (Abbildung 5).

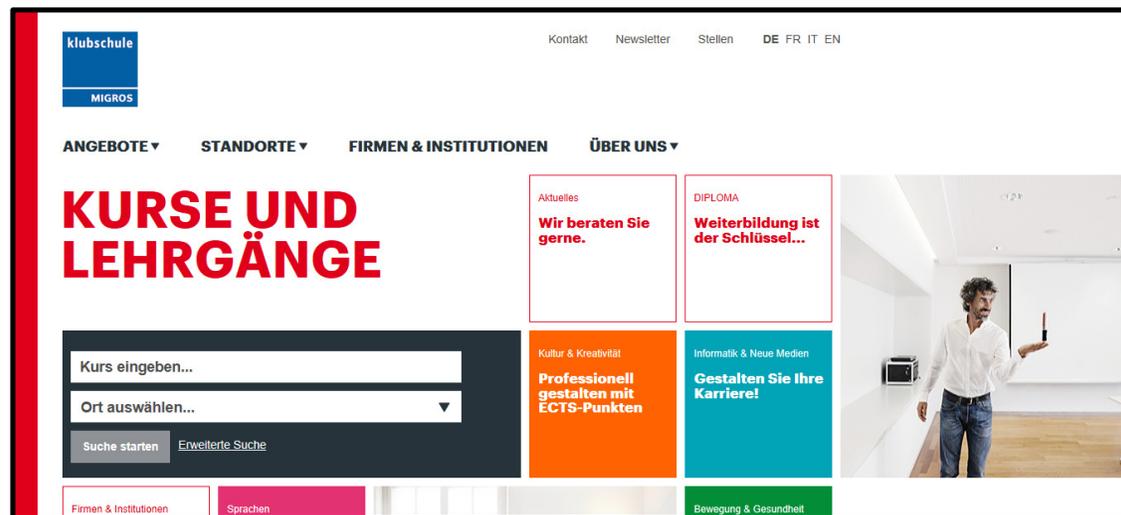


Abbildung 5: Startseite der Klubschule Migros (Klubschule Migros, 2014c)

Diese Website eignete sich für die Untersuchung, weil sie eine breite Benutzergruppe anspricht, welche für die Rekrutierung einer Stichprobe von 40 Testpersonen von Vorteil war. Die Zielgruppe der Klubschule Migros ist vielfältig hinsichtlich Bildung, Herkunft und Alter. Das Credo „Weiterbildung ist zugänglich für alle“ (Klubschule Migros, 2014b) gehört zum Bildungsleitbild der Klubschule Migros. Ein weiterer Grund für die Wahl der Website ist die umfangreiche Suchfunktion. Wie eine Studie zur Suchfunktion bei Online-Shops ergab, gewöhnten sich die Nutzer infolge Google daran, einfach über die Suchfunktion zu suchen und sich nicht lange durch die Navigation zu klicken. Die Suchfunktion ist eine der am häufigsten genutzten Funktionen und hat immens an Bedeutung gewonnen (E-Commerce-Leitfaden, 2014).

Bedingungsvariation

Die Testpersonen wurden in zwei Gruppen eingeteilt. Die Testleitung führte mit der Hälfte der Testpersonen den klassischen Usability-Test (Test A) und mit der anderen Hälfte den Usability-Walkthrough (Test B) durch. Jede Testperson durchlief nur eine Bedingung (Test A oder B) ohne Messwiederholung. Um die Testergebnisse vergleichen zu können, mussten alle Testpersonen dieselben Aufgaben lösen. Somit wurde für die Untersuchung ein Between-groups design gewählt. Dies bedeutete, dass für jede Bedingung eine andere Gruppe von Testpersonen untersucht wurde.

Gruppe A → UV- Bedingung 1 (Test A: Klassischer Usability-Test) → Messung der AV's

Gruppe B → UV- Bedingung 2 (Test B: Usability-Walkthrough) → Messung der AV's

Statistische Hypothese

In der vorliegenden Arbeit wurde die statistische Hypothese (Tabelle 5) untersucht.

Tabelle 5: Statistische Hypothese

Statistische Hypothese	
Nullhypothese (H0)	Es gibt keine statistische Evidenz, dass der Testkoeffizient strikt grösser oder kleiner als Null ist. Es bestehen keine signifikanten Unterschiede zwischen Test A und Test B.
Alternativhypothese (H1)	Es bestehen signifikante Unterschiede zwischen Test A und Test B.

Die Regressionskoeffizienten wurden daraufhin getestet, ob sie sich signifikant von Null unterscheiden. So konnte bestimmt werden, ob es sich um signifikante, d.h. relevante Unterschiede zwischen den zwei Testvarianten handelte. Die Datenerhebung und -auswertung der Untersuchung diente dazu, die Hypothesen an- oder abzulehnen. Die Nullhypothese wurde verworfen (und somit die Alternativhypothese angenommen), wenn die Datenanalyse signifikante Unterschiede zwischen den beiden Testvarianten feststellte. Die Alternativhypothese wurde verworfen (und somit die Nullhypothese angenommen), wenn keine signifikanten Unterschiede beobachtet wurden.

Versuchsordnung

Die Bedingung, welche variiert wurde, war die Testvariante. Bei beiden Gruppen wurden die gleichen drei Variablen gemessen, welche für die Beantwortung der Fragestellung relevant waren. Mit der Untersuchung des Zusammenhangs der erhobenen Daten konnten die drei abhängigen Variablen (Tabelle 6) miteinander verglichen werden. Mit der Prüfung der Signifikanz wurde getestet, ob die Unterschiede bedeutsam waren.

Tabelle 6: Versuchsordnung

Test A		Test B
Gesamtwert FTE	←→	Gesamtwert FTE
Anzahl gelöster Testaufgaben	←→	Anzahl gelöster Testaufgaben
Anzahl gefundene Usability-Probleme	←→	Anzahl gefundene Usability-Probleme

4.1.1 Abhängige Variablen

Wie in Tabelle 7 ersichtlich, wurden mehrere abhängige Variablen untersucht. Diese werden im Folgenden beschrieben.

Tabelle 7: Abhängige Variablen (multivariat)

Abhängige Variablen	Operationalisierung zur Messung
Abhängige Variable 1	Gesamtwert „Fragen zum Testerleben“ (FTE)
Abhängige Variable 2	Anzahl gefundener Usability-Probleme
Abhängige Variable 3	Anzahl gelöster Testaufgaben

Gesamtwert Fragen zum Testerleben

Die quantitative Messung des Testerlebens wurde mittels des Fragebogens zum Testerleben (FTE) operationalisiert. Es handelte sich dabei um einen standardisierten Fragebogen (Hogrefe Testsystem 4, 2014). Mittels dieser Operationalisierung konnte erhoben werden, wie positiv oder negativ die Testpersonen den Test erlebt haben. Wie in Tabelle 8 ersichtlich, bestand der Fragebogen aus zehn Items.

Tabelle 8: Die 10 Items des Fragebogens zum Testerleben

Items
1. Es fiel mir leicht, den Test zu bearbeiten.
2. Zeitweise habe ich den Anweisungen nicht folgen können.
3. Ich hatte ziemliche Angst, die Aufgaben nicht zu schaffen.
4. Meine Ausdauer war während der ganzen Zeit gleichmässig gut.
5. Ich hatte zeitweise ziemliche Probleme mit der Konzentration.
6. Ich könnte noch länger solche Aufgaben lösen.
7. Zeitweise habe ich nicht verstanden, was am Computer ablief.
8. Während der Bearbeitung hatte ich das Gefühl, aufgeben zu wollen.
9. Ich bin mit den Leistungen zufrieden, die ich zeigen konnte.
10. Ich war durch die Aufgaben stark überfordert.

Die Testpersonen beantworteten die Fragen aufgrund dessen, wie zutreffend sie die jeweiligen Aussagen empfanden. Die Skala bestand aus sechs Ausprägungen (trifft nicht zu / wenig zu / etwas zu / ziemlich zu / überwiegend zu / ganz genau zu). Je nach Polung des Items wurde für die Antwortstufen ein Punktwert von 0 bis 5 bzw. 5 bis 0 vergeben. Anschliessend wurden die Punkte zu einem Gesamtwert zusammengerechnet. Je höher der Gesamtwert einer Testperson ausfiel, je positiver hat diese den Test erlebt.

Die qualitative Messung der Variable wurde mittels der Items in der Tabelle 9 (S.21) operationalisiert.

Tabelle 9: Items zum Testerleben für qualitative Auswertung

Dimension	Indikator	Item
Testerleben	Allgemeines Befinden	Wie haben Sie den Test erlebt?
	Testsituation im Labor	Wie haben Sie die Testsituation im Labor erlebt?
	Erleben des Aufzeichnens mit der Kamera	Wie war es für Sie, dass der Test und ihr Umgang mit dem System von einer Kamera aufgezeichnet wurden?
	Empfinden des lauten Denkens	Wie haben Sie das laute Denken erlebt?
	Testangst	Wie haben Sie den Umstand empfunden, dass Sie Aufgaben unter Beobachtung lösen mussten?
	Abwesenheit der Testleiterin (je nach Testvariante)	Wie haben Sie es erlebt, dass sich die Testleiterin während des Tests in einem anderen Raum aufgehalten hat?
	Anwesenheit der Testleiterin (je nach Testvariante)	Wie haben Sie es erlebt, dass die Testleiterin während des Tests neben Ihnen gesessen hat?
	Erneute Teilnahme an einem Usability-Test	Würden Sie in Zukunft wieder einmal an einem Usability-Test mitmachen?

Anzahl gefundener Usability-Probleme

Die Frage nach der Effektivität der Tests wurde durch die Variable Anzahl der gefundenen Usability-Probleme operationalisiert. Die gefundenen Usability-Probleme wurden festgehalten und gezählt. Um die Erhebung zu standardisieren, wurden Auswertungsleitlinien ausgearbeitet (Anhang E, S.78). Für die qualitative Auswertung wurde die Effektivität mittels der Art der Usability-Probleme operationalisiert.

Anzahl gelöster Testaufgaben

Die Frage nach der zeitlichen Effizienz wurde mittels der Anzahl gelöster Testaufgaben operationalisiert und gemessen. Um die Datenerhebung zu standardisieren, wurden Auswertungsleitlinien ausgearbeitet.

4.1.2 Unabhängige Variablen

Testvariante

Der Fokus dieser Arbeit lag in der Analyse der Unterschiede zweier Varianten des Usability-Tests. Die Testpersonen wurden in zwei Gruppen eingeteilt und durchliefen entweder den Test A oder B. Es wurde somit die Variable Test manipuliert.

Kontrollvariablen

Verschiedene Faktoren können die abhängigen Variablen beeinflussen. Ein möglicher Einfluss der Testleitung (Verhalten und äusseres Erscheinungsbild) auf die abhängigen Variablen wurde mittels eines Drehbuchs kontrolliert. Das **Alter** hat sehr wahrscheinlich einen Zusammenhang mit dem Internetumgang und der Technikakzeptanz der Testpersonen, was die abhängigen Variablen beeinflusst und deshalb kontrolliert werden musste. Im Test-Setup wurde davon ausgegangen, dass der Zusammenhang zwischen dem Alter und der jeweiligen abhängigen Variable linear ist. Das **Geschlecht** der Testpersonen könnte einen indirekten Einfluss auf die abhängige Variable haben, bspw. durch Sympathieeffekte im Zusammenhang mit der Testleitung. **Internetkenntnisse** und **Häufigkeit der Internetnutzung** könnten einen Einfluss auf die abhängigen Variablen haben, indem die Testpersonen, die sich gut im Internet zurechtfinden, beim Usability-Test auch weniger Schwierigkeiten haben, mehr Aufgaben lösen und sich beim Test aufgrund ihrer „guten Leistung“ wohlfühlen. Da an zwei verschiedenen Orten getestet wurde, einmal im stationären Labor der FHNW in Olten und einmal mit dem mobilen Labor in einem Schulzimmer einer Sekundarstufenschule in Schaffhausen, wurde ein möglicher Einfluss des **Versuchsortes** auf die Datenerhebung ausgeschlossen. Der Einfluss, ob die Testpersonen das **Testsystem** bereits kannten, wurde ebenfalls kontrolliert.

4.2 Rekrutierung der Testpersonen

Eine quantitative Auswertung der Daten verlangt nach Raithel (2006) eine genügend grosse Stichprobe, damit Schlüsse aus den Ergebnissen einer Untersuchung gezogen werden können. Im Hinblick auf die zeitlichen Ressourcen konnte diese Forderung nicht vollumfänglich erfüllt werden. Dennoch wurde versucht, diesen Mangel durch Balancierung bei der Rekrutierung der Testpersonen auszugleichen. Es wurde darauf geachtet, dass die Stichprobe aus 50% Männern und 50% Frauen bestand. Da gemäss Nielsen (1993) bereits eine Stichprobe von fünf Testpersonen ausreicht, um 85% der Usability-Probleme aufzudecken, kann eine Stichprobe von 40 Testpersonen als ausreichend betrachtet werden. Richter und Flückiger (2010) empfehlen Testpersonen aus der potentiellen Benutzergruppe des Systems einzuladen. Da die Benutzergruppe der Klubschule Migros breit ist, wurde bei der Rekrutierung darauf geachtet, Männer und Frauen verschiedenen Alters und mit unterschiedlichen Bildungshintergrund einzuladen. Es wurden per E-Mail Studenten des Studiengangs Angewandte Psychologie der Fachhochschule Nordwestschweiz angefragt. Damit sich die Chancen auf eine Teilnahme erhöhten, wurden die Teilnehmer persönlich mit Namen angeschrieben. Es meldeten sich rund 27 Personen, davon konnte 20 Personen eine persönliche Einladung per E-Mail gesandt werden. Da sich in Olten zu wenige Testpersonen fanden, wurde die Suche auf den Bekanntenkreis der Testleitung ausgeweitet. Diese Personen wurden mittels Telefon, E-Mail und Kurznachrichten kontaktiert. So konnten nochmals 20 Testpersonen rekrutiert werden.

Zuteilung der Testpersonen zu den Testbedingungen

Der mögliche Einfluss der Variable Versuchsort wurde balanciert, indem an beiden Orten gleich viele Testpersonen rekrutiert wurden. In Olten konnten zwölf weibliche und acht männliche Testpersonen zum Versuch eingeladen werden. In Schaffhausen konnten zwölf männliche und acht weibliche Testpersonen rekrutiert werden. Die weitere Zuteilung der Testpersonen zur Versuchsbedingung (Test A oder B) basierte auf einer Zufallsauswahl mittels Losziehung. Die zwölf Frauen in Olten wurden durch Ziehen den zwei Bedingungen zugeordnet, d.h. sechs Frauen nahmen am klassischen Usability-Test (A) und sechs Frauen am Usability-Walkthrough (B) im Usability-Labor in Olten teil. Das gleiche zufällige Auswahlverfahren wurde mit den Männern in Olten durchgeführt, vier Männer machten beim Test A und vier Männer beim Test B mit. Dasselbe Verfahren wurde für den zweiten Versuchsort angewandt. Die Abbildung 6 stellt die zufällige Zuteilung der Testpersonen zu den zwei Versuchsbedingungen schematisch dar.

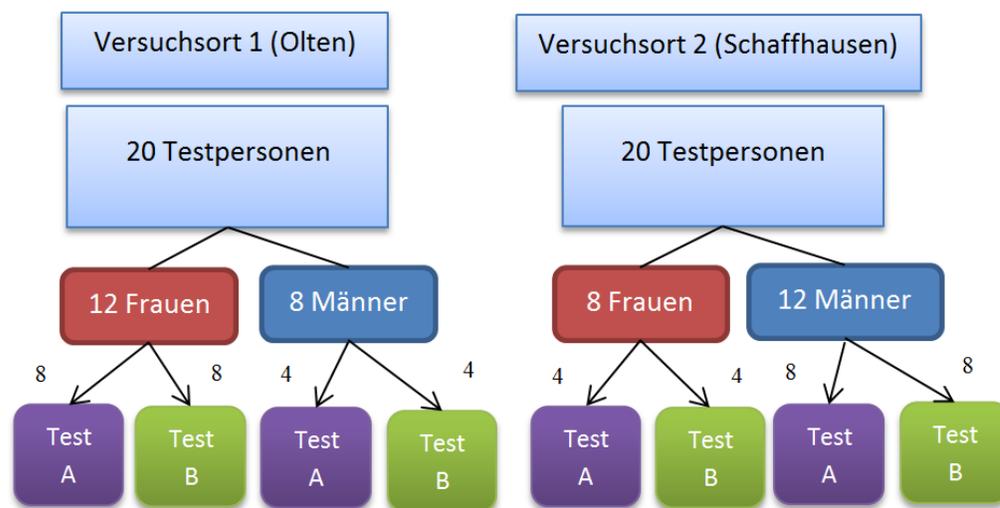


Abbildung 6: Zufällige Zuteilung der Testpersonen zu den Versuchsbedingungen

4.3 Beschreibung der Stichprobe

Zur Untersuchung eingeladen wurden 20 Frauen und 20 Männer. Obwohl bei der Rekrutierung darauf geachtet wurde, gleich viele Männer wie Frauen einzuladen, konnte dies aus organisatorischen Gründen nicht vollständig umgesetzt werden. Die Stichprobe bestand schlussendlich aus 21 Männern und 19 Frauen. Die Tabelle 10 (S.24) zeigt die Durchschnittswerte der Testpersonen.

Tabelle 10: Durchschnittswerte der Testpersonen

Durchschnittswerte	Anzahl Jahre
Durchschnittsalter aller Testpersonen	35.85
Durchschnittsalter Frauen	31.79
Durchschnittsalter Männer	39.52
Jüngste Testperson	16
Älteste Testperson	73
Median alle Testpersonen	29
Median Frauen	28
Median Männer	37

Wie in Abbildung 7 ersichtlich, war der höchste Bildungsabschluss bei sieben Testpersonen momentan die Sekundarstufe, fünf Testpersonen hatten eine Berufslehre absolviert, acht eine Berufsmatura, vier eine Gymnasiale Maturität, neun einen Bachelorabschluss einer (Fach-) Hochschule und sieben Testpersonen hatten einen Masterabschluss, ein Diplom oder Lizenziat einer (Fach-) Hochschule.

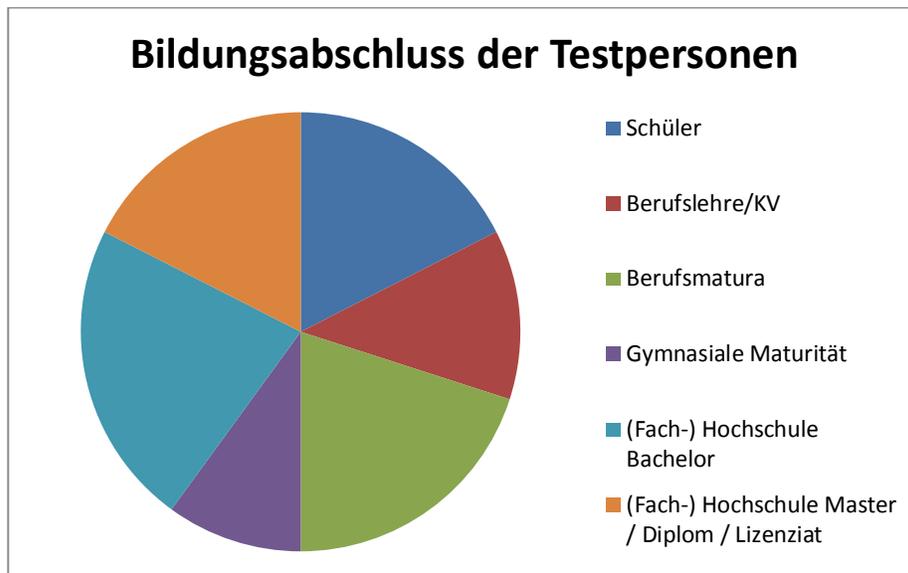


Abbildung 7: Bildungsabschluss der Testpersonen

Wie in Abbildung 8 (S.25) ersichtlich ist, benutzte eine Testperson das Internet sehr selten, sieben mehrmals wöchentlich, 15 täglich und 17 Testpersonen benutzten das Internet mehrmals täglich.

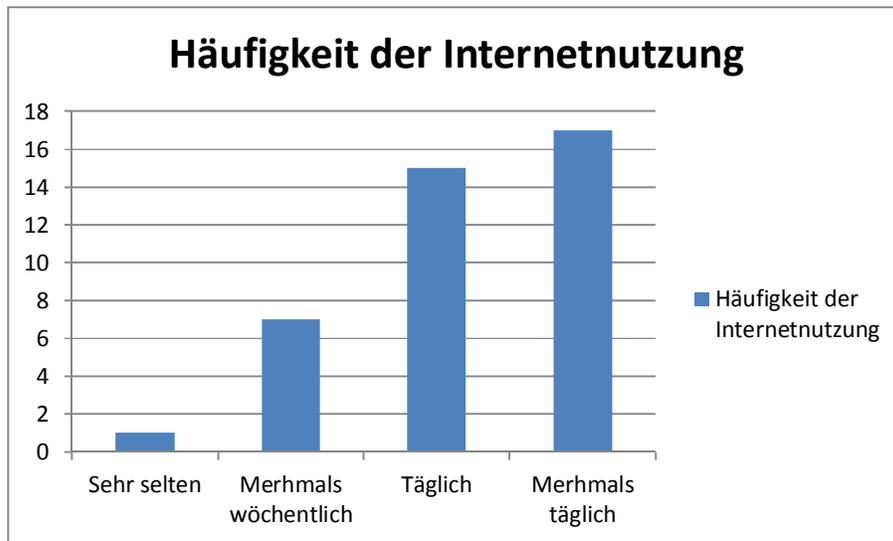


Abbildung 8: Häufigkeit der Internetnutzung der Testpersonen

4.4 Materialien und Setting

Für die Untersuchung wurde das Usability-Labor in Olten sowie die Ausrüstung des mobilen Labors der soultank AG benötigt. Benötigt wurden das Testleiterskript für den klassischen Usability-Test (Anhang D, S.65) und den Usability-Walkthrough (Anhang D, S.71), die Drehbücher für den klassischen Usability-Test (Anhang C, S.59) und den Usability Walkthrough (Anhang C, S.62), der Zeitplan und die Testaufgaben inklusive Vor- und Nachfragebogen für die Testpersonen (Anhang D, S.73). Eine Stoppuhr wurde verwendet, um die Testdauer von 30 Minuten anzuzeigen. Schokolade als Dankeschön und Wasser als Erfrischung für die Testpersonen wurden bereitgestellt.

Tests im stationären Usability-Labor

Das Usability-Labor der FHNW in Olten verfügt über drei Testräume. Im separaten Beobachtungsraum konnten die Tests koordiniert und aufgezeichnet werden. Beim klassischen Usability-Test konnte die Testleitung über die Monitore in die Testräume sehen und via Lautsprecher hören, was die Testpersonen sagten (Abbildung 9, S.26). Bei Bedarf konnten der Testperson Anweisungen über das Mikrophon gegeben werden. Beim Usability-Walkthrough sass die Testleitung im gleichen Raum wie die Testperson.

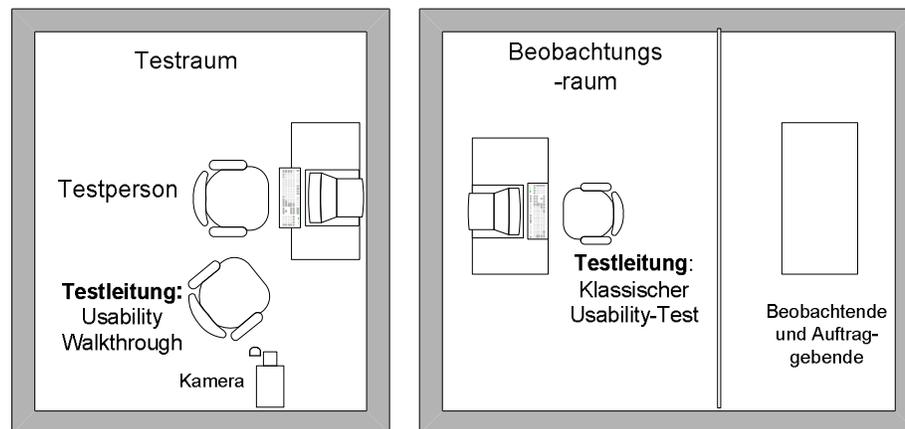


Abbildung 9: Klassischer Usability-Test und Usability-Walkthrough im stationären Labor in Olten

Das linke Foto der Abbildung 10 zeigt den Arbeitsplatz der Testleitung im Beobachtungsraum. Auf dem rechten Foto ist der Bildschirm zu sehen. Es zeigt einerseits den Bildschirm der Testperson, damit die Testleitung erkennen kann, wie die Testperson das System bedient, andererseits ist der Testraum zu sehen, welcher von der Kamera aufgezeichnet wird.

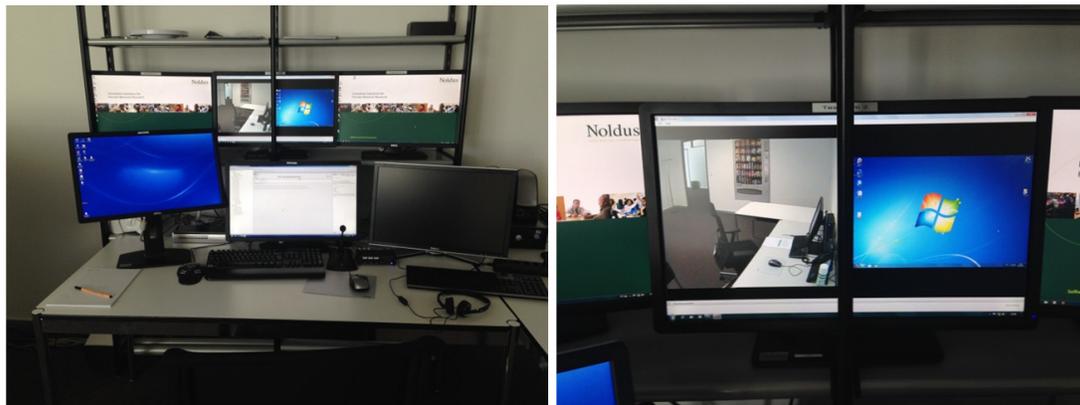


Abbildung 10: Arbeitsplatz der Testleitung im Usability-Labor in Olten

Tests mit dem mobilen Labor

Für die Durchführung des Versuchs in einer Schule der Sekundarstufe wurde die Ausrüstung des mobilen Labors von der soultank AG benötigt. Das mobile Labor bestand aus zwei Laptops, einem Monitor, einer Tastatur, einer Maus, einem Mikrophon und weiteren technischen Hilfsmitteln. Für den Versuch wurde ein Labor in einem freien Schulzimmer eingerichtet, das für die ganze Zeit der Versuchsdurchführung reserviert werden konnte. Die Testpersonen wurden nicht gestört und konnten in Ruhe arbeiten. Das Zimmer hatte eine Trennwand, so dass sich Testperson und Testleitung nicht sehen konnten. Für den klassischen Usability-Test saßen die Testpersonen alleine am Tisch und lösten selbstständig die Aufgaben. Die Testleitung sass im anderen Bereich des Raumes und verfolgte den Test auf dem Laptop. Mittels einer speziellen Software wurden die Aktivität mit dem System sowie das Bild der Testperson auf den Laptop der Testleitung übertragen und aufgezeichnet. Die

Testperson und die Testleitung konnten sich gegenseitig hören. Für den Usability-Walkthrough setzte sich die Testleitung neben die Testperson (Abbildung 11).

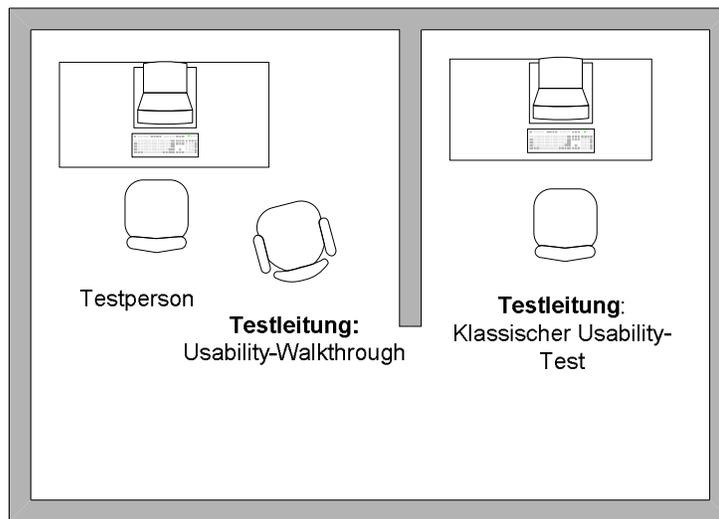


Abbildung 11: Klassischer Usability-Test und Usability-Walkthrough mit dem mobilen Labor

4.5 Durchführung

In diesem Kapitel wird die Versuchsdurchführung der beiden Testvarianten beschrieben.

Pretest

Um das Forschungsdesign, die Versuchsdurchführung, die Zeitplanung und die Verständlichkeit der Instruktionen zu testen, wurde im Vorfeld ein Pretest mit zwei Testpersonen durchgeführt. Mit einer Testperson wurde der klassische Usability-Test und mit der anderen den Usability-Walkthrough durchgeführt. Mithilfe der Rückmeldung der Testpersonen wurden die Testunterlagen und das Testleiterskript nach dem Pretest nochmals überarbeitet. Der Pretest diente ausserdem als Übung für die Testleitung, um den Ablauf und die Instruktion des Versuchs zu proben sowie die technische Ausrüstung bedienen zu können.

4.5.1 Durchführung des klassischen Usability-Tests

Die Testpersonen wurden vor dem Labor von der Testleitung begrüsst. Es wurde versucht, eine angenehme Atmosphäre zu schaffen, indem ein paar zwanglose Fragen gestellt wurden. Es wurde der Testperson für die Zeit und Teilnahme gedankt. Um die Testpersonen transparent über den Versuch zu informieren, wurde der Testperson der Beobachtungsraum gezeigt und erklärt, dass die Testleitung während des Tests in diesem Raum sitzt, dass aber über ein Mikrophon miteinander gesprochen werden kann und dass die Testleitung auf dem Bildschirm sehen kann, was die Testperson macht. Nach dieser Einführung wurde die Testperson von der Testleitung an den Arbeitsplatz im Testraum geführt. Die Testleitung erklärte die Laboreinrichtung und fragte nach Erlaubnis für die Video- und Audioaufzeichnung. Es wurde der Testperson Anonymität der erhobenen Daten zugesichert. Um die Instruktion und Erklärungen bei allen Testpersonen gleich zu halten, wurde ein Drehbuch verfasst.

Die Testleitung erklärte den Ablauf des Tests und die Spielregeln. Die Testperson füllte den Vorfragebogen aus, und die Testleitung ging in den Beobachtungsraum, wo das Aufnahmesystem gestartet wurde. Die Testperson hatte 30 Minuten Zeit, um Aufgaben zu lösen. Die Testperson musste allerdings nicht alle Aufgaben in dieser Zeit schaffen. Während die Testperson die Aufgaben löste, hielt sich die Testleitung so weit wie möglich mit Anmerkungen und Hilfestellungen zurück. Wenn die Testperson die Aufgabe nicht richtig verstanden hatte oder einen Teil der Aufgabe vergass, wies die Testleitung die Testperson darauf hin. Die vier Stufen der Hilfestellungen (Abbildung 12) wurden angewandt, allerdings nur, wenn es unumgänglich war. Das vierstufige Hilfestufenmodell diente der Testleitung als Anhaltspunkt bei Hilfeersuchen durch die Testpersonen. Für eine grösstmögliche Standardisierung und möglichst hohe Durchführungsobjektivität hielt sich die Testleitung an die Vorgaben des Drehbuches und des Hilfestufenmodells. Die Testleitung notierte sich im Testleiterskript, welche Aufgaben abgebrochen wurden und wieviel Hilfe angeboten wurde.

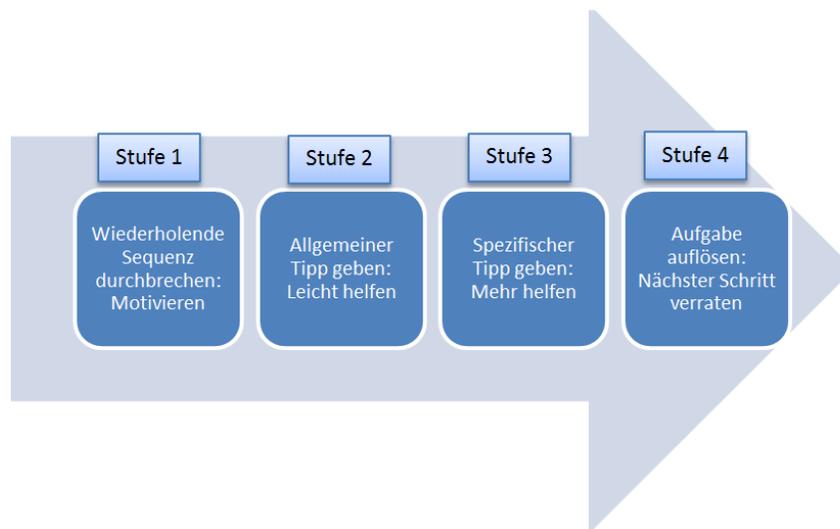


Abbildung 12: Vierstufiges Hilfsmodell (Eigene Darstellung, basierend auf Dumas & Loring, 2008)

Nach 30 Minuten wurde das Lösen der Aufgaben beendet, und die Testperson füllte den Nachfragebogen aus. Die Testleitung wartete eine Weile, kam in den Testraum zurück, setzte sich neben die Testperson und wartete, bis diese den Nachfragebogen fertig ausgefüllt hatte. Anschliessend stellte die Testleitung einige offene Fragen. Zum Schluss hatte die Testperson die Möglichkeit, ihrerseits offen gebliebene Fragen zu stellen. Der Testperson wurde für die Teilnahme gedankt und ein kleines Präsent als Dankeschön überreicht. Die Testleitung verabschiedete sich und stoppte die Aufnahme im Beobachtungsraum.

4.5.2 Unterschiede in der Durchführung vom klassischen Usability-Test zum Usability-Walkthrough

In diesem Kapitel werden die Bedingungen beschrieben, die im Usability-Walkthrough unterschiedlich zum klassischen Usability-Test gehandhabt wurden.

Anwesenheit der Testleitung

Der wesentliche Unterschied der beiden Testvarianten lag in der An- bzw. Abwesenheit der Testleitung. Im klassischen Usability-Test befand sich die Testleitung in einem anderen Raum, während die Testpersonen die Aufgaben selbstständig lösten. Im Usability-Walkthrough hingegen befand sich die Testleitung während des gesamten Versuchs neben den Testpersonen im gleichen Raum.

Anbieten von Hilfestellung

Beim klassischen Usability-Test wurde darauf geachtet, so wenig wie möglich ins Testgeschehen einzugreifen und nur minimale Hilfestellung anzubieten. Das heisst, dass die Hilfestufen im klassischen Usability-Test später angeboten wurden als im Usability-Walkthrough. Damit sollte der Einfluss der Testleitung im klassischen Usability-Test so gering wie möglich gehalten werden. Allerdings griff die Testleitung auch beim Usability-Walkthrough nicht unnötig ins Geschehen ein, sondern nur, wenn es nötig wurde und die Aufgabe zu einer Belastung für die Testperson wurde.

Kommunikation und Begleitung

Ein weiterer Unterschied lag in der vermehrten Kommunikation aufgrund der Begleitung der Testperson durch die Testleitung. Wenn bspw. eine Testperson ihren Unmut über eine unbrauchbare Systemfunktion bekundete, fragte die Testleitung nach Alternativen, z.B. „Was hätten Sie erwartet?“, „Wie würden Sie diese Funktion besser gestalten“? So wurden einerseits Verbesserungsvorschläge generiert, andererseits gab es durch die vermehrte Kommunikation zwischen Testleitung und Testpersonen mehr Unterbrechungen im Testlauf.

Zusammengefasst lässt sich festhalten, dass der schwerwiegendste Unterschied in der Versuchsanordnung der beiden Testgruppen die An- bzw. Abwesenheit der Testleitung war. Weitere Unterschiede lagen in der unterschiedlichen Zeitspanne, wann Hilfestellung angeboten wurde sowie in der vermehrten Kommunikation zwischen Testleitung und Testperson, hervorgerufen durch die Begleitung beim Usability-Walkthrough.

4.6 Auswertung der Daten

Dieses Kapitel beschreibt den Vorgang der quantitativen und qualitativen Auswertung.

4.6.1 Quantitative Auswertung

Die Unterfragestellungen wurden operationalisiert (vgl. Kapitel 4.1.1) und mit einer empirischen Untersuchung gemessen. Die vier abhängigen Variablen wurden statistisch ausgewertet. Nach Bühl (2009) dient die Regressionsanalyse zur Ermittlung der Art des Zusammenhangs. Mittels der Analyse kann der Wert einer (abhängigen) Variable aus den Werten anderer (unabhängiger) Variablen vorhergesagt werden (Bühl, 2009). Der Zusammenhang wurde mit einer einfachen linearen Regressionsanalyse ermittelt ($y = a + b \cdot x$). Dabei war b der Regressionskoeffizient und a der Ordinatenabschnitt. Die Resultate werden im Kapitel 5 dargestellt und erläutert.

4.6.2 Qualitative Auswertung

Testerleben

Die Variable Testerleben wurde qualitativ mittels eines kurzen Interviews zum Schluss des Tests erhoben und ausgewertet. Die Auswertung der Daten orientierte sich an der qualitativen Inhaltsanalyse, angelehnt an Mayring (2010). Die erhobenen Daten aus den Interviews flossen in die Indikatorenliste ein. Die Daten aus dem klassischen Usability-Test (Anhang E, S.80) und dem Usability-Walkthrough (Anhang E, S.82) wurden im ersten Schritt gesammelt. In einem weiteren Schritt wurde eine erste Priorisierung vorgenommen und wichtige inhaltliche Aussagen zur Beantwortung der Fragestellung fett markiert. In einem weiteren Materialdurchgang wurden die Daten nochmals reduziert, ohne dass relevante Aussagen verloren gingen. Die Endergebnisse werden im Kapitel 5.1.2 vorgestellt.

Art der Usability-Probleme

Die Abbildung 13 stellt die Vorgehensweise der Datenauswertung dar. Zunächst wurde der klassische Usability-Test ausgewertet. Die Probleme wurden beschrieben und zum besseren Verständnis mit Screenshots ergänzt. In einem zweiten Schritt wurden die Probleme einer Kategorie zugeteilt. Diese diente zur Übersicht und Einordnung der Art der Usability-Probleme (bspw. Navigation, Begrifflichkeit, Dateneingabe etc.). Anschliessend wurde bestimmt, gegen welche Kriterien der Dialoggestaltung nach DIN EN ISO 9241-10, 9241-110 die Probleme verstießen (vgl. Kapitel 2.3). Zum Schluss wurden die Probleme in der Tabelle (Anhang E, S.84) nach ihrem Inhalt geordnet und mit Zwischentiteln versehen (bspw. Design- und Layout-Probleme, Probleme der Navigation etc.). Der gleiche Prozess wurde bei der Auswertung des Usability-Walkthroughs angewandt. Es wurde gekennzeichnet, in welchen Testvarianten die Usability-Probleme gefunden wurden. Die Ergebnisse der Analyse werden im Kapitel 5.2.2 beschrieben.

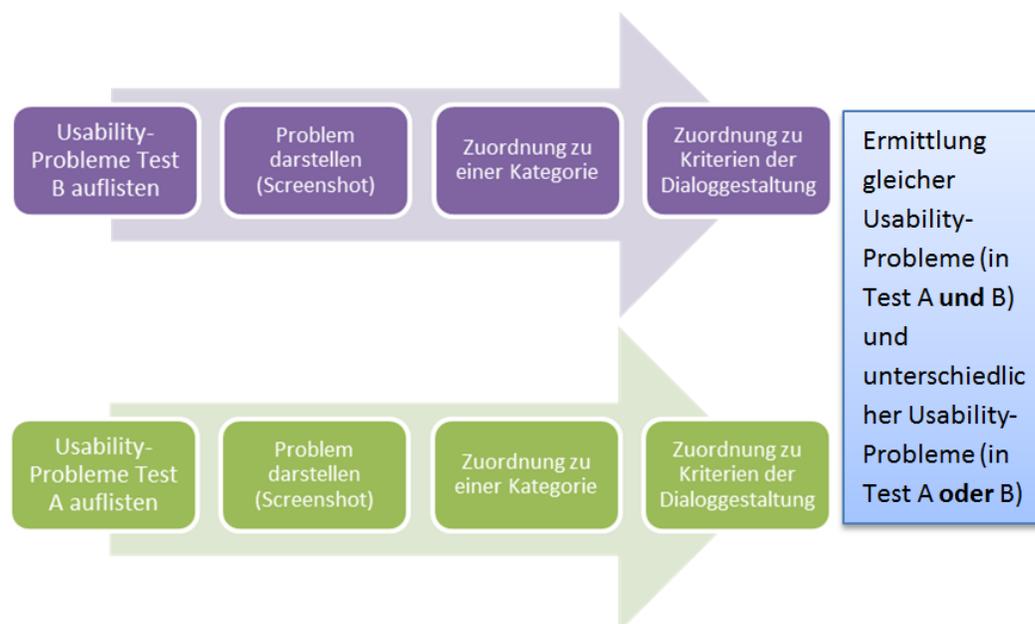


Abbildung 13: Vorgehensweise der Datenauswertung der Art der Usability-Probleme des klassischen Usability-Tests und des Usability-Walkthroughs

5 Ergebnisse

Im Folgenden werden die Ergebnisse der quantitativen und qualitativen Analyse erläutert. Im Kapitel 5.1.1 wird die statistische Auswertung der abhängigen Variable Testerleben dargestellt. Nachfolgend werden die Ergebnisse der qualitativen Interviews mit den Testpersonen zum Testerleben aufgezeigt. Die Effektivität der Testvarianten wird im Kapitel 5.2 anhand der Anzahl gefundener Usability-Probleme quantitativ ausgewertet und ergänzt mit der qualitativen Auswertung der Art der gefundenen Usability-Probleme. Im Kapitel 5.3 werden die Ergebnisse der Anzahl gelöster Testaufgaben vorgestellt und somit die Frage nach der zeitlichen Effizienz der zwei Usability-Tests beantwortet.

5.1 Testerleben der zwei Varianten des Usability-Tests

5.1.1 Quantitative Ergebnisse zum Gesamtwert FTE

In diesem Kapitel wird die Unterfragestellung eins mittels statistischer Auswertung der Daten beantwortet. Die Analyse bezieht den Gesamtwert aus dem Fragebogen zum Testerleben (FTE) ein. Je höher die Summe des Fragebogens einer Testperson ausfiel, desto positiver hat diese Person den Test erlebt (vgl. Kapitel 4.1.1).

Ausgeschlossene Variablen

Es wurden für die folgenden Regressionen nicht alle Variablen verwendet. Nachfolgend wird erklärt, welche Variablen aus dem Analyseprozess ausgeschlossen wurden und weshalb.

Die unabhängige Variable **Einschätzung der Benutzbarkeit** der Website wurde nicht in die Analyse miteinbezogen. Das Untersuchungsziel bestand darin, Aussagen über zwei Testmethoden zu machen mittels der Messung von Variablen, die einen grossen Bezug zur Benutzerfreundlichkeit der Website haben. Ein Einbezug dieser Variable wäre für die Beantwortung der Fragestellung folglich nicht sinnvoll. Die Variablen Häufigkeit der Internetnutzung und Einschätzung der Internetkenntnisse korrelierten sehr stark ($r=0.504$), da beide die Fähigkeit im Umgang mit dem Internet messen. Demnach reichte es aus, eine der beiden Variablen im Modell zu behalten. Es wurde angenommen, dass die Häufigkeit der Internetnutzung von den Testpersonen objektiver betrachtet wurde als die Selbsteinschätzung der Internetkenntnisse. Deshalb wurde die Variable **Einschätzung der Internetkenntnisse** aus dem Analyseprozess ausgeschlossen.

Einfache Regressionsanalyse

Für die Auswertung der ersten Unterfragestellung musste die Stichprobengrösse von 40 Testpersonen auf 39 reduziert werden, da ein Individuum nicht alle Fragen beantwortet hatte. Für die Analyse wurde eine einfache lineare Regression durchgeführt, was einem Mittelwertsvergleich entspricht. Als Grundlage dieser Berechnung diente die Formel $Gesamtwert\ FTE = \beta_0 + \beta_1\ test$. Wie in der Tabelle 11 (S.32) dargestellt, ist der Testkoeffizient von 3.816 mit 90%iger Konfidenz signifikant positiv. Das bedeutet, dass die Testpersonen den Test B (Usability Walkthrough) positiver als den Test A (klassischer Usability-Test) erlebt haben. Aussagen über diesen Effekt sind jedoch mit Vorsicht zu geniessen, da einflussreiche Variablen nicht miteinbezogen wurden. Zudem sind das R^2 (8.50%) und das korrigierte

R^2 (6.03%) sehr klein. Die Varianz der Testvariable erklärt folglich nur wenig von der Varianz der Variable Testerleben.

Tabelle 11: Effekt der Testvariante auf das Testerleben (einfache Regression)

Skala	Koeff.	Stand.f.	t	P> t	F(1,37)	3.44
Test	3.816	2.06	1.85	0.072	Prob > F	0.0717
Konst.	37.5	1.44	26.10	0.000	R ²	0.0850
					Korr. R ²	0.0603

Legende: Test (0)= Test A , Test (1)=Test B, Konst.= Konstante, Koeff.= Koeffizient, Stand.f.= Standardfehler, t= Testprüfgrösse des t-Tests, F (Anzahl Steigungsparameter, Anzahl Freiheitsgrade)= Prüfwert des F-Tests, P>|t|= P-Wert

Multiple Regressionsanalyse

Eine weitere Regressionsanalyse wurde mit den Variablen Test, Ort, Alter und Bildung durchgeführt. Wie in der Tabelle 12 nachgelesen werden kann, ist der Test-Koeffizient mit 3.824 positiv und auf einem 90%igen Konfidenzlevel signifikant. Dies bedeutet, dass die Testpersonen den Usability-Walkthrough signifikant positiver erlebt haben als den klassischen Usability-Test. Die Variablen Ort und Bildung haben keinen signifikanten Effekt, erklären jedoch einen nennenswerten Teil der Varianz der Variable Testerleben. Das Alter hat mit einem P-Wert von 0.096 mit 90%iger Konfidenz einen signifikant negativen Effekt (-0.155) auf das Testerleben. Dies bedeutet, dass die älteren Testpersonen den Test negativer erlebt haben als die jüngeren Testpersonen. Dies könnte darauf zurückzuführen sein, dass jüngere Testpersonen eher vertraut im Umgang mit dem Computer und dem Internet sind und an Tests mit dem Computer eher gewöhnt sind. Allerdings sind diese Erkenntnisse mit Vorsicht zu interpretieren. Das korrigierte R^2 ist mit 9.78% sehr tief. Das ganze Modell erklärt also wenig der Varianz der abhängigen Variable Testerleben. Zudem ist der F-Wert mit 2.03 gefährlich tief, und das ganze Modell müsste zugunsten der Nullhypothese (alle Steigungsparameter gleich null) auf einem 90%igem Konfidenzniveau knapp verworfen werden (da $\text{Prob}>F= 0.1122 > 0.1$). Die Annahme der Nullhypothese würde demnach bedeuten, dass die Mittelwerte der Tests sich nicht (oder nur zufällig) unterscheiden. Auf ein Miteinbeziehen von weiteren Kovariablen wurde verzichtet, da dies nur zu noch tieferen F-Werten führen würde.

Tabelle 12: Effekt der Testvariante auf das Testerleben (multiple Regression mit den Kovariablen Ort, Alter und Bildung)

Testerleben	Koeff.	Stand.f.	t	P> t	F(4,34)	2.03
Test	3.824	2.17	1.76	0.087	Prob > F	0.1122
Ort	-0.436	2.32	-0.19	0.852	R ²	0.1928
Alter	-0.155	0.09	-1.71	0.096	Adj. R ²	0.0978
Bildung	0.962	0.70	1.38	0.178		
Konst.	39.894	3.13	12.75	0.000		

Zwischenfazit

Zusammenfassend lässt sich aus der quantitativen Analyse feststellen, dass der Usability-Walkthrough von den Testpersonen als positiver als der klassische Usability-Test erlebt wurde. Ebenfalls signifikante Effekte konnten bei den Variablen Alter und Test festgestellt werden. Die älteren Testpersonen haben den Test eher negativer erlebt als jüngere Testpersonen. Allerdings sind die Ergebnisse aufgrund des tiefen korrigierten R^2 und F-Werts mit Vorsicht zu interpretieren.

5.1.2 Qualitative Ergebnisse zum Testerleben

In diesem Kapitel werden die Resultate der Analyse der Interviews mit den 40 Testpersonen zum Erleben des Tests dargestellt.

Erleben des klassischen Usability-Tests

Im Allgemeinen haben die Testpersonen den Test mehrheitlich als spannend und angenehm erlebt. Die Instruktionen der Testleitung waren verständlich. Der Schwierigkeitsgrad der Aufgabenstellungen wurde von den Testpersonen als gut eingeschätzt und die Dauer des Tests als angenehm. Viele Testpersonen sahen den Sinn der Usability-Tests. Die Testpersonen empfanden es mehrheitlich als angenehm, ab und zu mit der Testleitung kommunizieren zu können, bspw. bei Unklarheiten. Einige Testpersonen wussten nicht genau, wann eine Aufgabe fertig gelöst war, und konnten sich Bestätigung bei der Testleitung holen. Es wurde von den Testpersonen geschätzt, dass sie in der Nachbesprechung ihre individuelle Meinung und Ansicht zur Website der Klubschule Migros der Testleitung mitteilen konnten. Die Testsituation im Labor wurde von den Testpersonen als angenehm erlebt. Da die Testpersonen vor dem Test die Testleitung kennenlernten und ihnen der Beobachtungsraum und die Ausstattung gezeigt wurden, verringerte sich die Angst und Unsicherheit bei den Testpersonen. Die Gegensprechanlage stellte für die Testpersonen kein Problem dar. Die Kamera störte die Testpersonen nicht. Es wurde allerdings der Wunsch geäußert, die Kamera so zu verbergen, dass sie nicht mehr sichtbar ist. Für einige Testpersonen war das laute Denken ungewohnt und schwierig, oder es ging vergessen. Häufig war das laute Denken für die Testpersonen anstrengend, da sich die Personen einerseits auf die Aufgaben fokussieren und andererseits laut ihre Gedanken aussprechen mussten. Diese Problematik wurde bereits im Kapitel 3.5.3 erwähnt. Einige Testpersonen verspürten Testangst. Sie wollten die Aufgaben besonders gut machen. Eine schlechte Leistung wäre ihnen peinlich gewesen. Dass nicht alle Aufgaben gelöst werden mussten, hat allerdings vielen Testpersonen die Angst vor dem Test genommen. Für die Testpersonen war es kein Problem, dass sich die Testleitung in einem anderen Raum befand, da sie zuvor über die Situation aufgeklärt wurden und sich den Beobachtungsraum der Testleitung ansehen konnten. Alle Testpersonen würden in Zukunft wieder einmal an einem Usability-Test mitmachen.

Erleben des Usability-Walkthroughs

Diejenigen Ergebnisse, die gleich wie im klassischen Usability-Test waren, werden hier nicht explizit erwähnt.

Die Testpersonen haben den Test mehrheitlich als angenehm, interessant und spannend empfunden. Die Testpersonen empfanden den Testraum als angenehm, eine Person empfand den Raum als kahl

(ohne Blumen oder sonstige Dekorationsgegenstände). Das laute Denken wurde teilweise als etwas unangenehm und schwierig erlebt, vor allem bei anspruchsvolleren Aufgaben. Die Testpersonen fühlten sich nicht gestresst oder unter Druck, da ihnen die Angst zu Beginn durch die Instruktion genommen wurde. Der Usability-Walkthrough wurde von den Testpersonen nicht als Test empfunden. Die Testpersonen fanden die Anwesenheit der Testleiterin angenehm. Die persönliche Kommunikation wurde geschätzt. Der Usability-Walkthrough wurde eher als natürliches Gespräch denn als Test erlebt. Eine Testperson konnte sich durch die Anwesenheit der Testleiterin schlechter konzentrieren. Einigen Testpersonen war es unangenehm, dass sich die Testleitung beim Ausfüllen des Nachfragebogens nebenan befand und sehen konnte, welche Kreuze die Testperson setzte.

Fazit zu den Unterschieden des Testerlebens

Beide Testvarianten wurden im Allgemeinen als angenehm erlebt. Bei beiden Tests wurden die Versuchszeit und Aufgaben als angemessen angesehen. Die Problematik des lauten Denkens kam bei beiden Testvarianten gleichermaßen zutage. Einige Testpersonen hatten Mühe sich gleichzeitig auf die Aufgabenstellungen und auf das laute Verbalisieren ihrer Gedanken zu konzentrieren. Unterschiede zwischen den beiden Testvarianten konnten beim Indikator Testangst festgestellt werden. Während die Situation in den klassischen Usability-Tests eher als Test erlebt wurde, empfanden die Testpersonen den Usability-Walkthrough eher als natürliches Gespräch. Dies könnte darauf hinweisen, dass die Testpersonen das selbstständige Lösen von Testaufgaben alleine in einem Raum mental eher mit einer Testsituation verbinden, als wenn die Testleitung daneben sitzt und mehr kommuniziert wird. Einige Testpersonen empfanden es allerdings als ein wenig unangenehm, dass die Testleitung beim Ausfüllen des Nachfragebogens daneben sass. Zusammenfassend kann gesagt werden, dass beide Testvarianten von den Testpersonen als angenehm erlebt wurden, der klassische Usability-Test eher als Test und der Usability-Walkthrough eher als Gespräch.

5.2 Effektivität der zwei Varianten des Usability-Tests

5.2.1 Quantitative Ergebnisse zur Anzahl der Usability-Probleme

In diesem Kapitel wird die Unterfragestellung zwei mittels statistischer Analyse der Daten beantwortet. Anschliessend folgt die qualitative Auswertung der Unterfragestellung.

Einfache Regressionsanalyse

Der Zusammenhang der unabhängigen Variable Test und der abhängigen Variable Anzahl gefundener Usability-Probleme wurde mittels einer einfachen linearen Regression berechnet. Als Grundlage dieser Berechnung diente die Formel $Anzahl\ gefundener\ Usability-Probleme = \beta_0 + \beta_1 \text{ test}$. Wie in Tabelle 13 (S.35) ersichtlich, fällt der Test-Koeffizient mit -1.3 negativ aus. Dies bedeutet, dass beim Usability Walkthrough im Durchschnitt weniger Usability-Probleme gefunden wurden, wobei dieser Zusammenhang bei Weitem nicht signifikant ist. Insgesamt wurden im klassischen Usability-Test 338.5 Usability-Probleme gefunden und im Usability-Walkthrough 312.5. Demnach generierte der klassische Usability-Test absolut 26 Usability-Probleme mehr als der Usability-Walkthrough. Das R^2 fällt mit 1.85% sehr niedrig aus, weshalb die Variable Test wenig Information über die Anzahl der gefundenen

Usability-Probleme enthält. Das korrigierte R^2 ist mit -0.73% sogar negativ. Somit macht es keinen Sinn, dieses Modell auszuwerten, was durch den tiefen F-Wert von 0.72 bestätigt wird.

Tabelle 13: Effekt der Testvariante auf die Anzahl Usability-Probleme (einfache Regression)

Anz. U.-Probleme	Koeff.	Stand.f.	t	P> t	F(1,38)	0.72
Test	-1.3	1.53	-0.85	0.402	Prob > F	0.402
Konst.	16.93	1.09	15.59	0.000	R^2	0.0185
					Korr. R^2	-0.0073

Legende: Anz. U.-Probleme= Abhängige Variable“ Anzahl gefundener Usability-Probleme“

Neu einbezogene Variablen

Für die weitere Regression wurden nebst der unabhängigen Variable Test noch fünf weitere Variablen einbezogen. Da aufgrund des Rekrutierungsaufwandes an zwei verschiedenen Orten Tests durchgeführt wurden, musste die Variable **Ort** mit den unterschiedlichen Labors beachtet und kontrolliert werden. Ebenfalls neu wurde die Variable **Bildungsabschluss** in die Analyse miteinbezogen. Der Bildungsabschluss hat vermutlich einen grossen Einfluss auf die Internetkenntnisse und beeinflusst somit ebenfalls die abhängigen Variablen. Die Variable **Häufigkeit der Internetnutzung** hat, so die Annahme, einen grossen Zusammenhang mit den abhängigen Variablen zeitliche Effizienz und Anzahl Usability-Probleme. Aus diesem Grund musste die Variable Häufigkeit der Internetnutzung zwingend in die Analyse integriert werden. Die Variable **Kenntnis der Website** wurde in die Analyse miteinbezogen. Diejenigen Testpersonen, welche die Website der Klubschule Migros bereits kennen, sind im Umgang mit der Website vertraut, finden sich vermutlich schneller zurecht und beeinflussen deshalb die abhängigen Variablen zeitliche Effizienz und Anzahl Usability-Probleme. Hierfür musste statistisch kontrolliert werden. Zuletzt wurde die unabhängige Variable **Geschlecht** in die Analyse miteinbezogen, um auszuschliessen, dass das Geschlecht der Testpersonen einen Einfluss auf die abhängigen Variablen hat und dadurch die Resultate verzerren würde.

Multiple Regressionsanalyse

Durch die Kovariablen Ort, Bildung, Häufigkeit der Internetnutzung, Kenntnis der Website und Geschlecht gewinnt das Modell an Validität. Der F-Wert von 5.59 in der Tabelle 14 (S.36) ist erfreulich hoch. Das R^2 ist mit 50% und das korrigierte R^2 mit 41% sehr hoch, d.h. das Modell erklärt einen grossen Teil der Varianz der Anzahl Usability-Probleme. Das korrigierte R^2 von 41% ist sehr hoch, wenn die Stichprobengrösse beachtet wird und den Umstand, dass durch den Einbezug von sechs unabhängigen Variablen sechs Freiheitsgrade verloren gingen. Der Testkoeffizient ist wiederum negativ, aber nicht signifikant (P-Wert=0.571). Folglich hatte die Testvariante keinen Einfluss auf die Anzahl der gefundenen Usability-Probleme. Der Koeffizient Ort ist auf sämtlichen üblichen Konfidenzniveaus signifikant (P-Wert=0.000). Das zeigt, dass dessen Integration von beträchtlicher Bedeutung war. Der Grund dafür liegt wahrscheinlich darin, dass an einem Ort (Olten) die Untersuchung mehrheitlich mit Hochschulstudierenden durchgeführt wurde. Des Weiteren gibt es statistischen Beweis, dass der höchste Bildungsabschluss einen signifikant positiven Einfluss auf die Anzahl der gefundenen Usabili-

ty-Probleme hat (P-Wert=0.013). Dies bedeutet, dass Testpersonen mit höherem Bildungsabschluss mehr Usability-Probleme fanden als solche mit tieferem Bildungsabschluss. Dasselbe gilt für die Häufigkeit der Internetnutzung (P-Wert=0.008). Auch die Kenntnis der Homepage hat mit einem P-Wert von 0.097, zumindest auf einem 90%igen Konfidenzniveau, einen positiven Effekt auf die Anzahl gefundener Usability-Probleme. Bei Männern ist ein positiver Effekt auf die Anzahl Usability-Probleme feststellbar. Dieser ist allerdings bei Weitem nicht signifikant. Das Alter wurde für die lineare Regression auf Test, Ort, Bildung, Häufigkeit der Internetnutzung, Kenntnis der Website und Geschlecht nicht berücksichtigt, da das korrigierte R^2 ansonsten von 41.4% auf 39.6% fallen würde.

Tabelle 14: Effekt der Testvariante auf die Anzahl Usability-Probleme (multiple Regression mit den Kovariablen Ort, Bildung, Häufigkeit der Internetnutzung, Kenntnis der Website und Geschlecht)

Anz. U.-Probleme	Koeff.	Stand.f.	t	P> t	F(6,33)	5.59
Test	-0.720	1.26	-0.57	0.571	Prob > F	0.0004
Ort	-6.284	1.52	-4.13	0.000	R ²	0.5041
Bildung	0.851	0.33	2.61	0.013	Korr. R ²	0.4140
Häufigkeit der Internetnutzung	-2.478	0.88	-2.82	0.008		
Kenntnis der Website	-2.585	1.52	-1.71	0.097		
Geschlecht: (1=männlich)	1.366	1.29	1.06	0.297		
Konst.	28.234	4.33	6.52	0.000		

Zwischenfazit

Zusammenfassend liess sich kein signifikanter Zusammenhang zwischen Testvariante und Anzahl Usability-Probleme feststellen. Allerdings war der Testort höchst signifikant. Bildungsabschluss, Häufigkeit der Internetnutzung und Kenntnis der Website hatten einen signifikant positiven Einfluss auf die Anzahl der Usability-Probleme.

5.2.2 Qualitative Ergebnisse zur Art der Usability-Probleme

Als Ergänzung zur quantitativen Auswertung werden in diesem Kapitel die Ergebnisse der qualitativen Auswertung der Art der Usability-Probleme dargestellt. Zunächst werden zum besseren Verständnis einige Beispiele von Usability-Problemen der Website der Klubschule Migros aufgezeigt. Anschliessend werden die Usability-Probleme zum Vergleich der beiden Testarten in Kategorien eingeordnet. Zum Schluss des Kapitels werden die Usability-Probleme den Kriterien der Dialoggestaltung untergeordnet, um die Probleme der Website von einem höheren Standpunkt aus beurteilen zu können.

Exemplarische Beispiele einiger Usability-Probleme der Website

Folgend werden Beispiele von Problemen der Benutzbarkeit der Website von Klubschule Migros aufgezeigt. Die dargestellten Usability-Probleme tauchten bei nahezu allen Testpersonen auf, unabhängig von der Testvariante.

Problembispiel 1: System erfordert spezifische Kenntnisse

Nach der Auswahl der Region bspw. Aargau, Bern und Solothurn (Abbildung 14) listete das System automatisch mögliche Standorte in der gewählten Region auf (Abbildung 15).

Angebot verfeinern

Kursart	Kursort	Tage	Zeitpunkt
Bereich wählen ▼	Region wählen ▲	<input checked="" type="checkbox"/> Alle Tage <input checked="" type="checkbox"/> Montag	<input checked="" type="checkbox"/> Alle Zeiten <input checked="" type="checkbox"/> Vormittag <input type="checkbox"/> Mittag <input type="checkbox"/> Nachmittag <input type="checkbox"/> Abend
	Region wählen		Ergebnisse einschränken
	Aargau, Bern und Solothurn [20091]		<input type="checkbox"/> Informationsveranstaltungen <input type="checkbox"/> Garantierte Durchführung
	Basel [1946]		<input type="checkbox"/> Kurse mit Abschlüssen
	Ostschweiz [9926]		
	Tessin [2095]		
	Valais [110761]		

Abbildung 14: Region wählen (Klubschule Migros, 2014c)

Angebot verfeinern

Kursart	Kursort	Tage	Zeitpunkt
Bereich wählen ▼	Aargau, Bern und Solothurn [20091] ▼	<input checked="" type="checkbox"/> Alle Tage <input checked="" type="checkbox"/> Montag <input checked="" type="checkbox"/> Dienstag	
	Standort wählen ▲		
	Aarau [2876]		
	Baden [2059]		
	Bern [6487]		
	Biel/Bienne [1638]		
	Olten [1393]		

Abbildung 15: Standort wählen (Klubschule Migros, 2014c)

Wenn jedoch die Testpersonen nicht wussten, in welcher Region ein bestimmter Standort lag, konnte dieser nicht gefunden und dadurch die Aufgabe nicht oder nur durch langwieriges Ausprobieren gelöst werden. Das System verlangt von den Nutzern geografische Kenntnisse über die Schweiz, die aber unter Umständen nicht alle Nutzer mitbringen.

Problembispiel 2: System arbeitet gegen die Intuition der Nutzer

Bei der Suche „Angebot verfeinern“ waren die Filterfunktionen für Tage und Zeitpunkte bereits alle rot markiert und somit aktiv (Abbildung 16, S.38). Die Testpersonen aber haben jeweils diejenigen Filter angeklickt, die sie auswählen wollten, bspw. für Kurse am Abend klickten die Testpersonen auf den Filter Abend. Dieser wurde aber durch das Anklicken deaktiviert. Viele Testpersonen hatten dies nicht bemerkt und verstanden nicht, dass sie den Filter anstatt aktiviert, deaktiviert hatten, da das

System dazu auch keine Rückmeldung gab. Dieses System lief somit gegen die Erwartung und Logik der Nutzer. Die meisten Testpersonen hatten es genau umgekehrt erwartet.

Abbildung 16: Filter für Tage und Zeitpunkt (Klubschule Migros, 2014c)

Problembispiel 3: Onlineformular für Kursanmeldung

Viele Testpersonen waren verwirrt, dass beim Formular „Rechnung an abweichende Adresse“ (Abbildung 17) nach der Postleitzahl das Land und dann der Ort eingegeben werden musste. Dies entsprach nicht den mentalen Schemata der Nutzer.

Abbildung 17: Unlogische Reihenfolge im Datenformular (Klubschule Migros, 2014a)

Die meisten Testpersonen gaben beim Feld für das Postfach die Postleitzahl ein (Abbildung 18, S.39). Die Nutzer sind sich gewohnt, dass nach der Strasse die Postleitzahl und der Ort eingegeben werden müssen. Dies ist so mental verankert. Bei diesem Formular wird der Erwartung und Gewohnheit der Nutzer nicht entsprochen. Zudem verstanden die Testpersonen den Sinn von Postfach ohne Nummer und den Begriff Zusatz nicht.

Vorname*

Nachname*

Strasse und Nr.*

Postfach

Zusatz

PLZ / Ort*

Land* Schweiz ▼

Für die meisten Nutzer kommt hier die PLZ rein (falsche Reihenfolge)

Dies war den Testpersonen nicht verständlich

Abbildung 18: Onliniformular (Klubschule Migros, 2014a)

Diese ausgesuchten Beispiele von Usability-Problemen sollten einen Einblick in die Problematik der Benutzbarkeit der Website geben. Sie standen exemplarisch für einige schwerwiegende Usability-Probleme, welche nahezu alle Testpersonen sowohl im klassischen Usability-Test als auch im Usability-Walkthrough erkannten. Alle Usability-Probleme sind im Anhang E (S.84) ersichtlich.

Zuordnung der Usability-Probleme zu Kategorien

Wie die Analyse zeigte, wurden in beiden Testvarianten gleich viele Problemarten ermittelt. In beiden Testvarianten konnten 51 verschiedene Arten von Usability-Problemen unterschieden werden. Die Analyse der Usability-Probleme ergab, dass dieselben 41 Problemarten jeweils in beiden Testvarianten gefunden wurden. Im klassischen Usability-Test wurden zusätzlich zehn Arten von Usability-Problemen generiert. Im Usability-Walkthrough wurden zehn andere Arten von Problemen gefunden. Um detailliertere Aussagen über die Beschaffenheit der Usability-Probleme der Website in beiden Testvarianten machen zu können, wurden die Usability-Probleme in Kategorien eingeteilt. Damit konnte eine Aussage über die Art der Probleme, welche in beiden Testvarianten generiert wurden, gemacht werden. Wie in der Tabelle 15 (S.40) dargestellt ist, betrafen die meisten Usability-Probleme das Onlineformular und den Anmeldeprozess, die Suchfunktion nach den Kursen, die Missachtung der Erwartungen und Gewohnheiten der Nutzer sowie schlecht erkenntliche Hinweise und Funktionen. Weitere Probleme lagen bspw. in den Bereichen Scrollen, Layout, Navigation und Informationsarchitektur, Wording, Systemkomplexität und -transparenz. Diese Problemarten fielen allerdings nicht schwer ins Gewicht. Die Abbildung 19 (S.40) verdeutlicht die Anzahl der Problemarten hinsichtlich ihrer Kategorie. Die Auswertung zeigte, dass der klassische Usability-Test und der Usability-Walkthrough sehr ähnliche Usability-Probleme generierten. Geringe Unterschiede (Differenz: 1) gab es in den Bereichen nicht ersichtliche Informationen, Wording sowie im Onlineformular und Anmeldeprozess. Der grösste Unterschied (Differenz: 3) gab es im Bereich Suchfunktion. Zusammenfassend liessen sich nur sehr geringe Unterschiede hinsichtlich der Kategorien der Problemarten feststellen. Um die Problemarten von einer höheren Stufe aus beurteilen zu können, werden im Folgenden die Usability-Probleme den Kriterien der Dialoggestaltung zugeordnet.

Tabelle 15: Kategorien der Anzahl gefundener Usability-Probleme im klassischen Usability-Test und im Usability-Walkthrough

Kategorie	Anzahl Problemarten des Klassischen Usability-Tests	Anzahl Problemarten des Usability-Walkthroughs
Scrollen	3	3
Layout	2	2
Schlecht erkennbarer Hinweis oder Funktion	5	5
Informationen sind nicht ersichtlich	1	0
Navigation und Informationsarchitektur	3	3
Wording	2	1
Systemtransparenz	1	1
Fehlertoleranz	2	2
Erfordert spezifische Kenntnisse	1	1
Erwartung der Nutzer	6	6
Systemkonsistenz	2	2
Systemkomplexität	1	1
Suchfunktion	7	10
Onlineformular und Anmeldeprozess	15	14
Total	51	51

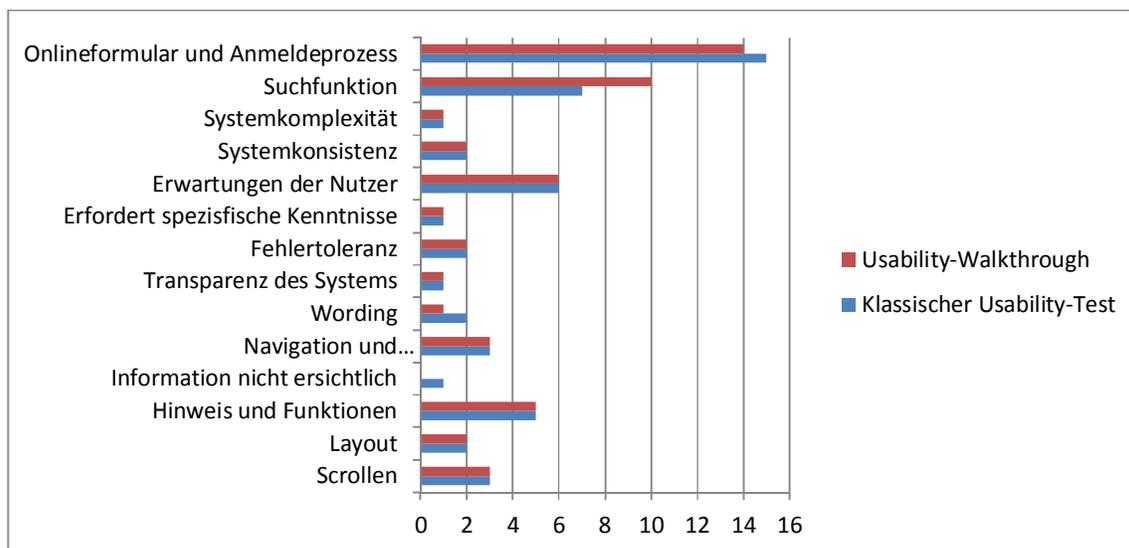


Abbildung 19: Kategorie der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough

Zuordnung der Usability-Probleme zu den Kriterien der Dialoggestaltung

Wie in Tabelle 16 (S.41) ersichtlich, lagen die meisten Usability-Probleme in den Bereichen Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit und Erwartungskonformität. Diese Probleme bereiteten den Testpersonen in beiden Testvarianten Mühe. Daher unterschieden sich die Ergebnisse

der beiden Tests nur sehr gering. Unterschiede liessen sich für die Kriterien Aufgabenangemessenheit und Individualisierbarkeit erkennen, wobei die Unterschiede sehr gering waren. Bei allen anderen Kriterien der Dialoggestaltung konnten jeweils zwischen gleich vielen Arten von Usability-Problem unterschieden werden. Die Abbildung 20 veranschaulicht die Ergebnisse.

Tabelle 16: Kriterien der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough

Kriterien der Dialoggestaltung	Anzahl Problemarten des Klassischen Usability-Tests	Anzahl Problemarten des Usability-Walkthroughs
Aufgabenangemessenheit	13	12
Selbstbeschreibungsfähigkeit	14	14
Steuerbarkeit	4	4
Erwartungskonformität	15	15
Fehlertoleranz	2	2
Individualisierbarkeit	2	3
Lernförderlichkeit	1	1
Total	51	51

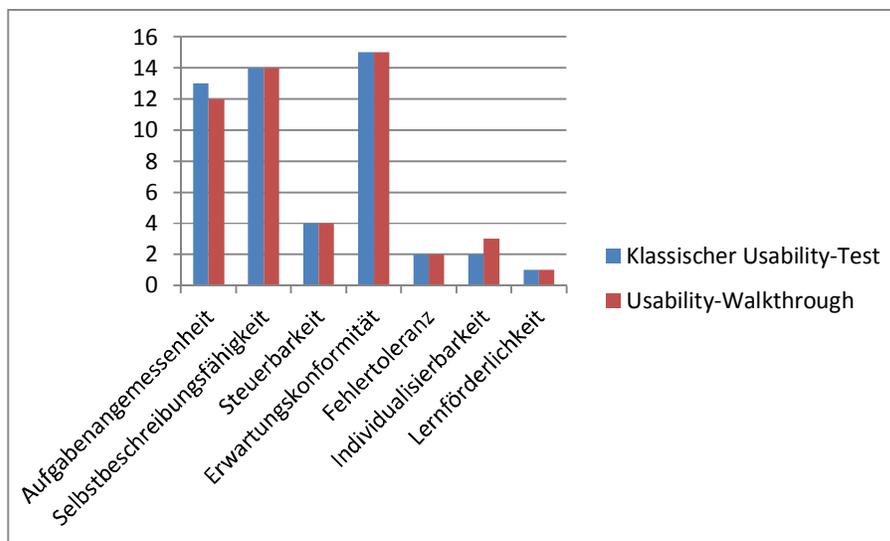


Abbildung 20: Kriterien der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough

Zwischenfazit

Zusammenfassend liessen sich zur Effektivität des klassischen Usability-Tests und des Usability-Walkthroughs nur sehr geringe Unterschiede feststellen. Die quantitative Datenauswertung stellte keinen signifikanten Zusammenhang hinsichtlich der Anzahl Usability-Probleme und Testvariante fest. Die qualitative Datenauswertung kam zum gleichen Schluss. In beiden Varianten des Usability-Tests wurden 51 verschiedene Arten von Usability-Problemen gefunden. Bei der Analyse nach der Zuordnung der Usability-Probleme zu Kategorien und Kriterien der Dialoggestaltung konnten ebenfalls nur sehr geringe Unterschiede festgestellt werden. Vor allem die offensichtlichen und schwerwiegenden Usability-Probleme wurden von den Testpersonen in beide Versuchsbedingungen eruiert.

5.3 Zeitliche Effizienz der zwei Varianten des Usability-Tests

In diesem Kapitel wird die Unterfragestellung drei mittels statistischer Auswertung beantwortet.

5.3.1 Quantitative Ergebnisse zur Anzahl gelöster Aufgaben

Einfache Regressionsanalyse

Für die Analyse der zeitlichen Effizienz der Tests, operationalisiert durch die Anzahl gelöster Testaufgaben, wurde eine lineare Regression durchgeführt. Als Grundlage dieser Berechnung diente die Formel $Anzahl\ gelöster\ Testaufgaben = \beta_0 + \beta_1 \cdot test$. Wie sich aus der Wurzel des R^2 berechnen liess, korrelieren die Variablen Anzahl gelöster Aufgaben und Test mit $r=0.25$. Wie in Tabelle 17 ersichtlich, ist der Test-Koeffizient mit 0.725 positiv, d.h. im Usability-Walkthrough lösten die Testpersonen im Durchschnitt mehr Testaufgaben als beim klassischen Usability-Test. Der P-Wert (0.12) verfehlt allerdings das 90%ige Konfidenzniveau knapp. Das korrigierte R^2 von 3.78 % ist sehr tief. Die Testvariante allein erklärt folglich nur wenig über die Anzahl gelöster Testaufgaben und somit über die zeitliche Effizienz der Tests.

Tabelle 17: Effekt der Testvariante auf die Anzahl gelöster Testaufgaben (einfache Regression)

Anz. gel. Aufg.	Koeff.	Stand.f.	t	P> t	F(1,38)	2.53
Test	0.725	0.46	1.59	0.120	Prob > F	0.1200
Konst.	4.600	0.32	14.27	0.000	R ²	0.0624
					Korr. R ²	0.0378

Legende: Anz. gel. Aufg.= Anzahl gelöster Aufgaben

Multiple Regressionsanalyse

Wie die Tabelle 18 (S.43) zeigt, ist der Test-Koeffizient mit 0.784 positiv. Der P-Wert von 0.062 ist auf einem 90%igen Konfidenzlevel signifikant. Somit kann mit statistischem Beweis gesagt werden, dass im Usability Walkthrough mehr Testaufgaben gelöst werden konnten als im klassischen Usability-Test. Dies könnte darauf zurückzuführen sein, dass beim Usability-Walkthrough Hilfe früher angeboten wurde als im klassischen Usability-Test. Die Hypothese, dass der Testort eine Auswirkung auf die Resultate hat, konnte im Zusammenhang mit der Anzahl gelöster Testaufgaben mit 90%iger Konfidenz verworfen werden (P-Wert=0.282). Die Variable Bildung hat mit einem P-Wert von 0.025 einen signifikant positiven Einfluss auf die Anzahl gelöster Testaufgaben, d.h., dass Testpersonen mit höherem Bildungsabschluss signifikant mehr Testaufgaben lösen konnten, als solche mit tieferem Abschluss. Dies könnte darauf hinweisen, dass diejenigen Testpersonen, die über einen höheren Bildungsabschluss verfügen, viel Zeit im Internet und im Umgang mit dem Computer verbringen und deshalb auch die Testaufgaben schneller und einfacher lösen konnten als Testpersonen mit einem tieferen Bildungsabschluss. Die Häufigkeit der Internetnutzung hat mit einem Koeffizienten von -0.382 im Durchschnitt einen negativen Effekt auf die Anzahl gelöster Testaufgaben. Dies bedeutet, dass Testpersonen, die wenig Zeit im Internet verbringen, durchschnittlich weniger Testaufgaben lösen konnten, was nicht weiter überraschend ist. Mit einem P-Wert von 0.183 ist dieser Effekt allerdings auf den gängigen Konfidenzlevels nicht signifikant. Das Alter der Testpersonen hat mit einem P-

Wert von 0.012 einen signifikant negativen Effekt auf die Anzahl gelöster Testaufgaben. Dies bedeutet, dass jüngere Testpersonen im Durchschnitt mehr Testaufgaben lösen konnten als ältere Testpersonen. Aufgrund der Tatsache, dass die jüngere Generation mit dem Internet eher vertraut ist und somit über bessere Internetkenntnisse verfügen dürfte, ist dieses Resultat nicht überraschend. Das Geschlecht der Testpersonen hat mit einem P-Wert von 0.242 keinen signifikanten Effekt auf die Anzahl gelöster Testaufgaben. Der F-Wert von 4.25 bestätigt die Validität dieses Modells. Auch das korrigierte R^2 ist mit 33.36% erfreulich hoch. Die Variable Kenntnis der Website wurde aus der Analyse ausgeschlossen, da der Verlust des Freiheitsgrads schwerer wiegt als der Informationsgewinn. Das korrigierte R^2 würde mit dem Einbezug der Variable von 33.36% auf 33.19% sinken.

Tabelle 18: Effekt der Testvariante auf die Anzahl gelöster Testaufgaben (multiple Regression mit den Kovariablen Ort, Bildung, Häufigkeit der Internetnutzung, Alter und Geschlecht)

Anz. gel. Aufg.	Koeff.	Stand.f.	t	P> t	F(6,33)	4.25
Test	0.784	0.41	1.93	0.062	Prob > F	0.0028
Ort	-0.620	0.57	-1.09	0.282	R^2	0.4362
Bildung	0.316	0.13	2.36	0.025	Korr. R^2	0.3336
Häufigkeit der Internetnutzung	-0.382	0.28	-1.36	0.183		
Alter	-0.047	0.02	-2.66	0.012		
Geschlecht: (1=männlich)	-0.489	0.41	-1.19	0.242		
Konst.	7.317	1.26	5.80	0.000		

Zwischenfazit

Zusammenfassend liess sich feststellen, dass beim Usability Walkthrough signifikant mehr Testaufgaben gelöst werden konnten als beim klassischen Usability-Test. Die Variable Bildung hatte ebenfalls einen signifikant positiven Effekt auf die Anzahl gelöster Testaufgaben. Einen signifikant negativen Effekt hatte die Variable Alter. Keinen Einfluss hatte die Variable Geschlecht.

6 Diskussion der Ergebnisse

In diesem Kapitel werden die Ergebnisse der Untersuchung zusammengefasst und mit den eingangs gestellten Fragestellungen in Zusammenhang gebracht. Anschliessend werden die Ergebnisse diskutiert, Stärken und Schwächen des Versuchs aufgezeigt und einen Ausblick auf offen gebliebene Fragen und Ideen für zukünftige Forschungsarbeiten gegeben.

6.1 Zusammenfassung der Ergebnisse

Testerleben der zwei Varianten des Usability-Tests

Die Unterfragestellung 1 „Wie werden die zwei Varianten des Usability-Tests von den Testpersonen erlebt?“ wurde mittels qualitativer und quantitativer Auswertung beantwortet. Die statistische Auswertung zeigte, dass der Usability-Walkthrough von den Testpersonen als signifikant positiver erlebt wurde als der klassische Usability-Test. Dies könnte darauf hinweisen, dass Testpersonen den persönlichen Kontakt zur Testleitung schätzen und sich dadurch wohl fühlen. Die Situation im klassischen Usability-Test, bei dem die Kommunikation über eine Gegensprechanlage verlief, könnte von den Testpersonen als künstlich wahrgenommen worden sein. Nach Köppel (2007) läuft der Vertrauensaufbau über persönliche Interaktionen. Bei fehlendem Face-to-Face-Kontakt, wie es während der Testphase im klassischen Usability-Test der Fall ist, ist ein Vertrauensaufbau zwischen Testperson und Testleitung gemäss Köppel (2007) schwieriger, was die Resultate erklären könnte.

Die qualitative Auswertung zeigte, dass beide Testvarianten von den Testpersonen als angenehm erlebt wurden. Unterschiede konnten festgestellt werden in der Art, wie die Testpersonen den Test empfanden. Während die Situation in den klassischen Usability-Tests eher als Test erlebt wurde, empfanden die Testpersonen den Usability-Walkthrough eher als natürliches Gespräch. Wie bereits erwähnt, könnte dies beim klassischen Usability-Test mit dem fehlenden persönlichen Kontakt zusammenhängen. Allerdings kann der persönliche Kontakt zwischen Testperson und Testleitung auch Gefahren bergen, wie bspw. der unerwünschte Testleitereffekt (vgl. Kapitel 3.8.1). Ausserdem empfanden einige Testpersonen die Anwesenheit der Testleitung beim Usability-Walkthrough beim Ausfüllen des Nachfragebogens als ein wenig unangenehm, da sie sich unter Beobachtung fühlten. Die Testleitung sollte sich der Gefahren und Chancen des persönlichen Kontaktes zu den Testpersonen bewusst sein und aufgrund des Testzieles entscheiden, welche Art von Test geeignet ist. Die Analyse zeigte ausserdem, dass es beim klassischen Usability-Test wichtig ist, die Testpersonen zuvor transparent über die Testsituation aufzuklären und ihnen die Situation im Labor zu erklären. Aufgrund der aufgedeckten Unterschiede muss die Nullhypothese, dass keine signifikanten Unterschiede zwischen den Tests bestehen, folglich zugunsten der Alternativhypothese verworfen werden.

Effektivität der zwei Varianten des Usability-Tests

Die Unterfragestellung 2 „Wie unterscheiden sich die zwei Varianten des Usability-Tests hinsichtlich ihrer Effektivität?“ wurde mittels qualitativer und quantitativer Auswertung beantwortet. Es konnte kein Zusammenhang zwischen der Testvariante und der Anzahl gefundener Usability-Probleme festgestellt werden. Das heisst, dass es für die Anzahl der Usability-Probleme keine Rolle spielte, ob der klassische Usability-Test oder der Usability-Walkthrough eingesetzt wird. Signifikante Einflüsse hat-

ten allerdings Faktoren, welche die Testpersonen betreffen, bspw. über welchen Bildungsabschluss die Testpersonen verfügen, wie viel Zeit sie im Internet verbringen und ob die Website der Klubschule Migros bereits bekannt war. Wie bereits im Kapitel 3.4.1 erwähnt, ist es von grosser Bedeutung, die richtigen Testpersonen zu rekrutieren. Daraus kann gefolgert werden, dass mehr die Testpersonen und nicht unbedingt die Testvariante eine wichtige Rolle für die Effektivität des Tests spielen.

Die qualitative Auswertung der Art der Usability-Probleme zeigte nur sehr geringe Unterschiede. In beiden Testvarianten wurden gleich viel verschiedene Arten von Usability-Problemen gefunden. Vor allem die schwerwiegenden Usability-Probleme wurden von den Testpersonen in beiden Testvarianten gefunden. Demnach konnten keine wesentlichen Unterschiede hinsichtlich der Effektivität zwischen dem klassischen Usability-Test und dem Usability-Walkthrough festgestellt werden. Die Nullhypothese, dass keine signifikanten Unterschiede zwischen den Testvarianten bestehen, wird folglich angenommen.

Zeitliche Effizienz der zwei Varianten des Usability-Tests

Die Unterfragestellung 3 „Wie unterscheiden sich die zwei Varianten des Usability-Tests hinsichtlich der zeitlichen Effizienz?“ wurde mittels quantitativer Analyse beantwortet. Das Ergebnis zeigte, dass beim Usability Walkthrough signifikant mehr Testaufgaben gelöst wurden als beim klassischen Usability-Test. Demzufolge wird die Nullhypothese zugunsten der Alternativhypothese verworfen. Die Gründe dafür könnten in der unterschiedlichen Art der geleisteten Hilfestellung liegen. Im klassischen Usability-Test hielt sich die Testleitung im Hintergrund und griff so wenig wie möglich ins Geschehen ein. Hilfestellung wurde so wenig wie möglich geleistet. Durch den persönlichen Kontakt im Usability-Walkthrough sprachen die Testpersonen automatisch mehr mit der Testleitung, als wenn diese sich nicht im gleichen Raum wie die Testpersonen aufgehalten hätte. Dies könnte die Ergebnisse erklären. Wie bereits bei der Untersuchung der Effektivität der Testvarianten festgestellt, hatten auch hier die Testpersonen einen Einfluss auf die Resultate. Beispielsweise hatte der Bildungsabschluss einen signifikant positiven Effekt auf die Anzahl gelöster Testaufgaben. Hingegen hatte das Alter einen signifikant negativen Effekt darauf, wie viele Testaufgaben gelöst werden konnten. Folglich erhärtete sich die Annahme, dass für Usability-Tests im Allgemeinen die Testpersonen von grösster Bedeutung für die Resultate des Tests sind und demnach mit grösster Sorgfalt im Hinblick auf die potentielle Benutzergruppe des Testsystems ausgewählt werden müssen.

6.2 Ergebnisdiskussion

Für eine quantitative Datenauswertung ist die Stichprobe von 40 Testpersonen eher klein. Um relevante Aussagen machen zu können und Schlüsse aus den Daten zu ziehen, dürfte die Stichprobe nicht mehr kleiner sein. Für die qualitative Datenauswertung ist die Stichprobe jedoch gross. Der Versuchsaufbau und die Datenauswertung sind statistisch korrekt. Das Forschungsdesign hat sich bewährt und hielt den Anforderungen stand. Obwohl diese Forschungsarbeit eine verhältnismässig kleine Stichprobengrösse aufweist, können die Ergebnisse dennoch als wegweisend für spätere Studien betrachtet werden. Mittels einer grösseren Stichprobe könnten zukünftig die Resultate auf einem höheren Signifikanzlevel repliziert und überprüft werden.

Die Durchführungsobjektivität kann trotz der standardisierten Testdurchführung mit Hilfe der Drehbücher nicht als vollumfänglich gegeben beurteilt werden. Obwohl die einzelnen Schritte und Phasen der Testdurchführung festgehalten wurden, bleibt dennoch der individuelle Einfluss durch die Testleitung bestehen. Hilfestellung anzubieten stellt ebenfalls eine Herausforderung dar. Es spielen individuelle Persönlichkeitsmerkmale und Ansichten der Testleitung eine Rolle, wann genau und in welchem Umfang Hilfestellung den Testpersonen angeboten werden. Um die Standardisierung zu erhöhen, war das Drehbuch sicherlich hilfreich. Allerdings können damit nicht alle individuellen Einflüsse durch die Testleitung ausgeschlossen werden. Den Versuch exakt zu replizieren, würde folglich eine Herausforderung darstellen.

Hingegen kann die Auswertungsobjektivität, vor allem bei der quantitativen Datenauswertung, als gegeben betrachtet werden. Die Resultate der Auswertung können jederzeit repliziert werden. Bei der qualitativen Datenauswertung kann die Auswertungsobjektivität als weniger gegeben betrachtet werden als bei der quantitativen. Dennoch kann anhand des Datenmaterials und der Auswertungsleitlinien die Auswertung repliziert werden. Es kann angenommen werden, dass möglicherweise minimale Unterschiede entstehen, jedoch der Kern der Aussagen auch bei einer erneuten Replizierung des Versuchs erhalten bleiben würde. Die Interpretationsobjektivität kann wiederum bei der quantitativen Datenauswertung als gegeben beurteilt werden. Bei der qualitativen Auswertung können individuelle Einflüsse der Person, welche den Versuch repliziert, nicht vollumfänglich ausgeschlossen werden. Allerdings werden die Einflüsse minimiert durch ein standardisiertes Forschungsdesign und eine transparente Vorgehensweise.

Die Reliabilität der Daten, d.h. die Zuverlässigkeit der Messergebnisse, kann als beinahe gegeben angesehen werden. Die abhängige Variable Testerleben wird mittels eines kontrollierten Fragebogens erhoben. Diese Messergebnisse lassen keinen Spielraum offen. Die abhängige Variable Anzahl gelöster Aufgaben lässt durch die Auswertungsleitlinien wenig Spielraum offen. Die Messung der abhängigen Variable Anzahl Usability-Probleme ist die kritischste Variable der gemessenen Daten. Mittels der Auswertungsrichtlinien wurde versucht, die Messung einheitlich vorzunehmen, dennoch könnten bei einer erneuten Messung dieser Variable kleine Abweichungen auftreten. Insgesamt kann die Reliabilität der Daten aber als ausreichend bezeichnet werden.

Eine hohe Validität des Usability-Tests bedingt nach Nielsen (1993) fundiertes Wissen über die Methode. Als Vorbereitung auf die Testdurchführung wurde fundierte Literatur zu Usability-Tests hinzugezogen (z.B. Nielsen, 1993; Sarodnick & Brau, 2011; Barnum, 2011; Dumas & Loring, 2008 u.w.). Um eine negative Beeinflussung der Validität durch Störvariablen zu minimieren, wurden mögliche Störvariablen ausgeschlossen oder kontrolliert. Wie Richter und Flückiger (2010) anmerken, kann die Generalisierbarkeit der Ergebnisse durch eine grosse Stichprobe verbessert werden. 40 Testpersonen sind für einen Usability-Test eine beachtliche Stichprobengrösse. Daher können den Ergebnissen eher vertraut werden als bei Usability-Tests mit kleineren Stichproben. Durch den Einsatz weiterer Evaluationsmethoden könnte die Validität allerdings weiter erhöht werden.

6.3 Stärken und Schwächen der Untersuchung

Im Folgenden wird zunächst auf die Stärken eingegangen und anschliessend auf die Herausforderungen der Versuchsanordnung mit Verbesserungsvorschlägen für zukünftige Untersuchungen.

Stärken

Die Versuchsanordnung wurde im Voraus geplant und mittels eines Pretests auf ihre Durchführbarkeit getestet. Es wurden anschliessend Verbesserungen vorgenommen. Das Zeitmanagement konnte eingehalten werden. Die technische Ausrüstung funktionierte reibungslos, da jeweils genügend Zeit im Voraus eingeplant wurde, um die Ausrüstung zu testen. Das Versuchsdesign hat funktioniert und konnte wie geplant durchgeführt werden. Die Daten wurden sorgfältig ausgewertet und beschrieben. Die Versuchspersonen wurden mit Respekt behandelt, transparent über den Test informiert und machten positive Erfahrungen mit Usability-Tests. Alle Testpersonen würden zukünftig wieder einmal an einer Usability-Evaluation teilnehmen.

Schwächen

Zu den Schwächen der Untersuchung gehört der Umstand, dass an zwei Orten mit zwei unterschiedlichen Labors getestet wurde. Eine höhere Standardisierung würde erreicht werden, wenn alle Tests am gleichen Ort und mit der gleichen Laborart stattgefunden hätten. So hätten alle Testpersonen die gleichen Konditionen gehabt. Aus organisatorischen Gründen war es jedoch nicht möglich, eine so grosse Stichprobe in der Zeit, die für die Bachelor Thesis zur Verfügung stand, an einem Ort zu rekrutieren.

Eine weitere Herausforderung der Versuchsanordnung bestand darin, dass der Einfluss der Testleitung bei Usability-Walkthrough höher war als beim klassischen Usability-Test. Vor allem beim Usability-Walkthrough kann die Testleiterunabhängigkeit als problematisch angesehen werden. Trotz der Einhaltung des Drehbuchs kann der Einfluss durch die Testleitung und die Auswirkung auf die Resultate nicht vollumfänglich ausgeschlossen werden.

Weiter kritisieren lassen sich die Operationalisierungen der abhängigen Variablen. Die Variable Testerleben wurde mittels des Gesamtwerts des Fragebogens zum Testerleben gemessen. Obwohl dieser Fragebogen ein standardisiertes Instrument ist, kann der Sinn einiger Items in Frage gestellt werden. Das siebte Item („Zeitweise habe ich nicht verstanden, was am Computer ablief“) musste teilweise von der Testleitung noch präzisiert werden. Am meisten irritierte das neunte Item („Ich bin mit den Leistungen zufrieden, die ich zeigen konnte“). Da den Testpersonen zuvor versichert worden war, dass es beim Usability-Test nicht um die Prüfung ihrer Fähigkeiten geht, sondern um die Aufdeckung von Schachstellen im System, verstanden einige Testpersonen den Sinn dieses Items nicht. Aus praktischer Sicht hätte dieses Item eliminiert oder zumindest umformuliert werden sollen. Allerdings verliert ein standardisierter Fragebogen an Validität, wenn er verändert wird. Die neuen Items müssten zunächst wieder in einem empirischen Verfahren überprüft werden. Aus diesem Grund wurde auf Veränderungen der Items verzichtet. Die Operationalisierung der Effektivität der Tests durch die Anzahl Usability-Probleme kann als herausfordernd betrachtet werden. Selbst Usability-Experten sind sich bei der Identifikation von Usability-Problemen nicht immer einig. Die Ausarbeitung von Auswer-

tungsleitlinien war zwar hilfreich, konnte aber die individuellen Einflüsse durch die Person, welche die Daten auswertete, nicht ganz eliminieren.

6.4 Schlussfolgerungen und Ausblick

Dieses Kapitel bildet den Abschluss der vorliegenden Arbeit. Es werden Schlussfolgerungen gezogen sowie Ideen für zukünftige Forschung vorgestellt.

6.4.1 Praxistransfer: Was bedeuten die Ergebnisse für die Praxis?

Diese Arbeit setzte sich intensiv mit der Thematik des Usability-Tests auseinander. Im Fokus stand die Hauptfrage, wie sich die zwei Testvarianten, der klassische Usability-Test und der Usability-Walkthrough, unterscheiden. Aus den Resultaten dieser Bachelor Thesis lassen sich für die soultank AG praxisrelevante Verbesserungsmöglichkeiten für den Einsatz des klassischen Usability-Tests und des Usability-Walkthroughs ableiten. Es wurde ein Merkblatt mit Tipps für die Testleitung (Anhang F, S.100) aus den essentiellen Erkenntnissen der Untersuchungsergebnisse zusammengestellt. Darin sind Informationen für die Testleitungen enthalten, wann welcher Test eingesetzt werden soll, was bei der Testvorbereitung zu beachten ist, was bei der Durchführung des klassischen Usability-Tests und des Usability-Walkthroughs zu berücksichtigen ist und welche Faktoren das Testen im stationären Labor oder mit dem mobilen Labor beeinflussen. Zur Verbesserung der Moderation stehen im Merkblatt unterstützend zehn Verhaltensregeln für die Testleitung sowie vier Stufen, wie Hilfestellung angeboten werden kann. Mithilfe der gewonnenen Erkenntnisse können möglicherweise zukünftige Usability-Tests verbessert werden.

6.4.2 Ideen für zukünftige Forschungen

Dieses Kapitel beschäftigt sich mit offen gebliebenen Fragen und Ideen für zukünftige Forschungen im Bereich Usability-Test.

Die Problematik des lauten Denkens

Bei der Auswertung der Daten wurde immer wieder auf die Problematik des lauten Denkens gestossen. Vor allem unerfahrene Testpersonen hatten teilweise Mühe und Hemmungen, ihre Gedanken laut auszusprechen. Viele Testpersonen überforderte es, sich gleichzeitig auf die Aufgaben, das Geschehen am Computer und das laute Verbalisieren zu konzentrieren. Diese Erfahrung wurde unabhängig von der Testvariante gemacht. Die Forschung könnte folglich zukünftig Alternativen zum lauten Denken untersuchen. Dazu wäre es interessant zu untersuchen, welche Effekte das laute Denken auf die Ergebnisse hat. Würde es andere Resultate generieren, wenn die Testpersonen nicht laut denken würden?

Der Testleitereffekt in Usability-Tests

Der Testleitereffekt ist in der Sozialwissenschaft ein häufiges und gut erforschtes Phänomen. Allerdings gibt es kaum Forschungen zum Testleitereffekt im Bereich Usability. So wäre es bspw. interessant zu untersuchen, ob es auf das Testerleben der Testpersonen oder auf die Ergebnisse einen Einfluss hat, ob der Test von einem Mann oder einer Frau durchgeführt wird? Welchen Einfluss hat das äussere Erscheinungsbild der Testleitung? Wie wirkt bspw. ein Laborkittel auf die Testpersonen und die Ergebnisse? Welchen Einfluss hat das Alter der Testleitung? Welche Einflüsse haben kulturelle

Aspekte? Wie muss sich die Testleitung verhalten, wenn der Usability-Test mit Testpersonen aus anderen Kulturen durchgeführt wird? Diese Fragestellungen werden in einer immer globaleren Arbeitswelt zukünftig von Bedeutung sein.

Umgang mit Testangst

Obwohl den Testpersonen beim Usability-Test mitgeteilt wurde, dass nicht ihre Fähigkeiten sondern das System getestet wird, erlebten viele Probanden die Situation im Usability-Test dennoch als Test-situation, bei der sie sich nicht blamieren wollten. Die Ergebnisse dieser Bachelor Thesis zeigten, dass vor allem die Situation im klassischen Usability-Test von den Testpersonen als Test erlebt wurde. Testangst ist hinderlich bei Usability-Tests, da sie die Probanden blockiert und das Lösen der Aufgabe sowie das Einlassen auf das System erschwert. Die Frage, wie sich die Testangst bei Usability-Tests reduzieren lässt, ist folglich bedeutend für zukünftige Untersuchungen und besitzt hohe Praxisrelevanz.

Die verschiedenen Varianten des Usability-Tests aus Sicht der Testleitung

Bei dieser Bachelorarbeit wurde der Fokus auf das Erleben der Testpersonen gelegt. Es wäre aber durchaus interessant zu untersuchen, wie die Testleitung die beiden Varianten des Usability-Tests erlebt hat. Welche Form des Usability-Tests ist für die Testleistung einfacher und angenehmer durchzuführen? Welche Unterschiede bezüglich des Empfindens der Testleitung bestehen zwischen dem klassischen Usability-Test und dem Usability-Walkthrough? Wie soll sich die Testleitung in der Situation des klassischen Usability-Tests oder des Usability-Walkthroughs verhalten? Aus den Erkenntnissen dieser Untersuchung könnten praktische Tipps und Verhaltensregeln für Testleitende abgeleitet werden.

Bezug zur gesamten Usability-Methodik

Die vorliegende Arbeit beschränkte sich auf die Thematik des Usability-Tests. Bei weiteren eingesetzten Usability-Methoden führt ein Usability-Experte Evaluationen mit Testpersonen durch, bspw. Interviews, Fokus-Gruppen, Card Sorting, Fragebogen etc. Bei diesen Methoden arbeitet der Usability-Verantwortliche mit Testpersonen zusammen. Es wirken demnach ähnliche Mechanismen wie auch im Usability-Test, wie bspw. der Testleitereffekt. Deshalb wäre eine Untersuchung bestimmter Wirkmechanismen auch in weiteren Usability-Methoden sinnvoll.

Schlusswort

Die vorliegende Arbeit zeigt auf, welche Unterschiede zwischen zwei Varianten des Usability-Tests bestehen. Der Unterschied, ob sich die Testleitung im gleichen Raum oder im Beobachtungsraum aufhält, hat einen Einfluss darauf, wie die Testpersonen den Test erleben. Hinsichtlich der Effektivität unterscheiden sich der klassische Usability-Test und der Usability-Walkthrough kaum. Unterschiede bestehen hinsichtlich der zeitlichen Effizienz und des Erlebens der Testpersonen. Aus den Resultaten der Untersuchung wurden praktische Schlüsse gezogen und ein Merkblatt zur Optimierung von Usability-Tests ausgearbeitet. Usability wird auch in Zukunft einen wichtigen Stellenwert einnehmen, um Kunden mit verbesserten Produkten zu begeistern.

7 Literaturverzeichnis

- Bahner, J. E., & Timpe, K.-P. (2005). Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis: Festschrift für Klaus-Peter Timpe (1.Aufl.). Düsseldorf: Symposion Publ.
- Barnum, C. M. (2011). Usability testing essentials: Ready, set-- test! Amsterdam, Boston: Morgan Kaufmann Publishers.
- Brinks, M. (2005). Aggression gegen Computer: Eine wissenschaftliche Untersuchung eines alltäglichen Phänomens. Stuttgart: ibidem-Verl.
- Bühl, A. (2009). PASW: Einführung in die moderne Datenanalyse ; [ehemals SPSS] (12.Aufl.). München, Boston [u.a.]: Pearson Studium.
- Dumas, J. S., & Loring, B. A. (2008). Moderating usability tests: Principles and practices for interacting. Amsterdam, Boston: Morgan Kaufmann/Elsevier.
- E-Commerce Leitfaden. Die Suchfunktion entscheidet über Kauf oder Nichtkauf. Verfügbar unter <http://www.ecommerce-leitfaden.de/qdie-suchfunktion-entscheidet-ueber-kauf-oder-nichtkaufq.html> [02.06.2014]
- Hamborg, K.-C. & Gediga, G. (2006). Methoden und Modelle für die Gestaltung gebrauchstauglicher Software. In N. Birbaumer, D. Frey, J. Kuhl, W. Schneider & R. Schwarzer (Hrsg.), Enzyklopädie der Psychologie. Band 2 Ingenieurspsychologie (S.495-529).Göttingen, Bern, Toronto, Seattle: Hogrefe, Verlag für Psychologie.
- Heinsen, S. (2003). Usability praktisch umsetzen: Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte. München, Wien: Hanser.
- Hembree, R. & College, A. (1988). Correlates, Causes, Effects, and Treatment of Test Anxiety. Review of Educational Research, 58 (1) 47-77.
- Himme, A.(2007). Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit. In S. Albers (Hrsg.), Methodik der empirischen Forschung (2. Aufl., S.375-390). Wiesbaden: Gabler.
- Hogrefe Testsystem 4. (2014). Fragen zum Testerleben. Verfügbar unter <http://www.unifr.ch/ztd/HTS/inftest/WEB-Informationssystem/de/4de001/b09c6f04b48749159fd4d6f60edf4da9/hb.htm> [02.06.2014]
- Jäger, R. S., & Kubinger, K. D. (2003). Schlüsselbegriffe der psychologischen Diagnostik: [Handbuch] (1.Aufl.). Weinheim [u.a.]: Beltz PVU.
- Köppel, P. (2007). Kulturelle Diversität in virtuellen Teams. In D. Wagner (Hrsg.), Diversity-Management als Leitbild von Personalpolitik (1.Aufl., Vol. 1, pp. 273–292). Wiesbaden: Dt. Univ.-Verl.
- Margraf, J. (Hrsg.). (2009). Lehrbuch der Verhaltenstherapie (3.Aufl.). Berlin [u.a.]: Springer.
- Mayring, P. (2010). Qualitative Inhaltsanalyse: Grundlagen und Techniken (11.Aufl.). Weinheim: Beltz.
- Klubschule Migros. (2014a). Anmeldeformular. Verfügbar unter https://www.klubschule.ch/OrderProcess/Step1/Ballett-fuer-Erwachsene-Aufbau--E_786959_1020 [02.06.2014]
- Klubschule Migros. (2014b). Bildungspolitik. Verfügbar unter <http://www.klubschule.ch/Ueberuns/Bildungspolitik#tab=tab2> [02.06.2014]

- Klubschule Migros. (2014c). Startseite. Verfügbar unter <http://www.klubschule.ch/> [02.06.2014]
- Nielsen, J. (1993). Usability engineering. Cambridge, Mass: AP Professional.
- Oberhuber, N. (2004). Mensch gegen Maschine. Zeit Online, 39. Verfügbar unter <http://www.zeit.de/2004/39/Computerfrust> [02.06.2014]
- Raithel, J. (2006). Quantitative Forschung: Ein Praxiskurs (1.Aufl.). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Richter, M., & Flückiger, M. D. (2010). Usability Engineering kompakt: Benutzbare Software gezielt entwickeln (2.Aufl.). Heidelberg: Spektrum Akad. Verl.
- Rubin, J., & Chisnell, D. (2008). Handbook of usability testing: How to plan, design, and conduct effective tests (2.Aufl.). Indianapolis, IN: Wiley Pub.
- Sarodnick, F., & Brau, H. (2006). Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung (1.Aufl.). Bern: Huber.
- Sarodnick, F., & Brau, H. (2010). Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung (2.Aufl.). Bern: Verlag Hans Huber.
- Schmidt-Atzert, L., & Amelang, M. (2012). Psychologische Diagnostik: Mit 82 Tabellen (5.Aufl.). Berlin [u.a.]: Springer.
- Schnepper, M. (2004). Robert K. Mertons Theorie der self-fulfilling prophecy: Adaption eines soziologischen Klassikers (1.Aufl., vol. 395). Frankfurt am Main, New York: Peter Lang.
- soultank AG. (2014a). Card Sorting. Verfügbar unter <http://soultank.ch/content/was-wir-tun/methoden/card-sorting/> [02.06.2014]
- soultank AG. (2014b). Mobiles Usability-Labor. Verfügbar unter <https://soultankag1.backpackit.com/pages/2398548-mobiles-usability-labor?term=mobiles+Labor> [02.06.2014]
- Stark, L., Vossius, G., & Young, L. R. (1962). Predictive Control of Eye Tracking Movements. Human Factors in Electronics, IRE Transactions on, 3(2), 52–57. Verfügbar unter <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4503342> [02.06.2014]
- Technische Universität Chemnitz. (2014). Usability. Verfügbar unter <http://www.tu-chemnitz.de/hsw/psychologie/professuren/allpsy1/usability.php> [02.06.2014]
- Wandmacher, J. (1993). Software-Ergonomie (Vol. 2). Berlin, New York: W. de Gruyter.
- Weikel, J. (2014). What is Usability? Verfügbar unter <http://www.weikeldesigns.com/what-is-usability/> [02.06.2014]
- Wenninger, G. (2001). Lexikon der Psychologie. Bd. 1-5: In fünf Bänden. Heidelberg: Spektrum.

8 Verzeichnisse

8.1 Abbildungsverzeichnis

Abbildung 1: Unbrauchbares Produkt	1
Abbildung 2: Überblick über Usability-Methoden	7
Abbildung 3: Beispielhafter Aufbau eines stationären Labors	13
Abbildung 4: Aufbau des Settings eines mobilen Testlabors	14
Abbildung 5: Startseite der Klubschule Migros	18
Abbildung 6: Zufällige Zuteilung der Testpersonen zu den Versuchsbedingungen	23
Abbildung 7: Bildungsabschluss der Testpersonen	24
Abbildung 8: Häufigkeit der Internetnutzung der Testpersonen	25
Abbildung 9: Klassischer Usability-Test und Usability-Walkthrough im stationären Labor in Olten.....	26
Abbildung 10: Arbeitsplatz der Testleitung im Usability-Labor in Olten	26
Abbildung 11: Klassischer Usability-Test und Usability-Walkthrough mit dem mobilen Labor	27
Abbildung 12: Vierstufiges Hilfsmodell	28
Abbildung 13: Vorgehensweise der Datenauswertung der Art der Usability-Probleme des klassischen Usability-Tests und des Usability-Walkthroughs	30
Abbildung 14: Region wählen	37
Abbildung 15: Standort wählen	37
Abbildung 16: Filter für Tage und Zeitpunkt	38
Abbildung 17: Unlogische Reihenfolge im Datenformular	38
Abbildung 18: Onliniformular	39
Abbildung 19: Kategorie der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough	40
Abbildung 20: Kriterien der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough	41

8.2 Tabellenverzeichnis

Tabelle 1: Grundsätze der Dialoggestaltung	5
Tabelle 2: Übersicht über die wichtigsten Gütekriterien	8
Tabelle 3: Die zehn goldenen Moderationsregeln.....	10
Tabelle 4: Fazit Vergleich zwischen klassischen Usability-Test und Usability-Walkthrough.....	17
Tabelle 5: Statistische Hypothese.....	19
Tabelle 6: Versuchsanordnung	19
Tabelle 7: Abhängige Variablen (multivariat).....	20
Tabelle 8: Die 10 Items des Fragebogens zum Testerleben	20
Tabelle 9: Items zum Testerleben für qualitative Auswertung	21
Tabelle 10: Durchschnittswerte der Testpersonen	24
Tabelle 11: Effekt der Testvariante auf das Testerleben (einfache Regression)	32
Tabelle 12: Effekt der Testvariante auf das Testerleben (multiple Regression mit den Kovariablen Ort, Alter und Bildung).....	32
Tabelle 13: Effekt der Testvariante auf die Anzahl Usability-Probleme (einfache Regression)	35
Tabelle 14: Effekt der Testvariante auf die Anzahl Usability-Probleme (multiple Regression mit den Kovariablen Ort, Bildung, Häufigkeit der Internetnutzung, Kenntnis der Website und Geschlecht).....	36
Tabelle 15: Kategorien der Anzahl gefundener Usability-Probleme im klassischen Usability-Test und im Usability-Walkthrough	40
Tabelle 16: Kriterien der Anzahl gefundener Usability-Problemarten im klassischen Usability-Test und im Usability-Walkthrough	41
Tabelle 17: Effekt der Testvariante auf die Anzahl gelöster Testaufgaben (einfache Regression)	42
Tabelle 18: Effekt der Testvariante auf die Anzahl gelöster Testaufgaben (multiple Regression mit den Kovariablen Ort, Bildung, Häufigkeit der Internetnutzung, Alter und Geschlecht).....	43

8.3 Abkürzungsverzeichnis

AV	abhängige Variable
bzw.	beziehungsweise
Bsp.	Beispiel
bspw.	beispielsweise
d.h.	das heisst
etc.	etcetera
evtl.	eventuell
ggf.	gegeben falls
sog.	sogenannt/e
UV	unabhängige Variable
u.w.	und weitere
vgl.	vergleiche
z.B.	zum Beispiel

9 Anhang

Anhang A	56
Glossar	56
Anhang B	57
Testplan	57
Ethische Aspekte zum Umgang mit Testpersonen.....	58
Anhang C	59
Drehbuch für den klassischen Usability-Test.....	59
Drehbuch für den Usability-Walkthrough	62
Anhang D	65
Testleiterskript für klassischen Usability-Test	65
Testleiterskript für Usability-Walkthrough.....	71
Testunterlagen für Testpersonen	73
Anhang E.....	78
Auswertungsleitlinien.....	78
Qualitative Auswertung Testerleben beim klassischen Usability-Test	80
Qualitative Auswertung Testerleben beim Usability-Walkthrough.....	82
Qualitative Auswertung Art der Usability-Probleme	84
Anhang F.....	100
Merkplatt für Testleitung	100

Anhang A

Glossar

Begriff	Bedeutung / Definition	Quelle
Card Sorting	Mit Card Sorting ist die Tätigkeit des Gruppierens von Karten gemeint. (...) Ein Card Sorting ist eine sehr gute Methode, wenn man nicht genau weiss, wie der Endnutzer des Produktes denkt und man sich in diesen einfühlen möchte, bevor man die Informationsarchitektur einer Webseite oder Software ausarbeitet.	soultank AG (2014a)
Eyetracker	Ist ein Instrument zur Messung der Augenbewegungen. Es ermöglicht die Aufzeichnung des Blickweges eines Menschen auf einem Bildschirm und wie lange die Augen jeweils an einem Punkt verweilen.	Stark, Vossius & Young (1962)
Fokusgruppen	(...) sowohl offene Gruppendiskussionen, strukturierte Gruppeninterviews als auch Workshops, in denen konkrete Lösungen für Probleme erarbeitet werden (...).	Heinsen (2003, S.138)
Guidelines	(...) eher generelle Richtlinien für die Verwendung und das Verhalten von (grafischen) User-Interface-Elementen.	Richter und Flückiger (2010, S.54)
Lautes Denken	Ist eine Methode zur Erfassung handlungsbegleitender (bewusstseinsfähiger) Kognitionen und Emotionen, die während der Nutzung (periaktional) einer Software verbalisiert werden (...)	Hamborg und Gediga (2006, S. 510)
Mobiles Labor	(...) eine mobile Einrichtung zur Durchführung von Usability-Tests vor Ort anstatt in einem Labor (...).	Richter und Flückiger (2010, S.66)
Objektivität	Objektivität bedeutet, dass die Ergebnisse eines diagnostischen Verfahrens unabhängig davon zustande kommen, wer die Untersuchung, die Auswertung und die Interpretation durchführt.	Schmidt-Atzert und Amelang (2012, S.133)
Prototyp	(...) wird im Usability-Engineering eingesetzt, um Aspekte der Benutzerschnittstelle zu entwerfen, zu evaluieren und zu verbessern, noch bevor ein lauffähiges System vorhanden ist.	Richter und Flückiger (2010, S.40)
Reliabilität	Die Reliabilität betrifft die Zuverlässigkeit und Stabilität eines Messinstruments. Das Kriterium bezieht sich auf die Frage, wie gemessen wird, und fordert, dass die Messergebnisse bei wiederholter Messung reproduzierbar sein sollten.	Himme (2007, S.375)
Simulation	Unter Simulation wird (...) das bewusste und absichtsvolle Vortäuschen von tatsächlich nicht vorhandenen körperlichen Beschwerden und kognitiven Beeinträchtigungen verstanden.	Jäger und Kubinger (2003, S.381)
Software Ergonomie	Software-Ergonomie ist eine Teildisziplin der Ergonomie (...). Gegenstand der Software-Ergonomie ist die Anpassung der Arbeitsbedingungen bei der Mensch-Computer-Interaktion an die sensumotorischen und kognitiven Fähigkeiten und Prozesse des Menschen.	Wandmacher (1993, S.1-2)
Styleguides	Konkrete Vorgaben für die visuelle Gestaltung und das Layout einer bestimmten Benutzeroberfläche. Styleguides beschreiben Aussehen und Verhalten (...) von User-Interface-Elementen, abhängig von der eingesetzten Technologie.	Richter und Flückiger (2010, S.54)
Testangst	Angstgefühl, das im Zusammenhang mit einer Testung auftreten und den Testwert verfälschen kann.	Wenninger (2001, S.309)
Usability	Die „Benutzbarkeit“ eines Systems	Richter und Flückiger (2010, S.3)
Usability Engineering	Usability Engineering umfasst Mittel und Techniken, um die Entwicklung neuer Software oder Produkte die angestrebte Usability zu erreichen.	Richter und Flückiger (2010, S.5)
Usability-Labor	(...) es ist die Umgebung, in der Usability-Tests durchgeführt werden und die mit der entsprechenden Ausstattung ein optimales Setting schafft, um ein möglichst natürliches Nutzungsverhalten zu beobachten und Daten für die weitere Analyse aufzuzeichnen.	Heinsen (2003, S. 206)
Validität	Die Validität bezieht sich auf die Gültigkeit und materielle Genauigkeit eines Messinstruments. Im Rahmen der Validitätsprüfung ist zu fragen, ob mit einem Messinstrument das gemessen wird, was gemessen werden soll.	Himme (2007, S.375)

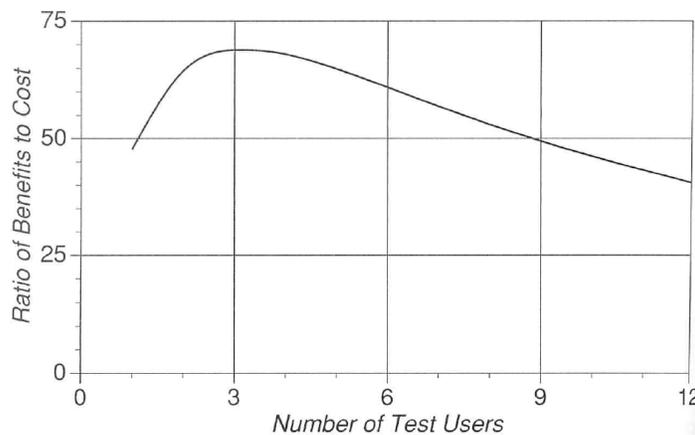
Anhang B

Testplan

Der Testplan wird in der Vorbereitungsphase geschrieben und enthält gemäss Nielsen (1993) folgende Aspekte:

- Testziel
- Zeit und Ort des Tests
- Länge jeder Testsequenz
- benötigter Computereinsatz, benötigte Software
- Stadium des zu prüfenden Systems, Bestimmung der Testleitung
- Auswahl und Anzahl benötigter Testpersonen
- Ausarbeitung der Aufgaben und festlegen, wann die Testpersonen die Aufgabe gelöst haben
- Auswahl der Hilfsmittel, die den Testpersonen zur Verfügung stehen werden
- Ausmass der Hilfestellung der Testleitung gegenüber den Testpersonen festlegen
- Sammlung und Auswertung der Daten bestimmen

Der Testplan soll ausserdem das Budget, das für den Usability-Test zur Verfügung steht, beinhalten (Nielsen, 1993). Um in Usability-Tests ein optimales Kosten-Nutzen-Verhältnis zu erlangen, entwickelte Nielsen (1993) die untenstehende Grafik. Auf der X-Achse wird die Anzahl an Testpersonen aufgelistet, die Y-Achse stellt das Verhältnis von Nutzen und Kosten dar. Wie in der Grafik ersichtlich ist, wird bereits nach einem Test ein Drittel der Usability Probleme aufgedeckt. Nach Nielsen (1993) kann nach der fünften Testperson der Usability-Test beendet werden, da 85% der Usability-Probleme aufgedeckt wurden.



Grafik: Das lohnende Verhältnis von Kosten und Nutzen für User-Tests (Nielsen, 1993, S.174, eigene Übersetzung)

Ethische Aspekte zum Umgang mit Testpersonen

1. Die Testpersonen sollen bei jedem Usability-Test mit grossem Respekt behandelt werden.
2. Die Testleitung soll sich bewusst sein, dass die Testpersonen sich unter Druck fühlen, eine „gute“ Leistung zu zeigen, obwohl sie wissen, dass das System und nicht ihre Fähigkeiten getestet werden.
3. Die Testpersonen machen zwangsläufig Fehler und arbeiten langsam, worauf sie sich leicht unzulänglich oder dumm fühlen können. Dass die Testpersonen während des Tests beobachtet und gefilmt werden, verschärft die Situation noch zusätzlich. Die Testleitung soll Ruhe ausstrahlen und kein zusätzlicher Druck auf die Testpersonen ausüben.
4. Die Testleitung hat die Verantwortung, dass sich die Testpersonen während und nach dem Test so wohl als möglich fühlen und darf sich niemals über die gezeigte Leistung lustig machen.
5. Die Testleitung sichert den Testpersonen Stillschweigen über ihre Leistung zu, sowie genereller Datenschutz zur Identität der Testpersonen (den Testpersonen Nummern zuteilen).
6. Die Testleitung soll eine möglichst entspannte Atmosphäre schaffen und sich Zeit nehmen für „Small-talk“.
7. Die Testleitung sollte einen Test in den Fällen abbrechen, wenn die Testperson sichtlich unglücklich ist und mit dem System überhaupt nicht zurechtkommt.
8. Die Testpersonen sollen nach dem Test die Möglichkeit haben, ihre Meinung zum System äussern zu können. Die Meinung der Testpersonen wird ernst genommen.

(Quelle: Nielsen, 1993)

Anhang C

Drehbuch für den klassischen Usability-Test

1. Vorbereitung: Vor dem Test

1.1 Äusseres Erscheinungsbild der Testleitung

Damit das Äussere der Testleitung konstant gehalten werden kann bei allen Tests werden folgende Regelungen eingehalten:

- **Brille/Kontaktlinsen:** Bei allen Tests trägt die Testleitung Kontaktlinsen (Konstant Haltung der Variable Brille/Kontaktlinsen)
- **Kleidung :** Bei allen Tests wird eine unauffällige Kleidung gewählt. Diese besteht aus Jeans und einem dunkeln Oberteil. Grelle Farben, sowie Röcke werden nicht getragen.
- **Schmuck:** Es wird kein auffälliger Schuck getragen
- **Frisur:** Die Haar werden nach hinten gebunden
- **Make-up:** Es wird kein Makeup getragen
- **Sprache:** Da alle Testpersonen Dialekt verstehen, wird bei allen Tests Mundart gesprochen
- **Umgang mit den Testpersonen:** Es wird ein neutraler, höflicher und respektvoller Umgang mit den Testpersonen gepflegt. Auf Fragen der TPs zum Testablauf und Durchführung wird eingegangen.

2. Durchführung: Testvorbereitung

2.1 Begrüssung der Testpersonen

- Die Testpersonen werden vor dem Labor begrüsst per Händeschütteln
- Die Testleitung fragt die Testperson, ob sie den Mantel abnehmen darf zum Aufhängen
- Die Testperson wird gefragt, ob sie ein Wasser möchte
- Es wird der Testperson für die Zeit und ihre Teilnahme gedankt

2.2 Laboreinrichtung

- Die Testleitung zeigt der Testperson den Beobachtungsraum und informiert die Testperson, dass die Testleitung während des Tests in diesem Raum sitzt, dass über Mikrophon miteinander gesprochen werden kann und dass die Testleitung sieht, was die Testperson macht.
- Die Testperson wird von der Testleitung an den Arbeitsplatz im Testraum geführt
- Die Testleitung erklärt der Testperson die Laboreinrichtung und fragt nach Erlaubnis den Test aufzunehmen. Die Testleitung versichert der Testperson, dass die Aufnahmen nur von ihr zu Auswertungszweck benutzt werden. Zusicherung der Anonymität.

2.3 Ablauf des Tests

- Die Testleitung informiert die Testperson über den Zweck von Usability der Untersuchung.
„Das Ziel von Usability ist, Systeme zu überprüfen, verbessern, um die Bedienung zu vereinfachen. Ich führe zwei verschiedene Varianten von Usabilitytests durch und vergleiche sie miteinander. Dies mache ich anhand eines Websitetests der Klubschule Migros“.
- Die Testleitung erklärt der Testperson den Ablauf des Tests:
- „Ich werde anschliessend in den Beobachtungsraum gehen. Bitte füllen Sie zuerst den Vorfragebogen aus. Anschliessend werde ich Ihnen sagen, dass Sie mit der 1.Aufgabe starten können. Für die Aufgaben sind 30 Minuten Zeit geplant. Sie müssen in dieser Zeit aber nicht alle Aufgaben lösen, sie bearbei-

- ten die Aufgaben soweit Sie können. Nach den 30 Minuten informiere ich Sie, dass die Zeit um ist. Ich bitte Sie dann den Nachfragebogen auszufüllen. Ich werde wieder zu Ihnen in den Testraum kommen und stelle noch einige abschliessende Fragen. Ist Ihnen der Ablauf klar? Haben Sie noch Fragen dazu?“
- Die Testleitung erklärt die Spielregeln:
 - „Sie dürfen den Test jederzeit unterbrechen, bzw. abbrechen“
 - „Es werden nicht Ihre Fähigkeiten beurteilt, sondern es sollen die Probleme der Webseite ermittelt werden“
 - „Bearbeiten Sie die Aufgaben, soweit Sie können. Wenn Sie bei einer Aufgabe nicht mehr weiterkommen, gehen Sie selbständig zur nächsten Aufgabe.“
 - „Während der Aufgabenbearbeitung bitte ich Sie, laut zu denken, das heisst jeden Schritt und Gedanken in Worte zu fassen. Somit geben Sie mir die Möglichkeit, mich an Ihren Überlegungen teilhaben zu lassen und Ihre Schritte besser nachvollziehen zu können.“
 - „Die Konversation verläuft kontrolliert über eine Gegensprechanlage“

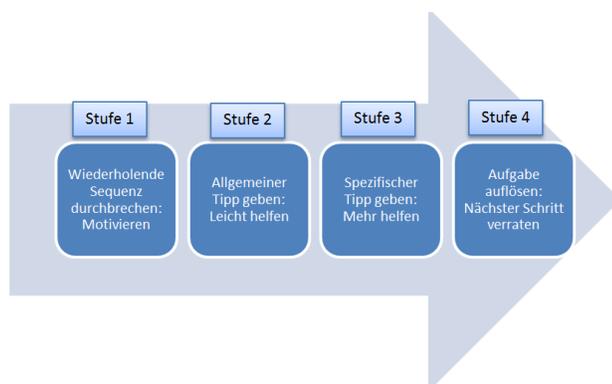
3. Durchführung: Websitetest

- Die Testleitung verlässt den Testraum und startet die Aufnahme im Beobachtungsraum
- Die Testperson füllt selbständig den Vorfragebogen aus
- Die Testperson beginnt den Test mit der ersten Aufgabe
- Die Testleitung startet die Stoppuhr (30 Minuten)

3.1 Umgang mit Fragen der Testperson: Hilfestellung

Wenn eine Testperson in einer Aufgabe nicht mehr weiterkommt und nach Hilfe fragt, reagiert die Testleitung folgendermassen:

- Wichtig ist, dass die Testleitung so wenig als möglich ins Testgeschehen eingreift
- Die Testleitung wartet länger als beim Usability-Walkthrough und bietet die Hilfe dann in vier Stufen an:
- **1. Stufe:** Motivieren. Z.B. „Was denken Sie was hier vorgeht?“, „Lesen Sie bitte nochmals die Aufgabe“
- **2. Stufe:** Einen allgemeinen Tipp geben. Z.B. „Erinnern Sie sich, wie Sie die Aufgabe begonnen haben? Sie waren nahe dran“, „Sie waren bereits bei der gesuchten Option“, „Lehnen Sie sich zurück und betrachten Sie den ganzen Bildschirm“, „Die gesuchte Option befindet sich in einem Menu, dass Sie bis jetzt noch nicht geöffnet haben“.
- **3. Stufe:** Einen spezifischen Tipp geben. Z.B. „Die Option befindet sich im Menu X“.
- **4. Stufe:** Aufgabe abschliessen und nächsten Schritt verraten. Z.B. „Klicken Sie auf den dritten Button“.
- Die Testleitung notiert sich im Beobachtungsbogen, welche Aufgaben zu welchem Zeitpunkt abgebrochen werden



3.2 Bei Abbruch einer Aufgabe

- -Die Testleitung notiert die abgebrochene Aufgabe
- -Die Testleitung notiert die Zeit, welche die TP für die Aufgabe aufgebracht hat
- -Die Testleitung notiert die Teilschritte, welche in der Aufgabe erreicht wurden

3.2 Aufforderung zum lauten Danken

Wenn die Testleitung merkt, dass die Testperson still ist und ihre Gedankengänge nicht laut mitteilt, soll die Testleitung intervenieren und die Testperson bitten „laut zu denken“. Da für die meisten Testpersonen das „Think aloud“ eine neue und ungewohnte Erfahrung ist, soll darauf geachtet werden, dass die Testperson mit der Aufforderung möglichst wenig zusätzlich gestresst werden. Tests stellen für die Testperson eine unnatürliche, belastende und stressige Situation dar. Dabei können sich negative frühere Erfahrungen, sowie Leistungsdruck, Versagensangst und kognitive sowie psychische Beanspruchung auf die Testperson auswirken. Die Testleitung möchte eine mögliche negative Abwärtsspirale vermeiden und fordert deshalb die Testperson in ruhigen Tonfall die Testperson auf „laut zu denken“. Die folgenden Aufforderungen zum „Laut denken“ erfolgen bei allen Tests immer in dieser Reihenfolge:

1. „Welche Gedanken gehen Ihnen im Moment durch den Kopf?“
2. „Was denken Sie, wenn Sie diese Seite betrachten?“
3. „Was denken Sie, was diese Meldung zu bedeuten hat?“
4. „Was denken Sie, was passieren wird?“
5. „Welche Reaktionen haben Sie erwartet?“

4. Abschluss des Tests

- Die Testleitung gibt das Ende der Zeit an und bittet die Testperson den Nachfragebogen selbständig auszufüllen
- Die Testleitung kommt ins Labor zurück und setzt sich neben die Testperson
- Die Testleitung stellt abschliessende offene Fragen
- Die Testleitung fragt die Testperson, ob diese noch Fragen zum Versuch hat
- Dank und Überreichung eines Geschenks
- Die Testleitung verabschiedet die Testperson mit einem Händedruck
- Die Testleitung stoppt die Aufnahme

Drehbuch für den Usability-Walkthrough

1. Vorbereitung: Vor dem Test

1.1 Äusseres Erscheinungsbild der Testleitung

Damit das Äussere der Testleitung konstant gehalten werden kann bei allen Tests werden folgende Regelungen eingehalten:

- **Brille/Kontaktlinsen:** bei allen Tests trägt die Testleitung Kontaktlinsen (Konstant Haltung der Variable Brille/Kontaktlinsen)
- **Kleidung :** bei allen Tests wird eine unauffällige Kleidung gewählt. Diese besteht aus Jeans und einem dunkeln Oberteil. Grelle Farben, sowie Röcke werden nicht getragen.
- **Schmuck:** Es wird kein auffälliger Schuck getragen
- **Frisur:** Die Haar werden nach hinten gebunden
- **Make-up:** Es wird kein Makeup getragen
- **Sprache:** Da alle Testpersonen Dialekt verstehen, wird bei allen Tests Mundart gesprochen
- **Umgang mit den Testpersonen:** Es wird ein neutraler, höflicher und respektvoller Umgang mit den Testpersonen gepflegt. Auf Fragen der TPs zum Testablauf und Durchführung wird eingegangen.

2. Durchführung: Testvorbereitung

2.1 Begrüssung der Testpersonen

- Im Beobachtungsraum wird die Aufnahme gestartet, sobald die Testperson zum Test erscheint
- Die Testpersonen werden vor dem Labor begrüsst per Händeschütteln
- Die Testleitung fragt die Testperson, ob sie den Mantel abnehmen darf zum Aufhängen
- Die Testperson wird gefragt, ob sie ein Wasser möchte
- Es wird der Testperson für die Zeit und ihre Teilnahme gedankt

2.2 Laboreinrichtung

- Die Testperson wird von der Testleitung an den Arbeitsplatz im Testraum geführt
- Die Testleitung erklärt der Testperson die Laboreinrichtung und fragt nach Erlaubnis den Test aufzunehmen. Die Testleitung versichert der Testperson, dass die Aufnahmen nur von ihr zu Auswertungszweck benutzt werden. Zusicherung der Anonymität.

2.3 Ablauf des Tests

- Die Testleitung informiert die Testperson über den Zweck von Usability der Untersuchung.
„Das Ziel von Usability ist, Systeme zu überprüfen, verbessern, um die Bedienung zu vereinfachen. Ich führe zwei verschiedene Varianten von Usabilitytests durch und vergleiche sie miteinander. Dies mache ich anhand eines Websitetests der Klubschule Migros“.
- Die Testleitung erklärt der Testperson den Ablauf des Tests:
- „Bitte füllen Sie zuerst den Vorfragebogen aus. Anschliessend werde ich Ihnen sagen, dass Sie mit der 1.Aufgabe starten können. Für die Aufgaben sind 30 Minuten Zeit geplant. Sie müssen in dieser Zeit aber nicht alle Aufgaben lösen, sie bearbeiten die Aufgaben soweit Sie können. Nach den 30 Minuten informiere ich Sie, dass die Zeit um ist. Ich bitte Sie dann den Nachfragebogen auszufüllen. Anschliessend werde ich Ihnen noch einige abschliessende Fragen stellen. Ist Ihnen der Ablauf klar? Haben Sie noch Fragen dazu?“
- Die Testleitung erklärt die Spielregeln:

- „Sie dürfen den Test jederzeit unterbrechen, bzw. abbrechen“
- „Es werden nicht Ihre Fähigkeiten beurteilt, sondern es sollen die Probleme der Webseite ermittelt werden“
- „Bearbeiten Sie die Aufgaben, soweit Sie können. Wenn Sie bei einer Aufgabe nicht mehr weiterkommen, gehen Sie selbständig zur nächsten Aufgabe.“
- „Während der Aufgabenbearbeitung bitte ich Sie, laut zu denken, das heisst jeden Schritt und Gedanken in Worte zu fassen. Somit geben Sie mir die Möglichkeit, mich an Ihren Überlegungen teilhaben zu lassen und Ihre Schritte besser nachvollziehen zu können.“

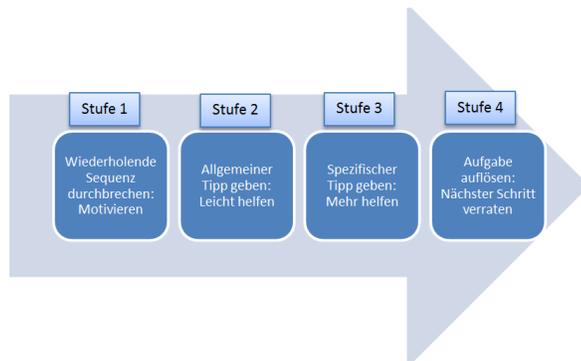
3. Durchführung: Websitetest

- Die Testperson füllt selbständig den Vorfragebogen aus
- Die Testperson beginnt den Test mit der ersten Aufgabe
- Die Testleitung startet die Stoppuhr (30 Minuten)

3.1 Umgang mit Fragen der Testperson- Hilfestellung

Wenn eine Testperson in einer Aufgabe nicht mehr weiterkommt und nach Hilfe fragt, reagiert die Testleitung folgendermassen:

- Wichtig ist, dass die Testleitung so wenig als möglich ins Testgeschehen eingreift
- Die Testleitung wartet weniger lange als beim Usability-Walkthrough und bietet die Hilfe dann in vier Stufen an:
- **1. Stufe:** Motivieren. Z.B. „Was denken Sie was hier vorgeht?“ „Lesen Sie bitte nochmals die Aufgabe“
- **2. Stufe:** Einen allgemeinen Tipp geben. Z.B. „Erinnern Sie sich, wie Sie die Aufgabe begonnen haben? Sie waren nahe dran“, „Sie waren bereits bei der gesuchten Option“, „Lehnen Sie sich zurück und betrachten Sie den ganzen Bildschirm“, „Die gesuchte Option befindet sich in einem Menu, dass Sie bis jetzt noch nicht geöffnet haben“.
- **3. Stufe:** Einen spezifischen Tipp geben. Z.B. „Die Option befindet sich im Menu X“.
- **4. Stufe:** Aufgabe abschliessen und nächsten Schritt verraten. Z.B. „Klicken Sie auf den dritten Button“.
- Die Testleitung notiert sich im Beobachtungsbogen, welche Aufgaben zu welchem Zeitpunkt abgebrochen werden



Das Hilfestufenmodell dient der Testleitung als Anhaltspunkt und Vorgehen bei Hilfesuchen der Testperson. Damit die Tests so weit als möglich standardisiert durchgeführt werden können und die Durchführungsobjektivität möglichst hoch gehalten werden kann, hält sich die Testleitung in jedem Fall an die Vorgaben des Drehbuches. Wichtig ist, dass die Testleitung nur eingreift, wenn es nötig ist und erst nach einer Wartezeit von einigen Minuten. Wenn die Testperson nach der Hilfestellung zwei und einer weiteren Wartezeit trotzdem nicht mehr

weiterkommt, wird die Aufgabe von der Testleitung abgebrochen und aufgelöst. Die Testperson löst die nächste Aufgabe.

3.2 Bei Abbruch einer Aufgabe

- Die Testleitung notiert die abgebrochene Aufgabe
- Die Testleitung notiert die Zeit, welche die TP für die Aufgabe aufgebracht hat
- Die Testleitung notiert die Teilschritte, welche in der Aufgabe erreicht wurden

3.3 Aufforderungen zum lauten Denken

Wenn die Testleitung merkt, dass die Testperson still ist und ihre Gedankengänge nicht laut mitteilt, soll die Testleitung intervenieren und die Testperson bitten „laut zu denken“. Da für die meisten Testpersonen das „Think aloud“ eine neue und ungewohnte Erfahrung ist, soll darauf geachtet werden, dass die Testperson mit der Aufforderung möglichst wenig zusätzlich gestresst werden. Tests stellen für die Testperson eine unnatürliche, belastende und stressige Situation dar. Dabei können sich negative frühere Erfahrungen, sowie Leistungsdruck, Versagensangst und kognitive sowie psychische Beanspruchung auf die Testperson auswirken. Die Testleitung möchte eine mögliche negative Abwärtsspirale vermeiden und fordert deshalb die Testperson in ruhigen Tonfall die Testperson auf „laut zu denken“. Die folgenden Aufforderungen zum „Laut denken“ erfolgen bei allen Tests immer in dieser Reihenfolge:

6. „Welche Gedanken gehen Ihnen im Moment durch den Kopf?“
7. „Was denken Sie, wenn Sie diese Seite betrachten?“
8. „Was denken Sie, was diese Meldung zu bedeuten hat?“
9. „Was denken Sie, was passieren wird?“
10. „Welche Reaktionen haben Sie erwartet?“

3.4 Kommunikation

Im Unterschied zum klassischen Usability-Test darf die Testleitung mehr moderieren und auf Bemerkungen der Testperson eingehen. Bei einem Ausruf oder wenn eine Testperson ihren Unmut über eine Problemstelle im System kundtut, kann die Testleitung nachfragen. Z.B. „Was hätten Sie erwartet, was passieren würde?“ oder „Wie würden Sie diese Funktion gestalten?“.

5. Abschluss des Tests

- Die Testleitung gibt das Ende der Zeit an
- Die Testperson füllt den Nachfragebogen aus
- Die Testleitung stellt abschliessende Fragen
- Die Testleitung fragt die Testperson, ob diese noch Fragen zum Versuch hat
- Dank und Überreichung eines Geschenks
- Die Testleitung verabschiedet die Testperson mit einem Händedruck
- Die Testleitung stoppt die Aufnahme

Anhang D

Testleiterskript für klassischen Usability-Test

Test:

TP-Nr.:

1. Testperson vor dem Labor begrüßen, Wasser anbieten
2. Testperson ins Labor führen und die Laboreinrichtung erklären. „Ich als Testleiterin sitze in diesen Raum und kann sie auf dem Bildschirm sehen und verstehen. Sie hören mich auch, wir können während des Tests miteinander sprechen. Sie sitzen hier während des Tests“.
3. Information zu Usability: „Ich führe meine Bachelor Thesis im Bereich Usability durch. Das Ziel von Usability ist, Systeme zu überprüfen, verbessern, um die Bedienung zu vereinfachen“.
4. Information zum Thema der Bachelor Thesis: „Ich führe zwei verschiedene Varianten von Usabilitytests durch und vergleiche sie miteinander. Dies mache ich anhand eines Websitetests der Klubschule Migros“.
5. Testablauf und Spielregeln erklären
Ablauf:
 - Vorfragebogen ausfüllen, 30 Minuten Zeit für Aufgaben lösen. „Es müssen nicht alle Aufgaben gelöst werden, lösen Sie die Aufgaben soweit Sie können“. Nachfragebogen ausfüllen, abschliessende Fragen der Testleiterin, Ende.**Spielregeln:**
 - „Es werden nicht ihre Fähigkeiten beurteilt, sondern es sollen die Probleme der Webseite ermittelt werden!“
 - „Sie können den Test jederzeit unterbrechen, bzw. abbrechen“
 - „Bearbeiten Sie die Aufgaben, soweit Sie können, Sie können selbständig zur nächsten Aufgabe gehen“
 - „Ich bitte Sie laut zu denken, damit ich ihre Gedankengänge nachvollziehen kann“
 - „Haben Sie noch Fragen? Ist Ihnen alles klar?“
6. Website Klubschule Migros zeigen
7. Die Testleiterin geht in den Beobachtungsraum
8. Die Testperson liest erste Seite und füllt Vorfragebogen aus
9. Die Testperson beginnt nach Absprache mit der Testleiterin die Aufgaben zu lösen.
10. Aufnahme starten
11. Zeit stoppen 30 Minuten

Anhang

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>1</p> <p>Sie möchten sich gerne sportlich betätigen und haben dabei an die Klubschule der Migros gedacht. Bitte suchen Sie auf der Website der Klubschule einen Rückentrainingskurs aus, der abends in Olten angeboten wird.</p>							

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>2</p> <p>Bitte gehen Sie zurück zur Startseite der Migros Klubschule Website. Sie sprechen schon fließend Englisch, weshalb Sie ein Einsteiger-Kurs gar nicht mehr interessiert, sondern mehr ein etwas fortgeschrittener Kurs (mindestens Level B oder höher). Bitte finden Sie heraus, welche Englischkonversations-Kurse für Fortgeschrittene von der Migros Klubschule abends in Zug angeboten werden. Sie möchten das Budget von 500 CHF nicht überschreiten, aber dennoch möglichst viele Lektionen besuchen können.</p>							

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>3</p> <p>Sie möchten sich im Umgang mit dem Internet etwas weiterbilden und zwar haben Sie in einem Prospekt von einem Kurs mit dem lustigen Namen „... und täglich grüsst das Internet“ gelesen. Bitte versuchen Sie den Kurs jeweils am Dienstagabend in Luzern zu buchen. Benutzen Sie für die Anmeldung folgende Angaben: Max Muster Musterstrasse 12 8200 Schaffhausen 052 624 12 00 max.muster@gmx.ch Sie wünschen die Korrespondenz und die Rechnung per E-Mail „Weiter“ „Weiter“ Achtung: Nicht auf „Angaben senden“ klicken.</p>							

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>4</p> <p>Eine Arbeitskollegin hat ihnen stolz ein Diplom gezeigt, das sie nach eigenen Angaben bei der Migros Klubschule in Zug gemacht habe.</p> <p>a) In welchen Bereichen können Diplome bei der Klubschule erlangt werden?</p> <p>b) Bitte wählen Sie einen beliebigen Diplommkurs am Standort Zürich aus und finden Sie heraus, wer das Zielpublikum dieses Angebotes ist.</p>							

Anhang

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>5</p> <p>In ihrer Firma hat jeder Mitarbeiter ein Budget von 1000 CHF für Aus- und Weiterbildung. Sie wollen sich nun informieren, welche Angebote mit einem Diplom oder Zertifikat es an der Klubschule gibt.</p> <p>Bitte suchen Sie ein Angebot im Bereich Führung in Winterthur aus und buchen Sie ein entsprechendes Angebot auf Kosten der Firma. Sie können die gleichen Angaben wie bei Aufgabe 3 verwenden. Die Adresse der Firma ist wie folgt:</p> <p>soultank AG Bahnhofplatz Postfach 4744 6304 Zug</p> <p>Nach der Eingabe der Daten können Sie stoppen, klicken Sie nicht auf „Weiter“.</p>							

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>6</p> <p>Sie möchten sich und ihrer Kollegin etwas Gutes tun und zwar wollen Sie gemeinsam einen Massage-Kurs bei der Migros Klubschule besuchen.</p> <p>Bitte suchen Sie einen Sportmassage-Kurs, der abends stattfindet. Der Kursort spielt keine Rolle.</p>							

Anhang

Aufgaben	Vorgehen der Testperson	Usability-Probleme			Aufgabe gelöst		
		Beschreibung	Anzahl	Art	Ja	Nein	abgebr.
<p>7</p> <p>Sie sind Kursleiterin/Kursleiter bei der Klubschule und Weiterbildung ist auch für Sie ein Thema. Nun interessiert Sie, ob die Klubschule auch Angebote für Sie parat hält. Bitte versuchen Sie ein auf Sie als Kursleiterin/Kursleiter zugeschnittenes Angebot zu finden und wählen Sie davon ein Angebot aus. Teilen Sie dem Testleiter mit, für welches Angebot Sie sich entscheiden würden.</p>							

12. Bei 30 Minuten stoppen und die Testperson informieren.
13. Die Testleiterin kommt nach einiger Zeit in den Testraum und setzt sich neben die Testperson. Währenddessen füllt die Testperson den Nachfragebogen fertig aus.
14. Mündliche Fragen an die Testperson:
 1. Was hat Ihnen an der Website der Klubschule Migros gut gefallen?
.....
.....
 2. Was hat Ihnen an der Website der Klubschule Migros nicht gefallen? Was würden Sie ändern?
.....
.....
 3. Wie haben Sie den Test erlebt, wie ist es Ihnen dabei ergangen?
.....
.....
- 3.1 Wie haben Sie die Testsituation im Labor erlebt?
- 3.2 Wie war es für Sie, dass der Test und ihr Umgang mit dem System von einer Kamera aufgezeichnet wurden?
- 3.3 Wie haben Sie den Umstand empfunden, dass sie Aufgaben lösen mussten unter Beobachtung?
- 3.4 Wie haben Sie es erlebt, dass sich die TL während des Tests in einem anderen Raum aufgehalten hat?
- 3.5 Würden Sie in Zukunft wieder einmal an einem Usability-Test mitmachen?
15. Dank und Geschenk überreichen
16. Testperson hinausbegleiten
17. Wenn Testperson gegangen ist:
 - Aufnahme stoppen und eine neue Observation einrichten
 - Usability-Probleme nochmals durchgehen, Anzahl und Art aufschreiben
 - Raum lüften, Wasser bereit stellen, Internetoptionen: Browserverlauf löschen, Formulardaten löschen

Testleiterskript für Usability-Walkthrough

Hinweis: Da es sich um die gleichen Aufgaben beim klassischen Usability-Test und beim Usability-Walkthrough handelt, werden hier nur noch die erst und letzte Seite vom Testleiterskript gezeigt, die sich unterscheiden.

Test: TP-Nr.:

1. Testperson vor dem Labor begrüßen, Wasser anbieten
2. Aufnahme starten
3. Testperson ins Labor führen
4. Information zu Usability: „Ich führe meine Bachelor Thesis im Bereich Usability durch. Das Ziel von Usability ist, Systeme zu überprüfen, verbessern, um die Bedienung zu vereinfachen.“
5. Information zum Thema der Bachelor Thesis: „Ich führe zwei verschiedene Varianten von Usabilitytests durch und vergleiche sie miteinander. Dies mache ich anhand eines Websitetests der Klubschule Migros.“

6. Testablauf und Spielregeln erklären

Ablauf:

- Vorfragebogen ausfüllen, 30 Minuten Zeit für Aufgaben lösen. „Es müssen nicht alle Aufgaben gelöst werden, lösen Sie die Aufgaben soweit Sie können“. Nachfragebogen ausfüllen, abschliessende Fragen der Testleiterin, Ende

Spielregeln:

- „Es werden nicht ihre Fähigkeiten beurteilt, sondern es sollen die Probleme der Webseite ermittelt werden!“
- „Sie können den Test jederzeit unterbrechen, bzw. abbrechen“
- „Bearbeiten Sie die Aufgaben, soweit Sie können, Sie können selbständig zur nächsten Aufgabe gehen“
- „Ich bitte Sie laut zu denken, damit ich ihre Gedankengänge nachvollziehen kann“
- „Haben Sie noch Fragen? Ist Ihnen alles klar?“

7. Website Klubschule Migros zeigen

8. Die Testperson liest die erste Seite und füllt den Vorfragebogen aus
9. Die Testperson beginnt nach Absprache mit der Testleiterin die Aufgaben zu lösen.
10. Zeit stoppen 30 Minuten
11. Bei 30 Minuten stoppen und Testperson informieren.
12. Die Testperson bitten, den Nachfragebogen auszufüllen.
13. Mündliche Fragen an die Testperson:

1. Was hat ihnen an der Website der Klubschule Migros gut gefallen?

.....
.....

2. Was hat Ihnen an der Website der Klubschule Migros nicht gefallen? Was würden Sie ändern?

.....
.....

3. Wie haben Sie den Test erlebt, wie ist es Ihnen dabei ergangen?

.....
.....

3.1 Wie haben Sie die Testsituation im Labor erlebt?

3.2 Wie war es für Sie, dass der Test und ihr Umgang mit dem System von einer Kamera aufgezeichnet wurden?

3.3 Wie haben Sie den Umstand empfunden, dass sie Aufgaben lösen mussten unter Beobachtung?

3.4 Wie haben Sie es erlebt, dass sich die TL während des Tests in einem anderen Raum aufgehalten hat?

3.5 Würden Sie in Zukunft wieder einmal an einem Usability-Test mitmachen?

14. Dank und Geschenk überreichen

15. Testperson hinausbegleiten

16. Wenn die Testperson gegangen ist:

- Aufnahme stoppen und speichern
- Usability-Probleme nochmals durchgehen, Anzahl und Art aufschreiben
- Raum lüften, Wasser bereit stellen, Internetoptionen: Browserverlauf löschen, Formulardaten löschen

Testunterlagen für Testpersonen

Willkommen zum Usability-Test für meine Bachelor Thesis

Im folgenden Usability-Test werden Sie verschiedene Aufgaben auf der Website der Klubschule Migros durchführen und beurteilen.

Die Gesamtdauer des Tests beträgt 45 Minuten, davon sind 30 Minuten zum Lösen der Aufgaben gedacht. Es müssen in dieser Zeit nicht alle Aufgaben bearbeitet werden.

Wenn Sie Anregungen und Kommentare haben, können Sie diese jederzeit der Testleiterin mitteilen oder auch auf den entsprechenden Blättern notieren.

Was mir wichtig ist

Es geht bei diesem Test nicht darum, Ihre Fähigkeiten und Ihr Verständnis im Umgang mit der Website, sondern die Schwachstellen der Website aufzudecken.

Herzlichen Dank für Ihre Teilnahme.

Ich wünsche Ihnen viel Spass!

Stefanie Friedli

Vorgängige Fragen

Test: TP-Nr.:

Ihr Jahrgang:

Ihr Geschlecht? Mann Frau

Ihr höchster Bildungsabschluss?

- | | |
|---|---|
| <input type="checkbox"/> Schüler | <input type="checkbox"/> (Fach-) Hochschule Bachelor |
| <input type="checkbox"/> Schulabschluss | <input type="checkbox"/> (Fach-) Hochschule Master / Diplom / Lizenciat |
| <input type="checkbox"/> Berufslehre / KV | <input type="checkbox"/> Doktorat |
| <input type="checkbox"/> Berufsmatura | <input type="checkbox"/> Andere, welche? |
| <input type="checkbox"/> Matura | |
-

Wie gut schätzen Sie Ihre Internetkenntnisse ein?

Sehr schlecht Sehr gut

Wie häufig nutzen Sie das Internet?

- Nie
 - Sehr selten
 - Einmal wöchentlic
 - Mehrmals wöchentlic
 - Täglich
 - Mehrmals täglich
-

Kennen Sie die Website www.klubschule.ch? Ja Nein

Aufgabe 1

Sie möchten sich gerne sportlich betätigen und haben dabei an die Klubschule der Migros gedacht. Bitte suchen Sie auf der Website der Klubschule einen **Rückentrainingskurs** aus, der **abends** in **Olten** angeboten wird.

Aufgabe 2

Bitte gehen Sie zurück zur Startseite der Klubschule Migros Website.

Sie sprechen schon fließend Englisch, weshalb Sie ein Einsteiger-Kurs gar nicht mehr interessiert, sondern mehr ein etwas fortgeschrittener Kurs (mindestens **Level B oder höher**).

Bitte finden Sie heraus, welche **Englischkonversations-Kurse** für Fortgeschrittene von der Klubschule Migros **abends** in **Zug** angeboten werden. Sie möchten das Budget von **500 CHF** nicht überschreiten, aber dennoch möglichst viele Lektionen besuchen können.

Aufgabe 3

Sie möchten sich im Umgang mit dem Internet etwas weiterbilden und haben in einem Prospekt von einem Kurs mit dem lustigen Namen „... **und täglich grüsst das Internet**“ gelesen.

Bitte versuchen Sie den Kurs jeweils am **Dienstagabend** in **Luzern** zu buchen.

Benutzen Sie für die Anmeldung folgende Angaben:

- Vorname Name: Max Muster
 - Strasse und Nr.: Musterstrasse 12
 - Ort: 8200 Schaffhausen
 - Telefon: 052 624 12 00
 - E-Mail: max.muster@gmx.ch
 - Sie wünschen die Korrespondenz und die Rechnung per E-Mail
 - Klicken Sie auf „Weiter“
 - Akzeptieren Sie die AGBs und klicken Sie auf „Weiter“
 - **Achtung: Nicht auf „Angaben senden“ klicken**
-

Aufgabe 4

Eine Arbeitskollegin hat ihnen stolz ein Diplom gezeigt, das sie nach eigenen Angaben bei der Klubschule Migros in Zug gemacht habe.

a) In **welchen Bereichen** können **Diplome** bei der Klubschule erlangt werden?

b) Bitte wählen Sie einen **beliebigen Diplomkurs** am Standort **Zürich** aus und finden Sie heraus, **wer das Zielpublikum** dieses Angebotes ist.

Aufgabe 5

In ihrer Firma haben alle Mitarbeitende ein Budget von **1000 CHF** für Aus- und Weiterbildung. Sie wollen sich nun informieren, welche Angebote mit einem Diplom oder Zertifikat es an der Klubschule gibt.

Bitte suchen Sie ein Angebot im Bereich **Führung** in **Winterthur** aus und buchen Sie ein entsprechendes Angebot auf Kosten der Firma. Sie können die gleichen Angaben wie bei Aufgabe 3 verwenden. Die Adresse der Firma ist wie folgt:

soultank AG
Bahnhofplatz
Postfach 4744
6304 Zug

Nach der Eingabe der Daten können Sie stoppen, klicken Sie nicht auf „Weiter“.

Aufgabe 6

Sie möchten sich und ihrer Kollegin etwas Gutes tun und zwar wollen Sie gemeinsam einen Massage-Kurs bei der Klubschule Migros besuchen.

Bitte suchen Sie einen **Sportmassage-Kurs**, der **abends** stattfindet. Der Kursort spielt keine Rolle.

Aufgabe 7

Sie sind Kursleiterin/Kursleiter bei der Klubschule und Weiterbildung ist auch für Sie ein Thema.

Nun interessiert Sie, ob die Klubschule auch Angebote für Sie parat hält.

Bitte versuchen Sie ein auf Sie als Kursleiterin/Kursleiter zugeschnittenes Angebot zu finden und wählen Sie davon ein Angebot aus. Teilen Sie der Testleiterin mit, für welches Angebot Sie sich entscheiden würden.

Herzlichen Dank

Dieser Teil des Tests ist nun abgeschlossen. Bitte beantworten Sie die nachfolgenden Fragen.

Nachträgliche Beurteilung

Wie beurteilen Sie die Benutzungsfreundlichkeit der Website der Klubschule Migros im Allgemeinen?

Sehr schlecht Sehr gut

Mittel

Fragen zum Test

Bitte kreuzen Sie jeweils an, wie Sie den Test erlebt haben. Bitte setzen Sie in jeder Zeile nur ein Kreuz.

	Trifft nicht zu	Trifft wenig zu	Trifft etwas zu	Trifft ziemlich zu	Trifft überwiegend zu	Trifft ganz genau zu
1. Es fiel mir leicht, den Test zu bearbeiten.						
2. Zeitweise habe ich den Anweisungen nicht folgen können.						
3. Ich hatte ziemliche Angst, die Aufgaben nicht zu schaffen.						
4. Meine Ausdauer war während der ganzen Zeit gleichmässig gut.						
5. Ich hatte zeitweise ziemliche Probleme mit der Konzentration.						
6. Ich könnte noch länger solche Aufgaben lösen.						
7. Zeitweise habe ich nicht verstanden, was am Computer ablief.						
8. Während der Bearbeitung hatte ich das Gefühl, aufgeben zu wollen.						
9. Ich bin mit den Leistungen zufrieden, die ich zeigen konnte.						
10. Ich war durch die Aufgaben stark überfordert.						

Herzlichen Dank für Ihre Teilnahme.

Die Testleiterin wird Ihnen abschliessend noch einige Fragen stellen.

Anhang E

Auswertungsleitlinien

1. Die Daten werden in das Excel-Input Datensheet eingegeben.
2. In die Datensheet werden folgende Daten eingegeben:

Aus dem Aufgabenblatt der Testpersonen:

- Testpersonennummer (1-40)
- Ort des Tests (0,1)
- Test (A oder B)
- Jahrgang
- Geschlecht (0,1)
- Höchster Bildungsabschluss (0-7)
- Einschätzen der Internetkenntnisse (0-5)
- Häufigkeit der Nutzung des Internets (0-5)
- Kennen Sie die Website bereits? (0,1)
- Beurteilung der Benutzerfreundlichkeit der Website (0-6)
- 10 Items des Fragebogens zum Testerleben (jeweils 0-5)

Aus dem Testleiterskript:

1. Anzahl gelöster Aufgaben

Richtlinien für Bestimmung Anzahl gelöster Aufgaben

Wertung	Fall	Begründung der Wertung
0	Eine Aufgabe wird zu Beginn oder Mitte aus Zeitgründen abgebrochen	Die Testperson hat die Aufgabe erst begonnen und aus Zeitgründen nicht fertig bearbeitet.
0	Eine Aufgabe muss abgebrochen werden und die Auflösung der Aufgabe durch die Testleitung erfolgt (Hilfestellung 4: Aufgabe auflösen).	Die Testperson kann auch nach längerem Probieren und mit Hilfestellung der Testleitung die Aufgabe nicht lösen, ist deshalb nun genervt, aufgeregt, gestresst oder unmotiviert weiter zu probieren, löst die Testleitung die Aufgabe auf und die Testperson geht zur nächsten Aufgabe. Die Testperson konnte die Aufgabe nicht lösen, weshalb es keine Punkte gibt.
0.5	Bei Aufgabe 4 wird nur eine Teilaufgabe gelöst, entweder a oder b	Da die Aufgabe 4 aus zwei Teilaufgaben a und b besteht, werden für jede fertig gelöste Teilaufgabe 0.5 Punkte berechnet. Wenn die ganze Aufgabe 4 (a und b) gelöst wird, wird 1 Punkt berechnet.
0.5	Eine Aufgabe wird gelöst mit Hilfe von Testleitung (Hilfestellung 1-2)	Die Testperson kann die Aufgabe nicht alleine und auch nach einiger Zeit und mehreren Versuchen nicht lösen. Mit Hilfestellung kann die Testperson die Aufgabe lösen. Da die Aufgabe zeitlich und inhaltlich nicht alleine gelöst werden konnte, wird nicht die volle Punktezahl berechnet.
0.5	Wenn eine Aufgabe fast fertig gelöst wird und aus Zeitgründen (Ende der 30 Minuten) abgebrochen werden muss	Die Aufgabe wird selbständig von der Testperson gelöst, muss aber aus Zeitgründen abgebrochen werden. Die Testperson hätte

		die Aufgabe mit mehr Zeit fertig lösen können. Es werden 0.5 Punkte berechnet.
1	Wenn eine Aufgabe gelöst wird mit wenig Hilfe vom Testleitung (Hilfestellung 1: Motivieren)	Die Testperson kann die Aufgabe mit einer motivierenden Frage der Testleitung selbstständig lösen. Es wird 1 Punkte berechnet.
1	Eine Aufgabe wird selbständig und ohne Hilfe gelöst	Die Testperson kann zeitlich und inhaltlich die Aufgabe selbstständig lösen.

2. Anzahl Usability-Probleme

Richtlinien für Bestimmung Anzahl Usability-Probleme

Wertung	Fall	Beispiele
0	Wenn das gleiche Usability-Problem mehrmals gefunden wird, bekommt es das erste Mal einen Punkt, bei Wiederholung wird es nicht mehr gezählt.	Das Usability-Problem „Zeiteingabe“ wird mehrmals erwähnt als Problem, nervend etc.
0	Ausrufe und Fluchen von Testpersonen ohne eine klare Begründung des Ärgernisses werden nicht als Usability-Probleme gewertet.	„Mann“, „So einen Seich“, „das nervt“, „Hä“,
0.5	Etwas wird erst nach langem Suchen gefunden oder es wird erst beim wiederholten Mal wahrgenommen.	Der Button „zu den Angeboten“ wird erst spät gesehen.
0.5	Beim Anmeldeformular wird eine Angabe zunächst am falschen Ort eingegeben, es dann aber sogleich bemerkt und richtiggestellt wird	Bei der Formulareingabe wird beim Feld Postfach die Postleitzahl der Stadt eingegeben.
1	Ein bedeutendes Usability-Problem (ein immer wiederkehrendes Problem wird aufgedeckt)	Auf dem Anmeldeformular wird „Rechnung an eine andere Adresse“ auf der 1.Seite erwartet. Da diese Option erst auf der 2.Seite erscheint, brechen viele Testpersonen die Suche ab.
1	Ein mittleres Usability-Problem (ein oft wiederkehrendes Problem wird aufgedeckt)	Bei der Suche müssen zweimal Ort und Kurs eingegeben werden. Dies ist unnötig und macht die Testpersonen ungehalten.
1	Ein Vorschlag der Testperson (z.B. „ich hätte das an diesem Ort erwartet ..., ich würde eher das erwarten..., ich hätte dies lieber so...“)	Die Testperson erwartet bei Filter Tageszeit, dass diese aktiv werden beim Anklicken.

Qualitative Auswertung Testerleben beim klassischen Usability-Test

Dimension	Indikator	Item	Zusammenfassung der Antworten der Testpersonen
Testerleben	Allgemeines Befinden	Wie haben Sie den Test erlebt?	<p>Testperson hat es etwas speziell erlebt, aber ok. Hat durchs Mikrofon gehört, wie sich Testleiterin Notizen machte, hat sich ernst genommen gefühlt.</p> <p>Testperson hat es angenehm gefunden, dass Testleiterin ab und zu kommuniziert hat (mehrere Nennungen).</p> <p>Hat den Test als spannend empfunden (mehrere Nennungen) und die Aufgaben als gut und kurz, nicht zu leicht und nicht zu schwierig (mehrere Nennungen). Hat die Hilfestellung der Testleiterin als gut empfunden. Fand es gut, dass nach dem Test noch individuelle Meinung zur Website gefragt wurde. Empfand die Dauer des Tests als angenehm, genau richtig. (mehrere Nennungen).</p> <p>Testperson hat Test ein wenig merkwürdig erlebt.</p> <p>Hat Test es als angenehm empfunden. Fand es gute Instruktion der Testleiterin.</p> <p>Empfand den Test als gut vorbereitet.</p> <p>Hatte Mühe den Aufgaben zu folgen.</p> <p>Hat Test als spannend erlebt, wie ein „Informatikkurs“</p> <p>Hat sich im Vorfeld über Usability-Test bei Google informiert. Empfindet es wichtig, dass man Usability-Tests macht und Produkte einfacher bedienbar macht (da Benutzer und Entwickler nicht auf derselben Ebene sind) (mehrere Nennungen).</p> <p>Bei Problemen hätte Testperson zu Hause abgebrochen, wenn es nicht für die Testleiterin gewesen wäre.</p> <p>Hat die Testleiterin als ruhig und angenehm empfunden.</p> <p>Hat manchmal nicht gewusst, wann die Aufgabe zu ende war.</p>
	Testsituation im Labor	Wie haben Sie die Testsituation im Labor erlebt?	<p>Die Situation war ok, da sie die Testleiterin zuerst kennengelernt hat und das Labor gesehen hat, dass sie es sich vorstellen kann (mehrere Nennungen). Das hat ihr die Angst genommen. War hilfreich, dass Testperson Testleiterin bereits kannte (Vertrauen).</p> <p>Es war für Testperson ok über die Gegensprechanlage zu sprechen.</p> <p>Hat Laboreinrichtung als angenehm/ok erlebt (mehrere Nennungen).</p>
	Erleben des Aufzeichnens mit der Kamera	Wie war es für Sie, dass der Test und ihr Umgang mit dem System von einer Kamera aufgezeichnet wurden?	<p>Kein Problem, Kamera hat nicht gestört (mehrere Nennungen).</p> <p>Empfand es als komisch, dass sie gefilmt und beobachtet wurde.</p> <p>Testperson fand es gut, dass die Kamera auf der Seite war, aber noch besser wäre es, wenn sie gar nicht zu sehen wäre.</p>
	Empfinden des lauten Denkens	Wie haben Sie das laute Denken erlebt?	<p>Lautes Denken war speziell für Testperson.</p> <p>Lautes Denken war ok für Testperson (mehrere Nennungen).</p>

			<p>Hat es manchmal vergessen. Das Laute Denken war schwierig/anstrengend, hat viel Konzentration gebraucht, beides gleichzeitig zu machen. Aufgaben lösen und Gedanken dabei laut äussern, ungewohnt (mehrere Nennungen). Lautes Denken war ungewohnt (mehrere Nennungen). Lautes Denken war kein Problem, ist es sich gewöhnt von Telefonkonferenz, denkt manchmal auch laut wenn sie alleine vor PC ist (mehrere Nennungen).</p>
	Testangst	Wie haben Sie den Umstand empfunden, dass Sie Aufgaben lösen mussten unter Beobachtung?	<p>Testperson war zu Beginn nervös, nach einiger Zeit ging es. Nicht gestört, war ein wenig wie beim Film. Für eine Testperson war es anspruchsvoll, man möchte es ja gut machen. Hat es als Test empfunden, wenn sie es schlecht gemacht hätte, wäre es peinlich gewesen. Testperson fühlt sich nicht so sicher am PC, ist nervös, am Anfang leicht gestresst. Hat den Test als locker erlebt (mehrere Nennungen). Dass die Testperson nicht alle Aufgaben schaffen musste, hat Angst genommen. Testperson hat nicht alle Fragen genau gelesen, wollte vorwärtsmachen. Fand dass er als Testperson zu langsam war, wenig schnell vorwärtskam bei en Aufgaben. Testperson hat gehofft, dass sie nichts falsch gemacht hat, dass sie alles richtig gemacht hat. Es ist trotzdem eine Testsituation und sie hat Angst, sich blosszustellen Wenn die Testperson die Website nicht schon gekannt hätte, wäre sie vielleicht gestresst gewesen.</p>
	Abwesenheit der Testleitung	Wie haben Sie es erlebt, dass sich die Testleiterin während des Tests in einem anderen Raum aufgehalten hat?	<p>Kein Problem, da es zuerst erklärt und gezeigt wurde.(mehrere Nennungen). Hat es als angenehm erlebt, Situation war wie im Büro oder zu Hause. Testperson empfand es als gut, dass sie die Testleiterin gekannt hat, sonst wäre es evtl. komischer. Alleine am PC war kein Problem. War für Testperson kein Problem (mehrere Nennungen). Testperson hätte es lieber gehabt, wenn die Testleiterin nebendran gesessen hätte. Abwesenheit der Testleiterin war komisch, aber nicht schlimm.</p>
	Erneute Teilnahme an einem Usability-Test	Würden Sie in Zukunft wieder einmal an einem Usability-Test mitmachen?	<p>Ja(mehrere Nennungen).</p>

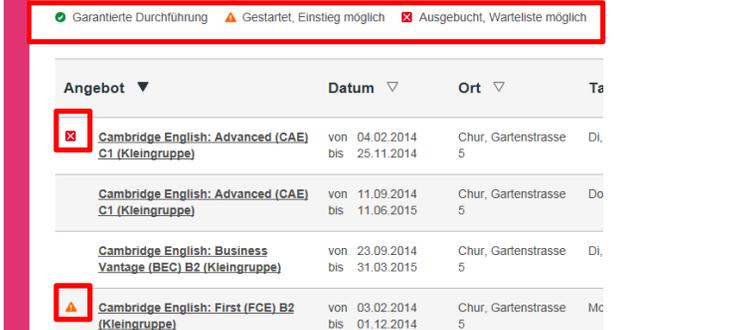
Qualitative Auswertung Testerleben beim Usability-Walkthrough

Dimension	Indikator	Item	Zusammenfassung der Antworten der Testpersonen
Testerleben	Allgemeines Befinden	Wie haben Sie den Test erlebt?	<p>Hat den Test als nicht schwierig und sehr angenehm erlebt (mehrere Nennungen).</p> <p>Hat den Test als spannend erlebt (mehrere Nennungen).</p> <p>Findet es wichtig, dass man Usability macht (mehrere Nennungen).</p> <p>Hat den Test als spannend erlebt (mehrere Nennungen).</p> <p>Hat den Test als interessant und toll empfunden (mehrere Nennungen).</p> <p>Hat manchmal nicht gewusst, wann die Aufgabe zu ende war.</p> <p>Hat etwas bei dem Test gelernt.</p> <p>Das Ziel der Testperson war alles zu geben und sich Mühe zu geben.</p> <p>Testperson hat sich wohl gefühlt.</p> <p>Der Test war neu, speziell für die Person.</p> <p>Testperson fand es schwierig, sich in die Aufgaben hineinzusetzen, es waren ja nicht ihre Ziele, sondern künstliche Aufgaben.</p> <p>Testperson hat tlw. die Aufgaben nicht genau gelesen.</p> <p>Empfand die Aufgaben als gut und verständlich.</p> <p>Die Instruktionen waren klar verständlich.</p>
	Testsituation im Labor	Wie haben Sie die Testsituation im Labor erlebt?	<p>Hat den Testraum als kahl empfunden, achtet auf den Arbeitsplatz.</p> <p>Hat den Testraum als ok empfunden.</p>
	Erleben des Aufzeichnens mit der Kamera	Wie war es für Sie, dass der Test und ihr Umgang mit dem System von einer Kamera aufgezeichnet wurden?	<p>Hat nicht gestört (mehrere Nennungen).</p>
	Empfinden des lauten Denkens	Wie haben Sie das laute Denken erlebt?	<p>Hat zu Beginn Lautes Denken als etwas unangenehm erlebt, hat nicht gewusst, wann und was sie sagen soll.</p> <p>Empfand das laute Denken als schwierig (aber sinnvoll). Wusste nicht genau, was man sagen soll (mehrere Nennungen).</p> <p>Lautes Denken war ok (mehrere Nennungen).</p> <p>Testperson hat lautes Denken als unüblich empfunden.</p> <p>Hat lautes Denken manchmal vergessen, weil Testperson konzentriert auf die Aufgaben war.</p> <p>Lautes Denken war für die Testperson schwieriger, wenn sie sich auf vieles konzentrieren musste.</p>
	Testangst	Wie haben Sie den Umstand empfunden, dass Sie Aufgaben lösen mussten unter Beobachtung?	<p>Hat sich nicht gestresst gefühlt (mehrere Nennungen).</p> <p>Hat sich entspannt gefühlt, die Angst wurde ihr genommen und sie konnte am PC ausprobieren.</p> <p>Versagensangst wurde der Testperson zu Beginn genommen mit In-</p>

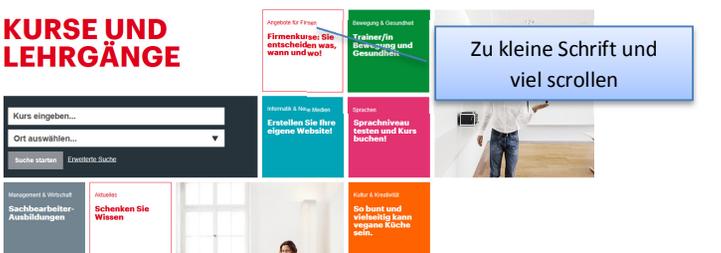
Anhang

			<p>struktion. Hat es nicht als „Test“ empfunden. Man gibt sich Mühe, man will fertig werden mit den Aufgaben. Der Testperson hat es geholfen, dass ihr mitgeteilt wurde, dass nicht sie, sondern das System getestet wurde. Testperson fühlte sich nicht unter Druck.</p>
	Anwesenheit der Testleitung	Wie haben Sie es erlebt, dass die Testleiterin während des Tests neben Ihnen gesessen hat?	<p>Testperson fand es angenehm mit der Testleiterin direkt sprechen zu können. Hat die Anwesenheit der Testleiterin als angenehm empfunden, es hat nicht gestört (mehrere Nennungen). Hat durch die Anwesenheit der Testleiterin den Test eher als Gespräch empfunden. Hätte erwartet, dass weniger gesprochen wird. Hat die Anwesenheit der Testleiterin als angenehm und natürlich empfunden. Testperson wäre konzentrierter gewesen, wenn die Testleiterin in einem anderen Raum gewesen wäre (Störung durch Gespräch, Einfluss der Testleiterin). Der Testperson war es wenig unangenehm, dass die Testleiterin beim Ausfüllen des Nachfragebogens danebensass und sehen konnte, welche Kreuze sie setzte (mehrere Nennungen). (Wenn Testleiterin nicht bekannt wäre, würde es die Testperson als unangenehm empfinden). Hat die Anwesenheit der Testleiterin nicht komisch empfunden (mehrere Nennungen) Die Testperson meint, dass die Anwesenheit der Testleiterin Einfluss hat, beeinflusst mit nonverbalen Signalen, objektiv zu sein ist wichtig.</p>
	Erneute Teilnahme an einem Usability-Test	Würden Sie in Zukunft wieder einmal an einem Usability-Test mitmachen?	<p>Ja, würde wieder mitmachen (mehrere Nennungen).</p>

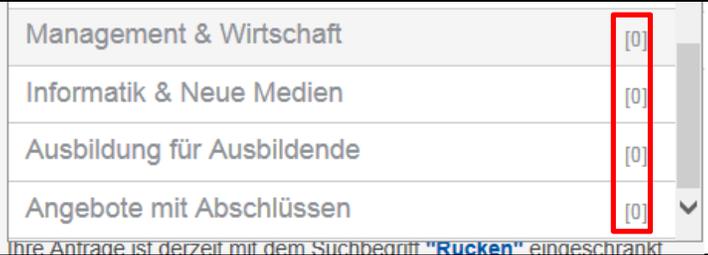
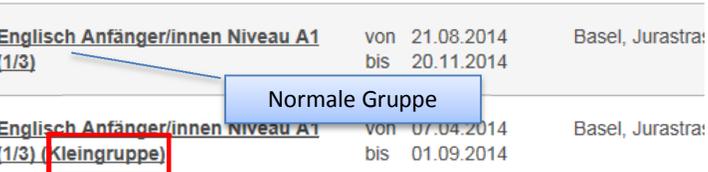
Qualitative Auswertung Art der Usability-Probleme

Art des Problems	Beispiel dazu: Screenshot	Zuordnung zur Kategorie	Verletzung der Kriterien der Dialoggestaltung	Problem gefunden bei Test A oder B
<p>Design und Layout Probleme</p> <p>Nachdem nach einem Kurs im Suchfeld gesucht wird, startet das System die Suche und auf der gleichen Ansicht steht unten im Blickfeld der Nutzer ein kleines Feld mit dem Hinweis „zu den Angeboten“ Dieses Feld wird aber von den Nutzern nicht wahrgenommen und die Kursangebote am unteren Ende der Webseite werden nicht gesehen, denn der Nutzer müsste dazu scrollen. Da der Nutzer über den Suchprozess nicht informiert wird, erkennt er auch nicht, wann die Suche abgeschlossen und die Angebote dargestellt werden.</p>		<p>Darstellungsproblem/Layout/Scrollen</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Da die Webpage bei der Darstellung der Kursangebote sehr viel scrollen verlangt, wird die Legende zu den Icons bei den Kursangeboten leicht übersehen. Einige Testpersonen haben beim langen scrollen den gesuchten Kurs übersehen und somit nicht gefunden. Dies frustriert die Nutzer.</p>		<p>Darstellungsproblem/zu viel scrollen</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Viele Testpersonen haben nicht gemerkt, dass man mit scrollen bei der Kursauswahl noch weitere Möglichkeiten zum Auswählen hat.</p>		<p>Scrollen ist nicht ersichtlich</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>

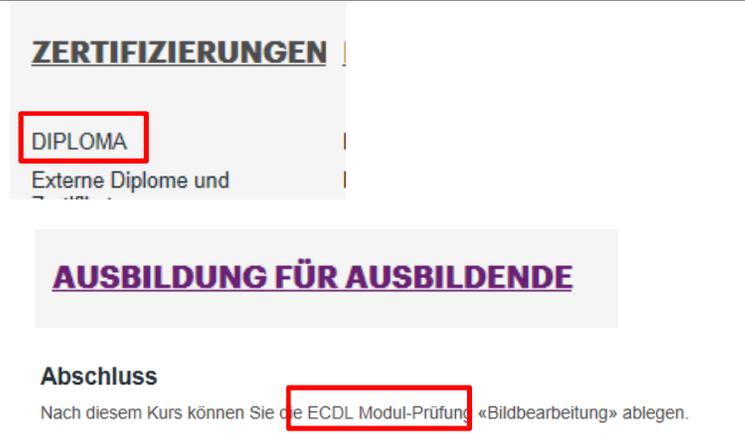
Anhang

<p>Viele Testpersonen haben den Sinn der Fotos nicht gesehen. Was soll dieses Bild für die Nutzer aussagen? Sie können nichts damit anfangen. Denn das Bild lässt sich auch nicht anklicken.</p>		<p>Sinnlose Bilder</p>	<p>Aufgabenange-messenheit</p>	<p>Test A und B</p>
<p>Einige Testpersonen fanden die Schrift auf der Startseite tlw. zu klein und es ist zu viel scrollen nötig. Die Nutzer wollen lieber die Startseite reduziert haben und mit einem Blick erfassen können.</p>		<p>Layout und Darstellung</p>	<p>Aufgabenange-messenheit</p>	<p>Test A und B</p>
<p>Die Pfeile neben der Legende zum Filtern wurden häufig von den Testpersonen nicht wahrgenommen.</p>		<p>Schlecht erkennbarer Hinweis</p>	<p>Selbstbeschrei-bungsfähigkeit</p>	<p>Test A und B</p>
<p>Obwohl die Filterfunktion „Angebot mit Abschlüssen“ bei einigen Testaufgaben hilfreich wäre, haben die Testpersonen den Filter nicht, oder erst sehr spät gesehen. Diese Funktion ist für die Nutzer nicht gut ersichtlich.</p>		<p>Filter schlecht erkennbar</p>	<p>Aufgabenange-messenheit</p>	<p>Test A und B</p>

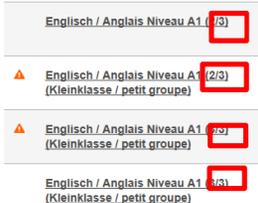
Anhang

<p>Die Testpersonen haben das kleine Null bei der Kurssuche und -auswahl häufig nicht gesehen und haben trotzdem diese Rubrik ausgewählt und weitere Rubriken, die aber keine Angebote hatten. Zum Schluss haben die Testpersonen keine Kurse gefunden. Die Null wird von den Nutzern schlecht gesehen und erkannt.</p>		<p>Schlecht erkennbare Hinweise</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Die Anzahl Lektionen sind nicht gut sichtbar und stehen unter dem Titel Preise. Einige Testpersonen haben die Anzahl Lektionen deshalb nicht gesehen.</p>		<p>Schlecht erkennbarer Hinweis</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Viele Testpersonen haben „Angebote mit Abschlüssen nicht gesehen“. Es ist für die Nutzer nicht gut ersichtlich, da es sehr klein und am unteren linken Ende im Blickfeld der Nutzer ist.</p>		<p>Schlecht erkennbarer Hinweis</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Bei einigen Kursen gibt es Kleingruppen und normale Gruppengrößen. Wie viele Personen dies sind, steht aber bei der Übersicht nicht. Dies erscheint erst bei den genaueren Informationen. Zudem ist die Teilnehmeranzahl nicht auf den ersten Blick ersichtlich.</p>		<p>Information nicht auf den ersten Blick ersichtlich</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A</p>

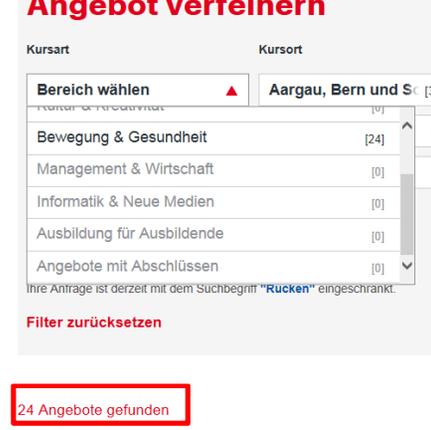
	<p>← Zurück zur Angebotsübersicht</p> <h3>Englisch Anfänger/innen Niveau A1 (1/3) (Kleingruppe)</h3> <p>Beschreibung Englischkurs für Anfänger/innen: Der leichte Einstieg in die englische Sprache. Sie erwerben Grundkenntnisse und trainieren in erster Linie die mündliche Kommunikation. So lernen Sie mit motivierenden Erfolgserlebnissen, sich im englischsprachigen Alltag zurechtzufinden.</p> <p>Inhalt ← Zurück zur Angebotsübersicht</p> <h3>Englisch Anfänger/innen Niveau A1 (1/3)</h3> <p>Beschreibung Englischkurs für Anfänger/innen: Der leichte Einstieg in die englische Sprache. Sie erwerben Grundkenntnisse und trainieren in erster Linie die mündliche</p> <div style="border: 1px solid gray; padding: 5px; margin-top: 10px;"> <p>Übersicht</p> <p>Kurs/Lehrgang Englisch Anfänger/innen Niveau A1 (1/3) (Kleingruppe)</p> <p>Kursnummer E 787872</p> <p>Max. Teilnehmer 8</p> <p>Dauer 14 Kurstag(e) Total 28 Lektion(en)</p> </div> <div style="border: 1px solid gray; padding: 5px; margin-top: 10px;"> <p>Übersicht</p> <p>Kurs/Lehrgang Englis Anfan Nivea</p> <p>Kursnummer E 797</p> <p>Max. Teilnehmer 15</p> </div>			
<p>Probleme der Navigation und Informationsarchitektur</p>				
<p>Wenn die Testpersonen einen bestimmten Kurs gesucht haben, bspw. einen Rücken- trainingskurs wussten sie nicht, ob dieser Kurs zu den Fitness-, Sport-, Gymnastik-, oder Entspannungskursen gezählt wurde. Also mussten die Testpersonen eines nach dem anderen ausprobieren, was einige Zeit dauerte und als mühsam empfunden wurde. Das gleiche Problem bestand bei anderen Testaufgaben (die Beispiele werden nicht zusätzlich aufgeführt, da es sich um das gleiche Grundproblem handelt).</p>		<p>Komplizierte Navigation, Informationsarchitektur nicht verständlich</p>	<p>Aufgabenange- messenheit</p>	<p>Test A und B</p>
<p>Einige Testpersonen erwarten in der Haupt- navigation eine eigene Rubrik für Diplome. Jetzt sind sie unter „über uns“, „Zertifizie- rungen“, „Diploma“ zu finden.</p>		<p>Informationsarchitektur</p>	<p>Erwartungskon- formität</p>	<p>Test A und B</p>

<p>Die einen Testpersonen hätten die Diplome unter den Angeboten in der Hauptnavigation erwartet.</p>		<p>Informationsarchitektur</p>	<p>Erwartungskonformität</p>	<p>Test A</p>
<p>Testperson erwartet bei über uns Rubrik „Klubschule als Arbeitgeber“ mit allen Informationen und Eigenmarketing (warum soll man bei der Klubschule arbeiten?).</p>		<p>Informationsarchitektur</p>	<p>Erwartungskonformität</p>	<p>Test B</p>
<p>Probleme der Verständlichkeit des Systems</p>				
<p>Einige Testpersonen haben den Begriff „DIPLOMA“ nicht genau verstanden. Die Nutzer würden eher „Diplome“ verstehen. Der Begriff ECDL war nicht für alle Testpersonen verständlich. Der Begriff „Ausbildung für Auszubildende“ war ebenfalls nicht für alle Testpersonen verständlich. Die Testpersonen würden eher den Begriff „Weiterbildung für Kursleitende“ erwarten.</p>		<p>Wording</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A</p>
<p>Neben den Diploma gibt es Kurs bei denen die Teilnehmer ein „Attest der Klubschule“ erhalten. Den Begriff und die Bedeutung war nicht allen Testpersonen klar verständlich. Der Begriff „Sommerkurs“ war nicht allen Testpersonen klar.</p>		<p>Wording</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test B</p>

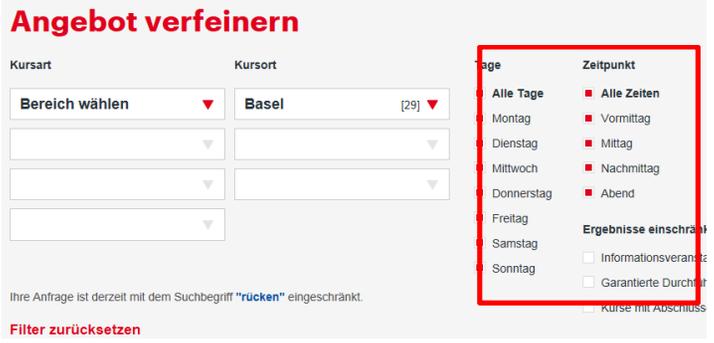
Anhang

<p>Für die Testpersonen war nicht verständlich, was (1/3, 2/3 oder 3/3) bedeutet. Es ist auch nirgendwo beschrieben.</p>		<p>Keine Erklärung</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A</p>
<p>Für viele Testpersonen ist das System nicht transparent. Sie wussten nicht, ob die Eingaben gespeichert werden, ob sie wieder alles neu eingeben müssen, wenn sie einen Schritt zurückgehen. Das System ist für die Nutzer nicht transparent genug.</p>	<p>Kein Screenshot</p>	<p>Zuwenig Transparenz des Systems</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Probleme der Fehlertoleranz</p>				
<p>Bei verschiedenen Suchaktionen und Ausprobieren der Testpersonen fielen die bereits getätigten Eingaben wieder raus. Dies verärgerte die Testpersonen, da die Suche wieder von vorn getätigt werden musste.</p>	<p>Keinen Screenshot</p>	<p>Geringe Fehlerrobustheit</p>	<p>Fehlertoleranz</p>	<p>Test A und B</p>
<p>Obwohl bei der Hauptsuche der Kurs und Ort eingegeben wurde, muss beim zweiten Schritt der Suche zunächst die Region und anschliessend den Standort gewählt werden. Bsp. Region Aargau, Bern und Solothurn, Standort Olten. Wenn nun aber die Testperson nicht wusste, in welcher Region Olten liegt, konnte der Standort nicht gefunden werden. Das gleiche Problem hatten viele Testpersonen bei Winterthur (Region Zürich oder Ostschweiz?) oder einige Testpersonen wussten nicht, dass Zug in der Zentralschweiz liegt. Hier werden geografische Kenntnisse gefordert, die aber unter Umständen nicht alle Nutzer mitbringen. Die Nutzer wünschten sich zur Unterstützung bspw. eine Landkarte.</p>		<p>Erfordert geografische Kenntnisse über die Schweiz</p>	<p>Lernförderlichkeit</p>	<p>Test A und B</p>

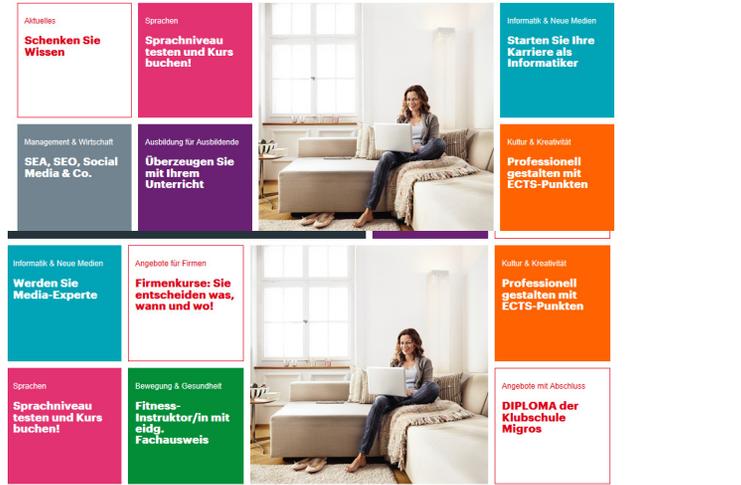
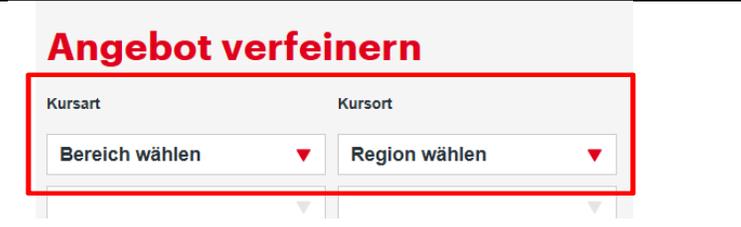
Anhang

<p>Der Kursname muss richtig so eingegeben werden, sonst findet ihn das System nicht. Bei Rechtschreibfehlern gibt es keinen Hinweis darauf. Das System macht keine Vorschläge (wie bspw. Google: „meinten Sie...“, „Vorschläge für...“). Die Fehlermeldung (Screenshot oben) bringt die Nutzer nicht weiter in ihrem Ziel.</p>		<p>Geringe Fehlerrobustheit</p>	<p>Fehlertoleranz</p>	<p>Test A und B</p>
<p>Probleme der Missachtung der Erwartungen und mentales System der Nutzer</p>				
<p>Einige Testpersonen klickten auf den Begriff „X Angebote gefunden“. Es passiert aber nichts. Hier wird die Erwartung der Nutzer, auf die gefunden Angebote zu klicken und so zu den Kursen zu gelangen nicht unterstützt. Stattdessen müssen die Nutzer weiter hinunterscrollen.</p>		<p>Erwartung wird nicht unterstützt</p>	<p>Erwartungskonformität</p>	<p>Test A</p>
<p>Einige Testpersonen klickten auf die Mitte ins Feld, dann passierte nichts. Man muss auf den Pfeil rechts klicken.</p>		<p>Erwartung der Nutzer nicht unterstützt</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Die Nutzer sind sich durch Google gewöhnt Vorschläge bei der Eingabe zu erhalten (Auto-Suggest). Die Suche nach Kursen der Klubschule unterstützt die Nutzer nicht mit der Auto-Suggest Funktion, obwohl sich die Testpersonen dies gewünscht haben.</p>		<p>Keine Auto-Suggest Funktion</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>

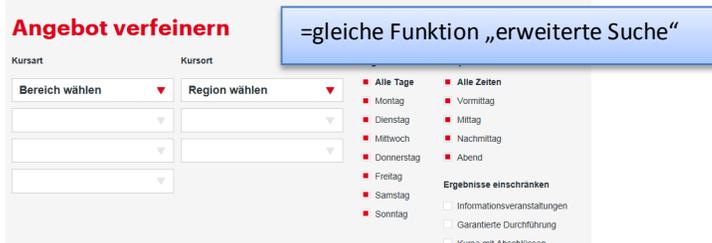
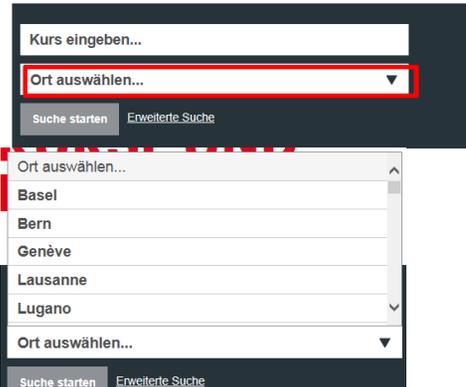
Anhang

<p>Einige Testpersonen möchten das Sprachniveau manuell eingeben können. Das ist aber nicht möglich. Zudem haben sie die Auswahl des Niveaus nicht wahrgenommen.</p>		<p>Manuelle Eingabe nicht möglich, Auswahl nicht wahrgenommen</p>	<p>Steuerbarkeit</p>	<p>Test A und B</p>
<p>Die Nutzer verwirrt es, dass die Dropdown-Auswahl der Ortschaften nicht alphabetisch geordnet ist, sondern zuoberst kommen zunächst fett gedruckt die grösseren Kursorte und unten folgen dann die übrigen Kursorte. Deshalb muss der Nutzer viel scrollen und länger suchen, als wenn es alphabetisch der Reihe nach geordnet wäre.</p>		<p>Gewohnheit der alphabetischen Darstellung</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Die Funktionen bei den Filtern für Tage und Zeitpunkte sind bereits alle rot markiert und somit aktiv. Die Testpersonen haben diejenigen Filter angeklickt, die sie auswählen wollten. Diese wurden aber durch das Anklicken deaktiviert. Viele Testpersonen haben dies nicht bemerkt und verstanden nicht, dass sie den Filter anstatt aktiviert, deaktiviert haben, da das System dazu auch keine Rückmeldung gab. Dieses System läuft gegen die Erwartung und Logik der Nutzer. Die meisten Testpersonen haben es genau anders erwartet.</p>		<p>Mentales System der Filteraktivierung wird missachtet</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Da sich die Testpersonen gewöhnt sind, über eine Suchfunktion zum Ziel zu gelangen, erwarten sie dies auch bei dieser Website. Eine Suche über die gesamte Website gibt es aber nicht.</p>	<p>Kein Screenshot</p>	<p>Suchfunktion</p>	<p>Erwartungskonformität</p>	<p>Test B</p>

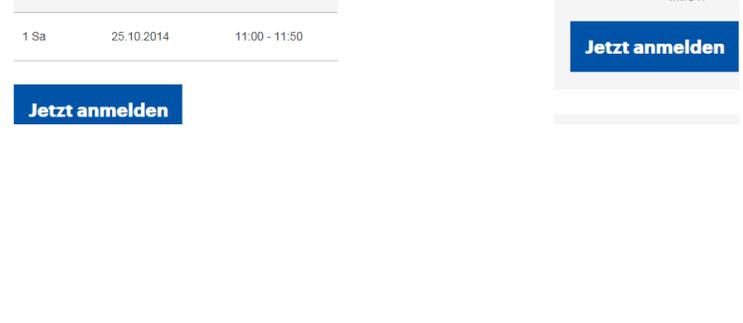
Anhang

Probleme der Inkonsistenz des Systems				
<p>Die Navigationshilfe „Home“ ist nicht auf allen Website von Migros Klubschule vorhanden. Die Testpersonen erwarten aber auf allen Seiten den Navigationspfad.</p>		<p>Fehlende Konsistenz</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Die Kacheln auf der Startseite der Migros Klubschule wechseln immer wieder den Platz. Aber die Testpersonen haben sich den Ort einer Kachel gemerkt, z.B. „Diploma“ und haben es auch wieder an diesem Platz erwartet. Als sie zurück auf die Startseite klickten, war die Kachel „Diploma“ aber an einer anderen Stelle, wodurch sie von den Testpersonen teilweise nicht mehr gefunden wurden. Dies verwirrte und verärgerte die Testpersonen.</p>		<p>Inkonsistenz auf der Startseite (Wechseln der Kacheln)</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
Probleme der Komplexität des Systems				
<p>Viele Testpersonen waren überfordert mit den vielen Angeboten. Es sind auch zu viele Angebote auf einer Seite angezeigt, dass scrollen erfordert. Wenn immer weiter Angebote geladen werden, wird die Seite sehr lange.</p>		<p>Komplexität</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>
Probleme der Suchfunktion				
<p>Nachdem bei der Hauptsuche der Ort und den Kurs eingegeben wurden, erscheint wiederum eine neue Suche (Angebote verfeinern) bei dem nochmals der Kurs und der Ort eingegeben werden muss. Dies verstimmt die Testpersonen. Die Nutzer wollen nicht zwei mal den gleichen Prozess durchlaufen, sie sehen darin keinen Sinn.</p>		<p>Über Umwege zum Ziel, keine Beeinflussbarkeit durch den Benutzer</p>	<p>Steuerbarkeit</p>	<p>Test A und B</p>

Anhang

<p>Es gibt keinen Filter für die Eingabe eines Budgets (z.B. Preis bis 500 CHF). Dies wurde jedoch von vielen Testpersonen gewünscht. So musste die Liste durchgesehen werden nach Kursen, die nicht zu teuer waren.</p>	<p>Keinen Screenshot</p>	<p>Fehlender Filter für Preisbeschränkung</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>
<p>Angebote mit Abschlüssen können einerseits als Filter angewählt werden (Screenshot links), andererseits können sie beim Dropdownmenu (Screenshot rechts) ausgewählt werden. Dies verwirrte die Testpersonen teilweise.</p>		<p>Mehrere gleiche Funktionen</p>	<p>Aufgabenangemessenheit</p>	<p>Test A</p>
<p>Dieser Kurs kostet 0 CHF. Dies verwirrte die Testpersonen, sie merkten tlw. nicht, dass es sich nur um eine Informationsveranstaltung handelt.</p>		<p>Sinnloses Angebot</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>
<p>Für einige Testpersonen waren es zu viele Optionen bei der erweiterten Suche. Zudem ist die Funktion „erweitere Suche“ genau die gleiche wie „Angebote verfeinern“. Dies machte aber keinen Sinn für die Testpersonen.</p>		<p>Zwei mal die gleiche Funktion</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>
<p>Die Nutzer möchten bei der Hautsuche auf der Startseite den Ort lieber manuell eingeben. Dies ist aber nicht möglich, es ist eine Dropdown Funktion.</p>		<p>Bevorzugen der manuellen Eingab</p>	<p>Steuerbarkeit</p>	<p>Test A und B</p>

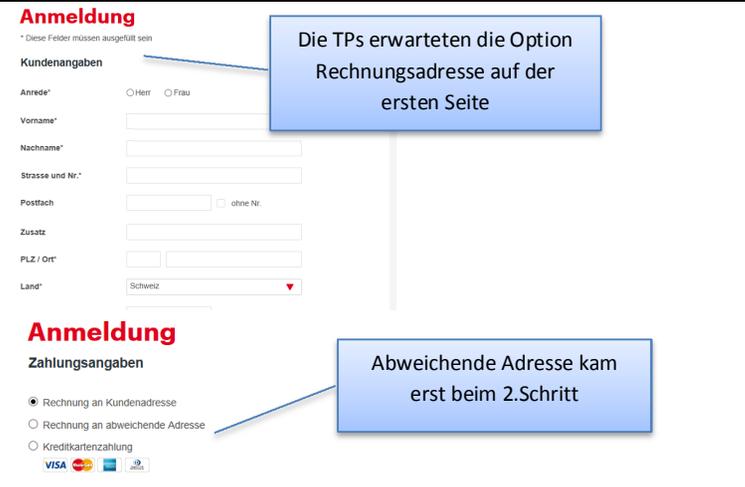
Anhang

<p>Die Nutzer erwarten, dass man beim Draufklicken auf eine Kachel (z.B. Diploma) direkt zur Webpage mit den Informationen über Diplome weitergeleitet wird. Stattdessen wendet sich die Kachel auf die Rückseite (Screenshot rechts). Die Nutzer müssen dann erneut auf den Begriff „DIPLOMA“ klicken, um zu den Informationen zu gelangen. Dies stellt für die Nutzer einen unnötigen Zwischenschritt dar.</p>		<p>Unnötiger Zwischenschritt</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Für einige Testpersonen war es nicht klar verständlich, welchen Preis gezahlt werden muss bei einem Kurs, der bereits gestartet ist. Es ist nicht klar, ob der angegebene Preis der reduzierte, oder der Gesamtpreis ist.</p>		<p>Ungenügende Beschreibung</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test B</p>
<p>Einige Testpersonen erwarteten bei der erweiterten Suche noch mehr Filtermöglichkeiten wie bspw. Zielgruppe, Niveau, Preis.</p>	<p>Kein Screenshot</p>	<p>Mehr Filtermöglichkeiten</p>	<p>Individualisierbarkeit</p>	<p>Test B</p>
<p>Testperson möchte gerne mehrere Sprachniveaus als Filter auswählen können. Dies ist aber nicht möglich.</p>		<p>System erlaubt kein Spielraum</p>	<p>Individualisierbarkeit</p>	<p>Test B</p>
<p>Der Anmeldebutton ist zwei mal vorhanden, einmal unten links und rechts nochmals. Einige Testpersonen fanden dies überflüssig.</p>		<p>Mehrere gleiche Funktionen</p>	<p>Aufgabenangemessenheit</p>	<p>Test B</p>

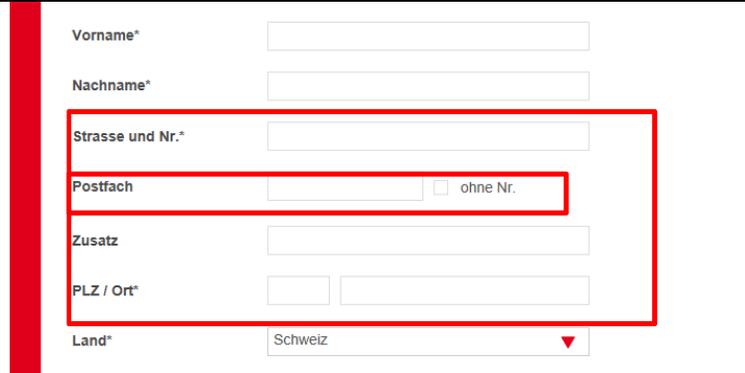
Onlineformular zur Kursanmeldung				
<p>Testperson erwartet zur besseren Erkennung bei der Angabe der Korrespondenz Icons als Hilfestellung.</p>	<p>Korrespondenz</p> <p>Wie möchten Sie die Korrespondenz der Klubschule Migros erhalten? *</p> <p><input type="radio"/> Ich wünsche sämtliche Korrespondenzen inkl. Rechnungen für E-Banking per Mailversand an die angegebene E-Mail-Adresse.</p> <p><input type="radio"/> Ich wünsche Briefe per Mailversand an die angegebene E-Mail-Adresse aber Rechnungen per Post.</p> <p><u>Vorteile der Korrespondenzen per Mailversand</u></p> <p><input type="radio"/> Ich wünsche sämtliche Korrespondenzen per Postversand</p>  	<p>Einsatz von Icons</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test B</p>
<p>Wenn die AGB's angeklickt werden, wird das Kästchen rot. Für einige Testpersonen war dies nicht gut ersichtlich, da sie finden es sei zu wenig kontrastreich.</p>	<p>Zahlung des Kursgeldes*</p> <p>Nachdem Sie sich für einen Kurs angemeldet haben, erhalten Sie von uns einen Teilnehmer-Ausweis, der Ihnen gleichzeitig als Einzahlungsschein dient. Der dort aufgeführte Zahlungstermin ist verbindlich. Die Anmeldung verpflichtet Sie zur Zahlung des Kursgeldes. Das Nichtbezahlen des Kursgeldes gilt nicht als Abmeldung. Die Bezahlung ist im Klubschulsekretariat und zum Zeitpunkt der Anmeldung über unsere Webseite (nur Kreditkarte) möglich.</p> <p>AGBs als PDF downloaden</p> <p><input checked="" type="checkbox"/> Ja, ich habe die AGB gelesen und bin damit einverstanden.</p>	<p>Schlecht erkennbarer Hinweis</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test B</p>
<p>Die Testpersonen erwarten ein komplett neues Eingabefeld für die Firmenadresse. Jetzt sind die persönlichen Daten bereits eingetragen und müssen gelöscht werden, dass bspw. die Firmenadresse eingeben werden kann. Dies empfinden die Nutzer als mühsam und sinnlos.</p>	<p>Rechnungsadresse</p> <p>Firma <input type="text"/></p> <p>Anrede * <input checked="" type="radio"/> Herr <input type="radio"/> Frau</p> <p>Vorname* <input type="text" value="Sabine"/></p> <p>Nachname* <input type="text" value="Muster"/></p> <p>Strasse und Nr.* <input type="text" value="Musterstrasse 12"/></p> <p>Postfach <input type="text"/> <input type="checkbox"/> ohne Nr.</p> <p>PLZ* <input type="text" value="8215"/></p> <p>Land* <input type="text" value="Schweiz"/></p> <p>Ort* <input type="text" value="dfsdf"/></p> <p>E-Mail* <input type="text" value="sabine.muster@gmx.ch"/></p> <p>AGB</p> <div data-bbox="961 789 1289 1016" style="border: 1px solid blue; background-color: #e6f2ff; padding: 5px;"> <p>Die Felder sind bereits ausgefüllt und müssen zuerst gelöscht werden, um Firmenadresse hinzuschreiben.</p> </div>		<p>Aufgabenangemessenheit</p>	<p>Test B</p>
<p>Bevor die Anmeldung gesendet wird, möchten die Testpersonen die Informationen ausdrucken können. Dies ist aber noch nicht möglich.</p>	<p>Kein Screenshot</p>	<p>Ausdrucken der Kursanmeldung nicht möglich vor Absendung</p>	<p>Individualisierbarkeit</p>	<p>Test A</p>

Anhang

<p>Bei der Dateneingabe im Anmeldeformular verwirrte es die Testpersonen, dass sie bei Rechnung an abweichende Adresse nochmals einen Vor- und Nachnamen angeben mussten. Sie hatten dies bereits im ersten Schritt gemacht und möchten dies nicht.</p>	<p>Rechnungsadresse</p> <p>Firma <input type="text"/></p> <p>Anrede * <input checked="" type="radio"/> Herr <input type="radio"/> Frau</p> <p>Vorname* <input type="text"/></p> <p>Nachname* <input type="text"/></p> <p>Strasse und Nr.* <input type="text"/></p> <p>Postfach <input type="text"/> <input type="checkbox"/> ohne Nr.</p> <p>PLZ* <input type="text" value="8200"/></p> <p>Land* <input type="text" value="Schweiz"/></p>	<p>Unnötige Eingabe</p>	<p>Aufgabenangemessenheit</p>	<p>Test A</p>
<p>Bei der Kursanmeldung ist die Angabe einer Telefonnummer zwingend. Dies störte aber einige Testpersonen, da sie nicht gerne ihre Nummer herausgeben.</p>	<p>Telefon* Bitte geben Sie mindestens eine Telefonnummer an</p>	<p>Herausgabe sensible Daten</p>	<p>Aufgabenangemessenheit</p>	<p>Test A</p>
<p>Bei der Dateneingabe ins Anmeldeformular haben einige Testpersonen einen Schritt vergessen, da es ihrer Ansicht nach sehr viele Angaben sind.</p>	<p>Kein Screenshot</p>	<p>Komplexität</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>
<p>Viele Nutzer klicken in das Feld „Bitte geben Sie mindestens eine Telefonnummer an“. Dies ist aber kein Eingabefeld. Die Nutzer müssen die Nummer unten bei Privat, Mobil oder Geschäft eingeben.</p>	<p>Telefon* Bitte geben Sie mindestens eine Telefonnummer an</p> <p>Privat <input type="text"/></p> <p>Mobil <input type="text"/></p> <p>Geschäft <input type="text"/></p>	<p>Formular Dateneingabe</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Einige Testpersonen hatten erwartet, dass wenn sie die Postleitzahl eingeben, automatisch der Ort erscheint. Diese Auto-Funktion gibt es beim Anmeldeformular der Klubschule nicht.</p>	<p>PLZ / Ort* <input type="text" value="8200"/></p> <p>Keine automatische Eingabe des Ortes</p>	<p>Eingabe der Nutzer wird nicht unterstützt</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>

<p>Bei der Anmeldung fanden die Testpersonen die Option für Rechnung an eine andere Adresse senden nicht. Sie erwarteten dies auf der ersten Seite und hatten Hemmungen auf „weiter“ zu klicken. Die Rechnung schicken ist aber erst auf der folgenden Seite. Dies bereitete den Testpersonen Mühe. Vielen schreiben die Rechnungsadresse bei m Feld für Bemerkungen hinein. Einige Testpersonen haben den Button „Rechnung an abweichende Adresse“ sogar übersehen</p>		<p>Erwartung der Nutzer nicht beachtet</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Die Nutzer haben keine Möglichkeit mehrere Personen bei einem Kurs anzumelden, jede Anmeldung muss einzeln getätigt werden. Die Testpersonen haben erwartet, mehrere Personen anmelden zu können (bspw. bei einem Geschenk).</p>	<p>Kein Screenshot</p>	<p>System erlaubt kein Spielraum</p>	<p>Individualisierbarkeit</p>	<p>Test A und B</p>
<p>Einige Testpersonen wollten bei der Anmeldung im Fortschrittsbalken oben wieder zurückklicken. Dies ist aber nicht möglich. Für die Nutzer ist es aber wichtig, jederzeit in den Dialogprozess eingreifen zu können.</p>		<p>Keine Möglichkeit zur Steuerung des Anmeldeprozesses</p>	<p>Steuerbarkeit</p>	<p>Test A und B</p>
<p>Viele Testpersonen waren verwirrt, dass beim Formular Rechnung an abweichende Adresse nach der Postleitzahl das Land und dann der Ort eingegeben werden muss. Dies entspricht nicht dem mentalen Schema der Nutzer.</p>		<p>Missachtung mentales Schemata</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>

Anhang

<p>Die meisten Nutzer geben beim Postfach die Postleitzahl ein. Die Nutzer sind sich gewohnt, dass nach der Strasse die Postleitzahl und der Ort eingegeben werden muss. Dies ist so mental verankert. Hier wird der Erwartung und Gewohnheit der Nutzer nicht entsprochen.</p>		<p>Dateneingabe nicht nach der gewohnten mentalen Landkarte gestaltet</p>	<p>Erwartungskonformität</p>	<p>Test A und B</p>
<p>Bei der Kursanmeldung fehlt die Information über den Kursort und die Adresse.</p>		<p>Fehlende Informationen</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Die Testpersonen verstanden den Sinn von Postfach ohne Nummer und Zusatz nicht.</p>		<p>Unverständliche/sinnlose Formulardaten</p>	<p>Aufgabenangemessenheit</p>	<p>Test A und B</p>

Anhang

<p>Viele Testpersonen mussten sehr lange lesen bei der Auswahl der Korrespondenz. Für die Nutzer ist dieser Text zu lange und zu kompliziert.</p>	<p>Korrespondenz</p> <p>Wie möchten Sie die Korrespondenz der Klubschule Migros erhalten? *</p> <p><input type="radio"/> Ich wünsche sämtliche Korrespondenzen inkl. Rechnungen für E-Banking per Mailversand an die angegebene E-Mail-Adresse.</p> <p><input type="radio"/> Ich wünsche Briefe per Mailversand an die angegebene E-Mail-Adresse aber Rechnungen per Post.</p> <p><u>Vorteile der Korrespondenzen per Mailversand</u></p> <p><input type="radio"/> Ich wünsche sämtliche Korrespondenzen per Postversand</p> <p>Weiter</p>	<p>Umständliche Formulierung</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A und B</p>
<p>Bei der Kursanmeldung wird nicht informiert, wann und wie weitere Informationen zum Kurs folgen.</p>	<p>Kein Screenshot</p>	<p>Fehlende Informationen</p>	<p>Selbstbeschreibungsfähigkeit</p>	<p>Test A</p>

Anhang F

Merkplatt für Testleitung

Aus den Erkenntnissen der vorliegenden Arbeit wurden Schlüsse für den Einsatz der Usability-Tests in der Praxis gezogen. Diese werden im Folgenden vorgestellt.

1. Wichtige Regeln bei der Vorbereitung

- Was ist das Ziel?
- Welches Produkt soll überprüft werden?
- Wer ist die Benutzergruppe?
- Es ist sehr wichtig, die „richtigen“ Testpersonen einzuladen, aus der potentiellen Benutzergruppe. Alter, Geschlecht, Soziodemografie, Persönlichkeitsprofil und Nutzerverhalten soll die spätere Benutzergruppe repräsentieren.

2. Entscheidung, welche Methode eingesetzt wird

Wann soll der klassische Usability-Test eingesetzt werden?

Wenn ein System sich bereits schon im Endstadium befindet und der Kunde erfahren möchte, wie potentielle Nutzer damit umgehen, eignet sich der klassische Usability-Test eher. Die Testpersonen sind auf sich gestellt und lösen die Aufgaben selbständig. Es gibt weniger Gespräche und Inputs. Die Testleitung hat weniger Einfluss auf das Resultat.

Wann soll der Usability-Walkthrough eingesetzt werden?

Zunächst muss entschieden werden, ob der Usability-Walkthrough oder der klassische Usability-Test geeigneter ist, um das System zu prüfen. Bei noch nicht vollständig ausgereiften Prototypen bietet sich der Usability-Walkthrough an. Die Testleitung hat die Möglichkeit den Prozess mit den Testpersonen durchzulaufen, bei Schwierigkeiten sofort zu reagieren, das Modell evtl. umzubauen, anderer Fragestellungen hinzuzunehmen und die Meinung der Testpersonen abzuholen. Änderungen am System und Prototyp sind in dieser Phase einfacher, als bei einem ausgereiften System.

3. Was muss beim klassischen Usability-Test und beim Usability-Walkthrough beachtet werden?

Was muss beim klassischen Test beachtet werden?

Beim klassischen Usability-Test sitzt die Testleitung nicht im gleichen Raum, wie die Testperson, und beobachtet diese durch eine Kamera oder einen Einwegspiegel. Die Untersuchung ergab, dass die Testpersonen den klassischen Usability-Test mehr als Test erleben, als den Usability-Walkthrough. Es ist schwieriger Vertrauen aufzubauen, wenn kein persönlicher Face-to-Face-Kontakt besteht. Testangst tritt eher auf, als beim Usability-Walkthrough. Demnach ist es wichtig, dass sich die Testleitung vor dem Test Zeit nimmt, die Testperson begrüsst, transparent über den ganzen Testdurchführung und den Ablauf informiert und sicherstellt, dass sich die Testperson wohl fühlt. Es soll mit Nachdruck betont werden, dass nicht die Fähigkeiten der Person, sondern das System getestet werden soll. Es ist auch wichtig, dass die Testperson ab zu Rückmeldung oder Bestätigung von der Testleitung erhält in Form von „mmmh“ etc. Ebenfalls vorteilhaft ist es, wenn die Testperson weiss, dass Fragen zur Klärung der Testaufgaben gestellt werden dürfen. Für die Testperson ist es wichtig zu wissen, dass die Testleitung da ist, auch wenn sie nicht zu sehen ist. Für die Nachbesprechung sollte die Testleitung wieder anwesend sein, um so die Meinung der Testperson ernst zu nehmen. Im persönlichen Kontakt fällt es möglicherweise leichter, seine Ansicht zu vertreten.

Was muss beim Usability-Walkthrough beachtet werden?

Beim Usability-Walkthrough sitzt die Testleitung neben der Testperson. Die Testperson empfindet dies als natürliche Situation und weniger als Test. Allerdings könnte es einigen Testpersonen peinlich sein, wenn die Testleitung beim Ausfüllen des Fragebogens nebenan sitzt und somit sieht, was die Testperson ankreuzt. Es emp-

zieht sich deshalb während der Zeit des Ausfüllens ein wenig Abstand zu halten oder evtl. aufzustehen und im Raum etwas vorzubereiten. Weiter muss sich die Testleitung im Klaren sein, dass sie unbewusst Einfluss auf die Testpersonen ausübt. Dies ist als sog. Testleitereffekt bekannt. Als Anhaltspunkt unterstützen die goldenen Regeln für die Moderation (vgl. Kapitel 3.5.1). Die Testperson wird automatisch mehr sprechen mit der Testleitung, wenn sie sich im gleichen Raum aufhält. Häufig fragen die Testpersonen auch mehr um Hilfe. Die Testleitung muss sich dieser Herausforderungen bewusst sein, und sich so weit wie möglich an die Regeln halten.

4. Im stationären Labor oder mit dem mobilen Labor testen?

Was muss bei Tests im stationären Labor beachtet werden?

Beim stationären Labor ist zu beachten, dass die Testpersonen zuvor transparent aufgeklärt werden, die Einrichtung erklärt (Kameras etc.) wird und wie die Kommunikation mit der Testleitung abläuft. Denn der Einwegspiegel kann bei Testpersonen Skepsis oder Stress auslösen. Im stationären Labor können Störeinflüsse, wie Lärm oder Unterbrechungen ausgeschaltet werden.

Was muss bei Tests mit dem mobilen Labor beachtet werden?

Mit dem mobilen Labor können Tests überall durchgeführt werden und der aktuelle Kontext, indem das Produkt verwendet wird (z.B. Arbeitsplatz, Lichtverhältnisse etc.), kann beobachtet werden. Eine Herausforderung ist es beim mobilen Testen die Umgebung zu kontrollieren, bspw. Lärm und Störung durch Mitarbeitende, was sich negativ auf die Konzentration der Testpersonen, sowie auf die Validität der Daten auswirken kann. Deshalb sollte die Testleitung bei Tests mit dem mobilen Labor auf Störungen und Lärm vorbereitet sein und diese wenn möglich minimieren (z.B. den Test an einem ruhigen Ort durchführen, an die Türe ein Schild „Bitte nicht stören“ hängen, Telefon und Handy ausschalten etc.).

5. Zur Problematik des lauten Denkens

Bei beiden Testvarianten bekundeten einige Testpersonen Mühe mit dem lauten Denken. Es war vor allem für unerfahrene Testpersonen schwierig und es ging immer wieder vergessen. Häufig ist das laute Denken für die Testpersonen anstrengend und benötigt viel Konzentration, da sich die Personen einerseits auf die Aufgaben fokussieren und andererseits laut ihre Gedanken aussprechen mussten. Diese Doppelbelastung aus Aufgabenbearbeitung und dem Äussern der Gedankengänge kann zudem die Bearbeitungsgeschwindigkeit senken. Die Testleitung muss aufgrund des Testzieles entscheiden, ob die Methode des lauten Denkens oder eine Alternative dazu eingesetzt wird. Eine Alternative stellt die Variante des postaktionalen Lauten Denkens dar. Dabei wird zunächst das Verhalten der Testpersonen ohne Aufforderung zum Lauten Denken auf Video aufgenommen, erst bei der anschliessenden Betrachtung der Aufzeichnung werden die Testpersonen dazu aufgefordert, laut zu denken. Beide Methoden haben Vor- und Nachteile, die Testleitung muss aufgrund des Testzieles entscheiden, welche für ihren Test die Geeignetste ist.

6. Moderation und Hilfestellung

Für den Einsatz beider Testvarianten gelten die gleichen Regeln (vgl. goldene Moderationsregeln Kapitel 3.5.1). Hilfestellung anzubieten stellt eine Herausforderung dar. Hilfestellungen können in vier Schritten angeboten werden (vgl. Kapitel 3.5.2). Der Prozess der Hilfestellung ist bei beiden Testvarianten derselbe. Allerdings könnte die Hilfestellung, aufgrund der Zurückhaltung der Testleitung beim klassischen Usability-Test später zum Einsatz kommen als beim Usability-Walkthrough. Dieses Merkblatt soll als Hilfestellung für eine qualifizierte Durchführung von Usability-Tests in der Praxis dienen. Durch weiterführende Untersuchungen könnte diese noch differenziert werden.