

Einfluss von Tonalität bei Fehlermeldungen auf die User Experience

Welche Tonalität wird bei Fehlermeldungen eines Hotelbuchungsformulars erwartet und inwiefern lässt sich damit die User Experience positiv beeinflussen?



Autoren:

Dorota Przygienda

Theodor Huber

Manuel Sanchez

Betreuer:

Markus Stolze, Prof. Dr.

Co-Referent:

Marcel B.F. Uhr, Dr. sc. techn. ETH, CEO Soutank AG

Master of Advanced Studies in Human Computer Interaction Design

Universität Basel, Fakultät für Psychologie

Technische Hochschule Rapperswil (HSR)



Impressum

Autoren:

Dorota Przygienda

Theodor Huber

Manuel Sanchez

Betreuer:

Markus Stolze, Prof. Dr.

Co-Referent:

Marcel B.F. Uhr, Dr. sc. techn. ETH, CEO Soultank AG

Danksagung

An dieser Stelle möchten wir ganz herzlich unserem Coach Prof. Dr. Markus Stolze danken, der uns stets kompetent und geduldig beraten und mit viel Inspiration unterstützt hat. Des Weiteren möchten wir uns bei unseren Probanden, HCID-Kolleg(inn)en, Freunden und Bekannten bedanken, welche an unserer UX-Studie teilgenommen haben. Unseren Experten Marietta Haller, Hans Schmid, Wolfram Scherer und Tom Klemm möchten wir ebenfalls danken für ihre lehrreichen Hinweise und Unterstützung.

Stilistische Anmerkung

In der vorliegenden Arbeit gilt die männliche Form als geschlechtsneutral. Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung weiblicher und männlicher Schreibformen verzichtet.

Abstract

Wenn Tonalität in Fehlermeldungen wichtig wäre, welchen Teil der User Experience würde sie beeinflussen? Lässt sich eine negativ beeinflusste User Experience wieder «glätten» und wenn ja wie? Bedeutet eine Fehlermeldung in jedem Fall eine schlechte User Experience? Dies sind Kernfragen, die wir uns zu Beginn dieser Masterarbeit gestellt haben und die uns auch durch das gesamte Projekt begleitet haben. Das Thema Fehlermeldungen in Kombination mit User Experience ist ein Gebiet, welches noch wenig erforscht wurde. Die Menge an Fachliteratur zum Thema ist entsprechend klein, die beschreibt, wie Fehlermeldungen unsere Emotionen beeinflussen und wie sie sich generell auf die Mensch-Maschine Interaktion auswirken. Diese Ausgangslage gab uns die nötige Inspiration und Motivation, dieses Gebiet genauer zu untersuchen. In unserer Masterarbeit haben wir uns auf die Tonalität bei Fehlermeldungen konzentriert, da wir der Ansicht sind, dass diese einen grossen Einfluss auf die User Experience hat. Mithilfe unseres selbst entwickelten Online-Tests welcher die User Experience misst, konnten wir mit mehr als 100 Probanden ein Experiment durchführen und die Daten entsprechend auswerten. Dabei zeigte sich, dass grössere Unterschiede zwischen Mann und Frau bestehen, aber auch zwischen den Älteren und den Jüngeren. Bei der expliziten Beurteilung der Tonalität zeigt sich, dass die persönliche Einschätzung der einzelnen Tonalitäten überhaupt nicht übereinstimmt mit der Beurteilung der Anwendung unter Einfluss von spezifischer Tonalität. Z.B. haben Männer höfliche Sätze sehr hoch bewertet, hingegen war die User Experience unter Einfluss von höflichen Fehlermeldungen eher schlecht bewertet. Umgekehrt verhält es sich bei den Frauen mit unpersönlicher Tonalität (kurze Sätze ohne weitere Information zur Behebung des Fehlers). !!!!

Selbständigkeitserklärung

Wir bestätigen,

- dass wir die vorliegende Arbeit ohne fremde Hilfe durchgeführt haben durchgeführt haben, ausser denjenigen, welche explizit beschrieben sind,
- dass wir sämtliche verwendeten Quellen erwähnt und gemäss gängigen wissenschaftlichen Regeln korrekt zitiert und
- dass wir keine durch Copyright geschützten Materialien (z.B. Bilder) in unerlaubter Weise genutzt haben.

Dorota Przygienda

Birmensdorf, 31. Januar 2014



Manuel Sanchez

Ettiswil, 31. Januar 2014



Theodore Huber

Zürich, 31. Januar 2014



Inhalt Verzeichnis

Ausgangslage	1
1.1 Ziel der Arbeit	2
1.2 Abgrenzung der Arbeit.....	2
1.3 Praktischer Nutzen der Arbeit	2
1.4 Struktur der Arbeit.....	4
2 Was ist User Experience (UX)?	5
2.1 User Experience als Teil der ISO-Norm 9241-210	6
2.2 Literaturrecherche: Ausprägungen der User Experience und Vorgehensmodelle	7
2.2.1 Jakob Nielsen und Don Norman: Definition der User Experience	7
2.2.2 Alan Cooper - About Face.....	7
2.2.3 Jesse James Garrett - Die Elemente der User Experience.....	8
3 Kommunikation und Interaktion	9
3.1 Kommunikation Mensch-Maschine vs. Mensch-Mensch?.....	9
3.1.1 Mensch-Mensch-Dialog.....	9
3.1.2 Mensch-Maschine-Dialog.....	9
3.1.3 Kann die Mensch-Maschine-Kommunikation von der Mensch-Mensch Kommunikation profitieren?	9
3.1.4 Wie unterscheidet sich die Mensch-Mensch-Kommunikation von der zwischen Mensch-Maschine?	10
3.2 Tonalität.....	10
4 Fehlermeldungen	12
4.1 Definition und Rolle	12
4.2 Anforderungen und Guidelines	13
4.3 Wissenschaftliche Grundlagen	14
4.4 Benutzersicht	16
5 Umfang der Studie.....	17
5.1 Hypothesen.....	17
5.2 Ziel der Studie	18
5.3 Abgrenzung der Studie	18
5.4 Erhoffter Nutzen der Studie	18
6 Vorgehen	19
6.1 Prozess ISO 9241-210	19

6.2	Analyse.....	20
6.2.1	Konzeption.....	20
6.2.2	Visualisieren	21
6.2.3	Überprüfen	21
6.3	Projektplan, Organisation und Risiken	22
6.3.1	Projektplan	22
6.3.2	Risiken.....	26
6.3.3	Organisation	27
7	Methoden.....	29
7.1	Prototyping	29
7.2	Expert Review	29
7.3	Usability Tests.....	30
7.4	Benutzerforschung: Standardisierter Fragebogen (QUIS).....	31
8	Umsetzung.....	32
8.1	Konzept der Tonalität	33
8.2	Auswahl Test Scenario - mySwitzerland.ch	35
8.3	Fehlermeldungen – Definition und Validierung	36
8.4	Prototyp.....	37
8.5	Iterative Entwicklung finaler User Experience Test.....	39
8.5.1	Formular Aufbau	39
8.5.2	Fragebogen.....	41
8.5.3	Datenbank	43
8.5.4	Rekrutierung von Test Personen und Wettbewerb	44
8.5.5	Prototyp Validierung - Qualitative UX Test	44
9	Durchführung der Studie und Datenerhebung	47
9.1	Quantitative Online UX Tests.....	47
9.2	Daten Auswertung und Analyse	48
9.2.1	Daten	48
9.2.2	Analyse	49
10	Ergebnisse.....	54
10.1	Zusammenfassung	54
10.1.1	Frauen.....	54
10.1.2	Männer	54
10.1.3	35 Jahre und jünger.....	54

10.1.4	45 Jahre und älter.....	54
10.2	Erste Eindrücke	55
10.3	Erfahrungsgrad.....	58
10.3.1	Beurteilung der Tonalitäten	61
10.4	Beurteilung der User Experience über die gesamte Testgruppe.....	62
10.5	Gruppierung nach Kapitel des QUIS Fragebogen.....	63
10.5.1	Kapitel 3: Gesamteindruck	64
10.5.2	Kapitel 4: Bildschirm (Layout).....	77
10.5.3	Kapitel 5: Terminologie und Systeminformationen	84
10.5.4	Kapitel 6: Lernfortschritt (Unterstützung bei Problemlösung)	102
10.5.5	Kapitel 7: Systemeigenschaften (Komplexität des Problemlösungsprozesses) 110	
10.5.6	Kapitel 9: Online-Anleitung (Effektivität der Hilfestellung)	116
10.5.7	Kapitel 10: Multimedia Komponenten (Visuell-informative Unterstützung) ..	125
10.6	Beurteilung der Tonalitäten durch Gruppen	129
11	Reflexion.....	132
12	Ausblick.....	134
13	Glossar	135
14	Abbildungsverzeichnis.....	139
15	Literaturverzeichnis.....	143
16	Anhang.....	147

AUSGANGSLAGE

“Error messages aren’t big-time fun for designers or customers. But dealing with them and other forms of messaging within a form is a critical part of the user experience, and handling them well or badly may make a huge difference in the overall success of a form.”

NICE

(Wroblewski 2008, Seite 194)

Fehlermeldungen werden nicht besonders gern gesehen, man wünschte sich, es gäbe sie gar nicht. Es wird ihnen wenig Beachtung geschenkt, sowohl von Benutzer als auch von Software-Entwickler-Seite. Wenn Fehlermeldungen bei der Arbeit mit einer Software auftauchen, lösen sie manchmal negativ behaftete Emotionen aus. Wie oft haben wir selber solche Fehlermeldungen gesehen, die unverständlich, ärgerlich und einfach mühsam waren. In unserem Arbeitsfluss gestört, haben wir kaum genug Geduld und Zeit, um genau zu lesen, was das Problem ist. So werden diese störenden Fenster und Dialogboxen als Erstes weggeklickt. Nur wenn es wirklich nicht weiter geht, schauen wir genauer hin und lesen.

Aber mit welchen Gefühlen und Emotionen? Sind wir froh um die Hilfe oder sind wir verärgert wegen der Störung? Oder es ist uns peinlich, einen Fehler verursacht und eine Mahnung vom System bekommen zu haben? Vielleicht sind wir so verärgert, gestresst oder verzweifelt, dass wir alles verwerfen und abbrechen.

Diese Art von Beobachtungen und Berichten oder eigene Erfahrung sind die Inspiration für folgende Überlegungen: **Startpunkte?**

1. Welche Aspekte und Attribute einer Fehlermeldung könnten die grösste Wirkung auf User Experience haben?
2. Wie wichtig sind Formulierung und Tonalität einer Fehlermeldung, für Empfinden und Erfahrung in einer Fehlersituation?
3. Was erwarten die Benutzer von einer Fehlermeldung und wie wollen sie über einen Fehler informiert werden?
4. Lassen sich die negativen Emotionen beim Auftreten eines Fehlers in positive Erfahrung umwandeln, und wie?
5. Bedeutet heutzutage eine Fehlermeldung nur eine schlechte Erfahrung?

Das führt zu folgender Forschungsfrage: **Welchen Einfluss hat die Tonalität bei Fehlermeldungen auf User Experience?**

In einem Online-Test werden die Benutzer mit Fehlermeldungen konfrontiert und anschließend zu ihrer User Experience befragt. Ein Prototyp für einen Online-UX-Test wird entwickelt und die Daten werden mit einem standardisierten Fragebogen erhoben.

1.1 ZIEL DER ARBEIT

Motivation:

- Tonalität bei Fehlermeldungen wurde bis jetzt noch nicht empirisch untersucht
- Aufbauend auf HCID Wissen und Vorgehensweise neue Erkenntnisse zu gewinnen und die der Allgemeinheit zur Verfügung zu stellen.
- Bedürfnis nach mehr Wissen zum Thema UX als Ergänzung des HCID Studium
- Neuland betreten

Ziel:

- den Zusammenhang von Fehlermeldungen und Benutzerzufriedenheit als Teil der UX zu untersuchen
- einen Prototyp zu entwickeln zur Durchführung der Studie und Daten Erhebung
- anhand der gesammelten Daten Aussagen treffen, wie sich die Tonalität bei Fehlermeldungen auf die UX auswirkt

1.2 ABGRENZUNG DER ARBEIT

Die vorliegende Arbeit ist auf die Tourismus Branche und Hotel Buchungssysteme bezogen. Andere Branchen und Anwendungen wurden hier nicht in Betracht gezogen. Des Weiteren bezieht sich diese Arbeit auf Einsatz von Web Formularen und uns dazu bekannte Publikationen und wissenschaftliche Studien, z.B. von S. Pauwels (2009), J.Bargas-Avila (2011), M.Seckler (2012), M. Hassenzahl (2004) und weitere. Keinesfalls kann diese Arbeit als eine abschliessende und vollständige Analyse auf dem Gebiet gesehen werden.

1.3 PRAKTISCHER NUTZEN DER ARBEIT

Diese Arbeit soll die Bedeutung von Tonalität der Fehlermeldungen aufzeigen und sowohl die Software- Entwickler und -Designer als auch die Auftragsgeber für das Thema sensibilisieren.

Wie wirkt die Formulierung der Fehlermeldungen auf die Benutzer und wie wird dadurch die User Experience beeinflusst? Diese Fragen sind der Leitmotiv der empirischer Studie, in der als Beispiel fünf verschiedenen Tonalitäts-Varianten verwendet und untersucht werden. Die gewonnenen Erkenntnisse sollen einen Beitrag in einem bislang wenig empirisch erforschten Gebiet leisten. Sie sollen auch als eine Art Empfehlung und Wegweiser bei der Entwicklung von ähnlichen Webformularen gesehen werden

NICE



1.4 STRUKTUR DER ARBEIT

Diese Masterarbeit ist in folgende Teile gegliedert:

Ausgangslage und Theorie

Das Ziel und der Nutzen der Arbeit sowie unsere Motivation, sich mit dem Thema Fehlermeldungen zu befassen, sind im Kapitel 1 beschrieben. Hier befindet sich eine Zusammenfassung aus der Literatur Recherche über die wichtigsten theoretischen Grundlagen, Ergebnisse aus bekannten wissenschaftlichen Studien und derzeit geltender Standards zu folgenden Themen:

User Experience (Kapitel 2), inkl. Definitionen, Normen, Meinungen von bekannten UX Experten

Kommunikation (Kapitel 3) zwischen Menschen sowie zwischen Mensch und Maschine. Gibt es bei den beiden Dialog Typen Ähnlichkeiten, gelten die gleichen Regeln bei der Gestaltung?

Fehlermeldungen (Kapitel 4), inkl. Erkenntnisse aus Wissenschaft, sowie Empfehlungen von UX Experten und wichtige Guidelines.

Umfang der Studie

Das Ziel, der erhoffte Nutzen und die Fragestellung der empirischer Studie, sowie Abgrenzungen werden im Kapitel 5 dargestellt

Vorgehen und Methodik

Auswahl und Beschreibung des Vorgehensmodells befindet sich im Kapitel 6. Über benutzte Methoden wird im Kapitel 7 geschrieben.

Empirische Studie

Die Umsetzung der Studie wird im Kapitel 8 ausführlich beschrieben. Dazu gehören Aufbau des Prototyps mit fünf Tonalitäten von Fehlermeldungen und dessen Validierung in qualitativen UX Testing. Die Durchführung der empirischen Studie und die Datenerhebung werden im Kapitel 9 beschrieben.

Ergebnisse

Die Ergebnisse aus der Datenanalyse und die Überprüfung der Forschungsfrage werden im Kapitel 10 behandelt. Anschliessend folgt die Reflexion über die Masterarbeit (Kapitel 11) und Ausblick (Kapitel 12) auf mögliche, zukünftige Untersuchungen.

2 WAS IST USER EXPERIENCE (UX)?

User Experience vs. Usability



Abbildung 1: User Experience vs. Usability

Der Begriff User Experience (Benutzererfahrung) hat in den letzten Jahren viel an Aufmerksamkeit gewonnen und wird oft als Oberbegriff für verschiedene Design- und Usability-Disziplinen verwendet, welche alle ihren Beitrag zur Entwicklung von besseren Produkten, Systemen oder Services leisten. Unser Fokus bestand darin herauszufinden, wie die User Experience gemessen werden kann und wie in dieser Disziplin mit Fehlermeldungen umgegangen wird. Nach eingehender Literaturrecherche haben wir jedoch festgestellt, dass wir den Begriff User Experience wegen seinen unterschiedlichen Definitionen und Ausprägungen neben seiner Definition in der ISO-Norm genauer betrachten müssen: Nielsen versteht unter UX das gesamtheitliche Benutzererlebnis zwischen einer Firma und dem Kunden, während Garrett die UX eher im Webbrowser beleuchtet und Goodwin damit Aspekte der Aktivitäten des Interaction Designers unterstreicht. Hassenzahl und Tractinsky (2006) setzen sich mit Emotion und Affekten der Interaktion von Mensch und Computer auseinander. Dabei untersuchen sie den Einfluss der Emotionen auf das Entscheidungsverhalten im Kontext der UX. Als Fazit der Literaturrecherche konnten wir folgendes festhalten:

- es gibt unterschiedliche Definitionen der UX
- UX Definitionen sind unterschiedlich detailreich
- in den UX-Definitionen gibt es unterschiedliche Bezugsräume

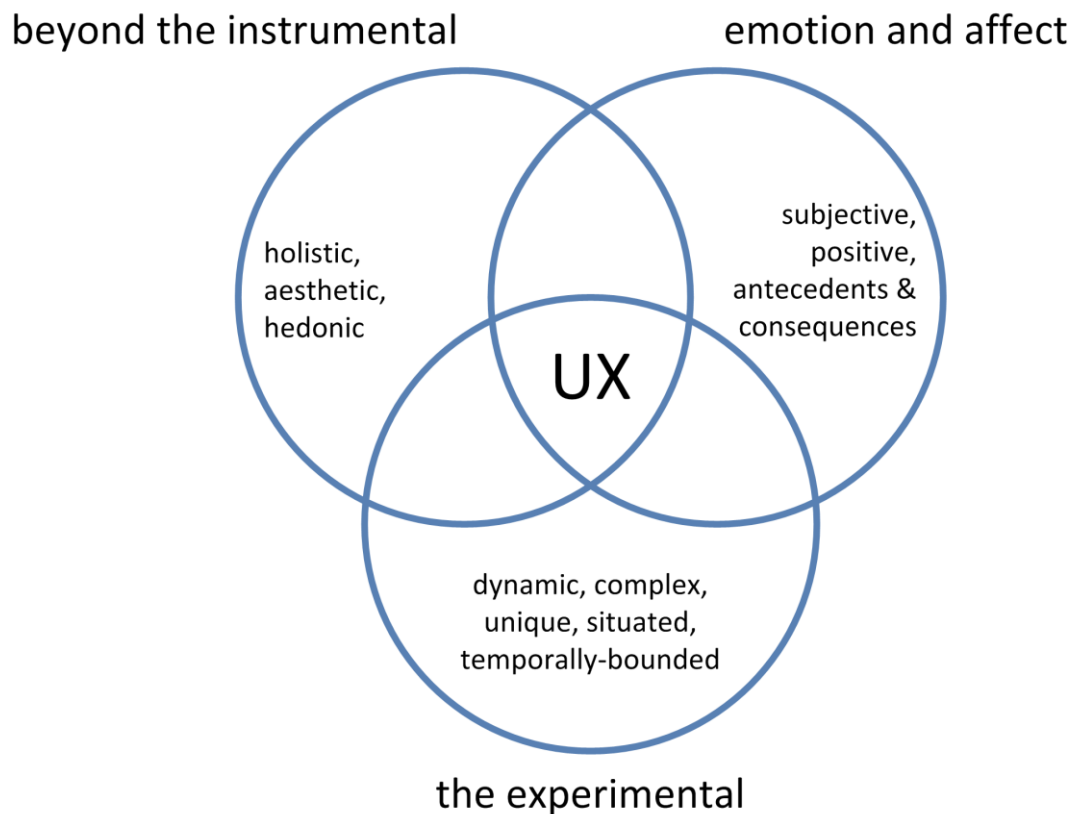


Abbildung 2: Facetten der UX nach Hassenzahl und Tractinsky (2006, S. 95)

2.1 USER EXPERIENCE ALS TEIL DER ISO-NORM 9241-210

Die ISO-Norm 9241-210 steht für benutzerorientiertes Vorgehen in Entwicklungsprojekten und richtet sich an Projektleiter, die systematisch Usability-Engineering-Aktivitäten im Entwicklungsprojekt planen und anwenden wollen. Die Norm grenzt die Konzepte Usability und User Experience voneinander ab. Im März 2010 wurde die Disziplin User Experience zum ersten Mal in der neuen ISO-Norm 9241-210 (Internationale Organisation für Normung, 2010) mit aufgenommen und zum internationalen Standard definiert. Sie besagt folgendes:

Die Wahrnehmungen und Reaktionen einer Person, die aus der Verwendung und / oder erwarteten Nutzung eines Produktes, Systems oder einer Dienstleistung erfolgt.

Faktoren der User Experience als Teil der Definition

Des Weiteren umfasst sie die beim Benutzer ausgelösten Emotionen, seine Vorstellungen, Vorlieben, Empfindungen, physischen und psychischen Reaktionen, Verhaltensweisen und erreichten Leistungen, die vor, während und nach dem Gebrauch auftraten.

Der Benutzer und seine Erfahrung mit dem Produkt werden beeinflusst durch:

das Marken-Image und wie sich die Marke dem Kunden präsentiert, die Funktionalität des Produktes und System-Performance, das interaktive Verhalten und die unterstützenden Funktionen des interaktiven Systems. Ebenfalls spielen der körperliche Zustand des

Benutzers und dessen frühere Erfahrungen, Ansichten, Kompetenzen sowie Persönlichkeit und Kontext der Verwendung eine wichtige Rolle.

Der Begriff Usability, welcher in der DIN EN ISO-Norm 9241-110 definiert ist, gilt als Mass für die Benutzbarkeit eines technischen Systems. Er enthält auch die Art der Wahrnehmung und der emotionalen Aspekte, die in der Regel mit User Experience verbunden sind. Usability-Kriterien können daher zur UX-Beurteilung verwendet werden.

2.2 LITERATURRECHERCHE: AUSPRÄGUNGEN DER USER EXPERIENCE UND VORGEHENSMODELLE

Um unsere Forschungsfrage einzugrenzen, klärten wir die Ausprägungen der Disziplin User Experience in bewährter und aktueller Literatur (digitale Medien sind berücksichtigt). Relevantes Material zur Definition haben wir in den Büchern von Alan Cooper, Jesse James Garrett und im Internet-Auftritt der Nielsen Group (Jakob Nielsen und Don Norman) gefunden. Anhand der unterschiedlichen Resultate zeigte sich ein gewisses Spannungsfeld bezüglich Ausprägungen und Vorgehensweisen für eine gute User Experience.

Überblick:

2.2.1 JAKOB NIELSEN UND DON NORMAN: DEFINITION DER USER EXPERIENCE

«User experience encompasses all aspects of the end-user's interaction with the company, its services, and its products.»

Nielsen und Norman betrachten User Experience als ein Konzept. Zielführend ist:

- die exakten Bedürfnisse der Kunden zu kennen
- Einfachheit und Eleganz des Produktes (es macht Freude, es zu besitzen und zu benutzen)

Des Weiteren erwähnen sie, wie wichtig es ist, dass alle zusammen arbeitenden Disziplinen im Projekt zusammengeführt werden:

- Engineering
- Marketing
- Grafik- und Industriedesign
- Interface Design

«As an example, consider a website with movie reviews. Even if the UI for finding a film is perfect, the UX will be poor for a user who wants information about a small independent release if the underlying database only contains movies from the major studios.» (Nielsen, 2013)

2.2.2 ALAN COOPER - ABOUT FACE

Hier wird der Terminus «Experience Design» für digitale Produkte verwendet, um die User Experience zu gestalten. Auch das Interaktionsdesign kommt zum Zuge, welches sich auf das

Design von Verhalten konzentriert und dabei die Beziehung zu Form und Inhalt berücksichtigt. Im Gegensatz zu Nielsen und Norman wird auf die gute Zusammenarbeit der Designdisziplinen hingewiesen. Cooper teilt diese in Form, Inhalt und Verhalten auf.

Das Technikerteam, Marketing und Management werden hier im Gegensatz zu Nielsen nur unterstützend erwähnt.

Das Vorgehensmodell «Goal-Directed-Design» von Cooper besagt: Wenn ein Benutzer mit einem Produkt seine Ziele erreichen kann, wird er zufrieden und glücklich sein, mit dem Produkt effizient arbeiten, gerne dafür zahlen und es weiterempfehlen.

Zum Thema Fehlermeldungen erwähnt Cooper, dass diese oft

- schlecht formuliert
- wenig hilfreich
- unhöflich

sind und leider nie dazu verwendet werden, um Fehler im Vorhinein zu vermeiden. Wann immer möglich, sollten Fehlermeldungen jedoch vermieden werden.

2.2.3 JESSE JAMES GARRETT - DIE ELEMENTE DER USER EXPERIENCE

Das Buch von Garrett ist 2011 in zweiter, überarbeiteter Auflage erschienen und Coopers «About Face» übrigens bereits in der dritten. Garrett begründet seine Überarbeitung mit der rasanten Entwicklung in den letzten 10 Jahren. Er führt aus, dass sich sowohl die Elemente wie auch die User Experience selbst verändert hätten. In der Erstauflage habe sich die Bezeichnung «User-Experience-Design» nur auf den begrenzten Rahmen des Web oder interaktiver, bildschirmgestützter Anwendungen bezogen. Das Spektrum der User Experience habe sich inzwischen vergrößert.

Im Vergleich zum Vorgehensmodell ISO 9241-210 welches benutzerzentriertes und iteratives Vorgehen verdeutlicht, ist das Vorgehensmodell namens 5-S Modell das erstmals 2002 von J.J. Garrett vorgestellt wurde, wesentlich anders. Die Konzeptidee basiert darauf, dass die Konzeption eines User Interface in Schichten gemacht werden sollte. Diese Schichten werden von Garrett als Elemente bezeichnet und in 5 Ebenen aufgeteilt:

- Surface (Oberfläche)
- Skeleton (Raster)
- Structure (Struktur)
- Scope (Umfang)
- Strategy (Strategie)

Garrett erwähnt nur nebenbei, dass bei seinem Vorgehensmodell auch iterativ vorgegangen werden soll.

Zum Thema Fehler weist Garrett darauf hin, dass diese möglichst zu vermeiden sind und mit jeder zusätzlichen Massnahme zur Fehlerbehandlung in der Interaktionsgestaltung sich der Anteil von Nutzern mit positiven Produkterfahrungen erhöht.

3 KOMMUNIKATION UND INTERAKTION

3.1 KOMMUNIKATION MENSCH-MASCHINE VS. MENSCH-MENSCH?

Symmetrische Kommunikation ist das Streben nach Gleichheit und Vermeidung von Unterschieden. Symmetrische Kommunikation kommt nur in Mensch-Mensch-Kommunikation vor. Mensch-Maschine-Kommunikation ist immer asymmetrisch. Unsere Mensch-Maschine-Kommunikation beschränkt sich nachfolgend auf Webapplikationen ohne Berücksichtigung akustischer Sprachausgabe.

3.1.1 MENSCH-MENSCH-DIALOG

Höflichkeit wird in der Kommunikation als Strategie der Konfliktvermeidung angesehen. Sie ist das Sicherheitsnetz der Kommunikation (Kienpointner, 2005). Höflichkeit drückt gegenseitigen Respekt und Toleranz aus, verhindert Dialogblockaden und sorgt für eine gute Stimmung in der Kommunikation.

Gegenseitiges Unterbrechen wird oft als unhöflich betrachtet. Bei Paaren, welche längere Zeit zusammen sind, wird das Unterbrechen nicht als störend empfunden, sondern als Zeichen der Anteilnahme.

3.1.2 MENSCH-MASCHINE-DIALOG

Der Mensch-Maschine Dialog wird als wechselseitiger Informationsaustausch zur Erledigung einer Aufgabe verstanden (Andreea Niculescu, Mensch-Maschine Kommunikation, 2002). Diese Art der Kommunikation hat mit einem asymmetrischen Verhältnis zu tun. Auf der einen Seite der Mensch als kognitives Wesen, welches sich mühelos in sprachlicher Interaktion fließend ausdrücken kann, und auf der anderen Seite die vorgegebene leistungsbegrenzte Maschine.

3.1.3 KANN DIE MENSCH-MASCHINE-KOMMUNIKATION VON DER MENSCH-MENSCH KOMMUNIKATION PROFITIEREN?

Gelten für die Mensch-Maschine-Kommunikation dieselben Regeln, wie bei der zwischenmenschlichen Kommunikation? Wir haben erfahren, dass bei einer Mensch-Mensch-Kommunikation Konfliktvermeidung im Vordergrund steht. Kann dieses Ziel auch bei Maschinen verfolgt werden? Wohl kaum. Wie sieht es umgekehrt aus? Was, wenn die Maschine sich vermehrt in Konfliktvermeidung versucht? Sind wir in der Lage, das Gegenüber, in diesem Falle eine Maschine, als vollwertigen Kommunikationspartner anzusehen? Dafür müssten wir uns im Klaren sein, welches Kommunikationsziel wir in Anwendungen erreichen möchten. Gemäss Prof. Dr. Andreas M. Heinecke (2004) ist das Kommunikationsziel in einer sicherheitskritischen Anwendung, den Benutzer zu warnen und vor Gefahren zu schützen. Diese Aussage hat ihre Präferenz ganz klar in der Effektivität und nicht in der Motivation. Sie ist auch nicht in der Konfliktvermeidung angesiedelt. Die Frage ist nun: kann die User Experience profitieren, wenn man anstelle einer rein aufgabenorientierten, der Effizienz und Effektivität verschriebenen Kommunikation zwischen Mensch-Maschine, eine konfliktvermeidende, motivationsbasierte Sprache verwendet?

3.1.4 WIE UNTERSCHIEDET SICH DIE MENSCH-MENSCH-KOMMUNIKATION VON DER ZWISCHEN MENSCH-MASCHINE?

Hier ist die grundsätzliche Frage, ob die Erwartungen abhängig von den aktuellen technischen Möglichkeiten sind, oder ob es zwischen Mensch-Maschine nie eine Ähnlichkeit zum zwischenmenschlichen Dialog geben wird. Kann der User eine emphatische Beziehung zur Maschine pflegen? Bei einer zwischenmenschlichen Kommunikation gibt es immer zwei Partner welche jeweils auf die Situation eingehen und auf das Feedback des Partners reagieren können. Die Art und Weise, wie eine Maschine mit dem User kommunizieren soll, ist abhängig vom Kontext der Anwendung. Die Tonalität bestimmt, welche Dialoge der Maschine innerhalb eines bestimmten Kontextes glaubwürdig erscheinen. Da die Anwendung nicht auf die aktuelle Stimmung des Users Rücksicht nehmen kann, sollte die Tonalität mit dem Kontext und der Aufgabe des Users abgestimmt sein.

3.2 TONALITÄT

Der Ausdruck Tonalität stammt ursprünglich vom französischen tonalité ab. Duden und Wikipedia definieren ihn mit Aspekten aus der Musik: jegliche Beziehung zwischen Tönen, Klängen und Akkorden. Wir beziehen uns jedoch auf die englische Variante tonality, welche im Bereich Werbung/Marketing angewendet wird. Gemäss Wikipedia ist die englische Variante ein Begriff aus der Kommunikation, welcher bei Werbebotschaften den Stil und die Atmosphäre an den Empfänger vermittelt. Die Tonalität wird beim Werbetext auch als «tone of voice» bezeichnet. Es geht also nicht nur darum, dass das Werbeobjekt beschrieben wird, sondern wie es präsentiert wird. Der «Grundton» der Werbung ist demnach die Tonalität. Sie ist verantwortlich für die Aufmerksamkeit und Identifikation bei den Konsumenten. Der so genannte Grundton, welcher auch als «atmosphärische Verpackung» bezeichnet wird, ist abgestimmt mit der Umgebung, in der die Werbung präsentiert wird: Raum, Outfit, Look.

Das Ziel der Tonalität ist:

- die Zielgruppe über eine emotionale Ansprache zu erreichen, so dass eine positive Wirkung erzielt werden kann
- eine Verbindung herzustellen zwischen dem Produkt und positiven Emotionen

Auf einer Website wäre die Tonalität demnach das Gefühl, welches dem Leser des Inhalts vermittelt wird. Nicht alle Leser reagieren gleich: Wie bereits erwähnt, wird eine Tonalität abgestimmt auf eine bestimmte Zielgruppe. Es folgen 4 stichwortartige Beispiele, wie Zielgruppen mit unterschiedlichen Interessen angesprochen werden können:

- Überzeugung: Zahlen und Fakten werden belegt und perfekt auf den Punkt präsentiert
- Sicherheit: Traditionen, Kontinuität, konservativ, keine Veränderungen und Visionen
- Erlebnis: neue Abenteuer, Progressive Sprache, Intuition, Fantasie, Metaphern
- Direkter Kontakt: offen auf Menschen zugehen, emotionale Brücke aufbauen, rhetorische Fragen, direkte persönliche Ansprache

Jede dieser Varianten erfüllt eine Funktion und spricht unterschiedliche Menschen an, indem es bestehende Denkmuster und Werte verstärkt. In unserer Masterarbeit haben wir 5 verschiedene Tonalitäten in den Fehlermeldungen angewendet, welche später genauer beschrieben werden.



4 FEHLERMELDUNGEN

4.1 DEFINITION UND ROLLE

Die Interaktion zwischen Mensch und Maschine verläuft nicht immer reibungslos. Es können dabei Missverständnisse, technische Probleme oder andere Fehler vorkommen, die den Benutzer daran hindern, seine Aufgabe abzuschliessen. Über diese Vorkommnisse wird der Benutzer mit einer Fehlermeldung informiert.

«Eine Fehlermeldung, englisch error message, ist eine Reaktion eines Computerprogramms auf einen erkannten Programmfehler oder eine bekannte Fehlersituation.» (Wikipedia, Fehlermeldung)

Die Rolle der Fehlermeldungen ist nicht nur, den Benutzer zu warnen und über ein Problem und dessen Ursachen zu informieren, sondern auch Hinweise für weiteres Vorgehen oder Hilfestellung anzubieten (Wroblewski 2008, Wilska 2013).

Aus diesen Gründen sind die Fehlermeldungen eine wichtige Komponente einer Software-Anwendung und gelten als eine Art Dialog zwischen Mensch und Maschine. Somit unterliegen sie auch den Vorgaben der ISO Norm 9241-110 Grundsätze der Dialoggestaltung, wo auch das Fehlertoleranz Prinzip definiert ist:

«Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann.» (ISO 9241-10)

Typen von Fehlern werden je nach deren Ursache und Auswirkung bezeichnet. Aus der Softwareentwicklungs-Perspektive werden diese Fehlertypen genannt:

- System-Fehler: ein technisches Problem, wie Netzwerk- oder Hardware- Störungen
- URL-Fehler: eine Ressource konnte nicht gefunden werden, HTTP-Server Fehler Code «404 - not found» ist der Fehler, der im Internet am meisten vorkommt
- Daten Fehler:
 - ungültige Eingabe
 - fehlende Eingabe
 - falsches Datenformat

Aus der Benutzer Perspektive gibt es vor allem folgende zwei Arten von Fehlern (Norman 2002, Seite 105):

- Ausrutscher (eng. slip): automatische Handlung, die unbeabsichtigt das Ziel verfehlt, z.B. ein Tippfehler
- Irrtum (eng. mistake): eine bewusste Handlung, die aber aus einer falschen Annahme hervorgeht

Die Vergangenheit hat den Fehlermeldungen nicht besonders viel Beachtung geschenkt und es war die Aufgabe des Programmierers, sie zu behandeln. Dabei wurde nach einem

Implementierungs-Modell (Cooper 2010) vorgegangen, das auf dem Ansatz einer fehlerfreien, «schuldlosen Software» basiert (Cooper 2010, Seite 317). Probleme, die während der Benutzung entstehen können, sind meistens durch falsche Anwendung verursacht. Dieses Modell schreibt dem Benutzer die Schuld für Misslingen einer Aktion zu, informiert ihn darüber und fordert oft eine Korrektur ein. Dabei wird der Benutzer in seinem Arbeitsfluss gestört und vom Ziel abgehalten. Mit dem Designprinzip: «Unterbrechen Sie den Arbeitsfluss nicht durch idiotische Dialogfelder» (Cooper 2010, Seite 225) plädiert Cooper für ein Umdenken und bessere Unterstützung des Benutzers. Seiner Meinung nach sind die typischen Fehlermeldungen überflüssig und sollten abgeschafft werden. Durch gutes Design und gezielten Einsatz von spezifischen GUI-Elementen, sollte der Benutzer gar keine Möglichkeit haben, einen Fehler zu machen.

Fehlervorbeugung propagiert auch Don Norman, der bekannte und angesehene UX-Experte: *«Eliminate error messages from your system. Always turn an error into an opportunity to help.» (Norman 2013)*

Das ist auch im Sinne von anderen bekannten UX-Experten: Wroblewski (2008) , J.J.Garrett (2012, Seite 86), Wilska (2013). Schon 1995 hat Jakob Nielsen in seinen bekannten «10 Usability Heuristics for User Interface» diese Meinung vertreten:

«Even better than good error messages is a careful design which prevents a problem from occurring in the first place.» (Nielsen 1995)

4.2 ANFORDERUNGEN UND GUIDELINES

Nicht immer lässt sich jeder Fehler vermeiden. Deswegen sind Fehlermeldungen in einer Software-Anwendung nach wie vor präsent. Damit sie gut funktionieren und ihre Rolle optimal erfüllen, sollten sie nach Jakob Nielsens (2001) Empfehlung so sein:

- **Explicit** - eindeutig
klar signalisieren, dass etwas passiert ist. Die schlimmsten Fehlermeldungen sind solche, bei denen ein Fehler vorkommt und der Benutzer kein Feedback erhält
- **Human-readable** - sprachlich verständlich
ohne unklare Fehlercodes und Abkürzungen, nicht wie *«an error of type 2 has occurred.»*
- **Polite** - höflich
höflich hinweisen, aber ohne Schuldzuweisung an den Benutzer und ohne ihn dummzustellen, wie bei der Meldung *«illegal command.»*
- **Precise** - genau
klare Beschreibung des Problems und nicht eine unbestimmte Generalisierung verwenden, wie *«syntax error.»*
- **Constructive advice** - konstruktiver Hinweis
informieren, wie man das Problem beheben oder umgehen kann
- **Visible** - sichtbar

die Fehlermeldung klar und deutlich darstellen und das betroffene GUI-Element ebenfalls

Diese Grundsätze wurden Anfangs dieses Jahrhunderts zusammengefasst und schon damals waren sie lange etabliert, wie es auch Nielsen feststellte:

*«The guidelines for creating effective error messages have been the same for 20 years»
(Nielsen 2001)*

Sind diese Regeln wirklich zeitlos und gelten sie heute immer noch in dieser Form? Oder haben sich die Anforderungen an die benutzerfreundlichen Fehlermeldungen im Laufe der Zeit verändert? Heutzutage werden diese Regeln teilweise bestätigt, überarbeitet und erweitert. Der bekannte und oft zitierte UX Experte Luke Wroblewski (2008) in seiner aktuellen Ausgabe zum Thema empfiehlt:

Ergänzende Praktiken für Formalarmeldungen gemäss Wroblewski

- Fehlermeldungen deutlich kommunizieren. Sie sind beim Erscheinen das wichtigste Element in einem Formular.
- Für die schnelle Korrektur die Fehler im Kontext platzieren
- Hilfestellungen anbieten, damit die Benutzer die Fehler schnell beheben können.
- Top-Level-Fehler zeigen an, dass es einen Fehler gibt und wie dieser behoben werden kann. Mehrere Fehler werden im Top-Level-Bereich aufgelistet.
- Bei Fehlern in Eingabefeldern muss mit deutlichen, farblichen Hervorhebungen sichergestellt werden, dass sich diese von den restlichen Feldern unterscheiden.
- Jedes Formularelement mit einem Fehler muss mit der Top-Level Fehlermeldung assoziiert werden und klar kommunizieren, dass der Fehler zuerst behoben werden muss, bevor es weitergehen kann.
 - Roter Text und Warnung-Ikonen sind für Fehlermeldungen reserviert.
 - Erfolgsmeldungen so im Kontext platzieren, dass diese die weiteren Arbeitsschritte nicht blockieren.
 - Bei kürzeren Formularen ist es möglich, auf die Top-Level-Meldungen zu verzichten und nur die Felder deutlich hervorzuheben, welche für die Fehler verantwortlich sind. Diese Wahl verlangt jedoch viel Sorgfalt.

4.3 WISSENSCHAFTLICHE GRUNDLAGEN

Aufbauend auf den Grundsätzen von Nielsen, Norman und Cooper werden heutzutage weitere Studien und Forschungsprojekte durchgeführt mit dem Ansatz, die Benutzer noch besser zu unterstützen, deren Zufriedenheit zu steigern und die Fehlerquote zu reduzieren.

Das Fehlervorbeugungs-Prinzip inspiriert weiterhin UX-Designer und Wissenschaftler, um neue, bessere Gestaltungs-Ideen und Ansätze zu finden. Eine potenziell grosse Fehlerquelle sind die **Pflichtfelder** im Bereich von Online-Formularen welche von Benutzern schnell wahrgenommen und ausgefüllt werden müssen. Solche Eingabefelder werden oft mit einem Asterix (*) hervorgehoben, was sich gut etabliert hat. Eine andere Methode der visuellen

Gestaltung für Pflichtfelder hat Stefan L. Pauwels et al. (2009) untersucht. In der Studie wurden die Pflichtfelder farblich markiert und somit waren sie für die Benutzer schneller und einfacher wahrnehmbar. Ergebnisse haben bestätigt, dass diese Methode zur schnelleren Erfassung des Formulars, weniger Fehler und mehr Benutzerzufriedenheit beigetragen hat.

Ebenso wichtig sind Angaben zum Datenformat bei den Eingabefeldern mit Format-Restriktionen. Der Benutzer soll einfach und verständlich darüber informiert werden, in welcher Form er die Daten eingeben soll. Diesen Aspekt hat Javier A. Bargas-Avila et al. (2011) in seiner Studie untersucht und 4 Varianten verglichen:

- *ohne jeglichen Hinweis,*
- *mit einem Eingabe-Beispiel,*
- *mit einer Format-Beschreibung*
- *Kombination von Eingabe-Beispiel und Format-Beschreibung.*

Wie die Ergebnisse gezeigt haben, war jede Art von Hinweis für die Benutzer hilfreich und hat sie unterstützt, Fehler zu vermeiden. Interessant dabei war, dass ein Eingabebeispiel keine bedeutende Rolle gespielt hat.

Die Platzierung von Fehlermeldungen hat auch eine grosse Bedeutung bei der Gestaltung von Online-Formularen. Diesen Aspekt hat Mirjam Seckler et al. (2012) in ihrer Studie untersucht und die 100 meistbenutzten Online-Shops diesbezüglich verglichen. Aus der Studie geht hervor, dass die Fehlermeldungen in den meisten Fällen (45,5%) am Anfang des Formulars angezeigt werden. Erstaunlich viel (21,8%) der untersuchten Shops verwenden die Fehlerdialogbox (alert box), obwohl das dem Designprinzip von Cooper widerspricht und den Ergebnissen der Studie von Bargas-Avila (2007). Die Fehlerdialogboxen stören bei der Aufgabe und werden entweder von den Benutzern möglichst schnell weggeklickt oder führen zu weiteren Fehlern. Eine wichtige Erkenntnis dieser Studie war, dass Fehlermeldungen in der Nähe des Eingabefeldes besser funktionieren, als solche am Ende oder Anfang des Formulars. Als intuitiv wurden sie empfunden, wenn sie rechts von dem fehlerhaften Eingabefeld platziert waren. Dies führte zu besserer Performanz und Effizienz beim Ausfüllen des Formulars, aber auch zu mehr Benutzerzufriedenheit.

Der Zeitpunkt der Validierung von Daten in Web-Formularen ist von grosser Bedeutung für Effizienz und Benutzerzufriedenheit. Wann soll das Fehler-Feedback dem Benutzer angezeigt werden, damit es nicht als störend, sondern als hilfreich empfunden wird? Dieser Frage ist Javier A. Bargas-Avila (2011) in seinen Recherchen nachgegangen und er konnte bestätigen, dass eine Validierung und Anzeige der Fehlermeldungen am Ende der Formularerfassung am besten funktioniert. Auf Grund dieser Beobachtung hat er folgende Theorie aufgestellt: der Benutzer befindet sich bei solchen Aufgaben in zwei verschiedenen Arbeitsmodi:

- Erfassungsmodus - Completion Mode
- Validierungsmodus - Revision Mode

Zuerst ist der Benutzer im Erfassungsmodus und auf seine Aufgabe und das Ziel konzentriert, das Formular möglichst schnell auszufüllen und abzuschicken. Fehlermeldungen, die in dieser Phase vorkommen, werden oftmals übersehen oder sogar ignoriert. Erst nachdem das gesamte Formular ausgefüllt ist, gelangt der Benutzer in den Validierungsmodus und kann alle Eingaben überprüfen und Fehlerkorrekturen vornehmen.

Diese neusten Recherchen liefern uns wichtige Erkenntnisse über das Verhalten der Benutzer und deren Umgang mit den Fehlern. Seit der Publizierung von Nielsen Guidelines und Heuristiken, hat es technologische Entwicklungen gegeben. Vielfalt und das Angebot der Online-Anwendungen sind gewachsen und eine neue, junge Generation der Web-Benutzer hat sich gebildet. Trotzdem sind die meisten Empfehlungen immer noch aktuell und werden vervollständigt und erweitert, um dem Geist der Zeit besser zu entsprechen. Usability von Fehlermeldungen hat Javier A. Bargas-Avila (2010) in seinen **Guidelines** so definiert:

- Eine Format-Spezifikation ohne zusätzliches Beispiel soll im Voraus dem Benutzer dort kommuniziert werden, wo ein bestimmtes Eingabe-Format erwartet wird (Guideline 13)
- Fehlermeldungen sollen höflich sein und den Benutzer in verständlicher Sprache informieren, dass ein Fehler aufgetreten ist und sich eventuell für den Fehler entschuldigen. (Guideline 14)
- Die fehlerhafte Eingabe darf nicht gelöscht werden (Guideline 15)
- Fehlermeldungen sollen erst am Ende, nach dem Ausfüllen und Abschicken des Formulars angezeigt werden. (Guideline 16)
- Fehlermeldungen sollen bemerkbar sein durch den Einsatz von Farbe, Icons und Text und so klar auf den Problembereich hindeuten. Sie sollen auf verständliche Art und Weise das Problem erklären und eine Lösung vorschlagen. (Guideline 17)

4.4 BENUTZERSICHT

„Der User fühlt sich beschämt, wenn ihm die Software mitteilt, er habe einen Fehler gemacht“. (Cooper 2010, Seite 489)

Das Designprinzip von Cooper beschreibt, wie die Fehlermeldungen aus Sicht der Benutzer wahrgenommen werden. Ein negatives Feedback des Systems über das Mislingen einer Aktion wird gleichzeitig als Schuldzuweisung empfunden.

Für die Benutzer sind Fehlermeldungen ein Störfaktor, da sie von ihrer ursprünglichen Aufgabe abgelenkt werden. Diese Unterbrechungen können zu Frustrationen und Irritationen führen oder sogar zum Abbruch der Aufgabe (Fischer 2009)

Zahlreiche Recherchen geben Auskunft darüber, wie die Benutzer über Fehler informiert werden sollen, zum Teil im Kapitel 4.3 beschrieben. Das gibt uns zwar einen guten Überblick über die Usability von Fehlermeldungen, aber weckt gleichzeitig auch den Eindruck, dass die Empfindungen und Emotionen der Benutzer zu wenig Gewichtung bekommen haben. Auf welche Art und Weise und in welchem Ton möchten die Benutzer über Fehlerfälle informiert werden, damit nach dem Abschluss des Formulars ein gutes Gefühl bleibt und trotz der Fehlermeldungen eine positive User Experience entsteht? Das wollen wir in der Studie unter die Lupe nehmen.

5 UMFANG DER STUDIE

Das Hauptgewicht dieser Studie liegt darin, den Einfluss von Tonalität bei Fehlermeldung auf die User Experience zu untersuchen. Die Art und Weise, wie die Fehlermeldungen formuliert sind und deren Tonalität scheint uns von grosser Bedeutung zu sein und ist noch wenig erforscht.

Für diese empirische Studie wird ein Test-Fall und ein Webformular Prototyp entwickelt, basierend auf einer bekannten Online-Hotelbuchungs-Plattform mySwitzerland.com, deren Usability den heutigen Standards entspricht. Dabei werden 5 verschiedene Varianten der Tonalität von Fehlermeldungen entwickelt und im Prototyp eingebaut. Die Probanden werden eine bestimmte Tonalität zu Beginn des UX Testes zugeteilt bekommen. In der Studie wird gewährleistet, dass jede Tonalität gleich oft getestet wird und ein Proband nur ein Mal den Test durchführt.

Die freundliche Ausdrucksweise bei der Formulierung von Fehlermeldungen bezeichnet Wilska (2013) als sehr wichtig:

«Providing the information the user needs—what's the problem, and how can I fix it?—in a straightforward, friendly manner is one of the most important things you can do to make your error messages effective and usable.»

Wir gehen davon aus und stellen eine weitere Frage:

Kann eine geschickte Wortwahl und benutzerfreundliche Tonalität bei Fehlermeldungen zu besserer User Experience beitragen?

5.1 HYPOTHESEN

Hypothese: Tonalität der Fehlermeldungen hat Einfluss auf die User Experience.

Weitere Aspekte der Studie, die durch die Datenauswertung beantwortet werden könnten, die aber nicht in vorliegender Arbeit genauer untersucht werden und nicht im Fokus sind:

- Lange Fehlermeldungen sind mühsam zu lesen und kommen bei den Benutzern schlechter an als kurze und prägnante Meldungen.
- Humorvolle Fehlermeldungen beeinflussen Emotionen auf eine positive Art und Weise
- Die Tonalität soll dem Umgebungskontext (Hotel-, Reise-Typ) angepasst werden. Auf diese Weise wirkt sie glaubwürdig und authentisch.
- Tonalität soll dem Zielpublikum angepasst werden und die gleiche Sprache verwenden.
- Fehlermeldungen müssen höflich formuliert werden, «Bitte...» wird als Standard Ansprache erwartet.
- Das Ungewohnte verwirrt den Benutzer und braucht mehr Zeit, verstanden zu werden.

5.2 ZIEL DER STUDIE

Diese Studie hat zum Ziel zu untersuchen, wie sich die Tonalität bei Fehlermeldungen auf die User Experience auswirkt. Mit Hilfe eines selbst entwickelten Online-UX-Test und eines standardisierten Fragebogens wurden die Testpersonen zu ihrer User Experience befragt. Die Auswertung der erhobenen Daten soll folgende Aspekte beleuchten:

- die Wichtigkeit der Tonalität von Fehlermeldungen
- Erwartungshaltung und Reaktion der Benutzer betreffend Tonalität und Fehlermeldungen.
- Einstimmung der Tonalität mit dem Benutzer-Kontext und der Aufgabe.
- Vorstellung und Evaluation von möglichen Tonalität Varianten

5.3 ABGRENZUNG DER STUDIE

Diese Studie bezieht sich auf ein bestimmtes Forschungsgebiet, wie bereits im Punkt 1.2 erwähnt. Zusätzlich sind noch folgende Abgrenzungen relevant:

- Der Bezug zur Website mySwitzerland.com soll die Reisetimmung bei den Probanden hervorbringen und somit den gewünschten Benutzer-Kontext erzeugen.
- Die Usability von mySwitzerland.com, die den heutigen Standards entspricht, war ebenfalls ein wichtiges Kriterium dafür, diese Website als Basis-Plattform für die Studie zu wählen.
- mySwitzerland.com darf keinesfalls als Objekt der Studie gesehen werden. Es ist nicht unsere Absicht, diese zu testen oder zu bewerten.
- Der Kontakt und die Zusammenarbeit mit mySwitzerland.com wurde zwar anfänglich in Betracht gezogen, jedoch gezielt unterlassen, um diese Studie unabhängig durchführen zu können.
- Die Studie kann nicht als repräsentativ gesehen werden. Die demografische Struktur der Testpopulation ist keine Abbildung der Bevölkerungs-Struktur in der Schweiz.
- Aus Datenschutzgründen werden die persönlichen Daten der Testpersonen, wie Name, Wohn- und Email-Adresse, vertraulich behandelt und nicht weitergegeben.

5.4 ERHOFFTER NUTZEN DER STUDIE

Anhand der Resultate soll es möglich sein, Aussagen darüber zu treffen, welche Tonalität sich positiv auf eine bestimmte Altersgruppe auswirkt und bei welchem Aspekt der UX. Dies kann hilfreich sein z.B. für weitere Hotelbesitzer aus der Branche, die online ihre Dienste mithilfe einer angepassten Tonalität an die entsprechende Altersgruppe verbessern möchten.

Das Vorgehen dieser Studie könnte als Beispiel für weitere Untersuchungen dienen, bei denen die Messung von User Experience angewandt werden muss.

6 VORGEHEN

Wir haben uns bei der Auseinandersetzung mit der Methodenwahl immer wieder folgende Kernfragen gestellt:

- Wie messen wir die User Experience?
- Wie lässt sich das, was beim Benutzer emotional im Kopf vorgeht verständlich als Daten festhalten und auswerten?
- Wie stellen wir sicher, dass wir den Einfluss der Tonalität testen?

Um die letzte Frage zu beantworten war schnell klar, dass wir unser eigenes Werkzeug bereitstellen, bzw. entwickeln mussten.

Zu Beginn der Masterarbeit waren wir uns einig, dass wir viele User-Experience-Daten mithilfe unserer Probanden erheben müssen und der Benutzer somit eine zentrale Rolle spielt. Deshalb wählten wir ein UCD-Vorgehensmodell, da bei diesem vom Benutzer ausgegangen wird und es zum Ziel hat, dass das System «Mensch-Computer» möglichst gut funktioniert. Als Vorgehensmodelle standen zur Wahl:

- Goal Directed Design™ von Cooper et al.
- 5-S Modell von J. J. Garrett
- ISO 9241-210 (Human-centred design for interactive systems)

Alle 3 Modelle berücksichtigen die User Experience. Für das ISO-Modell haben wir uns entschieden, da es bei der Entwicklung unseres Tests die nötige Flexibilität bot. Zur Risikominimierung schien dieses UCD-Modell mit dem iterativen Vorgehen geeignet, um Neuland zu betreten.

6.1 PROZESS ISO 9241-210

Am Anfang des Projektes wird der Nutzungskontext analysiert und definiert. Auf dessen Grundlage werden die Requirements (Nutzungsanforderungen) bestimmt, welche in der nächsten Phase "Konzeption" verwendet werden, um Ideen und Entwürfe zu entwickeln. In der dritten Phase "Visualisieren" werden erste Prototypen gestaltet. In der vierten und letzten Phase werden diese mit einem repräsentierenden Endbenutzer getestet.

Gemäss der Beschreibung von diesem Vorgehensmodell auf der Website von Dr. Marcel Uhr (www.soultank.ch, 8.12.2013), kann ein Projekt in folgende Phasen aufgeteilt werden:

- Initialisierung
- Analyse
- Entwurf
- Realisierung
- Einführung
- Nutzung

Innerhalb dieser einzelnen Phasen soll nun iterativ (lat. iterare, «wiederholen») vorgegangen werden. Jedes einzelne iterative Vorgehen wird ebenfalls in 4 Phasen aufgeteilt: "Analyse, Konzeption, Visualisieren, Überprüfen". Sinn und Zweck des mehrfachen Durchlaufens ist, dass der nächste Schritt erst angegangen wird, wenn das Resultat für die Benutzer zufriedenstellend ist, bzw. das System den Nutzeranforderungen entspricht.

6.2 ANALYSE

In dieser Phase wird analysiert, wer die Benutzer sind und welche Aufgaben sie zu lösen haben. Ebenfalls werden der Nutzungs-Kontext (Requirements Engineering) und das Umfeld analysiert, in welchem das System genutzt wird. Diese Analysen können mit verschiedenen Methoden durchgeführt werden, z.B. mit Interviews und Workshops. Aus den gesammelten Informationen lassen sich anschliessend Personas, Use Cases und Flussdiagramme ableiten.

Als Ergänzung zur ISO-Norm haben wir in der Analyse-Phase die Einflussfaktoren der Usability berücksichtigt (Shackel, 2001): Benutzer, Aufgabe, System, Umfeld.

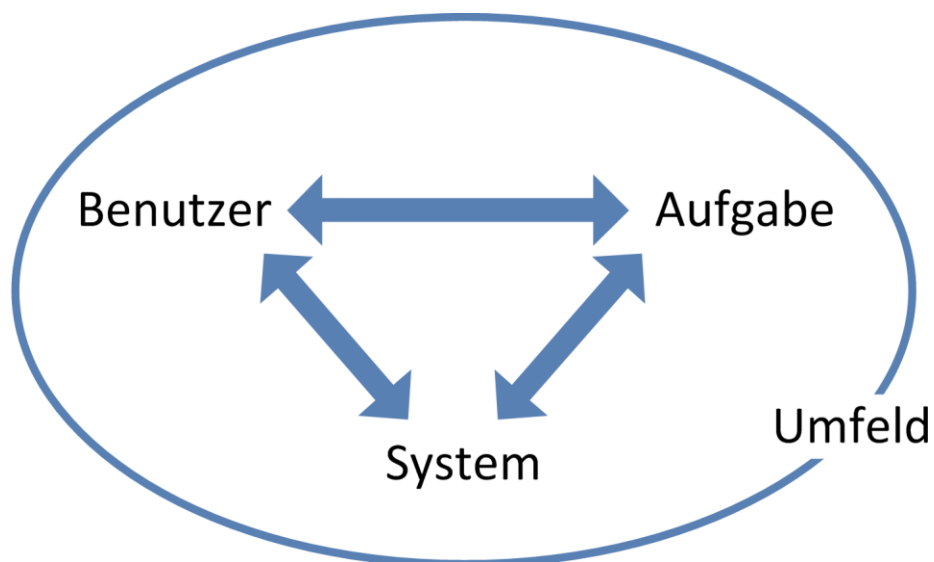


Abbildung 3: Einflussfaktoren der Usability (Shackel, 2001)

In Bezug auf Tonalität haben wir untersucht, in welchen Formen sie im Webformular angewandt wird und welche Bedingungen diese mit sich bringt. Hier mussten wir jedoch auf Anraten einer Fachperson einen Werbetexter als Berater hinzuziehen, was äusserst lehrreich war.

6.2.1 KONZEPTION

Wenn das Requirements-Engineering abgeschlossen ist, kann das Systemkonzept festgelegt werden. Ein Teil davon ist die Informationsarchitektur (welche benutzerrelevante Information befindet sich wo) und das Interaktionskonzept (wie der Benutzer zwischen Seiten navigiert). Das Interaktionskonzept kann in Form von ersten Wireframes,

Vorgehen

Flussdiagrammen oder Skizzen dargestellt werden. Auch ein Inhaltskonzept ist nötig. Es definiert, welche Informationen relevant sind, damit der Benutzer seine Aufgaben durchführen kann.

Diese Konzepte sind das Fundament des Systems und entscheidend für die nächsten Schritte der Visualisierung.

6.2.2 VISUALISIEREN

Hier wird das System gestaltet. Dies kann in Formen von Prototypen mit unterschiedlichen Detaillierungsgraden umgesetzt werden: Papier-Prototypen, klickbare HTML-Prototypen bis zu fertig gestalteten, funktionierenden Benutzeroberflächen, die dem System entsprechen. In diesen Schritten werden Styleguides und Guidelines entwickelt und dabei bestehende Designkonzepte berücksichtigt. Nun sind die Prototypen so weit entwickelt, dass sie mit repräsentativen Nutzern des Systems im nächsten Schritt überprüft werden können.

6.2.3 ÜBERPRÜFEN

Es gibt viele Methoden ein System zu überprüfen: Usability-Tests mit Benutzerbefragungen, Beobachtungen der Benutzer am Einsatzort des Systems, Evaluationen des Systems anhand von Checklisten und Normen. Diese Tests bringen wichtige, positive wie auch negative Aspekte der Usability, User Experience und Ergonomie des geprüften Systems hervor.



Dorota Przygienda bei der Besprechung der Projektplanung im Team.

6.3 PROJEKTPLAN, ORGANISATION UND RISIKEN

6.3.1 PROJEKTPLAN

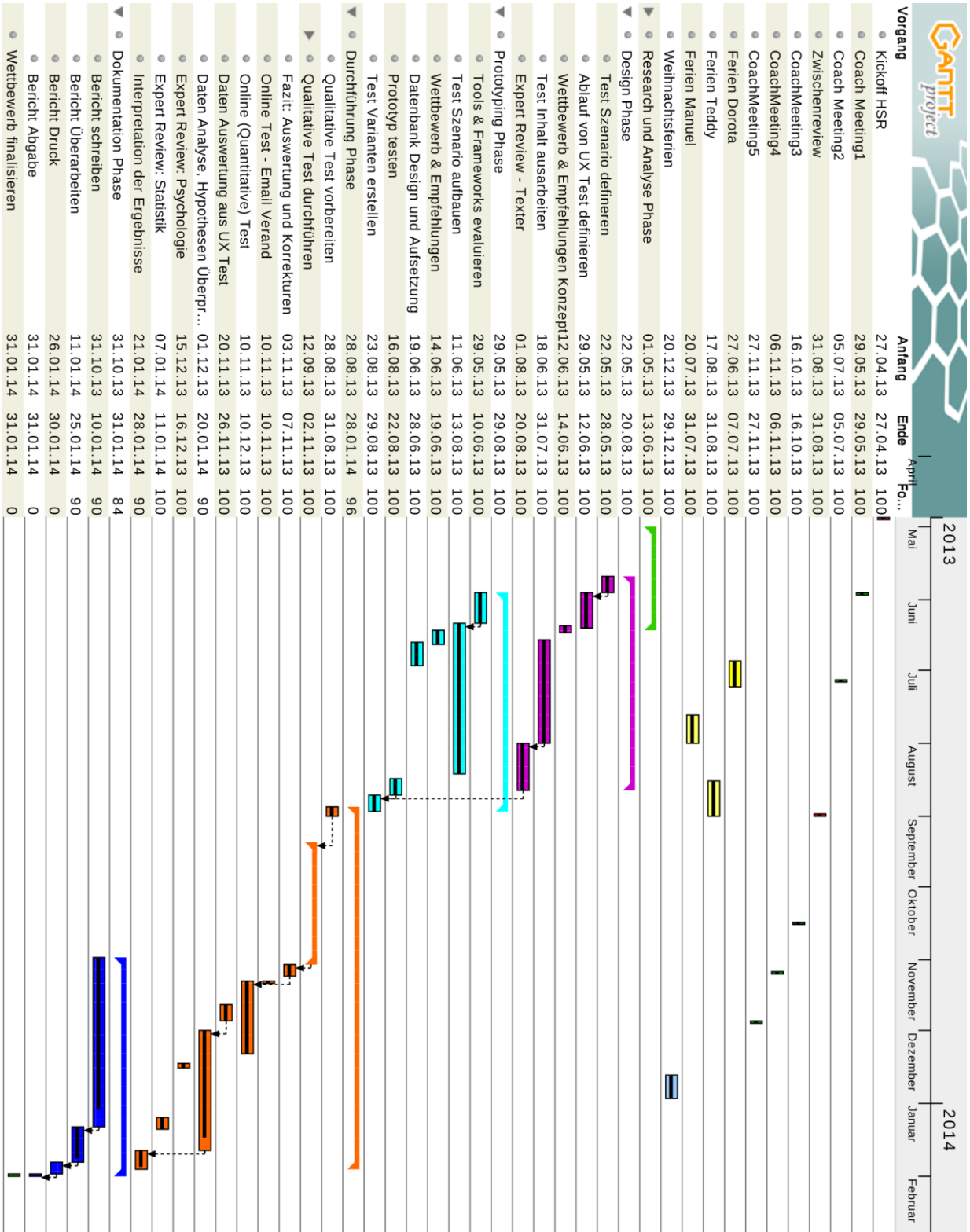


Abbildung 4: Projektplan

Projektphasen und Zwischenziele wurden am Anfang gemeinsam im Team diskutiert und in einem Projektplan festgehalten. Wichtige Meilensteine und geplante Ferien wurden im Projektplan erfasst (Anhang: A003). Diese Vorgehensweise basiert auf dem iterativen UCD Modell ISO 9241-210, beschrieben im Kapitel 1.2 und wird an die Bedürfnisse des Projektes wie folgt angepasst:

Prozess ISO 9241-210

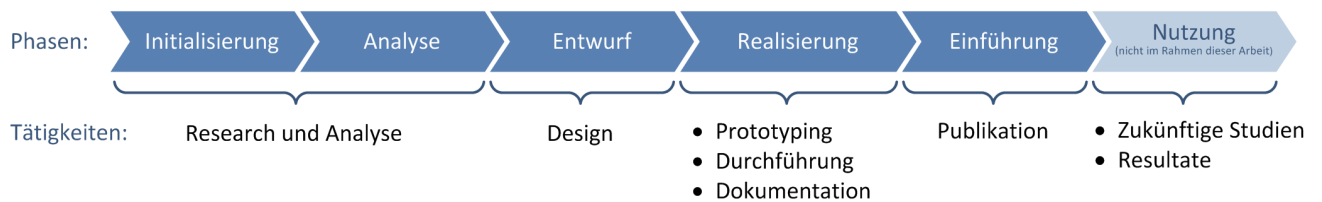


Abbildung 5: Projekt Tätigkeiten im Bezug auf ISO Standard Prozess

Research und Analyse Phase

Hauptgewicht in der Phase liegt in der Suche nach relevanten Publikationen und Knowhow Aufbau. Dank der Unterstützung von Prof. Dr. Markus Stolze sind gleich am Anfang relevante, wissenschaftliche Studien der Universität Basel (Kapitel 4, Literaturliste) für uns verfügbar. Durch die Literatur- und Web-Recherche entsteht eine gemeinsame Wissensgrundlage im Projekt Team, was zu Formulierung der Forschungsfragen führt. Des Weiteren wurde ein passendes Vorgehensmodell und Methodik ausgewählt und Ablauf einer wissenschaftlichen Studie grob festgelegt, sowie ein standardisierter Fragebogen evaluiert.

Design Phase

Im Fokus dieser Phase steht das Design der empirischen Studie. Zuerst werden das Test Szenario und der Ablauf der Studie definiert. Ein Wettbewerbs-Konzept wird entwickelt, um die Test Personen für Teilnahme an der Studie zu motivieren. In zweiter Iteration, ist das um einen Empfehlungs-Konzept erweitert, ähnlich wie bei einem Schneeball System. Der gesamte Inhalt des Online-UX-Tests wurde in einer zusätzlichen Iteration bearbeitet.

Ein weiteres Ergebnis dieser Phase ist die Entwicklung eines Tonalitäts-Konzept und das Verfassen von Fehlermeldungen zu jeder Tonalität. So definierte Fehlermeldungen werden in einem Expert-Review von einem Werbetexter validiert.

Entwicklung von Tonalität Konzept und Verfassen von Fehlermeldungen zu jeder Tonalität ist auch ein wichtiger Ergebnis dieser Phase. Fehlermeldungen in jeder Tonalität werden verfasst und in Expert-Review von einem Werbetexter validiert.

Prototyping Phase

In diese Phase steht der Aufbau des Prototyps im Vordergrund. In der ersten Iteration werden die verfügbaren Tools und Frameworks evaluiert und mit einer Machbarkeitsstudie geprüft. Darauf folgt der Aufbau des Test Szenarios basierend auf MySQL Datenbank, PHP Server, HTML und Javascript. In zweiten Iterationen wurden der Wettbewerb und die Empfehlungs-Funktionalität umgesetzt. In einer weiteren Iteration werden die 5 Tonalität Varianten bei Fehlermeldungen angebunden und automatisch zum Test zugeteilt. Der fertige Prototyp wird vorab durch das Projekt Team getestet.

Durchführung Phase

Einer der wichtigen Aspekte in dieser Phase war die Validierung des Prototyps, in einer Reihe von qualitativen UX Tests, die jeweils in den Räumlichkeiten unserer Arbeitsgeber und am Flughafen Zürich stattgefunden haben. In mehreren Iterationen wurde der Prototyp verbessert und schlussendlich bei der Durchführung der online Studie eingesetzt. Darauf folgten die Datenauswertung und Analyse der Ergebnisse.





Vierter qualitativer User Experience Test mit Proband bei UBS Zürich.



Dritter qualitativer User Experience Test bei CSS Versicherung in Luzern.

Aus den Beobachtungen und Erkenntnissen entstanden einige Verbesserungsvorschläge, die in weiteren Iterationen in den Prototypen eingebaut und danach erneut validiert wurden. Als letzte Validierung des Prototyps wurde eine Reihe von qualitativen UX Test mit zufälligen Probanden vorgesehen, die an einem Tag am Flughafen Zürich in der Abflugzone durchgeführt wurde. Anschliessend begann die Etappe von quantitativen online UX Test. Die Einladung zur Teilnahme an der Umfrage wurde per E-Mail an unsere Bekannten und Freunde versendet. Die Datenerhebung dauerte einen Monat zwischen 10.Nov.2013 und 10.Dez.2013. Ursprünglich wurden nur 2 Wochen eingeplant. Jedoch dabei war das Risiko zu hoch, nicht genügend gültige Daten in so kurzer Zeit zu erheben. Dank der Hilfe unseres Coach Prof. Dr. Markus Stolze und der Platzierung des Links zur Umfrage auf Socialmedia Plattformen, konnten weitere Testpersonen für die Studie rekrutiert werden.

Mit der Analyse der erhobenen Daten wurde bereits während der Laufzeit der Umfrage begonnen. Nach Abschluss der Umfrage wurde die Hypothese mit statistischen Methoden überprüft. Die Korrektheit der ausgewählten statistischen Methoden wurde in einem Expert-Review durch einen Mathematiker validiert

Dokumentation Phase

Die Resultate aus der Studie wurden in zahlreichen Diagrammen dargestellt und dokumentiert. Das Masterarbeit Dokument wurde verfasst und ausgearbeitet.

6.3.2 RISIKEN

Zu Beginn der Arbeit haben wir einige Risiken erkannt, die uns zur Suche nach Gegenmassnahmen veranlasst haben. Das grösste Projektrisiko war, dass wir zu wenig Testpersonen für die Studie rekrutieren können. Ein kritischer Aspekt, der sich auf die Datenanalyse und schliesslich auf die Ergebnisse der Studie hätte auswirken können. Um dieses Projektrisiko zu minimieren und um möglichst viele Testpersonen für die Teilnahme zu motivieren, haben wir ein Wettbewerbs-Konzept entwickelt. Eine weitere Massnahme, die Rekrutierung zu unterstützen, war das Einbeziehen von privaten sozialen Netzwerken. Erstens wurden Personen aus dem eigenen Bekannten- und Freundeskreis angefragt und zur Teilnahme eingeladen, was zu einer grossen Rücklaufquote führte. Zweitens haben wir die Rekrutierung über die Social Media Plattformen (XING, Google+) genutzt und die Laufzeit der Studie verlängert. Als weitere Gegenmassnahme wurde ein Empfehlungs-Konzept entwickelt, mit dem jeder Teilnehmer bis zu 5 weitere Personen zur Teilnahme an der Studie einladen konnte. Ein Wettbewerbs-Konzept, mit einem iPad mini als Hauptpreis, war auch als Motivation der Testpersonen gedacht. Die Gewinnchancen stiegen mit jeder abgeschlossenen Empfehlungs-Teilnahme.

Ein weiteres Risiko für das Projekt war die Dauer der online Umfrage: Wie viel Zeit würden die Testpersonen bereit sein zu opfern, um an der Studie teilzunehmen? Der ausgewählte Fragebogen QUIS ist relativ lang und enthält viele Fragen, die für unsere Studie irrelevant sind. Als Massnahme wurde hier die Kürzung von QUIS um die irrelevanten Fragen und Kapitel ergriffen. Im Rahmen eines HCID Peer Review mit unseren Kollegen wollten wir die optimale Dauer einer solchen Umfrage in Erfahrung bringen.

In einem späteren Zeitpunkt haben wir weitere hohe Risiken im Bereich Datenauswertung und Statistik erkannt. Diese Themen waren Neuland für das gesamte Projektteam und das Wissen und die Vorgehensweise mussten wir zuerst lernen. Um das Risiko zu minimieren,

liessen wir uns zum Thema Statistik beraten und führten Expert-Reviews durch. Die ersten statistischen Auswertungen haben ergeben, dass die erhobenen Daten keine Normalverteilung aufzeigten. Aus diesem Grund mussten andere Methoden zur Untersuchung des Signifikanz Niveaus eingesetzt werden. Um das Risiko der Methodenwahl zu minimieren war die Beratung eines Profi-Statistikers im Rahmen eines Expert Reviews erforderlich.

Zu einem weiteren Risiko wurden statistische Analysen, da keine signifikanten Unterschiede über die gesamten Tonalitätsgruppen gefunden werden konnten. Methoden hierfür und der Einsatz geeigneter statistischer Tools wären zu komplex und das Vorgehen zu zeitaufwändig. Hier wählten wir die einfachere Option, die Analysen und Diagramme in MS Excel zu erstellen.

Die Operationellen Risiken in Bezug auf technische Infrastruktur und Datenverluste wurden zu Beginn des Projektes erkannt und mittels Backups, Datenabzug und Datenbank Kopie minimiert.

6.3.3 ORGANISATION

Eine optimale Grösse des Projektteams und unsere gute Kommunikation haben eine flexible Organisation und dynamische Zusammenarbeit ermöglicht. Je nach Standpunkt der Arbeiten und Situation war es möglich, sich abzusprechen und gemeinsam Entscheidungen zu treffen. Auf Grund örtlicher Entfernung war oft Skype und Telefon im Einsatz für Team Besprechungen. Auch andere online Tools wie DropBox, Google-Kalender und Google-Docs wurden verwendet für Daten Ablage und Organisation des Projektes.

Die Team Meetings und Projektarbeitstage waren je nach Bedarf geplant und fanden in den Räumlichkeiten unserer Arbeitgebern statt: AdNovum Zürich, CSS Luzern, UBS Zürich. Die Aufgabenverteilung geschah flexibel und wurde je nach Bedarf gemeinsam entschieden. Unsere Berufserfahrung und Knowhow kam dadurch gut zum Einsatz aber auch die eigenen Interessen und Entwicklungsbedürfnisse konnten berücksichtigt werden. Darum werden die Rollen im Team eher als Kernkompetenzen gesehen und jeder kann auf dem Gebiet entsprechend mitwirken.

























Thema	Manuel	Dorota	Teddy
Design & Grafik			
CSS			
Prototyping			
Online Test Aufbau			
PHP, Javascript, HTML			
UX Fragebogen			
Statistik			
Datenbank			
Datenauswertung			
Analyse der Ergebnisse			
Projekt Mgmt.			

Abbildung 6: Rollen und Aufgaben im Team

 Hauptverantwortung

 Mitarbeit

7 METHODEN

Um Daten analytisch mit echten Benutzern in einem Webformular zu erheben, gibt es heute verschiedene gratis Online Angebote. Wir haben uns für den Onlinedienst «Wufoo» (www.wufoo.com) entschieden, da er Teil des bekannten SurveyMonkey.com Dienstes ist und daher einen seriösen Eindruck auf uns machte.

7.1 PROTOTYPING

Wufoo ist ein Online Datenerhebungsdienst, welcher die Möglichkeit anbietet, mit bestehenden, bewährten Formularelementen und Design-Patterns ein modernes Formular innert kürzester Zeit zu erstellen und damit eine Online-Umfrage zu starten. Die gesammelten Daten können ebenfalls auf Wufoo ausgewertet, dargestellt oder in einem anderen Format in weitere Programme exportiert werden. Somit hatten wir genau das gewünschte Werkzeug, das wir in der Visualisierungsphase benötigten.

Mit relativ wenig Aufwand von ca. 4 Stunden Programmierarbeit ist es uns gelungen, das Formular von myschweiz.ch optisch originalgetreu nachzubauen und für aktuelle Browser bereitzustellen. Funktional ergaben sich einige Schwierigkeiten bei der Platzierung und der Wortwahl der Fehlermeldungen. Dies konnte jedoch mit zusätzlichem Aufwand mit CSS behoben werden. In einer weiteren Iteration der Entwicklung stellte sich heraus, dass wir das Design auf der Umfangebene ändern mussten: Es sollte möglich sein, an der gleichen Stelle im Formular, dem User unterschiedliche Fehler aufzuzeigen. Wufoo bietet grundsätzlich nicht an, dass Fehlermeldungen editiert werden können. Auch das Verhalten des Formulars kann nur beschränkt angepasst werden. Für uns war es wichtig, das Verhalten des Users mitverfolgen zu können. Dies führte im Team zur Entscheidung, den Online-Test von Grund auf selbst zu entwickeln. Wufoo hat uns jedoch geholfen, unsere Requirements genauer zu erheben und die Erkenntnisse daraus in die Entwicklung der finalen Prototypen einfließen zu lassen.

7.2 EXPERT REVIEW

Bei der Evaluations-Methode Expert Review (Goodwin, 2009) prüft ein erfahrener Design- oder Usability-Experte Schritt für Schritt das Produkt oder das Design einer Anwendung anhand etablierter Usability-Standards und Richtlinien. Dadurch werden Probleme entdeckt und Verbesserungspotentiale hervorgehoben. Diese Methode benötigt wenig Zeit und ist nicht teuer, das Review soll von einem Profi seines Faches durchgeführt werden.

In unserer Arbeit haben wir diese Methode bei den Tonalitäts-Fragen zusammen mit einem Werbetexter angewendet, um unsere definierten 5 Varianten der Tonalität und deren Anwendung bei Fehlermeldungen im Webformular zu besprechen und zu prüfen. Den gesamten Test haben wir von 2 Psychologen mit dieser Methode beurteilen lassen, um zu erfahren, ob der Tonalitäts-Aspekt tatsächlich mit dem QUIS-Fragebogen abgebildet werden kann. Ein weiterer Expert Review kam bei der statistischen Auswertung der Daten des QUIS-Fragebogens zum Zuge, um die Signifikanz unserer Daten und das Mapping QUIS nach User Experience zu prüfen.

7.3 USABILITY TESTS

Der Cognitive Walkthrough ist eine Usability-Inspektionsmethode und gehört zu den analytischen Evaluationsverfahren bei dem sich ein Experte in einen hypothetischen Benutzer versetzt und konkrete vorgegebene Handlungsabläufe analysiert (Clayton Lewis, John Rieman, and Cathleen Wharton, 1990).

Bei der benutzerorientierten Entwicklung unseres Tests haben wir uns an die die Methode Usability Walkthrough nach Richter und Flückiger gehalten. Dort wird die Methode als formativer Usability Test beschrieben, bei dem der Testleiter den Benutzer durch den Prototyp begleitet (Richter und Flückiger, 2010). Zusätzlich haben wir diese Methode ergänzt mit ethnographischen Interviews am Flughafen Zürich nach Cooper et al. (2010, S.84-93). Dabei werden die Benutzer bei ihren Aktivitäten beobachtet und in ihrer Umgebung befragt, bei der sie ihre Arbeit tagtäglich ausführen. Dies kann wichtige Details ihrer Verhaltensweisen ans Licht bringen. Wir wollten mit dieser Methode herausfinden, ob sich Unterschiede ergeben bezüglich der Akzeptanz der Tonalität, wenn sich die Benutzer allenfalls in Ferien- oder Reisetimmung befinden.



Manuel Sanchez und Theodor Huber bereiten die Technik für den Usability Walkthrough vor.

7.4 BENUTZERFORSCHUNG: STANDARDISIERTER FRAGEBOGEN (QUIS)

Der QUIS-Fragebogen (engl. Questionnaire For User Interaction Satisfaction) wurde 1987 entwickelt, um die subjektive Zufriedenheit von spezifischen Aspekten der Schnittstelle in Mensch-Computer-Systemen zu ermitteln. In der aktuellen Version 7.0 werden folgende Inhalte erfasst:

- Demographischer Angaben
- Die Bewertung der allgemeinen Zufriedenheit mit dem Gesamtsystem mittels 6 Items
- Neun Bewertungsbereiche, die jeweils die Zufriedenheit des Benutzers mit einer spezifischen Schnittstellenkomponente erfassen. Diese Bereiche sind:
 1. System-Erfahrungen
 2. bisherige Erfahrungen
 3. Bildschirm
 4. Terminologie und System-Informationen
 5. Lernfortschritt
 6. System-Eigenschaften
 7. Technische Dokumentation und Online-Hilfe
 8. Online-Anleitung
 9. Multimedia Komponenten

In unserer Masterarbeit war der QUIS-Fragebogen die Methode, um die User Experience zu messen und auszuwerten. Der Entscheid ist auf QUIS gefallen, da dieser Fragebogen bereits in ähnlichen wissenschaftlichen Arbeiten angewandt wurde, u.a. in der Arbeit "Error prevention in online forms" von Christian Hübscher et al (2009). Des Weiteren wird der Fragebogen im Buch «Quantifying the user experience» von Jeff Sauro und James R. Lewis (2012) als Standard-Fragebogen zur Messung der User Experience erwähnt.

8 UMSETZUNG



Abbildung 7: Anwendung der Tonalität bei Fehlermeldungen und Komponenten des Mensch-Maschine-Systems beim Online-UX-Test

8.1 KONZEPT DER TONALITÄT

Der Grundgedanke hinter dem Tonalitäts-Konzept war inspiriert durch die Erwartungshaltung der Hotelgäste. Wie möchten sie angesprochen werden? Gilt der gleiche Kommunikationsstil sowohl für die Hotelrezeption als auch beim Online-Buchungssystem? Können sich die Benutzer vorstellen, dass sie durch das GUI des Buchungssystems mit einer realen Hotelrezeption und echten Menschen kommunizieren? Falls ja, würden sie vielleicht eine passende Tonalität erwarten. Auf Grund der verschiedenen Hotel-Kategorien und Persönlichkeitstypen der Hotelgäste, wurden die Tonalitäts-Varianten «höflich», «humorvoll» und «motivierend» entwickelt. «Standard» entspricht den heutigen Normen in Webformularen und «unpersönlich» basiert auf der typischen Ausdrucksweise für Systemmeldungen. Alle 19 Fehlermeldungen, die im Online-UX-Test auftreten können, wurden in allen 5 Tonalität Varianten verfasst, iterativ verbessert und schliesslich in einem Expert Review durch einem Werbetexter validiert. (Anhang: A002)

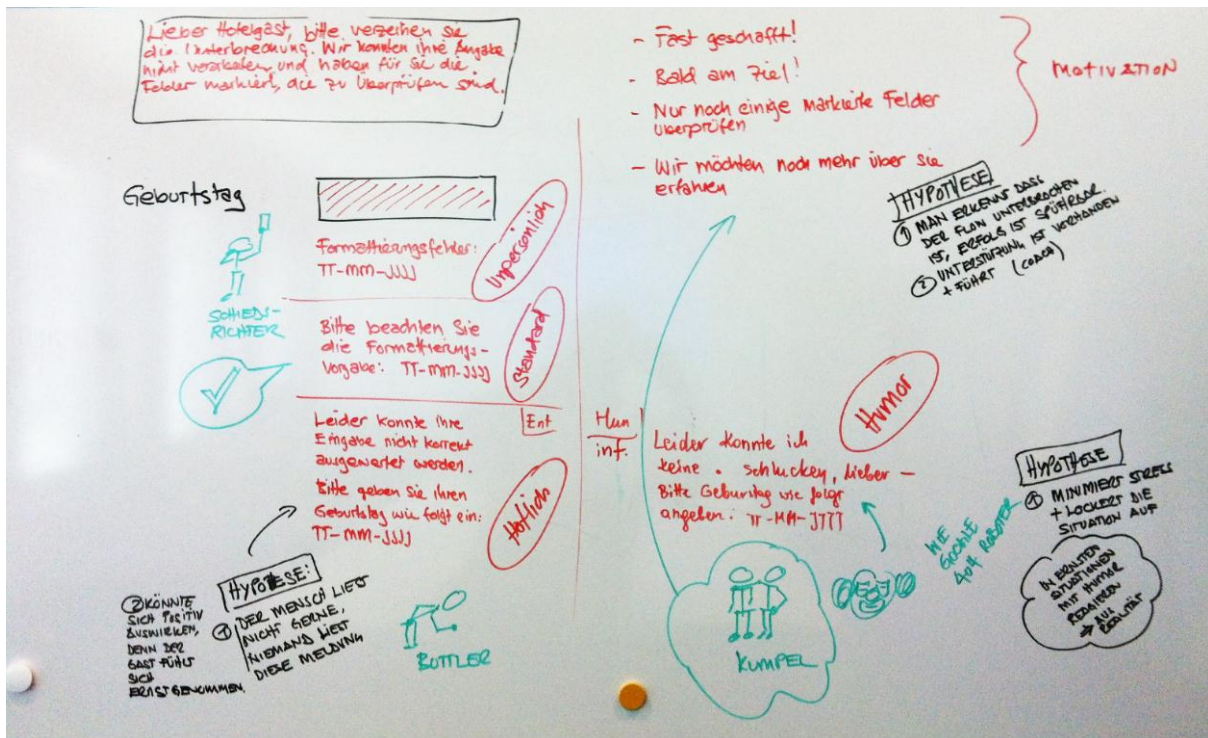


Abbildung 8: Erste Ideen zum Konzept der Tonalität wurden an der Wandtafel festgehalten und diskutiert






Tonalität	Beschreibung
	<h3 data-bbox="469 327 576 358">Höflich</h3> <p data-bbox="469 389 1318 651">Es ist eine sehr höfliche Ausdrucksweise, wie von einem Butler in einem Luxushotel oder wie sie in einem noblen Restaurant erwartet wird. Der Butler hat einen gehobenen Sprachstil, ist den Gästen sehr behilflich und zuvorkommend, damit sie sich wohl und königlich behandelt fühlen. Bei den Benutzern unseres Hotelbuchungsformulars soll diese Tonalität ähnliche Gefühle entstehen lassen.</p>
	<h3 data-bbox="469 712 627 743">Humorvoll</h3> <p data-bbox="469 770 1305 880">Witzig formulierte Fehlermeldungen sollen bei dem Benutzer für Auflockerung sorgen und eventuell die negativen Gefühle bei einem Fehler beseitigen.</p> <p data-bbox="469 902 1305 1014">Wie ein Clown, der mit Humor und gutem Taktgefühl, ein Lächeln hervorrufen und positive Gefühle wecken kann, sollte diese Tonalität bei den Benutzern ähnliche Wirkung haben.</p>
	<h3 data-bbox="469 1070 655 1102">Motivierend</h3> <p data-bbox="469 1128 1318 1283">Ein Sprachstil, wie man ihn von einem Coach erwarten würde, der trotz Missgeschicken und Fehlern Verständnis zeigt und motivieren kann, weiter zu machen. Ähnlich soll der Benutzer durch die Tonalität in der Fehlermeldung motiviert werden.</p>
	<h3 data-bbox="469 1395 608 1426">Standard</h3> <p data-bbox="469 1453 1305 1563">Diese Formulierung entspricht den heutigen Standards bei Fehlermeldungen. Eine höfliche Ausdrucksweise bei der «Bitte...» als selbstverständlich gilt.</p>
	<h3 data-bbox="469 1682 671 1713">Unpersönlich</h3> <p data-bbox="469 1740 1278 1883">Kurze und prägnante Fehlermeldungen, die unpersönlich und emotionslos sind, wie man sie von einem System oder Roboter erwarten würde. Diese Tonalität basiert auf dem alten Stil von Systemmeldungen, der heute immer noch verwendet wird</p>

Abbildung 9: Definition der Tonalitäten bei Fehlermeldungen

8.2 AUSWAHL TEST SCENARIO - MYSWITZERLAND.CH

Zu Beginn sammelten wir alle Aspekte der Usability und der User Experience die im Zusammenhang mit Webformularen für uns wichtig schienen. Diese stammten aus der Literaturrecherche und unserem Lehrmaterial und wurden im Team besprochen. Von da an konnten wir das Test-Szenario Konzept entwickeln und begannen mit der Suche nach einem geeigneten Formular, welches die folgenden Kriterien abdecken musste:

- Usability
- das Formular musste Teil einer Website mit hoher Benutzerfrequenz sein
- zeitgemässes Visual-Design
- Responsive Design berücksichtigen
- insgesamt den Ansprüchen aktueller Standards genügen (Best of Swiss Web)

Damit wir sicher sein konnten, dass unsere Test-Daten nicht durch fremde Einflüsse, wie z.B. mangelnde Usability, beeinflusst werden, musste unser Test-Szenario auf einem Webformular aufbauen, welches mit grosser Sorgfalt und dem aktuellen Wissen der Usability und User Experience entwickelt wurde. Doch wie konnten wir uns da sicher sein? Eine erste Anlaufstelle um mit der Suche zu beginnen war die Organisation «Best of Swiss Web» welches unsere Kriterien berücksichtigte und uns am Ende unserer Suche 3 Kandidaten zur Wahl lieferte:

- www.srf.ch (Schweizer Radio und Fernsehen)
- www.sbb.ch (Schweizerische Bundesbahnen)
- www.myswitzerland.com

Myswitzerland.com hat bei der Verleihung des Best of Swiss Web Awards 2012 1x Gold, 2x Silber und den 5. Masterplatz erhalten. Von daher konnten wir sicher sein, dass diese Tourismusplattform mit jährlich 24.3 Mio. Besucher aus aller Welt in die nähere Wahl kam. Sbb.ch gewann im Jahre 2012 den Titel «Master of Swiss Web» und www.srf.ch 2013 Gold im Bereich Usability. Nachdem wir alle 3 Plattformen untersucht und besprochen hatten, entschieden wir uns für die myswitzerland.com Plattform. Ihr Hotelbuchungsformular brachte uns sofort erste Ideen für mögliche Fehler-Szenarios die uns realistisch und umsetzbar vorkamen. Die gute Qualität der Formulare und die Umsetzung der Fehlermeldungen war ein weiterer Grund für diesen Entscheid. Auch das Responsive Design funktionierte tadellos bei ersten Tests mit einem iPad der ersten Generation.

8.3 FEHLERMELDUNGEN – DEFINITION UND VALIDIERUNG

Das Hotelbuchungsformular unseres Online-UX-Tests bestand aus 20 Formularelementen (Eingabefelder, Checkboxen etc.) bei dem die Probanden ihre Eingabe tätigen konnten. Insgesamt war es möglich, damit 19 Fehlermeldungen pro Tonalitäts-Variante hervorzurufen. Diese Texte haben in einer Excel-Tabelle festgehalten und in mehreren Iterationen aufgebessert. Da wir uns weder als Werbetexter noch als Experten der Deutschen Sprache bezeichnen konnten, wurde uns bei der Ausarbeitung der Fehlermeldungen klar, dass wir in weiteren Iterationen dieser Phase, den Text in einem Expert-Review begutachten mussten. Die ersten Varianten der geschriebenen Fehlermeldungen liessen wir von einer Expertin aus dem Bereich Technische Dokumentation gegenprüfen. Dabei konnten wir sicherstellen, dass die Texte einem Qualitäts-Standard entsprechen, die in einem technischen Umfeld angewendet werden können. Nachdem wir uns mehrfach mit dem Thema Tonalität beschäftigt hatten, stellte sich heraus, dass wir zusätzlich zur Expertin einen Werbetexter zum gegenprüfen der Tonalität benötigten. Tonalität gehört in der Werbung zum täglich Brot. Nach ein paar Tagen fanden wir dann einen geeigneten Werbetexter. Das Review war für beide Seiten (Webetexter und Team) sehr interessant und wir konnten konzeptionelle Fragen zum Online-UX-Test und der Anwendung von Tonalität klären. Der Werbetexter erhielt von uns beim ersten Treffen die Tonalität «höflich», «humorvoll» und «motivierend» zum Begutachten. Hier ein Beispiel unserer 5 finalen Tonalitäts-Varianten:



Höflich: Damit wir Sie bei uns gebührend willkommen heissen können, benötigen wir Ihre Ankunftszeit.



Humorvoll: Sagen Sie uns doch, wann Sie anreisen, damit wir den Willkommens-Apéro kühl stellen können. Santé!



Motivierend: Bitte nennen Sie uns Ihre Ankunftszeit, damit wir Sie noch besser bedienen können.



Standard: Bitte geben Sie die Ankunftszeit des Gastes an



Unpersönlich: Eingabe erforderlich

Abbildung 10: Fehlermeldungs-Beispiele in jeder Tonalität

8.4 PROTOTYP

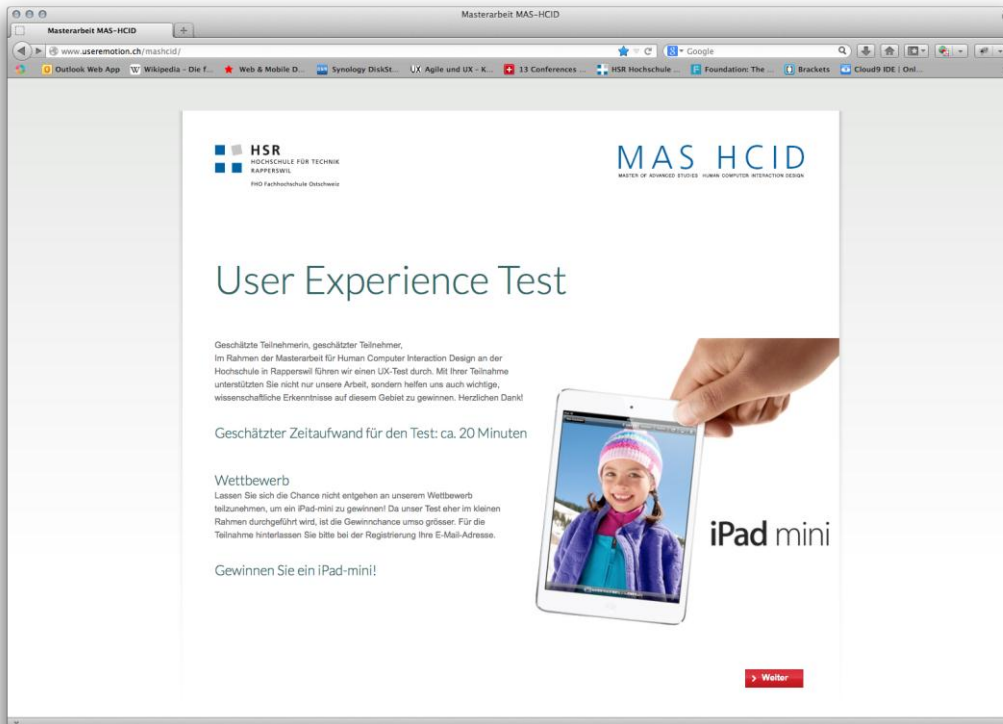


Abbildung 11: Einstiegs-Seite für den Online-UX-Test

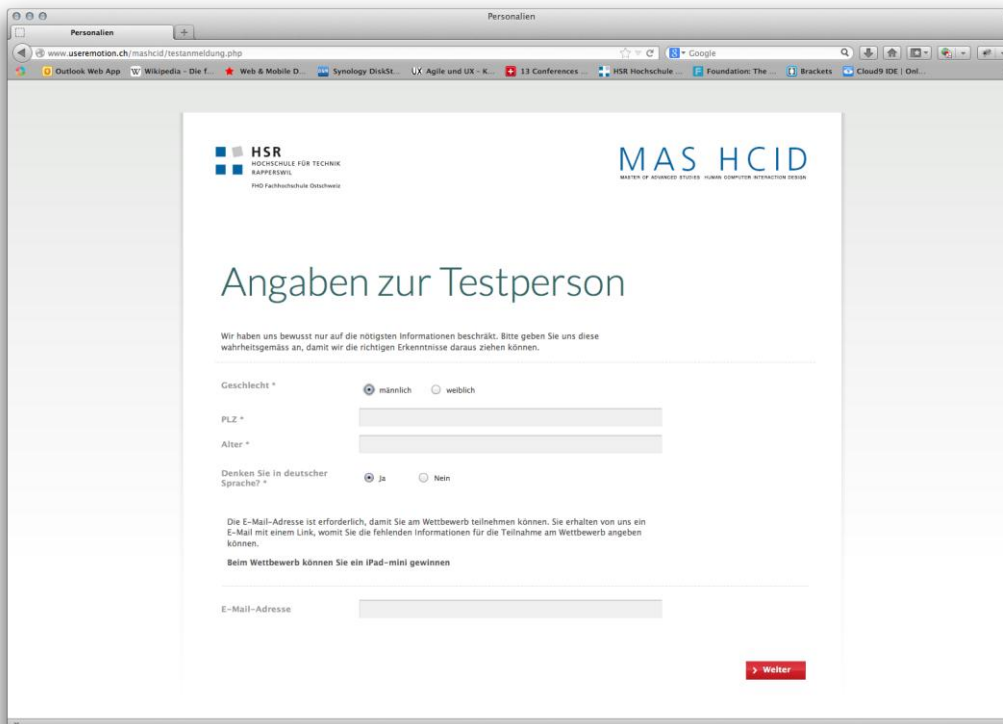


Abbildung 12: Personalien Erfassung im Online-UX-Test

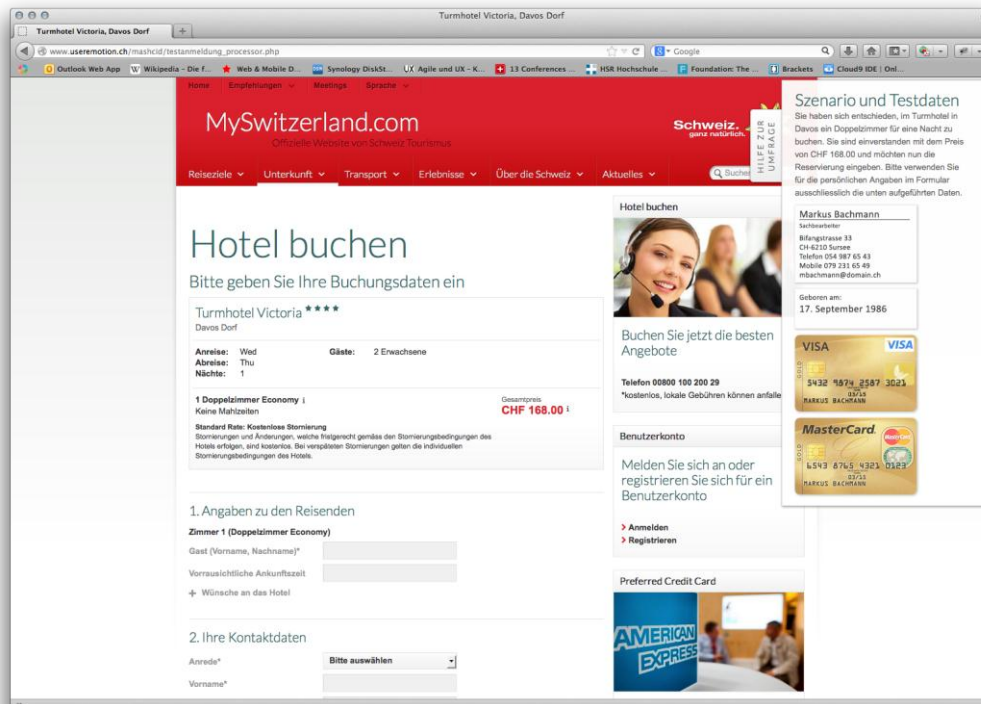


Abbildung 13: Formular und Hilfestellung (rechts) des Online-UX-Tests

Da wir im Team noch keine Erfahrung mit Wufoo hatten, entschieden wir uns zu Beginn der Prototyp-Phase als «Plan B» bereits mit der Entwicklung eines selbst programmierten Formulars zu beginnen und parallel dazu erste Wufoo-Entwürfe zu erstellen. Wufoo war von Anfang an ein attraktives Angebot, um schnell unser gewünschtes Formular online darzustellen und auch im Team zu besprechen. In unseren Diskussionen konnten wir mit Wufoo schnell eine gemeinsame Sprache für die einzelnen Komponenten entwickeln und gleichzeitig unser Konzept des Online-UX-Tests weiterentwickeln und prüfen. Ein weiterer Faktor für den Entscheid für Wufoo war, dass sich ein Wufoo-Formular praktisch in jede gestaltete Umgebung einbinden lässt, ohne dass für den Benutzer Wufoo erkennbar ist. Dies ermöglichte uns eine optisch identische Variante des mySwitzerland.com-Buchungsformulars zu erstellen. Auf der funktionellen Ebene sind wir dann doch nach kurzer Zeit an die Grenzen des machbaren gestossen und wechselten dann zu «Plan B», dem selbst entwickelten Online-UX-Test.

Beim selbst entwickelten Online-UX-Test handelt es sich clientseitig um HTML Seiten, welche serverseitig mit PHP Scripts arbeiten. Wir mussten vor allem Daten aus der Datenbank in die HTML Seite transportieren. Was ohne PHP und mit den Rahmenbedingungen eines Frameworks sehr schwierig geworden wäre. Ferner haben wir spezifische Fehlerkonstellationen welche es nötig machte, auf die Eingabe des Benutzers Rücksicht zu nehmen. Aus diesem Grund haben wir uns entschlossen auf Javascript zu setzen für die Validierung, und auf PHP für die serverseitige Aufbereitung und Speichern der Daten.

8.5 ITERATIVE ENTWICKLUNG FINALER USER EXPERIENCE TEST

8.5.1 FORMULAR AUFBAU

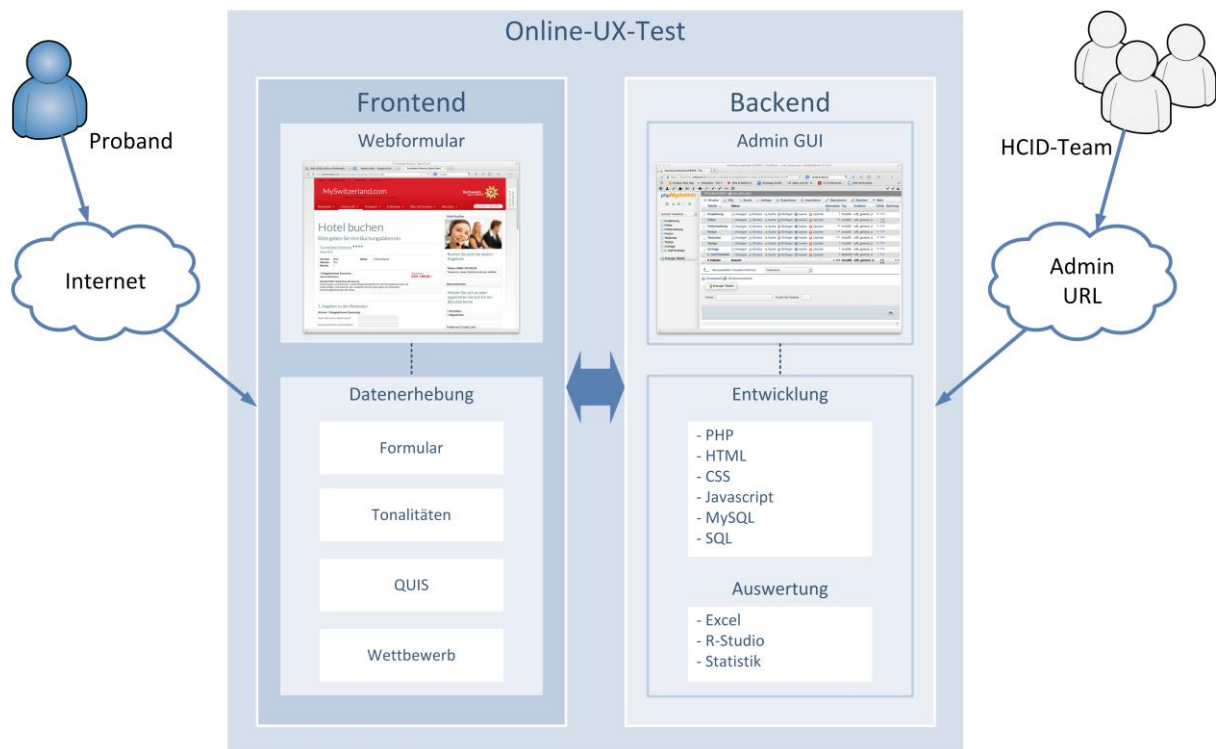


Abbildung 14: Komponenten und Aufbau des Online-UX-Tests

Technologiewahl

Durch die Prototyping Phase mit Wufoo sind wir schnell zur Erkenntnis gelangt, dass das Design zwar sehr ansprechend ist, aber auf der anderen Seite die Einschränkungen welche uns das Framework auferlegt, zu gross sind, um entsprechend allen Anforderungen im Test zu genügen. Aus diesem Grund sind wir zur Einsicht gekommen, dass eine Umsetzung mit PHP für unsere Zwecke ideal wäre. Dies lässt vielseitige Gestaltung zu, aber auch Themen wie Datenbankeinbindung sind sichergestellt.

Website clonen

Unsere Entscheidung, eine Website für unseren Test zu verwenden, welche einerseits allgemein akzeptiert und andererseits aus offizieller Stelle als gut befunden wurde, führte uns, wie bereits erwähnt, zu mySwitzerland.com. Es stellt sich nun die Frage, sollen wir die Seite komplett eigenständig neu aufbauen, visuell angelehnt an das Original, oder gibt es hier Hilfestellung dafür? Die Antwort erhalten wir bereits nach kurzer Recherche im Netz. WinHTTrack Website Copier 3.47-14 (Open Source offline Browser). Diese Gratis Software ermöglicht es uns, die gesamte Website von mySwitzerland.com (ab einer gewissen Stelle) herunter zu laden. Von nun an haben wir sämtliche HTML und CSS Files. Natürlich geht es nicht ohne manuelle Eingriffe. Aber dieser Website Kopierer hat uns sehr viel Arbeit gespart, wenn wir stattdessen die ganze Registrierungsseite hätten nachbauen müssen.

Formularstruktur

Das Formular von mySwitzerland.com ist mit einer aktiven Breite von 1000 Pixel zentriert dargestellt. Die aktive Breite ist in drei Drittel unterteilt. Zwei Drittel auf der linken Seite sind für die Formulareingabe und ein Drittel rechts ist für Werbung reserviert.

Die Label sind links des Eingabefeldes positioniert und auch links ausgerichtet. Die Eingabefelder werden in einheitlicher Grösse angezeigt (als Block). Einzig wenn auf einer Zeile zwei Eingaben erwartet werden (z.B. PLZ und Ort) teilen sich diese zwei Eingabefelder die einheitliche Breite auf. Meistens ergeben sich ein kleineres und ein grösseres Feld.

Fehlermeldungen werden bei mySwitzerland.com grundsätzlich keine dargestellt. Fehler aufgrund fehlender Eingabe zeigt das System, indem das Label rot eingefärbt wird, und die Hintergrundfarbe des Eingabefeldes mit einem hellem stark gebrochenem rot eingefärbt wird.

Ganz oben in der Seite wird ein Hinweis angezeigt, dass Fehler auf der Seite entstanden sind.

Die Darstellung und Positionierung der Fehlermeldungen orientiert sich an den wissenschaftlichen Grundlagen, wie bereits im Kapitel 4.3 erwähnt.

Manuelle Eingriffe in das bestehende Formular

Die Fehlerbehandlung ist für unsere Zwecke nicht genügend. Wir müssen eine Fehlermeldung anzeigen, damit wir mit verschiedener Tonalitäten testen können. Die Fehlermeldung wird jeweils unterhalb des Eingabefeldes links aligniert dargestellt. Der Fehlertext soll dasselbe Rot erhalten, wie mySwitzerland.com die Labels damit einfärbt. Die Hintergrundfarbe des Eingabefeldes bleibt im selben hellen Rot wie bei mySwitzerland.com.

Ferner mussten wir verschiedene Probleme bei der Seite von mySwitzerland.com lösen. Eine Unschönheit war, dass die Fehlermeldungen nicht alle einheitlich daher kamen. Der grösste Teil der Fehler ist geselltet, wie oben bereits beschrieben und bildet entspricht den Standard von mySwitzerland.com. Bei Fehler im Bereich der Kreditkarte wird jedoch stattdessen eine Alert Box angezeigt. Bei Comboboxen wird die Hintergrundfarbe nicht auf Fehler gesetzt. Diese unterschiedliche Art und Weise, wie auf fehlerhafte Eingabe reagiert wird, wollen wir nicht so belassen, und entscheiden uns alle Eingaben dem Standard anzugleichen.

Validierung

Für die Validierung wird eine Javascript Funktion erstellt, welche alle Eingaben überprüft. Wir entscheiden uns, die Fehler bei jedem Request in der Datenbank festzuhalten. Dabei wollen wir wissen, wie viele und welche Fehler pro Durchlauf entstanden sind und welche Wiederholungsfehler waren. Ebenfalls die Durchlaufzeiten zwischen den Commits werden in der Datenbank festgehalten. Vollständigkeitshalber entscheiden wir uns die Fehlermeldung jeweils pro Fehler ebenfalls festzuhalten. Von den fünf Tonalitäten wird jeder Proband jeweils nur mit einer Tonalität konfrontiert. Auf diese Weise sehen wir direkt beim geschriebenen Fehler in der Datenbank welche Ausprägung von Fehlermeldung dem Probanden angezeigt wurde, und allenfalls bei Wiederholungsfehler können auf diese Weise Rückschlüsse auf die Qualität der Fehlermeldung gezogen werden.

Vorprogrammierte Fehler für Online-UX-Test

Damit der User mit genügend Fehlermeldungen konfrontiert wird, werden künstliche Hürden in die Anwendung eingebaut, welche unabhängig von der Eingabe einer Fehlersituation entsteht. Folgende vorprogrammierte Fehler wurden vorgesehen:

- Geburtsdatum: Datumsformat TT-MM-JJJJ erforderlich. Wir gehen davon aus, dass niemand ohne Vorgabe das Geburtsdatum mit Bindestrichen eingibt.
- Telefonnummer: Bei der ersten Eingabe der Telefonnummer wird, falls die Länge der Telefonnummer korrekt ist, die vierte Ziffer herausgelöscht, und so eine falsche Eingabe vorgetäuscht.
- Kreditkartenservice: Bei der ersten Auswahl einer Kreditkarte wird ein Servicefehler vorgetäuscht. Der User wird aufgefordert eine andere Kreditkarte einzugeben.
- Kreditkartennummer: Bei der ersten Eingabe der Kreditkartennummer wird, falls die Länge der Kreditkartennummer korrekt ist, die vierte Ziffer herausgelöscht, und so eine falsche Eingabe vorgetäuscht.

Sollte eine zwingende Eingabe fehlen, wird ohnehin eine Fehlermeldung angezeigt. Diese ist jedoch nicht vom Test provoziert.

Die Idee für die künstlich fabrizierten Fehler bekamen wir während qualitativen UX-Tests welche wir insgesamt mit 5 Usern durchgeführt haben. Auf diese Weise konnten wir den User während der Durchführung des UX-Tests beobachten, und allenfalls Schwachstellen in der Benutzung des UX-Tests identifizieren und in der nächsten Iteration beheben. Eine der Beobachtungen, die wir gemacht haben, war die Tatsache, dass die User sich streng an die Vorgaben gehalten haben, und dementsprechend keine bis wenige Fehlermeldungen angezeigt bekommen haben. Beim Beantworten der QUIS Fragen mussten wir uns selbst eingestehen, dass die Anzahl der gezeigten Fehlermeldungen nicht ausreichend war um diese Fragen zu beantworten. Also haben wir uns entschlossen, künstlich Fehler zu produzieren, und zwar möglichst so, dass der User dies nicht bemerkt.

8.5.2 FRAGEBOGEN

Nachdem der User das Formular erfolgreich ausgefüllt hatte, gelangte er zum QUIS-Fragebogen, bei dem er zu seiner Zufriedenheit befragt wurde.

Die Nutzung des Standard-Fragebogens QUIS muss von der Universität Maryland lizenziert werden. Bis wir den Fragebogen so einsetzen konnten, dass er in unser Konzept des Online-UX-Tests passte, mussten wir einige Anpassungen daran vornehmen und viele Fragen abklären. Unsere Erwartung war für die 50\$ Studenten-Lizenz zu hoch gesetzt, da wir vergebens eine ausführliche Anleitung zu seiner Anwendung und zur Auswertung der Daten suchten. Um einige Anpassungen am Fragebogen vorzunehmen können, wollten wir bezüglich der Aspekte Copyright und Validität sicher sein, wie weit dies möglich ist. Eine schriftliche Nachfrage zum Thema Auswertung bei der Universität von Maryland ergab nur wenig bis gar keine Information. Bei der Befragung unseres Dozenten Christian Hübscher und Dr. Stefan Pauwels, welche QUIS bei deren Masterarbeit angewandt hatten, erhielten wir die Information, dass das Programm SPSS bei der Auswertung zum Einsatz kam, aber Excel auch möglich ist. Gemäss der Website der Universität von Maryland, sind Anpassungen

am Fragebogen möglich: „*The QUIS can be customized and sections or items dropped. Most people drop sections that are not appropriate. Sometimes additional questions are added as needed.*“ Unser Coach Prof. Dr. Markus Stolze gab uns dann die Bestätigung, dass wir mit der nötigen Vorsicht den Fragebogen kürzen und mit eigenen Fragen ergänzen dürfen.

QUIS ist ein sehr ausführlicher Fragebogen, der die verschiedensten Schnittstellen eines zu untersuchenden Systems abdeckt: Um die Menge der Fragen an unsere Probanden zu reduzieren, haben wir uns in Absprache mit unserem Coach darauf geeinigt, Fragen und Bereiche, die ausserhalb unseres Test-Umfelds liegen, aus dem Fragebogen zu streichen. Darunter fallen

- Bisherige Erfahrungen
- Telekonferenzen
- Systeminstallation

Durch diese Reduktion konnten wir sicherstellen, dass die Umfrage innerhalb 20 Minuten zu bewältigen war. Dadurch wurde das Risiko, dass unsere Probanden der quantitativen Tests frühzeitig abbrechen, erheblich reduziert. In den qualitativen Tests hat sich gezeigt, dass wir die Fragen an gewissen Stellen leicht anpassen mussten, wo es zu oft zu Missverständnissen oder Unklarheiten kam. Ein Beispiel war der Titel „Online Anleitung“ welchen wir nachfolgend anpassten: “Online Anleitung (Reiter rechter Rand ‘Hilfe zur Umfrage’)”. Experten Reviews von zwei Psychologen gaben uns weitere wichtige Hinweise zur Optimierung unseres Tests. Um herauszufinden, ob eine Tonalität den Probanden prägt, und wie er zu den 4 weiteren Tonalitäten steht, haben wir am Schluss des Fragebogens eine weitere Frage hinzugefügt, welche die 5 Tonalitäten aufzeigt und bewerten lässt.

UX-Mapping

Die Frage, wie die Daten ausgewertet und interpretiert werden können stand immer noch im Raum. Bezüglich der Operationalisierung der Daten waren wir uns zu diesem Zeitpunkt nicht sicher, ob wir mit QUIS alle Fragen beantworten können: Sagt uns die Frage “Die Zeichen auf dem Bildschirm waren schwer/leicht erkennbar” etwas über die Attraktivität, Designqualität oder Benutzungsqualität aus? Obwohl QUIS die Zufriedenheit des Benutzers mit einer spezifischen Schnittstellenkomponente in 9 Bewertungsbereichen erfasst und den Gesamteindruck mit 6 Items misst, hatten wir das Bedürfnis, unsere Daten nochmals in einem eigenen “Mapping” zu gliedern. Die Idee dahinter war, die Daten einfacher interpretieren zu können. Wir liessen uns dabei vom S-UEQ (Short User Experience Questionnaire) von SAP inspirieren. In einem Card-Sorting im Team versuchten wir dann die Bewertungsbereiche des QUIS auf die des S-UEQ Fragebogens zu gliedern, was sich als schwierig herausstellte. Mit einem gewissen Misstrauen dem Resultat gegenüber liessen wir unser Mapping von zwei unabhängigen Psychologen gegenprüfen und erhielten unterschiedliche Resultate. Es stellte sich heraus, dass es zu viele Unterschiede beim Zuweisen der Bewertungsbereiche gab. Beide Psychologen waren sich einig, dass wir diese Zuteilung nicht weiterverfolgen sollten, da ein Mapping von Bewertungsbereichen von zwei unterschiedlichen Fragebogen ohne Anpassungen nicht zufriedenstellend möglich ist. Somit haben wir uns im Team entschieden, den QUIS Fragebogen und seine Bewertungsbereiche wie folgt beizubehalten:

- Bildschirm-Faktoren,
- Terminologie und Systeminformationen,
- Erlernbarkeit,
- Systemcharakteristika,
- Handbücher und Onlinehilfen,
- Multimedia-Faktoren,
- Ton der Fehlermeldung (eigene Fragen)

8.5.3 DATENBANK

Während Entwicklung von Webformularen für unseren Online-UX-Test war auch gleichzeitig Modellierung und Entwicklung der Datenbank Teil im Gange. Die wichtigsten Anforderungen an das Datenbanksystem waren die Stabilität, Sicherheit sowie online Verfügbarkeit während Laufzeit der Studie. Deswegen haben wir die Dienstleistungen von Plattform MetaNet Plesk im Einsatz genommen und dort eine MySQL Datenbank aufgebaut. Zugriff auf die Daten und Tabellen war mit einem einfachen online Admin GUI möglich. Aus den Überlegungen, welche Daten für das Projekt eine zentrale Rolle spielen werden, haben wir folgendes logisches Datenmodell entworfen:

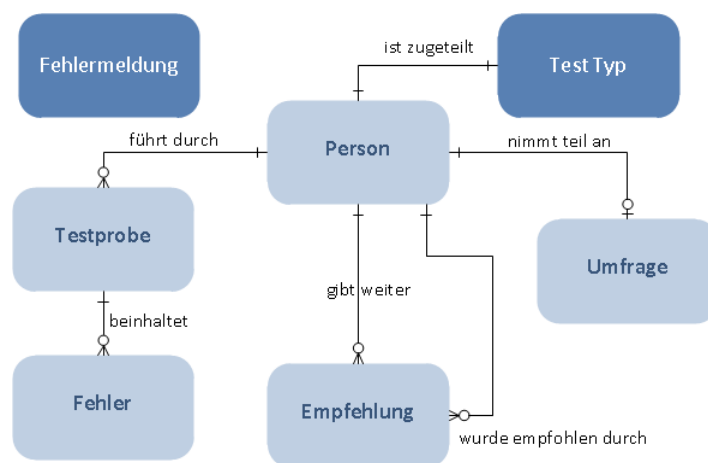


Abbildung 15: Datenmodell für Online-UX-Test

Das Datenmodell beinhaltet:

statische, Referenzdaten:

- Fehlermeldungen Texte - Fehlermeldung
- Tonalität Typen – Test Typ

dynamische, operative Daten aus der Umfrage:

- Antworten zu QUIZ Fragen - Umfrage
- persönliche Daten der Testpersonen - Person
- erfasste Empfehlungen - Empfehlung
- Angaben über Test Durchlauf und Anzahl der Fehler - Testprobe, Fehler

Durch die Anwendung von Datenbank Views und SQL Abfragen konnten wir einen Überblick über Datenqualität und Wachstum während Laufzeit der Studie behalten.

Eine zusätzliche Funktionalität hat für die automatische Zuteilung einer Tonalitäts-Variante für den nächsten Test gesorgt. Zuerst wurde ermittelt, welche Tonalität am wenigsten gültige Umfrage-Daten hat und damit war sie als nächste an der Reihe. Dadurch konnten wir sicherstellen, dass alle Tonalitäts-Gruppen gleich gross waren und miteinander verglichen werden konnten.

8.5.4 REKRUTIERUNG VON TEST PERSONEN UND WETTBEWERB

Zur Teilnahme an der Studie wurden vor allem Personen aus privaten Bekannten- und Freundes-Kreisen rekrutiert. Sie wurden zuerst persönlich angefragt und später per E-Mail zur Teilnahme eingeladen. Dadurch wurde eine hohe Rücklaufquote erreicht. Ein Teil der qualitativen UX Tests wurde mit zufälligen Probanden an Flughafen Zürich durchgeführt. Rekrutierung fand auch über die Nutzung von Socialmedia Plattformen wie XING und Google+, wo der Link <http://www.useremotion.ch/mashcid/> publiziert wurde.

Als Motivation für die Teilnahme an der Studie haben wir ein Wettbewerbs-Konzept entwickelt. Dabei konnte man einen iPad mini gewinnen und sogar eigene Gewinnchancen steigern, indem man die Studie an weitere Personen empfohlen hat und sie auch tatsächlich an der Studie teilgenommen haben. Das Empfehlung Konzept basierte auf dem Prinzip von Schneeball System und man konnte bis zu 5 E-Mail Adressen erfassen, an die eine Einladung E-Mail automatisch geschickt wurde.

8.5.5 PROTOTYP VALIDIERUNG - QUALITATIVE UX TEST

Um das Prototyp zu validieren haben wir eine Reihe von qualitativen UX Test durchgeführt, die jeweils in Räumlichkeiten von unseren Arbeitgeber: AdNovum Zürich, CSS Luzern und UBS Zürich stattgefunden haben. Dank der verfügbaren Infrastruktur und Technik hatten wir optimale Bedingungen für UX Testing aufbauen können. Der Proband war gebeten vor einem Computer zu sitzen und den Online-UX-Test auszufüllen, und dabei laut zu denken. Das Projekt Team war in den gleichen Raum anwesend und hatte folgende Rollen klar zugewiesen:

Test Begleitung und Interview: diese Person war für Einführung des Probanden in die Test Situation zuständig, gab Induktionen zur Aufgabe und war als Ansprechpartner während des Tests verfügbar, nach dem Test befragte den Probanden über sein Empfinden beim Test

Beobachter: eine passive Rolle im Hintergrund, Beobachtung des Testablaufs, Beachtung von Besonderheiten

Protokoll: relevanten Aussagen und Kommentare des Probandes wurden festgehalten.

Während des Tests wurden mit Erlaubnis des Probandes, Sprach- oder Video Aufnahmen gemacht, nachhinein ausgewertet und im Protokoll zusammengefasst. Wie der Proband die Aufgabe löste, konnten alle im Raum an der Leinwand beobachten

Das Vorgehen bei dem UX Test hat sich bewertet und wir haben sehr wertvolle Hinweise und Bemerkungen bei sowohl während dem Test als beim dem Interview bekommen. Wir konnten bei allen Probanden beobachten, dass sie sehr aufmerksam und vollständig das Formular ausfüllen und eigentlich nur 1 Fehler machen. Damit die Probanden mit mehr Fehler konfrontiert werden, wurden einige künstliche Fehler in Prototyp eingebaut. Die

Umsetzung

Beobachtungen bei QUIS Fragebogen, haben uns bestätigt, dass noch einige der Fragen überflüssig und irrelevant waren und einige Begriffe ganz verständlich. Das wurde ebenso im Prototyp angepasst.

Die letzte Etappe des qualitativen UX-Testing wurde am Flughafen Zürich mit zufälligen Test Personen an einem Samstag durchgeführt. Wir haben in der Abflugzone wartende Personen zur Teilnahme an der Umfrage angefragt. Auf diese Art und Weise wurden noch 6 Test Personen rekrutiert. Als Dankschön gab einen Kaffee oder eine kleine Packung von Schweizer Schokolade auf den Weg in die Ferne Länder.



Die gesammelte Beobachtungen und Erfahrungen haben uns bestätigt, unser Online-UX-Test funktioniert gut und die online Studie kann jetzt beginnen.



Manuel Sanchez und Dorota Przygienda mit Proband (Mitte) beim Qualitativen UX Test am Flughafen Zürich.



Manuel Sanchez und Dorota Przygienda und Proband beim Qualitativen UX Test am Flughafen Zürich.



Theodor Huber und Dorota Przygienda mit 2 Probanden (Mitte) gleichzeitig bei zwei Qualitativen UX Tests am Flughafen Zürich.

9 DURCHFÜHRUNG DER STUDIE UND DATENERHEBUNG

9.1 QUANTITATIVE ONLINE UX TESTS

Im Voraus haben wir in unserem Bekannten- und Freundeskreis nach Testpersonen gesucht. Dadurch konnten wir schon früh ein gewisses Interesse wecken und die Teilnahmebereitschaft sicherte sich damit. Die Einladung zur Teilnahme folgte per E-Mail Versand.



Das erste Email an einen möglichen Probanden wird verschickt.

Bereits in den ersten Tagen haben wir das Datenwachstum beobachten können und die ersten Auswertungen gemacht. Die meisten Daten waren korrekt erfasst, aber es gab auch solche, die unvollständig waren und keine Umfrage Daten enthielten. Diese wurden bereits am Anfang aussortiert und waren demnach für die Studie irrelevant.

Nach 2 Wochen Laufzeit der Studie wurde zusätzlich auf Social Media Plattformen Xing und Google+ der Link zur Online-UX-Test publiziert:

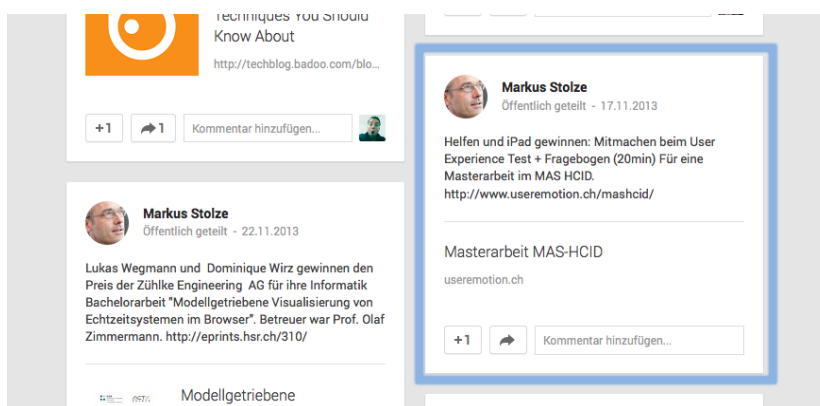


Abbildung 16: Veröffentlichung des Online-UX-Tests auf Google+

Die Teilnahme folgte durch den Link: <http://www.useremotion.ch/mashcid/> und nach einer kurzen Begrüssung und Einführung (Abbildung 11) folgte die Angabe der demografischen Daten, wie Alter, Geschlecht, Postleitzahl und optional die E-Mail Adresse und persönlichen Daten, die für den Wettbewerb nötig waren (**Fehler! Verweisquelle konnte nicht gefunden werden.**). Mit dem nächsten Schritt gelangte man zur eigentlichen Aufgabe, nämlich der Buchung eines vorausgewählten Hotelzimmers auf der Website mySwitzerland.com (Abbildung: 13). Die Beschreibung der Aufgabe und eine Hilfestellung sowie die Testdaten, waren allesamt auf dieser Formulareseite verfügbar. Während der Aufgabe wurde der Proband mit vorprogrammierten Fehlern konfrontiert. Die Tonalität der Fehlermeldungen wurde beim Einstieg des Tests automatisch zugeteilt, aber ohne den Proband darüber zu informieren.

Die Onlinestudie dauerte vom 10. November bis 10. Dezember 2013. Daran haben 110 Personen teilgenommen: 30 Frauen und 80 Männer. Der jüngste Teilnehmer war 19 und der älteste 65 Jahre alt. Die Probanden waren aus der Deutsch-Schweiz, meistens aus dem Grossraum Zürich.

Zusätzliche 70 Personen haben sich für die Teilnahme registriert, aber die Umfrage abgebrochen und wurden dadurch auch von der Auswertung ausgeschlossen. Durch den Empfehlungs-Prozess haben an der Studie zusätzlich 21 Testpersonen teilgenommen

9.2 DATEN AUSWERTUNG UND ANALYSE

Für die Analyse der Daten haben wir anfänglich die Mittelwerte gegeneinander verglichen, und in der Gegenüberstellung berücksichtigt. Aufgrund der Mittelwerte haben wir festgestellt, dass es wenig aussagt, auch wenn Differenzen ersichtlich sind. Mittelwerte sind nicht sehr robust gegenüber Ausreissern. Obwohl bei unseren Daten Ausreisser praktisch nicht möglich sind, da wir ordinale Werte zwischen 1 und 9 erhalten. Dennoch ist eine Darstellung und Mitberücksichtigung aller Quartile (Median als 50% Wert anstelle des Mittelwertes) realitätsnaher, und vor allem sicherer wenn es um Signifikanz geht.

Wir haben für die Vergleiche und Gegenüberstellung der Daten eine Gruppierung auf Tonalität vorgenommen. Diese Gruppierung haben wir für alle Auswertungen beibehalten.

Für die Auswertung haben wir festgestellt, dass es unerlässlich ist, jeweils die Daten zu filtern, sodass die Resultate heterogener wurden. Auswertungen über alle Daten hinweg haben ergeben, dass es kaum Unterschiede in der Beurteilung der User Experience auf Basis der Tonalität gibt.

9.2.1 DATEN

Insgesamt stehen 110 Datensätze für die Analyse zur Verfügung. Die Aufteilung auf die verschiedenen Tonalitäten ist wie folgt:

- Höflich (22 Datensätze)
- Humorvoll (22 Datensätze)
- Motivierend (22 Datensätze)
- Standard (22 Datensätze)

- Unpersönlich (22 Datensätze)

Durch den Algorithmus der Tonalitätszuteilung an Probanden konnte eine ausgewogene Verteilung erreicht werden.

Die Analyse ohne weitere Gruppierung der Daten hat am wenigsten Unterschiede zwischen den Tonalitäten bewirkt. Darüber hinaus wurden folgende Gruppierungen vorgenommen:

- Alle
- Frauen
- Männer
- 35 Jahre und jünger
- 45 Jahre alt und älter

Die Antworten aus dem QUIS Fragebogen sind alle ordinal skaliert mit Werten zwischen 1 und 9. Die Zahl 0 bedeutet, dass keine Aussage gemacht werden konnte. Diese 0-Werte werden in der Analyse nicht berücksichtigt.

Gruppiert werden die Daten beim Vergleich immer nach den Tonalitäten. Um eine Differenzierung zu erhalten, gruppieren wir die Daten nach den oben genannten Gruppen (immer in Kombination mit der Gruppierung der Tonalitäten) und erhalten dadurch signifikantere Werte.

9.2.2 ANALYSE

Wir sind folgendermassen vorgegangen um die Daten zu analysieren, und vor allem auch um die Gruppierungen zu finden, welche die grossen Unterschiede zwischen den Tonalitäten ausmachten.

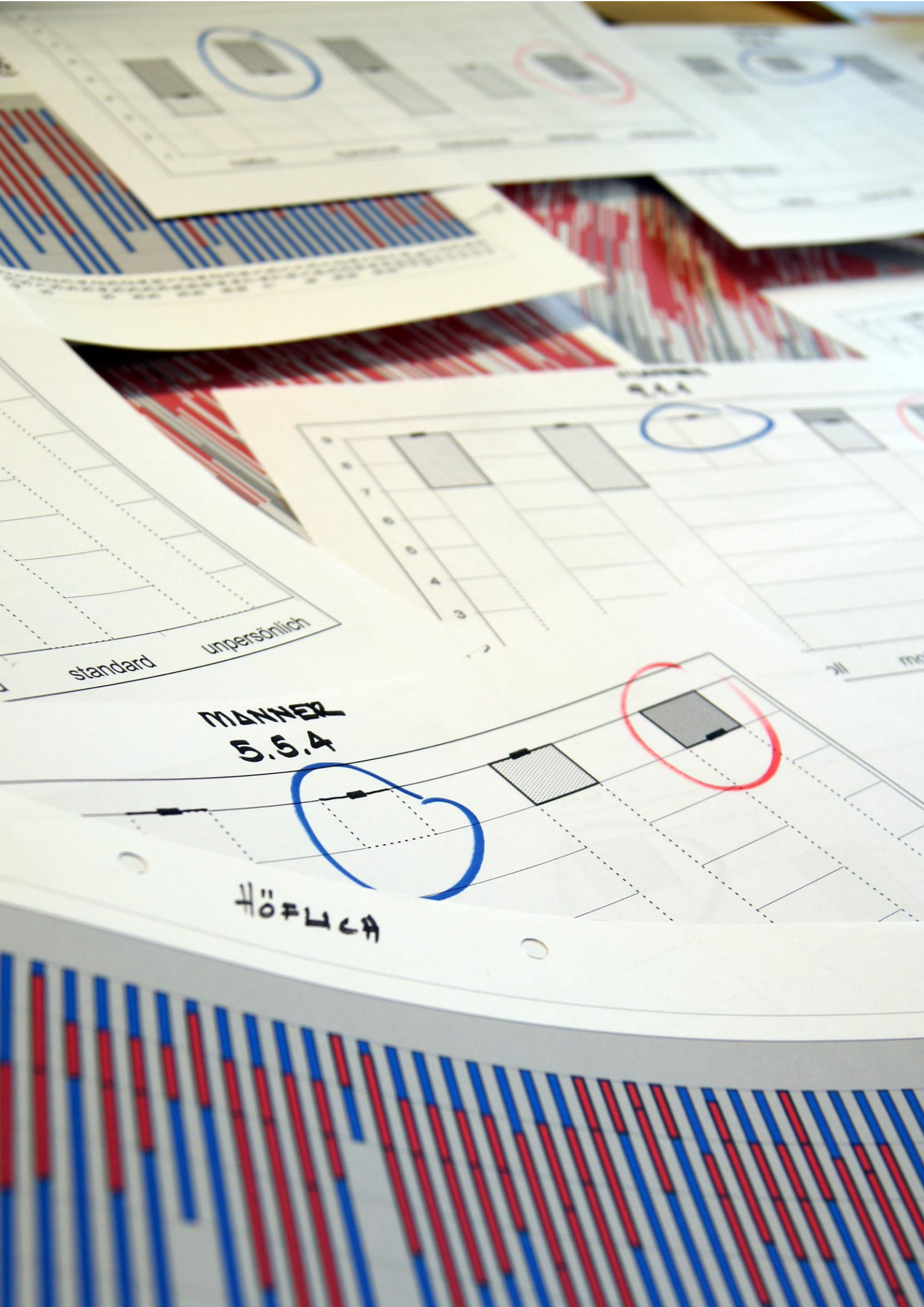
Wir haben als erstes einmal Diagramme erstellt, welche die Werte nach dem Mittelwert anzeigen. Dadurch konnten wir bereits einmal aufzeigen bei welchen Attributen die grossen Unterschiede zu Tage kommen. Z.B. weiblich / männlich trennen, Altersgruppen aus dem Ganzen herauslösen und für sich betrachten, Wohnort (nach PLZ) des Benutzers (was ist wenn man Zürcher und Luzerner trennt?).

Wir können es natürlich nicht dabei belassen, dass wir Mittelwerte miteinander vergleichen, und haben jeweils für alle Fragen Quartile erstellt, und mittels Plot Boxen (Box-Whisker-Plot) diese Quartile visualisiert.

Es gibt jedoch unzählige Möglichkeiten, welche nun visuell analysiert werden sollten, oder als Input im System eingegeben werden mussten.

Folgende Methode wendeten wir an, um schnell an die möglichen Stellen zu gelangen mit den grössten Unterschieden:

Pro Frage im QUIS haben wir nach allen Tonalitäten gefiltert, und jeweils die Quartile 25%, 50% und 75% gebildet. Daraus folgten 5 Wertegruppen welche pro Tonalität die einzelnen Quartile für eine Frage abbildeten.



MANNER
5.5.4

HÖFLICH

standard unpersönlich

211 mo

Beispiel:

Frage	4.3.1	4.3.1	4.3.1	4.3.1	4.3.1
Tonalität	höflich	humorvoll	motivierend	standard	unpersönlich
Maximalwert	9.00	8.00	9.00	8.00	9.00
Oberes Quartil	7.50	7.00	8.00	7.25	8.00
Median	6.00	7.00	7.00	7.00	8.00
Unteres Quartil	3.00	6.00	6.00	5.00	7.00
Minimalwert	0.00	3.00	5.00	3.00	7.00

Abbildung 17: Tabelle mit Quartilen pro Tonalität

Nun wird auf Basis von Formeln ermittelt, welche Tonalitäten eine möglichst kleine Überschneidung haben. D.h. wir schauen welche Werte aus der Reihe oberes Quartil kleiner sind oder gleich dem unteren Quartil (in unserem Beispiel wäre das humorvoll und unpersönlich [7.00]).

Die Visualisierung dieser Daten geschieht mittels Boxplot Diagramme (auch Box-Whisker-Plot genannt). Wir bilden jeweils 5 Boxes nebeneinander und vergleichen jeweils zu einer bestimmten Frage die 5 Tonalitäten ab. Bei den Boxen zeigen wir folgende Werte:

- Oberes Quartil (75% Wert)
- Median (50% Wert)
- Unteres Quartil (25% Wert)

Bei unserem Selektionsverfahren ermitteln wir Daten, welche sich zu 75% nicht überschneiden. Nun muss jedoch berechnet werden ob dieses nicht überschneiden bereits reicht, um signifikant zu sein.

Wir erlauben uns darüber hinaus verschiedene Gruppierungen zu betrachten. Eines ist die Bewertung nach einzelnen Fragen. Die einzelnen Fragen werden gegenüber allen Tonalitäten verglichen und aufgrund der Unterschiede untereinander eine Aussage getroffen. Zusätzlich wird mittels Kruskal Wallis Test statistisch nachgewiesen, dass innerhalb dieser Gruppierung einen signifikanter Unterschied besteht. Mit dem Wilcoxon-Test werden jeweils Paarvergleiche angestellt, und definitiv die 2 Tonalitäten ermittelt, welche einen signifikanten Unterschied untereinander haben.

Das Andere ist die Beurteilung zusammengefasst nach Kapitel. Wir berechnen zu diesem Zweck die Quartile über den ganzen Kapitel hinweg. Als Letztes berechnen wir die Gesamtbewertung, welche die Totale der Kapitel mittels Mittelwertberechnung zusammenfasst (über jedes einzelne Quartil wird der Mittelwert ermittelt).

Die einzelnen Kapitel haben inhaltlich folgende Semantik:

Kapitel 1: Dieses Kapitel wird bei der Bewertung nach Kapitel nicht berücksichtigt, da es sich um eine Frage der Erfahrung mit Online Buchungssystemen handelt. Diese Frage trägt nichts zur Beurteilung der User Experience bei. Dieses Kapitel wird jedoch in einer separaten Berechnung beigezogen, nämlich wenn es um die Frage geht: „hat die Erfahrung einen Einfluss auf die Beurteilung?“.

Kapitel 3: Gesamteindruck:
Für die Beurteilung des Gesamteindruckes werden folgende Fragestellungen verwendet:

- ist der Gesamteindruck angenehm?
- schafft der Gesamteindruck Zufriedenheit?
- ist der Gesamteindruck anregend oder eher frustrierend?
- ist die Anwendung insgesamt schwierig oder leicht
- ist die Anwendung insgesamt flexibel oder starr?

→ *Gesamteindruck bezüglich User Experience*

Kapitel 4: Bildschirm:
Unter Bildschirm werden Fragen gestellt wie:

- ist die Kommunikation am Bildschirm verständlich?
- ist die Auszeichnung von Texten hilfreich?
- ist die Menge an Informationen ausreichend?
- ist die Anordnung logisch?

→ *Layout*

Kapitel 5: Terminologie und Systeminformation
Unter diesem Kapitel werden folgende Themen beurteilt:

- Konsistenz der Ausdruckweise
- Aufgabenbezogenheit der Ausdrucksweise
- Konsistenz der Fehlermeldungen
- Klarheit der Fehlermeldungen
- Klarheit der Anweisungen zur Fehlerbehebungen
- Zustandsbezogene Informationen
- Vorhersehbarkeit der Aktionen
- Erträglichkeit der Wartezeiten
- Problembezogenheit der Fehlermeldungen
- Wie angenehm waren die Fehlermeldungen

→ *Effektivität der Rückmeldungen*

Kapitel 6: Lernfortschritt
Dieses Kapitel befasst sich mit dem Lösen von Problemstellungen:

- Konnten Probleme auf einfache Weise gelöst werden?
- folgten die Ausführungsschritte einer logischen Reihenfolge?
- Informiert das System über den Fortschritt der Problemlösung?

→ *Unterstützung bei Problemlösung*

Kapitel 7: Systemeigenschaften

Dieses Kapitel befasst sich mit der Schwierigkeit, Fehler zu beheben:

- war die Fehlerbehebung schwierig?
- war das Korrigieren von Tippfehlern komplex?
- konnte man mit zunehmender Erfahrung das Hotelbuchungssystem besser bedienen?

→ *Komplexität des Problemlösungsprozesses*

Kapitel 9: Online-Anleitung

Dieses Kapitel befasst sich mit der Effektivität der Online-Anleitung:

- war die Online-Anleitung hilfreich?
- war der Zugang zur Online-Anleitung leicht?
- war der Inhalt der Online-Anleitung zielführend?
- waren die Informationen in der Online-Anleitung komplett und informativ?
- waren die Informationen kurz und prägnant?
- waren die Anweisungen in der Online-Anleitung klar?
- war die Zeit zum Lösen der Aufgabe angemessen?

→ *Effektivität der Hilfestellung*

Kapitel 10: Multimedia Komponenten

Dieses Kapitel beschäftigt sich mit der Beurteilung der visuellen Aspekte:

- hat der Einsatz von Bildmaterial den Ablauf unterstützt?
- war der Einsatz von Farben natürlich?
- war die Menge von verwendeten Farben angemessen?

→ *Visuell-informative Unterstützung*

Für die Aussage, ob ein Paarvergleich signifikante Unterschiede hat, nehmen wir ein Statistik Tool zur Hilfe namens R-Editor. Mit Hilfe dieses Tools testen wir die Gruppen jeweils mit dem Kruskal-Wallis-Test, welcher mehrere Gruppen in die Berechnung mit einbezieht. Wir testen jeweils 5 Gruppen (verschiedene Tonalitäten jeweils zu einer Aussage) innerhalb dieses Tests. Das Resultat daraus ist ein p-Wert. Dieser Wert (p-value aus Kruskal-Wallis-Test) sagt aus, wie hoch die Wahrscheinlichkeit ist, dass es keine Unterschiede zwischen den Gruppen hat. Je kleiner der p-Wert ist, desto eher sind signifikante Unterschiede bei gewissen Paarvergleichen auszumachen.

Die Paarvergleiche werden mit dem Wilcoxon-Test (auch genannt Mann-Whitney-Test) gemacht. Dieser Test ist ein 2-seitiger Paarvergleichstest. Der Wilcoxon-Test liefert schliesslich die Gewissheit ob ein Paarvergleich den erwünschten Unterschied ausmacht. Auch hier ist ein p-Wert (später einfach p-value genannt), der die Wahrscheinlichkeit voraussagt, dass künftige Werte vergleichbar (also nicht unterschiedlich) sein werden. Je kleiner der Wert, desto grösser ist der Unterschied und auch seine Signifikanz. Wir halten für unsere Analysen an einem Signifikanzniveau von 0.05 fest. Als zweites Signifikanzniveau halten wir 0.1 als angemessen. Dieses sagt keine Signifikanz aus, aber eine gewisse Tendenz kann hier nicht unbeachtet bleiben. Werte über 0.1 werden nicht mehr berücksichtigt, da diese Unterschiede unter Umständen rein zufällig sind.

10 ERGEBNISSE

10.1 ZUSAMMENFASSUNG

Im aktuellen Kapitel werden Gruppierungen von Daten gruppenweise und paarweise getestet, und eine Aussage bezüglich ihrer Signifikanz gemacht. Folgende Besonderheiten sind uns speziell aufgefallen:

10.1.1 FRAUEN

- Frauen stehen eher auf die Tonalität **unpersönlich**. Diese Tonalität lässt die Anwendung für die Frauen die Anwendung als angenehmer erscheinen und finden die Ausdrucksweise sei konsistent.
- Die Tonalität **höflich** hilft bei den Frauen wenn es um Akzeptanz der Wartezeiten geht. Darüber hinaus verschafft diese Tonalität den Frauen das Gefühl einer leichteren Bedienung der Anwendung.
- Die Tonalität **motivierend** lässt bei den Frauen die Flexibilität bei einer Anwendung steigern.
- Die explizite Einschätzung der Tonalitäten durch Frauen hat ein komplett anderes Bild ergeben, als dies im Test ermittelt wurde. **Höflich** wurde viel zu hoch eingeschätzt, und **unpersönlich** viel zu tief. Hingegen **motivierend** haben sie richtig eingeschätzt.



10.1.2 MÄNNER

- Männer stehen grundsätzlich eher auf **humorvoll**. Diese Tonalität verschafft der Anwendung ein angenehmer und anregender Gesamteindruck.
- Die Tonalität **motivierend** hilft den Männern bei der Akzeptanz von Wartezeiten.
- Auch bei den Männern ist die explizite Einschätzung verschiedener Tonalitäten nicht der Realität entsprechend ausgefallen. Männer schätzen **höflich** viel zu hoch ein, und **humorvoll** viel zu tief. Hingegen **unpersönlich** haben sie korrekt eingeschätzt.



10.1.3 35 JAHRE UND JÜNGER

- Bei den Jüngeren führt die Tonalität **unpersönlich** zu einer besseren Beurteilung der Anwendung.
- Als besonders angenehm wird die Tonalität **humorvoll** empfunden.
- Die Tonalität **motivierend** hilft bei den Jüngeren die Anwendung als flexibler zu betrachten.
- Bei der expliziten Beurteilung der Tonalitäten wurden ebenfalls Diskrepanzen zur Realität festgestellt. **Höflich** wurde zu hoch und **unpersönlich** zu tief eingeschätzt.

10.1.4 45 JAHRE UND ÄLTER

- Bei den Älteren führt die Tonalität **motivierend** zu einer besseren Beurteilung der Anwendung.

- Die Tonalität **höflich** verhilft den Älteren die Anwendung anregender zu beurteilen.
- Die explizite Einschätzung der Tonalitäten ist bei den Älteren Realitätsnaher als bei allen anderen Gruppen. Aber auch hier gibt kleine Unterschiede vor allem im Bereich **höflich** und **unpersönlich**.

10.2 ERSTE EINDRÜCKE

Beim Betrachten der Daten (Diagramm: Gesamtwertung über alle Fragen) gruppiert nach Tonalitäten, fällt in erster Linie auf, dass die Unterschiede nicht sehr gross sind. Trotzdem fällt uns dabei eine Tonalität sofort ins Auge: motivierend. Diese Tonalität hat den höchsten Median, das obere- und untere Quartil ist ebenfalls am höchsten. Auf der anderen Seite ist der Minimalwert von motivierend am tiefsten. Trotzdem, die oberen 75% der Werte liegen leicht höher als die anderen, und nur die unteren 25% liegen tiefer. Die Visualisierung der Daten durch Whisker-Boxen liefert dem Betrachter ein Gesamtbild, welches folgende Informationen mitliefert:

- Maximalwert (100% Wert)
- Obere Quartil (75% Wert)
- Median (50% Wert)
- Unteres Quartil (25% Wert)
- Minimalwert (0% Wert)

Diese Betrachtungsweise lässt optisch bereits stärkere und schwächere Überlappungen erkennen oder zumindest erahnen.

Der grösste Unterschied ist zwischen den Tonalitäten höflich und motivierend auszumachen. Der Unterschied ist trotzdem nicht signifikant (p-value=0.1372). Der Wert liegt nicht unterhalb des Signifikanzniveaus von 0.05.

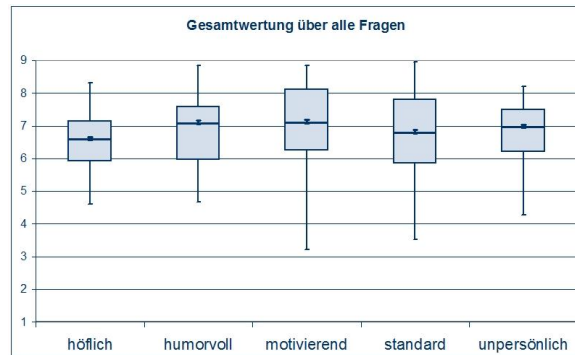


Abbildung 18: Gesamtwertung über alle Fragen

	höflich	humorv.	Motivier.	Standard	unpers.
75%	7.1428	7.5818	8.1198	7.8063	7.4982
50%	6.5640	7.0589	7.0838	6.7713	6.9365
25%	5.9021	5.9551	6.2377	5.8314	6.1869

Kruskal-Wallis Test: p-value = 0.5844

Wilcoxon- Test:

motivierend -höflich: p-value = 0.1372

motivierend -humorvoll: p-value = 0.5893

motivierend -standard: p-value = 0.3219

motivierend -unpersönlich: p-value = 0.3335

Zunächst möchten wir wissen, was unsere Probanden explizit zu einzelnen Tonalitäten aussagen bzw. wie sie diese bewerten. Fünf Fragen aus dem Fragebogen nehmen zu diesem Thema Stellung. Diese Beurteilung haben alle Probanden für alle Tonalitäten gemacht. Was hat der Benutzer dabei bewertet? Dem Benutzer wurde je eine Fehlermeldung in den unterschiedlichen Tonalitäten präsentiert, und er musste pro Fehlermeldung eine Bewertung abgeben.

Insgesamt haben 89 vom 110 Probanden in 166 Fällen mehrere Tonalitäten mit derselben Bewertung belegt. Alle anderen haben die Tonalitäten in eine Reihenfolge der Präferenz gesetzt.

Gemäss Gesamtbeurteilung der Tonalitäten ist motivierend die Präferierte Tonalität. Folgende Reihenfolge hat sich aus der Beurteilung ergeben:

1. motivierend
2. höflich
3. standard
4. humorvoll
5. unpersönlich

Es stellt sich die Frage, wie sieht die Auswertung nach den Grundtonalitäten aus. D.h. gibt es Unterschiede in der Bewertung, wenn die Probanden unterschiedliche Grundtonalitäten beim Test hatten, bzw. hat die Grundtonalität einen Einfluss auf die Beurteilung der unterschiedlichen Tonalitäten?

Wenn man die Grafik der Grundtonalität höflich anschaut, fallen folgende Dinge sofort auf:

- Probanden mit Grundtonalität höflich bewerten die eigene Tonalität schlechter als der Rest der Probanden.
- Die Bewertung für die Tonalität motivierend fällt besser aus.
- Die restlichen Tonalitäten wie humorvoll, standard und unpersönlich

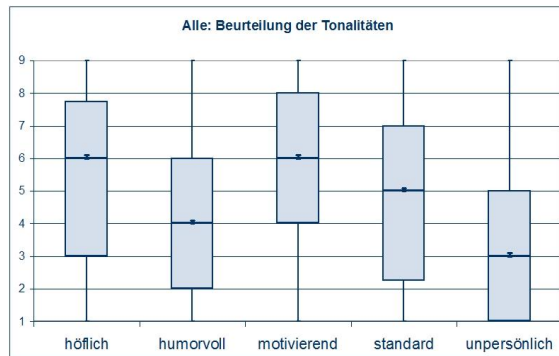


Abbildung 19: Beurteilung der Tonalität durch aller Probanden

	höflich	humorvoll	motivierend	standard	unpers.
75%	7.7500	6.0000	8.0000	7.0000	5.0000
50%	6.0000	4.0000	6.0000	5.0000	3.0000
25%	3.0000	2.0000	4.0000	2.2500	1.0000

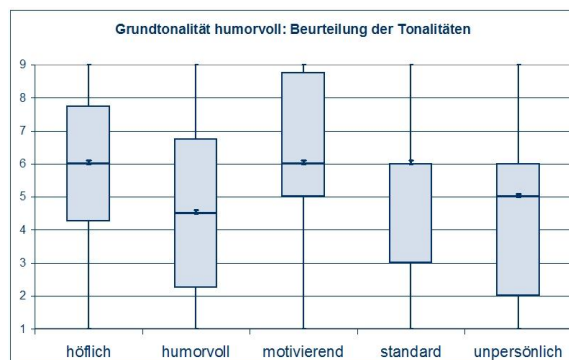


Abbildung 20: Beurteilung der Tonalität durch Probanden der Tonalität humorvoll

unpersönlich sind schlechter bewertet.

Wenn man die Grafik der Grundtonalität unpersönlich anschaut, ist es interessant zu sehen, dass die Probanden welche die Grundtonalität unpersönlich hatten, dieser Tonalität eine höhere Wertung einräumten.

Hauptsächlich wird von allen die Tonalität motivierend bevorzugt, ausser von den Probanden, welche mit dieser Tonalität getestet haben. Auf der anderen Seite muss man den Tester der Tonalität motivierend jedoch zugestehen, dass sie grundsätzlich der Anwendung eine gute User Experience attestiert haben. Kann es sein, dass der Benutzer durch eine bestimmte Tonalität eine bessere User Experience hat, auf der anderen Seite aber diese Tonalität bewusst als nicht ansprechend taxieren würde? Während der qualitativen Tests hatten wir eine Situation, wo der Benutzer aussagte: Er mag lieber Fehlermeldungen die kurz sind, jedoch bei der Bewertung der Tonalitäten (ca. 3 Minuten später) den langen Fehlermeldungen (höflich, humorvoll und motivierend) die bessere Bewertung zuteilte, als den kurzen Fehlermeldungen.

Grundsätzlich müssen wir davon ausgehen, dass die Beurteilung der Tonalitäten an sich, stark von der vorausgegangenen Erfahrung mit der eigenen, dem Probanden zufällig zugewiesenen Tonalität abhängig ist. Diesen Aspekt haben wir jedoch nicht weiter verfolgt. Das wäre ein Forschungsgebiet für weitere Untersuchungen, welche nicht innerhalb unseres Scopes liegen.

Wir werden weitere Untersuchungen anstellen, indem wir die Daten auf unterschiedlichste Weise gruppieren und filtern. Bis jetzt haben wir festgestellt, dass die Tonalität motivierend höhere Akzeptanz genießt als andere Tonalitäten.

Schauen wir im nächsten Teil an, ob allenfalls die Erfahrungsunterschiede der Probanden in der Gesamtwertung signifikante Unterschiede aufdeckt.

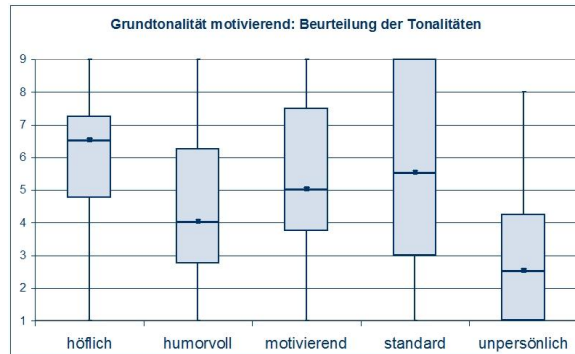


Abbildung 21: Beurteilung der Tonalität durch Probanden der Tonalität motivierend

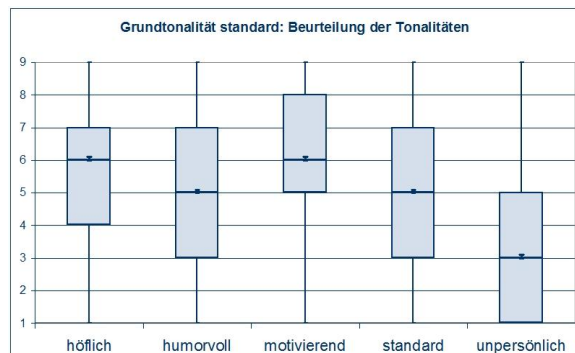


Abbildung 22: Beurteilung der Tonalität durch Probanden der Tonalität standard

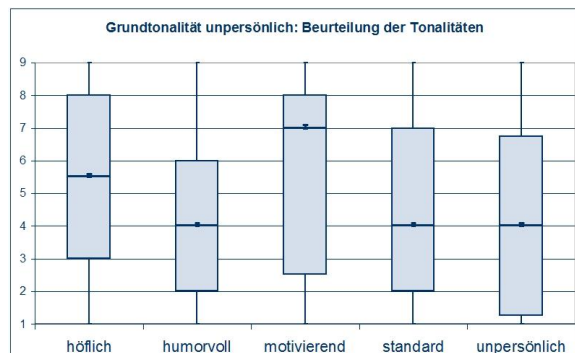


Abbildung 23: Beurteilung der Tonalität durch Probanden der Tonalität unpersönlich

10.3 ERFAHRUNGSGRAD

Eine der Fragen aus dem QUIS Fragebogen war: Wie oft verwenden Sie Online-Buchungssysteme für Hotels? Die Antworten (nominal skaliert) sind folgendermassen aufgebaut:

1. täglich
2. 1x pro Woche
3. 1x pro Monat
4. 1x alle 3 Monate
5. 1x alle 6 Monate
6. 1x pro Jahr
7. 1x alle 2 Jahre
8. 1x alle 3 Jahre
9. seltener als alle 3 Jahre

Die gesamthafte Erfahrungsverteilung ist beinahe normalverteilt.

- 7 Pers. 12x / Jahr = 6%
- 19 Pers. 4x / Jahr = 17%
- 41 Pers. 2x / Jahr = 37%
- 24 Pers. 1x / Jahr = 22%
- 6 Pers. 1x alle 2 Jahre = 5%
- 2 Pers. 1x alle 3 Jahre = 2%
- 11 Pers. Seltener als alle 3 Jahre = 10%

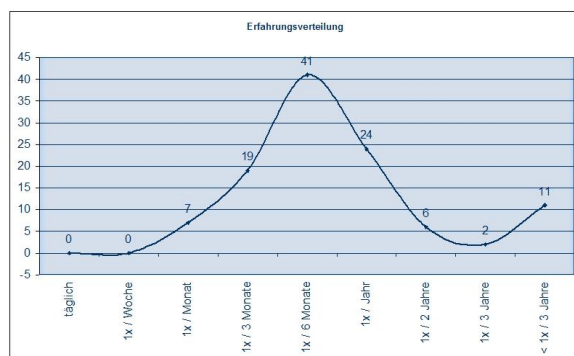


Abbildung 24: Verteilungskurve der Erfahrungsgruppen

Voraussetzung für diese Gegenüberstellungen ist, dass der Erfahrungsgrad der Benutzer über alle Tonalitäten gleichmässig verteilt ist.

Die Verteilung der Erfahrung pro Tonalität sieht wie folgt aus:

- höflich: Mittelwert=**5.18**, Quartile (25%=**4.25**, 50%=**5**, 75%=**6**)
- humorvoll: Mittelwert=**5.68**, Quartile (25%=**5**, 50%=**5**, 75%=**6**)
- motivierend: Mittelwert=**5.64**, Quartile (25%=**5**, 50%=**5**, 75%=**6**)
- standard: Mittelwert=**5.59**, Quartile (25%=**5**, 50%=**5**, 75%=**6**)
- unpersönlich: Mittelwert=**5.32**, Quartile (25%=**4**, 50%=**5**, 75%=**6**)

Alle Tonalitäten weisen durch den Median (50% Wert) eine 5 aus (1x alle 6 Monate). D.h. 50% aller Teilnehmer führen 2x pro Jahr eine Online-Hotelbuchung aus. Der Erfahrungsgrad mit Hotelbuchungssystemen ist demzufolge gleichmässig über alle Tonalitäten verteilt. Es kommen bei jeder Tonalität Personen vor, welche a) sehr oft, und b) sehr selten buchen.

Die Frage ist, hat die Systemerfahrung Einfluss auf die Bewertungen? Nachfolgend zeigen wir einige Auswertungen mit einzelnen Erfahrungsgruppen. Es können leider nicht alle Erfahrungsgruppen dargestellt werden, da zum Teil zu wenig Probanden zur Verfügung stehen.

Bei den erfahrenen Benutzern welche *4 mal pro Jahr* eine Hotelreservation online durchführen, ist **unpersönlich** die Tonalität von Fehlermeldungen, welche gesamthaft für eine gute User Experience sorgt. Auf der anderen Seite steht die Tonalität **höflich** welche von den 5 Tonalitäten für die schlechteste User Experience sorgt.

Die Differenzen zwischen der Tonalität höflich und unpersönlich ist **signifikant** (p-value=0.03333).

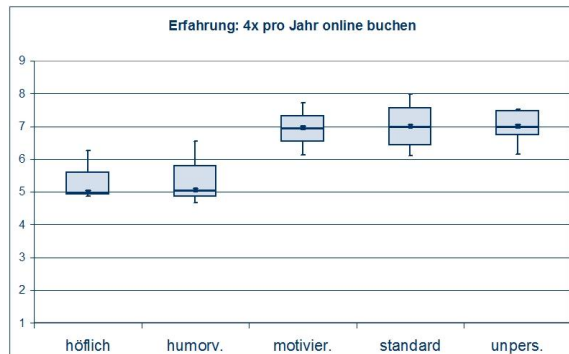


Abbildung 25: 4x pro Jahr online buchen

	höflich	humorv.	motivier.	standard	unpers.
75%	5.6065	5.7875	7.3145	7.5607	7.4716
50%	4.9571	5.0202	6.9194	6.9631	6.9619
25%	4.9090	4.8458	6.5243	6.4074	6.7339

Kruskal-Wallis-Test: p-value = 0.09014

Wilcoxon-Test: unpersönlich-höflich: p-value = 0.03333

Bei Benutzern welche *1 mal pro Jahr* eine Hotelreservation online durchführen, ist ebenfalls die Tonalität **unpersönlich** diejenige, welche die User Experience am besten beeinflusst. Die Tonalität **höflich** hat auch hier den schlechtesten Einfluss auf die User Experience. Aufgrund der Grafik rechts sollte man denken, die Tonalität **standard** ist diejenige welche den tiefsten Wert hat. Aufgrund des Rangsummentest (Wilcoxon-Test) ist der Vergleich zwischen unpersönlich und standard ($p\text{-value}=0.1429$) mit einer kleineren Signifikanz behaftet als der Vergleich zwischen unpersönlich und höflich ($p\text{-value}=0.06667$).

Dieser Test ist **nicht signifikant**.

Der Wert ist jedoch unterhalb des 2. Signifikanzniveau von 0.1, und demzufolge tendenziell zu beachten.

Bei unerfahrenen Benutzern welche seltener als alle 3 Jahren eine Hotelreservation durchführen, ist die Tonalität **humorvoll** die Art von Fehlermeldungen, welche die User Experience ansteigen lässt. Die Tonalität **unpersönlich** ist im Gegensatz dazu diejenige, welche die User Experience negativ beeinflusst.

Interessant ist, dass sich das Bild von erfahrene zu unerfahrene Benutzer komplett ändert. Leider ist der Vergleich der Tonalitäten bei den unerfahrenen Benutzern **nicht signifikant** ($p\text{-value}=0.2$).

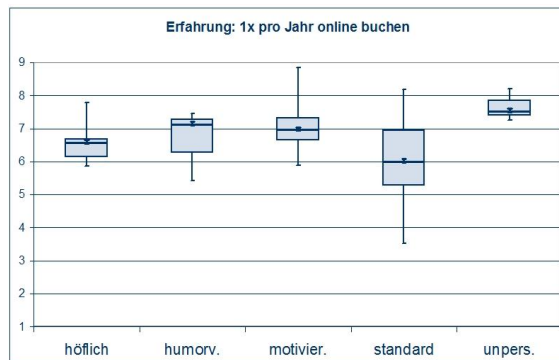


Abbildung 26: 1x pro Jahr online buchen

	höflich	humorv.	motivier.	standard	unpers.
75%	6.6791	7.2768	7.3276	6.9361	7.8571
50%	6.5568	7.1012	6.9457	5.9833	7.5048
25%	6.1292	6.2601	6.6408	5.2592	7.3850

Kruskal-Wallis-Test: $p\text{-value} = 0.2311$

Wilcoxon-Test:

unpersönlich-höflich: $p\text{-value} = 0.06667$

unpersönlich-standard: $p\text{-value} = 0.1429$

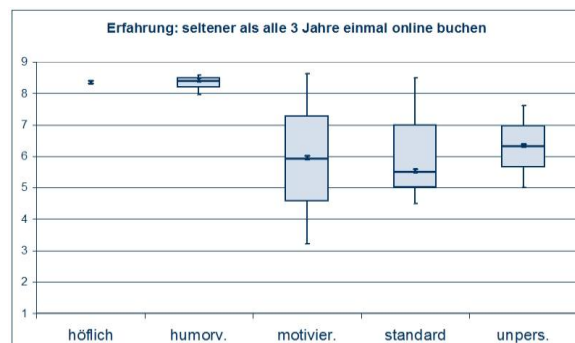


Abbildung 27: seltener als alle 3 Jahre einmal online buchen

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3119	8.4855	7.2696	6.9958	6.9578
50%	8.3119	8.3854	5.9155	5.4917	6.3036
25%	8.3119	8.1767	4.5613	4.9929	5.6495

Kruskal-Wallis-Test: $p\text{-value} = 0.7024$

Wilcoxon-Test: humorvoll-unpersönlich: $p\text{-value} = 0.2$

Je mehr Erfahrung eine Person mit Onlinebuchungssystemen hat, desto besser werden unpersönliche Fehlermeldungen akzeptiert. Bei der Tonalität humorvoll ist auffällig, dass je mehr Erfahrung ein Proband mit der Bedienung einer Anwendung hat, desto schlechter die User Experience bewertet wird. D.h. je erfahrener der Benutzer ist, desto schlechter kommt Humor in den Fehlermeldungen an.

Hingegen bei den motivierenden Fehlermeldungen (welche ja ebenfalls längere Fehlermeldungen sind), verhält es sich genau umgekehrt. Je mehr Erfahrung ein Benutzer in der Bedienung von Onlinebuchungssystemen hat, desto besser kommen motivierende Fehlermeldungen an.

Wie fällt die explizite Beurteilung der einzelnen Tonalitäten durch die verschiedenen Erfahrungsgruppen?

10.3.1 BEURTEILUNG DER TONALITÄTEN

Jede Testperson hat explizit eine Bewertung einzelner Fehlermeldungen in verschiedener Tonalitäten abgegeben. Hier werden wir ihren Eindruck gruppiert wiedergeben.

Beurteilung durch sehr erfahrene Benutzer

Im Gegensatz zur Beurteilung der User Experience durch sehr erfahrene Benutzer fällt die explizite Beurteilung der Tonalität ganz anders aus. Hier werden Sätze mit **höflicher und motivierender** Tonalität eher besser bewertet als **standard und unpersönlich**.

Man kann hier argumentieren, dass die Beeinflussung der Tonalität wie bereits erwähnt, beim erfahrenen Benutzer je kürzer und prägnanter die Sätze, oder allenfalls bei bereits bekannten Fehlermeldungen eine bessere User Experience resultiert. Jedoch wenn diese Probanden jeweils explizit einen Satz beurteilen müssen (in einer Gegenüberstellung der Sätze), dann „gefällt“ ihnen höflich und motivierend besser.

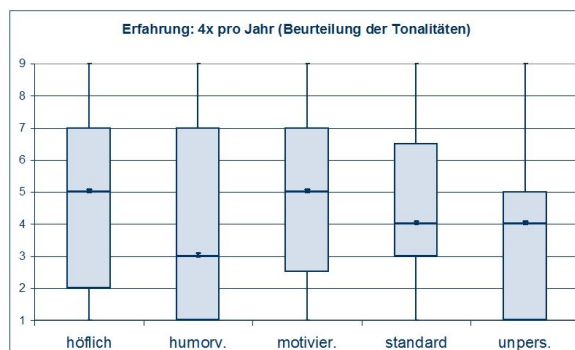


Abbildung 28: Beurteilung der Tonalität durch sehr erfahrene Benutzer

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	7.0000	7.0000	6.5000	5.0000
50%	5.0000	3.0000	5.0000	4.0000	4.0000
25%	2.0000	1.0000	2.5000	3.0000	1.0000

Beurteilung durch unerfahrenen Benutzer

Bei dem unerfahrenen Benutzer fällt auf, dass die Tonalität **humorvoll** viel besser bewertet wird, als bei erfahrenen Benutzern. Ebenso verhält es sich mit der Tonalität **unpersönlich**. Diese wird auch viel schlechter bewertet als bei den erfahrenen Benutzern. Fehlermeldungen mit unpersönlicher Ausprägung generieren mit höchster Wahrscheinlichkeit auch eine gewisse Unsicherheit bei unerfahrenen Benutzern.

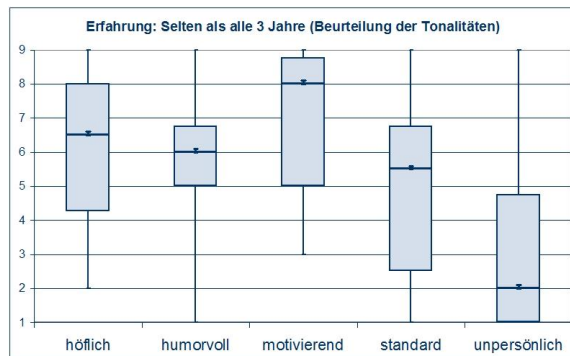


Abbildung 29: Beurteilung der Tonalität durch unerfahrene Benutzer

	höflich	Humorv.	Motivier.	standard	Unpers.
75%	8.0000	6.7500	8.7500	6.7500	4.7500
50%	6.5000	6.0000	8.0000	5.5000	2.0000
25%	4.2500	5.0000	5.0000	2.5000	1.0000

10.4 BEURTEILUNG DER USER EXPERIENCE ÜBER DIE GESAMTE TESTGRUPPE

Die User Experience ist zusammengefasst das Resultat des gesamten Fragebogens. D.h. Kapitel 3 bis Kapitel 10 geben ein Urteil ab bezüglich gesamter Nutzungserfahrung bzw. User Experience. Auf der anderen Seite gibt Kapitel 3 einen Gesamteindruck des Benutzers bezüglich Anwendung. Die aggregierten Resultate über alle Kapitel decken sich in der Gesamtaussage mit der Aussage aus Kapitel 3 (vergleiche die Diagramme rechts „Alle: Gesamtwertung über alle Fragen“ und „Alle: Kapitel 3“).

Die Unterschiede sind in Kapitel 3 etwas offensichtlicher als bei der Gesamtauswertung, aber die Reihenfolge der einzelnen Tonalitäten bleibt gleich.

Für den Nachweis, dass die Tonalität einer Fehlermeldung Einfluss in die User Experience hat, müssen wir nur zwei Tonalitäten finden, welche aufgrund der Wertung signifikante Unterschiede haben.

Der Wilcoxon-Test zwischen den Tonalitäten höflich und humorvoll liefert einen p-value

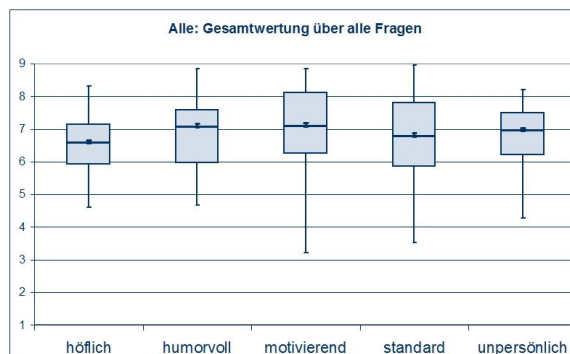


Abbildung 30: Gesamtwertung über alle Fragen

	Höflich	Humorv.	Motivier.	Standard	Unpers.
75%	7.1428	7.5818	8.1198	7.8063	7.4982
50%	6.5640	7.0589	7.0838	6.7713	6.9365
25%	5.9021	5.9551	6.2377	5.8314	6.1869

unterhalb des Signifikanzniveaus von 0.05. Der Unterschied ist somit signifikant ($p\text{-value}=0.04695$).

Somit zeigt sich, dass indem wir innerhalb eines Tests nur die Tonalität ändern, und die Beurteilung des Gesamteindruckes trotzdem so unterschiedlich ausfallen kann, dass die Tonalität eindeutig die User Experience beeinflusst.

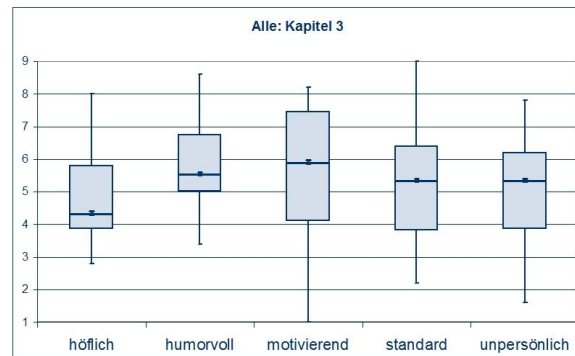


Abbildung 31: Beurteilung der User Experience aufgrund des Kapitel 3 von QUIS

	höflich	humorv.	motivier.	Standard	unpers.
75%	5.8000	6.7500	7.4500	6.4000	6.2000
50%	4.3000	5.5000	5.8750	5.3000	5.3000
25%	3.8500	5.0000	4.1000	3.8000	3.8500

Kruskal-Wallis-Test: $p\text{-value} = 0.3246$

Wilcoxon-Test: humorvoll -höflich: $p\text{-value} = 0.04695$

10.5 GRUPPIERUNG NACH KAPITEL DES QUIS FRAGEBOGEN

Interessant sind sicherlich die Vergleiche unter den Kapitel. Die Kapitel haben jeweils eine spezielle Bedeutung in unserer Umfrage. Die Bedeutung sehen wir in der nachfolgenden Liste:

- Kapitel 1: **Systemerfahrung** → Erfahrung mit Hotelbuchungssystemen
- Kapitel 3: **Gesamteindruck** → Gesamteindruck bezüglich User Experience.
- Kapitel 4: **Bildschirm** → Layout
- Kapitel 5: **Terminologie und Systeminfo** → Effektivität der Rückmeldungen
- Kapitel 6: **Lernfortschritt** → Unterstützung bei Problemlösung
- Kapitel 7: **Systemeigenschaften** → Komplexität des Problemlösungsprozesses
- Kapitel 9: **Online-Anleitung** → Effektivität der Hilfestellung
- Kapitel 10: **Multimedia Komponenten** → Visuell-informative Unterstützung

10.5.1 KAPITEL 3: GESAMTEINDRUCK

Auswertung über die Gesamte Testgruppe

Wir nehmen *höflich* als offensichtlich tiefsten Wert und vergleichen es mit allen anderen Tonalitäten mittels Wilcoxon-Test:

höflich-humorvoll: p-value = 0.04695
 höflich-motivierend: p-value = 0.1236
 höflich-standard: p-value = 0.5181
 höflich-unpersönlich: p-value = 0.431

Wie wir sehen, resultiert beim Vergleich höflich-humorvoll der kleinste p-value. D.h. *humorvoll* ist die Tonalität mit dem grössten Unterschied zu höflich. Demzufolge hat humorvoll die höchste Wertung, und *höflich* hat die tiefste Wertung.

Gegenbeweis für den tiefsten Wert. Gehen wir also vom höchsten Wert aus (humorvoll), und vergleichen es mit allen anderen:

humorvoll – höflich: p-value = 0.04695
 humorvoll – motivierend: p-value = 0.9251
 humorvoll – standard: p-value = 0.2644
 humorvoll – unpersönlich: p-value = 0.2846

Wie wir sehen, ist kein anderer Vergleich mit einem tieferen Wert belegt als die Verbindung humorvoll-höflich.

Ab hier werden nur noch die Höchst- und Tiefstwerte dokumentiert.

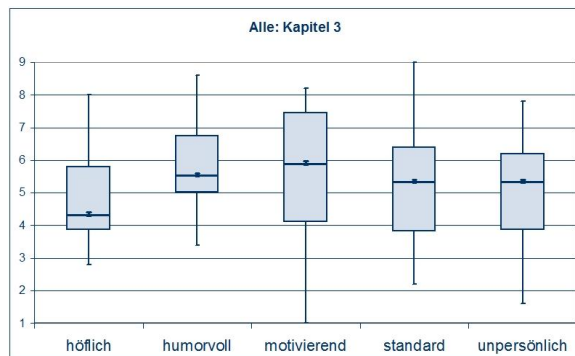


Abbildung 32: Alle: Kapitel 3

	höflich	humorv.	motivier.	standard	unpers.
75%	7.1428	7.5818	8.1198	7.8063	7.4982
50%	6.5640	7.0589	7.0838	6.7713	6.9365
25%	5.9021	5.9551	6.2377	5.8314	6.1869

Kruskal-Wallis-Test: p-value = 0.3246

Wilcoxon-Test:

höflich-humorvoll: p-value = 0.04695

höflich-motivierend: p-value = 0.1236

höflich-standard: p-value = 0.5181

höflich-unpersönlich: p-value = 0.431

humorvoll – höflich: p-value = 0.04695

humorvoll – motivierend: p-value = 0.9251

humorvoll – standard: p-value = 0.2644

humorvoll – unpersönlich: p-value = 0.2846

Wie sieht das Ganze aus, wenn wir uns auf Teile der Population konzentrieren. Z.B. Frauen und Männer oder jüngere und ältere Personen, oder regionale Unterschiede?

Auswertung über Gruppe Frauen

Durch die Gruppierung nach dem weiblichen Geschlecht erhalten wir eine Datenmenge n=30. D.h. wir haben gesamthaft 30 Frauen unter unseren Probanden. Das macht folgende Aufteilung pro Tonalität:

- Höflich: 5 Frauen
- Humorvoll: 6 Frauen
- Motivierend: 8 Frauen

- Standard: 4 Frauen
- Unpersönlich: 7 Frauen

Bei den Frauen scheint die Tonalität unpersönlich der Favorit zu sein. D.h. der Gesamteindruck wird am besten beurteilt, wenn die Nutzung unter dem Einfluss von unpersönlichen Fehlermeldungen steht.

Humorvolle Fehlermeldungen kommen bei den Frauen weniger an, verglichen mit der Gesamtpopulation. Alle anderen Tonalitäten ausser unpersönlich fallen bei den Frauen eher schlechter aus als bei der Gesamtpopulation.

Bedingt durch die kleine Menge an Daten pro Tonalität, ist der Rangsummenwert (W) nicht tief genug, um das gewünschte Signifikanzniveau zu erzielen

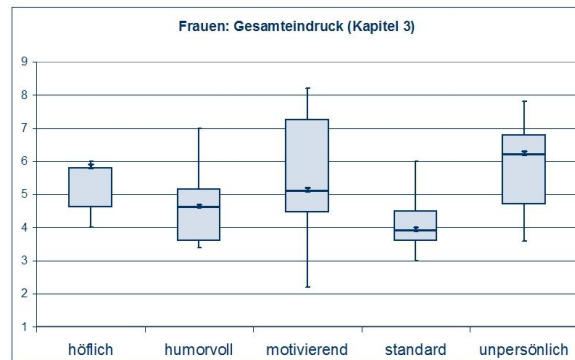


Abbildung 33: Frauen: Kapitel 3

	höflich	humorv.	motivier.	standard	unpers.
75%	5.8000	5.1500	7.2500	4.5000	6.8000
50%	5.8000	4.6000	5.1000	3.9000	6.2000
25%	4.6000	3.6000	4.4500	3.6000	4.7000

Kruskal-Wallis Test: p-value = 0.4936

Wilcoxon-Test: unpersönlich -standard: p-value = 0.1849

Frage 3.1: Wie angenehm ist die Anwendung?

Die Beantwortung der Frage, ob der Gesamteindruck eher angenehm ist oder unangenehm, finden die Frauen, welche die Grundtonalität unpersönlich hatten, die Anwendung am angenehmsten.

Durch den Wilcoxon-Vorzeichen-Rang-Test mit dem direkten Vergleich zwischen den Tonalitäten *standard* und *unpersönlich* haben wir ein Vergleich mit einem p-Wert von 0.06169. Der Unterschied zwischen standard und unpersönlich ist nicht signifikant. Der Wert liegt jedoch unterhalb des 2. Signifikanzniveau von 0.1 und ist somit für uns tendenziell ernst zu nehmen.

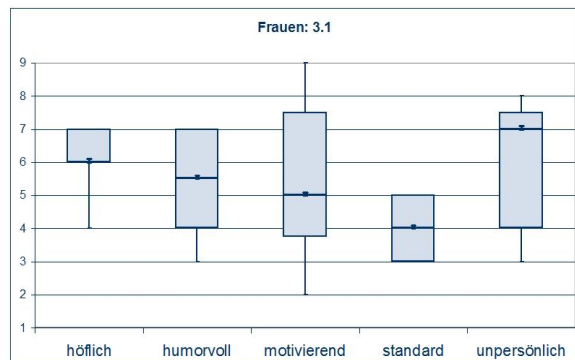


Abbildung 34: Frauen: Frage 3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	7.0000	7.5000	5.0000	7.5000
50%	6.0000	5.5000	5.0000	4.0000	7.0000
25%	6.0000	4.0000	3.7500	3.0000	4.0000

Kruskal-Wallis-Test: p-value = 0.5847

Wilcoxon-Test: höflich -standard: p-value = 0.06169



Frage 3.3: Wie anregend ist die Anwendung?

Hauptsächlich haben **unpersönliche** Fehlermeldungen bei den Frauen einen positiven Einfluss darauf, wie anregend eine Anwendung erscheint. Hingegen sorgt die Tonalität **humorvoll** eher nicht für Anregung.

Der Vergleich zwischen den Tonalitäten unpersönlich und humorvoll liefert mit dem Wilcoxon-Test einen p-value=0.03851 und ist damit **signifikant**.

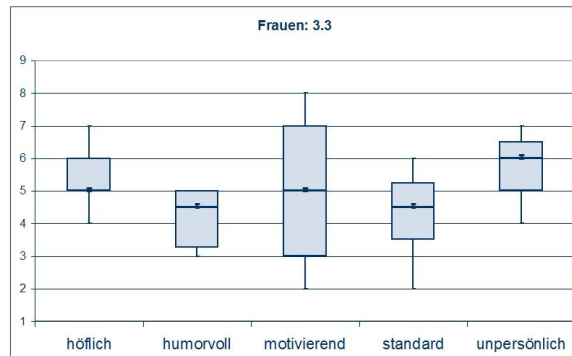


Abbildung 35: Frauen: Frage 3.3

	höflich	humorv.	motivier.	standard	unpers.
75%	6.0000	5.0000	7.0000	5.2500	6.5000
50%	5.0000	4.5000	5.0000	4.5000	6.0000
25%	5.0000	3.2500	3.0000	3.5000	5.0000

Kruskal-Wallis-Test: p-value = 0.3446

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.03851



Frage 3.4: Wie leicht ist die Anwendung zu bedienen?

Bei den Frauen lassen **unpersönliche** Fehlermeldungen die Anwendung leichter bedienbar erscheinen. Hingegen bei **höflichen** Fehlermeldungen erscheint dieselbe Anwendung schwieriger zu bedienen. Natürlich informieren Fehlermeldungen über das weitere Vorgehen, und damit lässt sich die Bedienung leichter gestalten. Jedoch im vorliegenden Fall verhält es sich so, dass die unpersönliche Fehlermeldung die Anwendung leichter bedienbar macht. Das erstaunt uns sehr, weil die Fehlermeldung unpersönlich keinerlei Informationen für die Behebung des Fehlers liefert (z.B. „ungültige Eingabe“).

Die Unterschiede im Paarvergleich der Tonalitäten höflich und standard sind **signifikant** (p-value=0.04814).

Der Unterschied zwischen der Tonalität unpersönlich und standard erscheint in der Grafik grösser, ist jedoch statistisch gesehen nicht aussagekräftig, da bei standard n=4 ist, und keine signifikante Aussage gemacht werden kann.

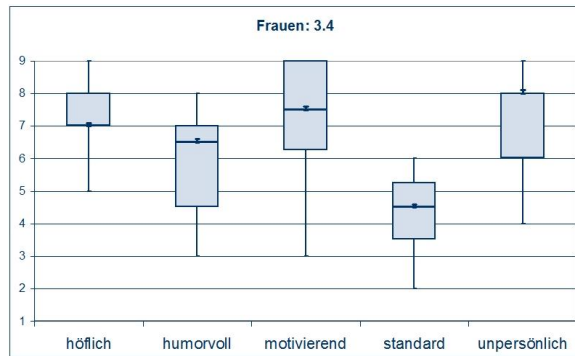


Abbildung 36: Frauen: Frage 3.4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	9.0000	5.2500	8.0000
50%	7.0000	6.5000	7.5000	4.5000	8.0000
25%	7.0000	4.5000	6.2500	3.5000	6.0000

Kruskal-Wallis: p-value = 0.1464

Wilcoxon-Test:

höflich-standard: p-value= 0.04814

unpersönlich-standard: p-value = 0.05364



Frage 3.6: Wie flexibel ist die Anwendung?

Hinsichtlich Flexibilität haben **motivierende** Fehlermeldungen bei den Frauen einen positiveren Einfluss als **humorvolle** Fehlermeldungen.

Der Unterschied zwischen der Tonalität humorvoll und motivierend ist **nicht signifikant** (p-value=0.151)

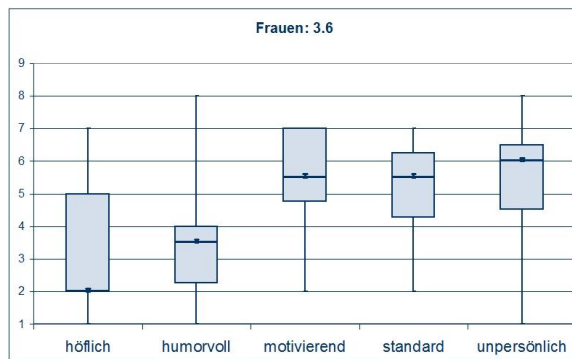


Abbildung 37: Frauen: Frage 3.6

	höflich	humorv.	motivier.	standard	unpers.
75%	5.0000	4.0000	7.0000	6.2500	6.5000
50%	2.0000	3.5000	5.5000	5.5000	6.0000
25%	2.0000	2.2500	4.7500	4.2500	4.5000

Kruskal-Wallis Test: p-value = 0.4072

Wilcoxon-Test: motivierend -humorvoll: p-value=0.151

Die Tatsache, dass bei den Frauen generell die Tonalität unpersönlich favorisiert wurde, war für uns eher überraschend. Es ist zu beachten, dass weder die Erfahrung mit online Buchungssystemen noch das Durchschnittsalter eine Rolle zwischen den Geschlechtern spielt, da die Erfahrungswerte bei beiden Geschlechtern gleich ist und das Durchschnittsalter bei beiden Geschlechtern bei 39 Jahre liegt.

Gemäss Gesamtbewertung gibt es jedoch bei den Frauen keine grossen Differenzen bei der Tonalität.

Kommentar einer Probandin zu den verschiedenen Sätzen der Tonalität:

so kurz wie möglich. bei Buchungen bin ich lese faul? ich will einfach so schnell wie möglich buchen können.



Auswertung der Gruppe Männer

Durch die Gruppierung nach Männer, erhalten wir eine Datenmenge $n=80$. D.h. wir haben gesamthaft 80 Männer unter unseren Probanden. Das macht folgende Aufteilung pro Tonalität:

- höflich: 17 Personen
- humorvoll: 16 Personen
- motivierend: 14 Personen
- standard: 18 Personen
- unpersönlich: 15 Personen

Bei den Männern hat die Tonalität **humorvoll** den grössten Einfluss auf den Gesamteindruck. Die Tonalität **unpersönlich** ist bei den Männer von der Reihenfolge her schlechter bewertet als bei den Frauen (siehe oben). Was sehr erstaunt, ist die Tatsache, dass die Tonalität höflich das Schlusslicht bildet.

Der Unterschied beim Vergleich zwischen den Tonalitäten höflich und humorvoll ist **hoch signifikant** ($p\text{-value}=0.009397$)

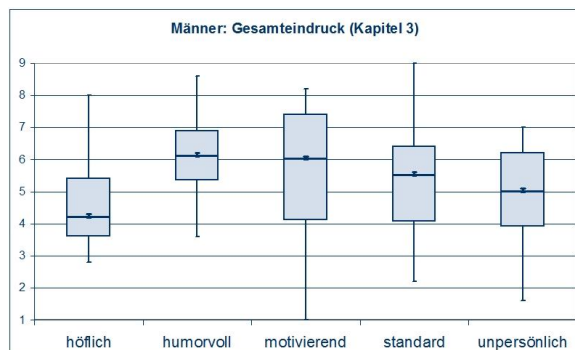


Abbildung 38: Männer: Kapitel 3

	höflich	humorv.	motivier.	standard	unpers.
75%	5.4000	6.9000	7.4000	6.4000	6.2000
50%	4.2000	6.1000	6.0000	5.5000	5.0000
25%	3.6000	5.3500	4.1000	4.0500	3.9000

Kruskal-Wallis-Test: $p\text{-value} = 0.08929$

Wilcoxon-Test:

humorvoll -höflich: $p\text{-value} = 0.009397$

humorvoll-unpersönlich: $p\text{-value} = 0.0341$



Frage 3.1: Wie angenehm ist die Anwendung?

Männer finden, dass **humorvolle** Fehlermeldungen die Anwendung angenehmer gestalten als **höfliche** Fehlermeldungen.

Der Unterschied zwischen höflichen und humorvollen Fehlermeldungen ist **signifikant** (p-value=0.0276). Demzufolge hat die Tonalität einen Einfluss auf die Frage, ob eine Anwendung als angenehm empfunden wird oder nicht.

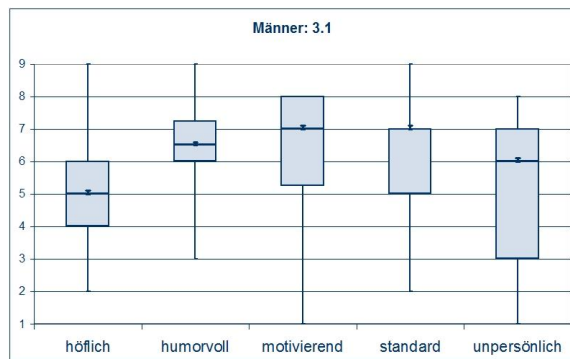


Abbildung 39: Männer: Frage 3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	6.0000	7.2500	8.0000	7.0000	7.0000
50%	5.0000	6.5000	7.0000	7.0000	6.0000
25%	4.0000	6.0000	5.2500	5.0000	3.0000

Kruskal-Wallis-Test: p-value = 0.1918

Wilcoxon-Test: humorvoll -höflich: p-value = 0.0276

Frage 3.3: Wie anregend ist die Anwendung?

Der Unterschied zwischen **humorvollen** Fehlermeldungen und **unpersönlichen** Fehlermeldungen ist bezüglich der Frage nach Anregung am höchsten. Der Unterschied ist **signifikant** (p-value=0.02423). D.h. humorvolle Fehlermeldung haben einen positiven Einfluss auf die Empfindung, wie anregend ein Mann die Anwendung findet.

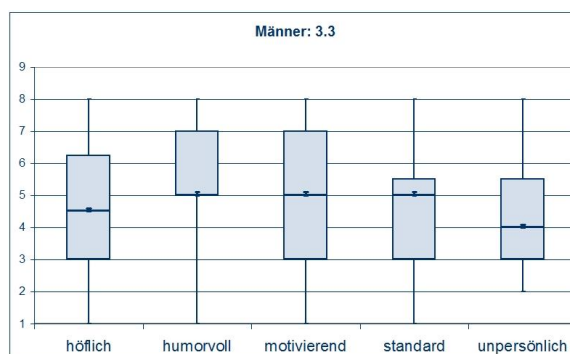


Abbildung 40: Männer: Frage 3.3

	höflich	humorv.	motivier.	standard	unpers.
75%	6.2500	7.0000	7.0000	5.5000	5.5000
50%	4.5000	5.0000	5.0000	5.0000	4.0000
25%	3.0000	5.0000	3.0000	3.0000	3.0000

Kruskal-Wallis-Test: p-value = 0.3123

Wilcoxon-Test: humorvoll-unpersönlich: p-value = 0.02423



Bei den Männern hat die Tonalität humorvoll und motivierend einen hohen positiven Einfluss auf den Gesamteindruck einer Anwendung. Der grosse Unterschied zwischen Mann

und Frau liegt offensichtlich darin, dass der Mann Fehlermeldungen mit einem hohen Mass an Emotionen besser findet, hingegen die Frau eher auf emotionslose Fehlermeldungen steht. D.h. es muss nicht zwingend mit der Emotionalität zu tun haben, es kann auch sein, dass die Länge der Fehlermeldungen hier eine Rolle spielt (wie bereits bei der Auswertung der Frauen erwähnt).

Auswertung über Gruppe 35 jährig und jünger

Durch die Gruppierung nach Probanden, welche ≤ 35 Jahre alt sind, erhalten wir eine Datenmenge $n=42$. D.h. wir haben gesamthaft 42 Personen unter unseren Probanden welche gleich oder jünger als 35 Jahre alt sind. Das macht folgende Aufteilung pro Tonalität:

- höflich: 7 Personen
- humorvoll: 13 Personen
- motivierend: 7 Personen
- standard: 8 Personen
- unpersönlich: 7 Personen

Die Tonalitäten *humorvoll* und *unpersönlich* lösen bei den jüngeren Personen (bis 35 Jahre alt) einen positiven Gesamteindruck der Anwendung aus. Hingegen die *höfliche* Tonalität hat einen eher negativen Einfluss auf den Gesamteindruck der Anwendung.

Der Unterschied zwischen höflich und humorvoll ist *signifikant* ($p\text{-value}=0.02835$). D.h. statistisch gesehen, hat die Tonalität einen Einfluss auf den Gesamteindruck der Anwendung.

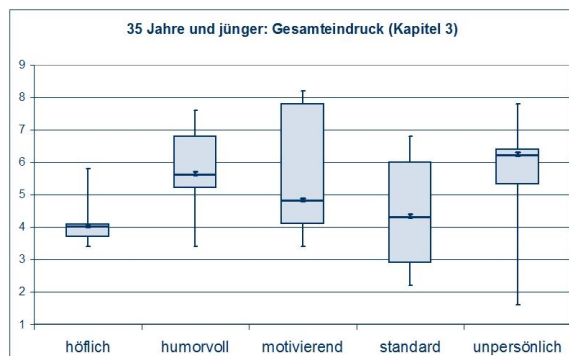


Abbildung 41: 35 Jahre und jünger: Kapitel 3

	höflich	humorv.	motivier.	standard	unpers.
75%	4.1000	6.8000	7.8000	6.0000	6.4000
50%	4.0000	5.6000	4.8000	4.3000	6.2000
25%	3.7000	5.2000	4.1000	2.9000	5.3000

Kruskal-Wallis-Test : $p\text{-value} = 0.1289$

Wilcoxon-Test:

humorvoll -höflich: $p\text{-value} = 0.02835$

unpersönlich -höflich: $p\text{-value} = 0.05368$

Frage 3.1: Wie angenehm ist die Anwendung?

Die jüngeren finden, dass **humorvolle** Fehlermeldungen die Anwendung viel angenehmer gestalten, als eine **höfliche** Ausdrucksweise. Prinzipiell liegt die Tonalität unpersönlich wie auch humorvoll als Präferenz ganz oben.

Statistisch gesehen ist die Differenz zwischen höflich und humorvoll grösser als die Differenz zwischen höflich und unpersönlich. Bei humorvoll ist n=13 und bei unpersönlich ist n=7. Dieser Unterschied macht es aus, dass der Vergleich höflich/motivierend signifikanter ist als höflich/unpersönlich.

Der Test ist jedoch **nicht signifikant** (p-value=0.07165) aber dadurch, dass p-value unter 0.1 liegt können wir trotzdem die Aussage wagen, dass die Tonalität tendenziell einen Einfluss hat auf die Beurteilung, wie angenehm die Anwendung ist.

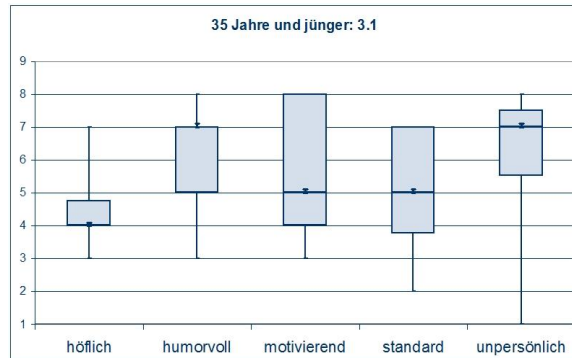


Abbildung 42: 35 Jahre und jünger: Frage 3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	4.7500	7.0000	8.0000	7.0000	7.5000
50%	4.0000	7.0000	5.0000	5.0000	7.0000
25%	4.0000	5.0000	4.0000	3.7500	5.5000

Kruskal-Wallis-Test: p-value = 0.3305

Wilcoxon-Test: humorvoll -höflich: p-value = 0.07165

Frage 3.2: Wie zufriedenstellend ist die Anwendung?

Welche Tonalität lässt grössere Zufriedenheit verspüren? **Humorvoll** hat gemäss Rangsummentest die grösste Differenz zu **höflich**. D.h. humorvoll lässt die Zufriedenheit bei der Bedienung der Anwendung am höchsten steigen. Hingegen trägt die Tonalität höflich am wenigsten zur Zufriedenheit bei.

Der Unterschied bei der Tonalität höflich und humorvoll ist **signifikant** (p-value=0.04137)

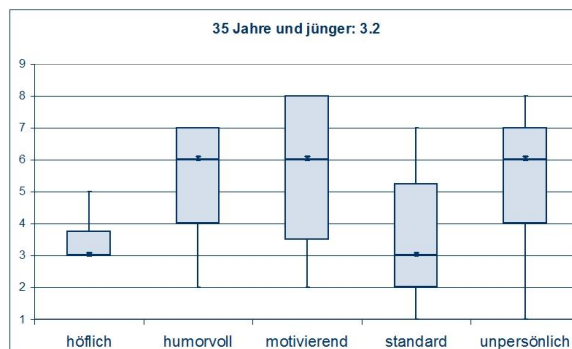


Abbildung 43: 35 Jahre und jünger: Frage 3.2

	höflich	humorv.	motivier.	standard	unpers.
75%	3.7500	7.0000	8.0000	5.2500	7.0000
50%	3.0000	6.0000	6.0000	3.0000	6.0000
25%	3.0000	4.0000	3.5000	2.0000	4.0000

Kruskal-Wallis-Test: p-value = 0.136

Wilcoxon-Test: humorvoll-höflich: p-value = 0.04137

Frage 3.3: Wie anregend ist die Anwendung?

Die höchste Anregung verschafft der Anwendung **humorvolle** Fehlermeldungen. Am wenigsten vermag die **höfliche** Art und Weise anzuregen. Die Differenz zwischen humorvoll und höflich ist **signifikant** (p-value=0.04755).

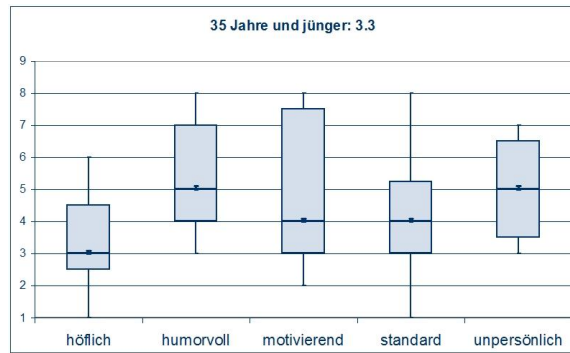


Abbildung 44: 35 Jahre und jünger: Frage 3.3

	höflich	humorv.	motivier.	standard	unpers.
75%	4.5000	7.0000	7.5000	5.2500	6.5000
50%	3.0000	5.0000	4.0000	4.0000	5.0000
25%	2.5000	4.0000	3.0000	3.0000	3.5000

Kruskal-Wallis-Test: p-value = 0.3522

Wilcoxon-Test: humorvoll -höflich: p-value = 0.04755

Frage 3.6: Wie flexibel ist die Anwendung?

Positiven Einfluss üben die Fehlermeldungen mit **motivierender** Tonalität auf den Gesamteindruck bezüglich Flexibilität aus. Negativen Einfluss hat die **höfliche** Tonalität. Die Unterschiede zwischen den Tonalitäten höflich und motivierend ist **signifikant** (p-value=0.01345).

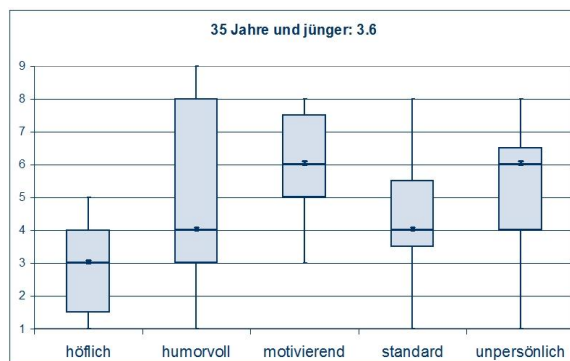


Abbildung 45: 35 Jahre und jünger: Frage 3.6

	höflich	humorv.	motivier.	standard	unpers.
75%	4.0000	8.0000	7.5000	5.5000	6.5000
50%	3.0000	4.0000	6.0000	4.0000	6.0000
25%	1.5000	3.0000	5.0000	3.5000	4.0000

Kruskal-Wallis chi-squared = 6.701, df = 4, p-value = 0.1526

Wilcoxon-Test: motivierend -höflich: p-value = 0.01345

Bei den Probanden im Alter von bis und mit 35 Jahren scheint die höfliche Form der Fehlermeldung nicht sehr ansprechend zu sein. Durch diese Tonalität wird der Gesamteindruck massiv schlechter bewertet als unter dem Einfluss anderer Tonalität wie z.B.

humorvoll und unpersönlich. Speziell ist auch die hohe Akzeptanz von unpersönlichen Fehlermeldungen, welche ja in einem starken Kontrast zu humorvoll steht.

Eine mögliche Erklärung für die Neigung zu unpersönlichen Fehlermeldungen könnte sein, dass es unter den jüngeren Probanden mehr erfahrene Benutzer gibt als unerfahrene.

Von den 42 Probanden buchen 10 Personen 4x pro Jahr und öfter, 15 Personen buchen 2x pro Jahr, 12 Personen 1x pro Jahr, 2 Personen alle 2 Jahre, und 3 Personen seltener als alle 3 Jahre.

Auswertung über Gruppe 45 jährig und älter

Durch die Gruppierung nach Personen, die 45 Jahre und älter sind, erhalten wir eine Datenmenge n=38. D.h. wir haben gesamthaft 38 Personen unter unseren Probanden, welche gleich oder älter als 45 Jahre alt sind. Das macht folgende Aufteilung pro Tonalität:

- höflich: 9 Personen
- humorvoll: 4 Personen
- motivierend: 8 Personen
- standard: 10 Personen
- unpersönlich: 7 Personen

Bei älteren Personen (ab 45 Jahre alt) wird der Gesamteindruck durch *motivierende* Fehlermeldungen positiv beeinflusst. Die grösste Differenz wurde bei der Tonalität motivierend und *unpersönlich* errechnet.

Der Unterschied zwischen der Tonalität motivierend und unpersönlich ist *nicht signifikant* (p-value=0.1462).

Es könnten sich durchaus bei den einzelnen Fragen aus diesem Kapitel signifikante Unterschiede bei der Tonalität ergeben.

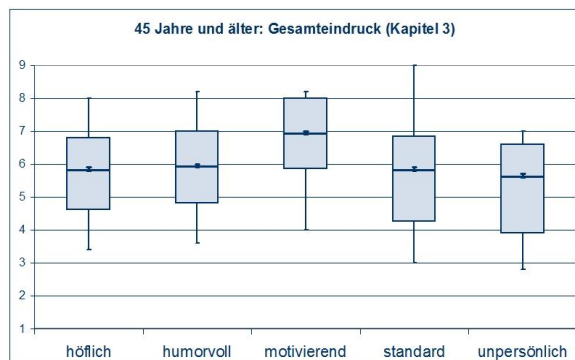


Abbildung 46: 45 Jahre und älter: Kapitel 3

	höflich	humorv.	motivier.	standard	unpers.
75%	6.8000	7.0000	8.0000	6.8500	6.6000
50%	5.8000	5.9000	6.9000	5.8000	5.6000
25%	4.6000	4.8000	5.8500	4.2500	3.9000

Kruskal-Wallis-Test: p-value = 0.5564

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.1462

Frage 3.1: Wie angenehm ist die Anwendung?

Einen angenehmen Gesamteindruck hinterlässt auch hier die Tonalität **motivierend**. Im Kontrast dazu steht die Tonalität **unpersönlich**. Unpersönlich hat statistisch gesehen die grösste Differenz zu motivierend. D.h. motivierend wird durch diese Altersgruppe am angenehmsten empfunden, hingegen unpersönlich am unangenehmsten. Die Differenzen sind jedoch bei dieser Frage und Altersgruppe **nicht signifikant** (p-value=0.1096).

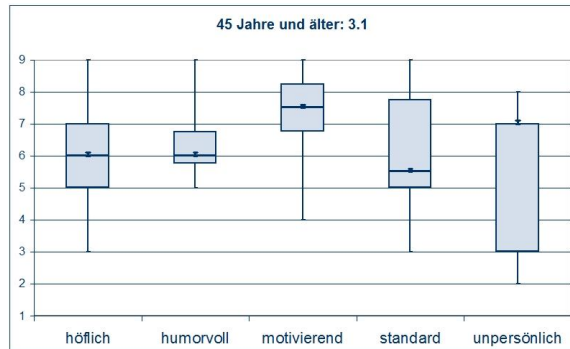


Abbildung 47: 45 Jahre und älter: Frage 3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	6.7500	8.2500	7.7500	7.0000
50%	6.0000	6.0000	7.5000	5.5000	7.0000
25%	5.0000	5.7500	6.7500	5.0000	3.0000

Kruskal-Wallis-Test: p-value = 0.481

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.1096

Frage 3.3: Wie anregend ist die Anwendung?

Die älteren Benutzer finden, dass die Tonalität **höflich** einen positiven Einfluss auf die Frage hat, wie anregend die Anwendung ist. Hingegen hat die Tonalität **standard** einen eher negativen Einfluss auf die Anregung.

Die Differenz zwischen höflich und standard ist **nicht signifikant** (p-value=0.0838), aber durch den tiefen Signifikanzwert unter 0.1 können wir auch hier von tendenziellem Einfluss sprechen.

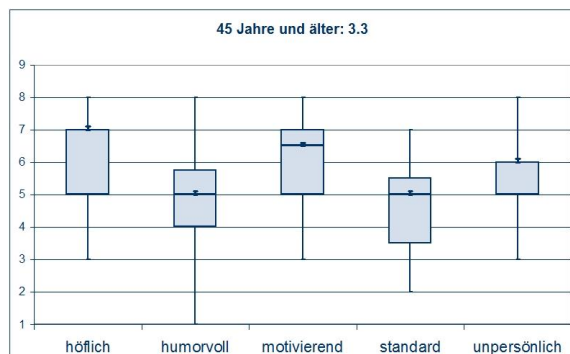


Abbildung 48: 45 Jahre und älter: Frage 3.3

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	5.7500	7.0000	5.5000	6.0000
50%	7.0000	5.0000	6.5000	5.0000	6.0000
25%	5.0000	4.0000	5.0000	3.5000	5.0000

Kruskal-Wallis-Test: p-value = 0.3625

Wilcoxon-Test: höflich-standard: p-value = 0.0838

Frage 3.6: Wie flexibel ist die Anwendung?

Motivierende Fehlermeldungen haben den besten Einfluss auf die Empfindung, wie flexibel eine Anwendung erscheint. Auffallend ist auch hier, dass bei dieser Altersgruppe **unpersönliche** Fehlermeldungen einen eher schlechteren Einfluss haben.

Der Unterschied zwischen der Tonalität motivierend und unpersönlich ist **signifikant** (p-value=0.03826).

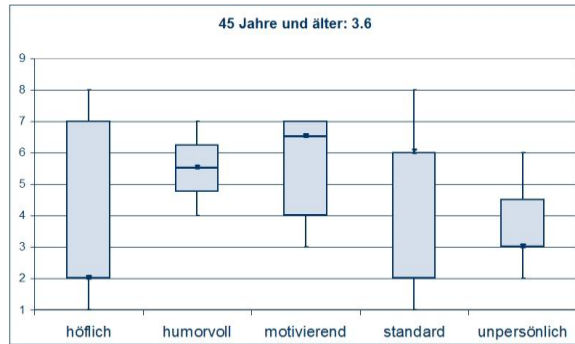


Abbildung 49: 45 Jahre und älter: Frage 3.6

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	6.2500	7.0000	6.0000	4.5000
50%	2.0000	5.5000	6.5000	6.0000	3.0000
25%	2.0000	4.7500	4.0000	2.0000	3.0000

Kruskal-Wallis-Test: p-value = 0.3273

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.03826

Bei den 45 jährigen und älter wird im Gegensatz zu den jüngeren eine motivierende Fehlermeldung favorisiert. D.h. der Gesamteindruck einer Anwendung wird durch die Verwendung einer motivierenden Fehlermeldung positiv geprägt.

10.5.2 KAPITEL 4: BILDSCHIRM (LAYOUT)

Die Gegenüberstellung der Tonalitäten zum Thema Bildschirm bzw. Layout, ist die Tonalität **motivierend** diejenige welche den besten Einfluss ausübt. **Höflich** ist die Tonalität welche den schlechtesten Einfluss auf die Beurteilung der Bildschirmfragen hat.

Der Vergleich zwischen der Tonalität motivierend und höflich ist **nicht signifikant** ($p\text{-value}=0.05675$). Der Wert liegt jedoch *unterhalb des 2. Signifikanzniveau*. Demzufolge kann man tendenziell davon ausgehen, dass diese Aussage doch ihre Richtigkeit hat.

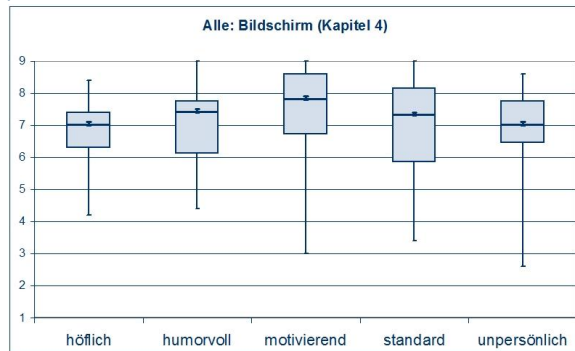


Abbildung 50: Alle: Kapitel 4

	höflich	humorv.	motivier.	standard	unpers.
75%	7.4000	7.7500	8.6000	8.1500	7.7500
50%	7.0000	7.4000	7.8000	7.3000	7.0000
25%	6.2750	6.1000	6.7000	5.8500	6.4500

Kruskal-Wallis-Test: $p\text{-value} = 0.3568$

Wilcoxon-Test: motivierend -höflich: $p\text{-value} = 0.05675$

Auswertung über alle Frauen

- Datenmenge n=30
- Höflich: 5 Frauen
- Humorvoll: 6 Frauen
- Motivierend: 8 Frauen
- Standard: 4 Frauen
- Unpersönlich: 7 Frauen

Bei den Frauen erscheint die Tonalität **unpersönlich** auch hier der Favorit zu sein. Der Vergleich zwischen den Tonalitäten **unpersönlich** und **standard** liefert die grösste Differenz.

Der Unterschied zwischen unpersönlich und **standard** ist jedoch **nicht signifikant** ($p\text{-value}=0.1074$).

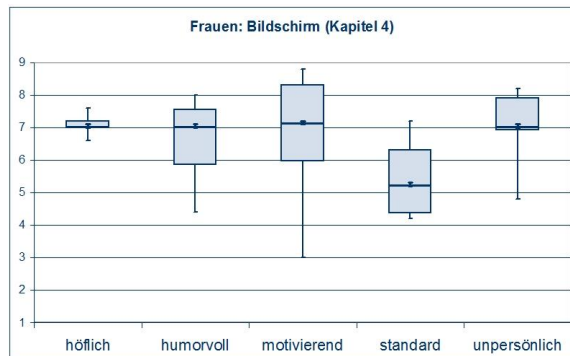


Abbildung 51: Frauen: Kapitel 4

	höflich	humorv.	motivier.	standard	unpers.
75%	7.2000	7.5500	8.3000	6.3000	7.9000
50%	7.0000	7.0000	7.1000	5.2000	7.0000
25%	7.0000	5.8500	5.9500	4.3500	6.9000

Kruskal-Wallis-Test: $p\text{-value} = 0.5307$

Wilcoxon-Test: unpersönlich -standard: $p\text{-value} = 0.1074$

Bei den Frauen sind keine einzelnen Fragen von der Auswertung her aufgefallen, sodass diese näher angeschaut werden müssten. Die Unterschiede sind in diesem Kapitel zu klein.



Auswertung über alle Männer

Durch die Gruppierung nach Männern, erhalten wir eine Datenmenge $n=80$. D.h. wir haben gesamthaft 80 Männer unter unseren Probanden. Das macht folgende Aufteilung pro Tonalität:

- Datenmenge $n=80$
- höflich: 17 Personen
- humorvoll: 16 Personen
- motivierend: 14 Personen
- standard: 18 Personen
- unpersönlich: 15 Personen

Bei den Männern ist auch beim Thema Bildschirm (Layout) die Tonalität **motivierend** die Art von Fehlermeldung, welche die höchste Akzeptanz bei der Anwendung bewirkt. Mit Abstand die schlechtesten Bewertungen haben die Tonalitäten **höflich und unpersönlich** erhalten.

Der Vergleich zwischen höflich und motivierend wie ist **signifikant** ($p\text{-value}=0.03497$) aber auch der Vergleich zwischen unpersönlich und motivierend ist **signifikant** ($p\text{-value}=0.02872$) bzw. noch deutlicher.

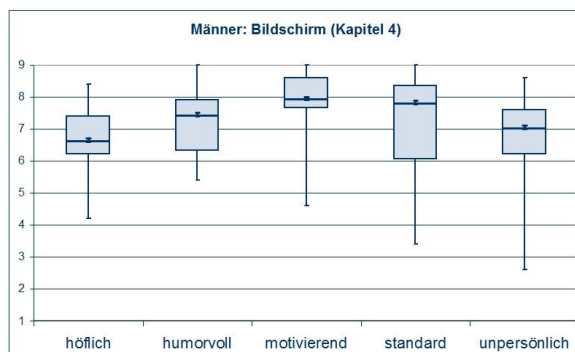


Abbildung 52: Männer: Kapitel 4

	höflich	humorv.	motivier.	standard	unpers.
75%	7.4000	7.9000	8.6000	8.3500	7.6000
50%	6.6000	7.4000	7.9000	7.7750	7.0000
25%	6.2000	6.3000	7.6500	6.0500	6.2000

Kruskal-Wallis chi-squared = 6.8373, df = 4, p-value = 0.1447

Wilcoxon-Test:

motivierend-höflich: p-value = 0.03497

motivierend-unpersönlich: p-value = 0.02872



Frage 4.1: Wie leicht waren die Zeichen auf dem Bildschirm erkennbar?

Männer mit **motivierenden** Fehlermeldungen waren der Meinung, die Zeichen auf dem Bildschirm wären besser erkennbar, als Männer welche **unpersönliche** Fehlermeldungen hatten. Der Vergleich zwischen den Tonalitäten **motivierend** und **unpersönlich** ist **hoch signifikant** (p-value=0.004889). Ähnlich, aber mit etwas kleinerem Unterschied verhält es sich mit dem Vergleich zwischen **motivierend** und **höflich**. Auch dieser Unterschied ist **signifikant** (p-value=0.0326).

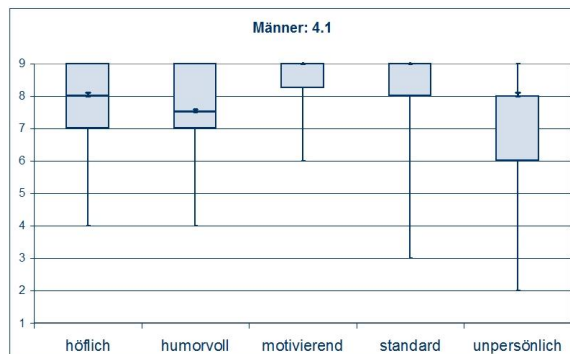


Abbildung 53: Männer: Frage 4.1

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	8.0000
50%	8.0000	7.5000	9.0000	9.0000	8.0000
25%	7.0000	7.0000	8.2500	8.0000	6.0000

Kruskal-Wallis-Test: p-value = 0.07017

Wilcoxon-Test:

motivierend- höflich: p-value = 0.0326

motivierend-unpersönlich: p-value = 0.004889



Auswertung über Personen 35 jährig und jünger

- Datenmenge n=42
- höflich: 7 Personen
- humorvoll: 13 Personen
- motivierend: 7 Personen
- standard: 8 Personen
- unpersönlich: 7 Personen

Der Favorit bei den jüngeren Probanden bezüglich Thema Bildschirm ist die Tonalität **motivierend**. Hingegen ist die Tonalität **höflich** diejenige welche die schlechteste Bewertung aufweist.

Bei diesem Thema ist die Differenz zwischen höflich und motivierend die grösste Differenz. Der Vergleich zwischen diesen beiden Tonalitäten ist jedoch **nicht signifikant** (p-value=0.09397). Der Wert liegt jedoch *unter dem 2. Signifikanzniveau* von 0.1. D.h. man kann tendenziell davon ausgehen, dass die Aussage hier korrekt ist

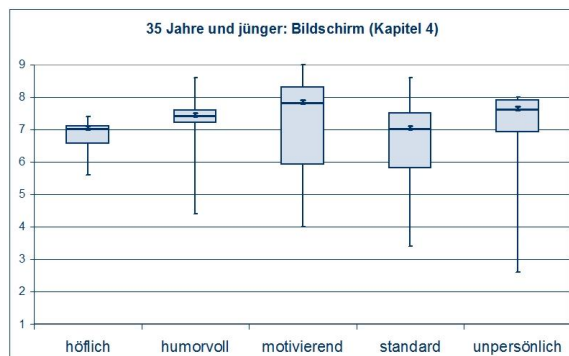


Abbildung 54: 35 Jahre und jünger: Kapitel 4

	höflich	humorv.	motivier.	standard	unpers.
75%	7.1000	7.6000	8.3000	7.5000	7.9000
50%	7.0000	7.4000	7.8000	7.0000	7.6000
25%	6.5500	7.2000	5.9000	5.8000	6.9000

Kruskal-Wallis chi-squared = 2.8749, df = 4, p-value = 0.579

Wilcoxon-Test: motivierend -höflich: p-value = 0.09397

Bezüglich Layout ist die Tonalität motivierend diejenige welche tendenziell von den jüngeren Probanden favorisiert wird. Hingegen bewirkt die Tonalität höflich genau das Gegenteil. Höfliche Fehlermeldungen lassen jüngere Probanden bei Fragen zu Bildschirm schlechtere Bewertung zu geben. Die Unterschiede sind bei dieser Thematik nicht signifikant, aber signifikant genug, um von Tendenzen zu sprechen.

Auswertung über Personen 45 jährig und älter

- Datenmenge n=38
- höflich: 9 Personen
- humorvoll: 4 Personen
- motivierend: 8 Personen
- standard: 10 Personen
- unpersönlich: 7 Personen

Beim Thema Bildschirm scheint die Tonalität **motivierend** Favorit gegenüber **unpersönlich** zu sein.

Der Unterschied zwischen motivierend und unpersönlich ist **signifikant** (p-value=0.0636).

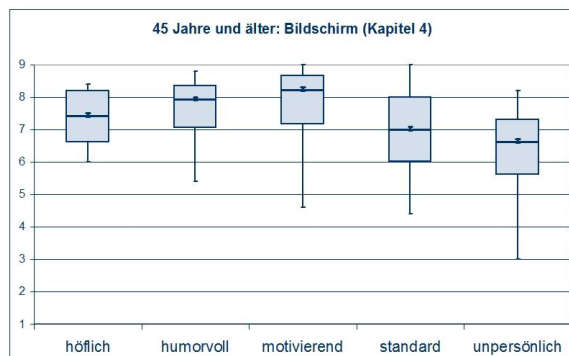


Abbildung 55: 45 Jahre und älter: Kapitel 4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.2000	8.3500	8.6500	8.0000	7.3000
50%	7.4000	7.9000	8.2000	6.9750	6.6000
25%	6.6000	7.0500	7.1500	6.0000	5.6000

Kruskal-Wallis-Test: p-value = 0.3213

Wilcoxon-Test: motivierend-unpersönlich; p-value = 0.0636

Frage 4.1: Wie leicht waren die Zeichen auf dem Bildschirm erkennbar?

Kann die Tonalität die Erkennbarkeit der Zeichen auf dem Bildschirm beeinflussen?

Die **motivierenden** Fehlermeldungen werden durch die älteren Probanden favorisiert, wenn es um die Beantwortung dieser Frage geht. D.h. die Probanden welche motivierende Fehlermeldungen hatten bewerteten diesen Punkt massiv besser als die Probanden mit der Tonalität **unpersönlich**.

Der Unterschied zwischen den Tonalitäten motivierend und unpersönlich ist **signifikant** (p-value=0.01972).

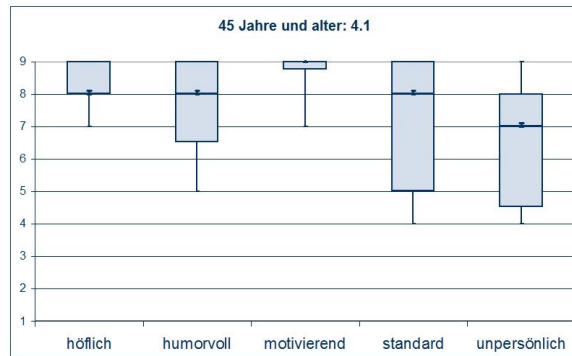


Abbildung 56: 45 Jahre und älter: Frage 4.1

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	8.0000
50%	8.0000	8.0000	9.0000	8.0000	7.0000
25%	8.0000	6.5000	8.7500	5.0000	4.5000

Kruskal-Wallis chi-squared = 7.0758, df = 4, p-value = 0.1319

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.01972

Frage 4.3: Wie hilfreich waren die Bildschirmlayouts?

Die favorisierte Tonalität ist auch hier **motivierend**. Auf der anderen Seite ist **höflich** die Tonalität, welche am wenigsten zu hilfreichen Layouts beitragen konnten.

Die Unterschiede zwischen motivierend und höflich sind **nicht signifikant** (p-value=0.07961). Der Wert liegt jedoch unterhalb des 2. Signifikanzniveau von 0.1 und sollte demzufolge tendenziell beachtet werden.

Die Unterschiede zwischen **motivierend** und **standard** sind noch **weniger signifikant** (p-value=0.1176).

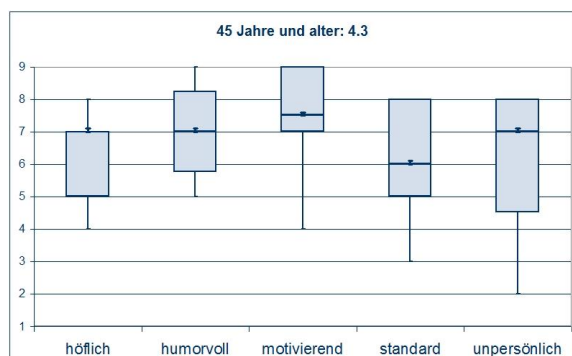


Abbildung 57: 45 Jahre und älter: Frage 4.3

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	8.2500	9.0000	8.0000	8.0000
50%	7.0000	7.0000	7.5000	6.0000	7.0000
25%	5.0000	5.7500	7.0000	5.0000	4.5000

Kruskal-Wallis-Test: p-value = 0.3897

Wilcoxon-Test:

motivierend-höflich: p-value = 0.07961

motivierend-standard: p-value = 0.1176

10.5.3 KAPITEL 5: TERMINOLOGIE UND SYSTEMINFORMATIONEN

Beim Kapitel Terminologie und Systeminformationen geht es inhaltlich um Systemfeedback, also auch um Fehlermeldungen.

Grundsätzlich haben Probanden, welche mit der Tonalität *motivierend* getestet haben, die beste Beurteilung abgegeben. Die grösste Differenz erzielte *motivierend* und *unpersönlich*. Demzufolge ist *unpersönlich* die Tonalität mit der schlechtesten Bewertung. Die Unterschiede zwischen den Tonalitäten sind *nicht signifikant* (p-value=0.2907).

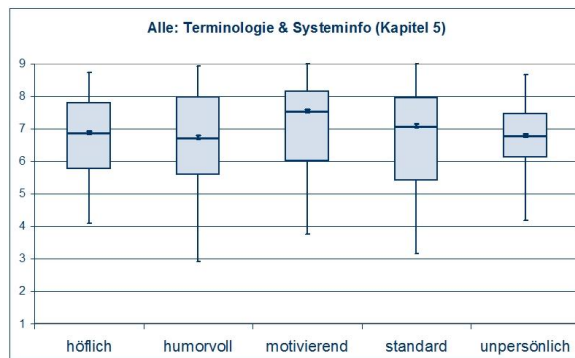


Abbildung 58: Alle: Kapitel 5

	höflich	humorv.	motivier.	standard	unpers.
75%	7.8011	7.9773	8.1572	7.9583	7.4583
50%	6.8333	6.6894	7.5000	7.0417	6.7500
25%	5.7625	5.5833	6.0000	5.3958	6.1042

Kruskal-Wallis-Test: p-value = 0.854

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.2907

Auswertung über Gruppe Frauen

- Datenmenge n=30
- höflich: 5 Personen
- humorvoll: 6 Personen
- motivierend: 8 Personen
- standard: 4 Personen
- unpersönlich: 7 Personen

Zum Thema Terminologie und Systeminformationen wurde bei den Frauen die höchste Differenz zwischen den Tonalitäten *unpersönlich* mit bester Bewertung und *humorvoll* mit schlechtester Bewertung festgestellt. Der Unterschied zwischen unpersönlich und humorvoll ist ganz knapp *nicht signifikant*. Der Wert liegt *unterhalb des 2. Signifikanzniveau von 0.1* und demzufolge sollte auch diese Aussage tendenziell Beachtung geschenkt werden.

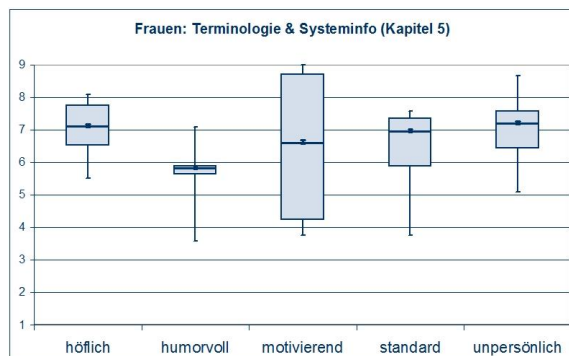


Abbildung 59: Frauen: Kapitel 5

	Höflich	humorv.	motivier.	standard	unpers.
75%	7.7500	5.8958	8.7083	7.3504	7.5833
50%	7.0833	5.7917	6.5833	6.9280	7.1667
25%	6.5000	5.6250	4.2292	5.8750	6.4167

Kruskal-Wallis-Test: p-value = 0.4276

Wilcoxon-Test: unpersönlich - humorvoll: p-value = 0.05128



Frage 5.1: Wie konsistent war die Ausdrucksweise im Formular?

Generell werden bei Frauen unpersönliche Fehlermeldungen favorisiert. Bei dieser Fragestellung ist die Tonalität höflich leicht besser bewertet als unpersönlich.

Höfliche Fehlermeldungen vermitteln dem Benutzer am ehesten eine konsistente Ausdrucksweise im Formular. Benutzer finden bei **humorvollen** Fehlermeldungen die Ausdrucksweise am inkonsistentesten von allen.

Der Vergleich zwischen höflicher und humorvoller Tonalität ist **signifikant** (p-value=0.04384)

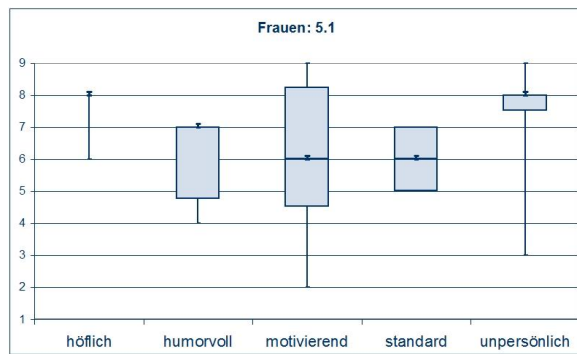


Abbildung 60: Frauen: Frage 5.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.2500	7.0000	8.0000
50%	8.0000	7.0000	6.0000	6.0000	8.0000
25%	8.0000	4.7500	4.5000	5.0000	7.5000

Kruskal-Wallis-Test: p-value = 0.1909

Wilcoxon-Test: höflich-humorvoll: p-value = 0.04384

Frage 5.1.2: Wie konsistent war die aufgabenbezogene Ausdrucksweise?

Auch bei der Frage nach aufgabenbezogenen Ausdrucksweise war die Meinung der Probandinnen, dass eine **höfliche** Tonalität den grössten Einfluss auf eine hohe Konsistenz hat. Im Gegensatz dazu steht die **humorvolle** Tonalität, welche den grössten Einfluss für eine tiefe Konsistenz hat.

Der Unterschied zwischen höflich und humorvoll ist **hoch signifikant** (p-value=0.004288). Der Vergleich zwischen **höflich** und **standard** ergibt ebenfalls einen **signifikanten** Unterschied (p-value=0.01041)

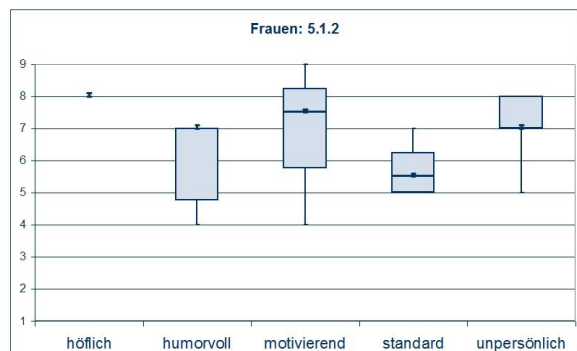


Abbildung 61: Frauen: Frage 5.1.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.2500	7.0000	8.0000
50%	8.0000	7.0000	6.0000	6.0000	8.0000
25%	8.0000	4.7500	4.5000	5.0000	7.5000

Kruskal-Wallis-Test: p-value = 0.03765

Wilcoxon-Test:

höflich-humorvoll: p-value = 0.004288

höflich-standard: p-value = 0.01041



Frage 5.2: Wie oft passte die Ausdrucksweise zur auszuführenden Arbeit?

Frauen sind der Meinung, dass die **höfliche** Tonalität besser zum Hotelbuchungssystem passt, und sie sind sich auch einig, dass **humorvoll** im Gegensatz dazu, eben nicht passt.

Der Unterschied zwischen höflich und humorvoll ist **signifikant** (p-value=0.04225).

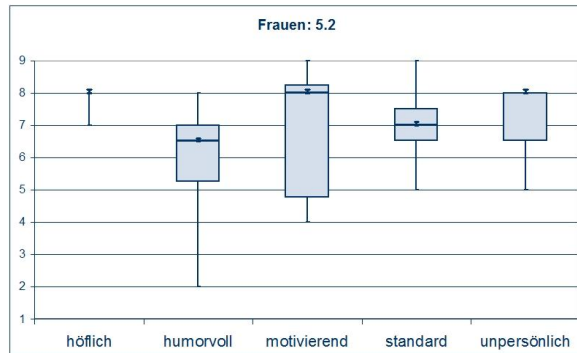


Abbildung 62: Frauen: Frage 5.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.2500	7.5000	8.0000
50%	8.0000	6.5000	8.0000	7.0000	8.0000
25%	8.0000	5.2500	4.7500	6.5000	6.5000

Kruskal-Wallis-Test: p-value = 0.4435

Wilcoxon-Test: höflich-humorvoll: p-value = 0.04225

Frage 5.3: Wie konsistent waren die Fehlermeldungen?

Interessant ist, dass Konsistenz, wenn es sich explizit auf die Fehlermeldungen bezieht, nicht mehr höflich die präferierte Tonalität ist, sondern **unpersönlich**. Frauen welche die unpersönliche Tonalität getestet haben, waren der Meinung, die Fehlermeldungen seien in einem hohen Masse konsistent. Hingegen die Tonalität **standard** wird im Gegensatz dazu als eher inkonsistent erachtet.

Der Unterschied zwischen der Tonalität standard und unpersönlich ist **nicht signifikant** (p-value=0.08098). Tendenziell sollte diesem Paar (unpersönlich-standard) eine gewisse Beachtung geschenkt werden zu diesem Thema.

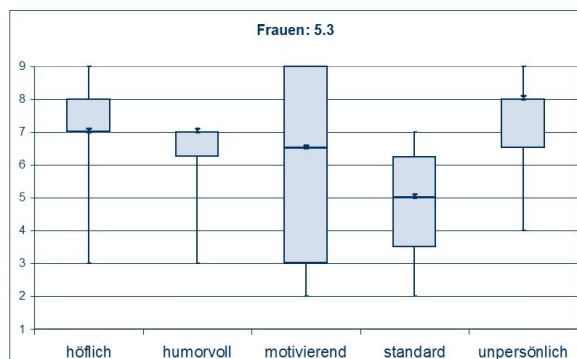


Abbildung 63: Frauen: Frage 5.3

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	9.0000	6.2500	8.0000
50%	7.0000	7.0000	6.5000	5.0000	8.0000
25%	7.0000	6.2500	3.0000	3.5000	6.5000

Kruskal-Wallis-Test: p-value = 0.3909

Wilcoxon-Test: unpersönlich -standard: p-value = 0.08098



Frage 5.5.2: Wie oft führte die Ausführung einer Operation zu einem vorhersehbarem Ergebnis?

Die **höfliche** Tonalität vermittelt bei den Frauen eine höhere Vorhersehbarkeit der Ergebnisse als die **humorvolle** Tonalität.

Die Unterschiede zwischen höflich und humorvoll ist **signifikant** (p-value=0.03083).

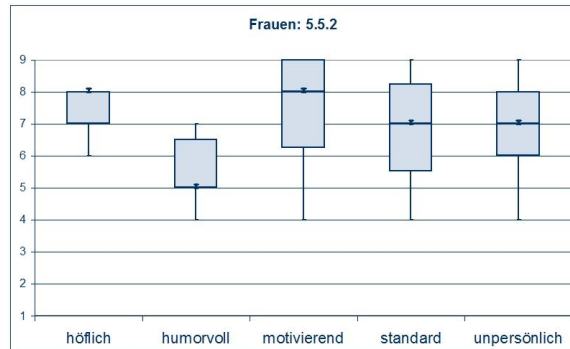


Abbildung 64: Frauen: Frage 5.5.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	6.5000	9.0000	8.2500	8.0000
50%	8.0000	5.0000	8.0000	7.0000	7.0000
25%	7.0000	5.0000	6.2500	5.5000	6.0000

Kruskal-Wallis-Test: p-value = 0.3267

Wilcoxon-Test: höflich-humorvoll: p-value = 0.03083

Frage 5.5.4: Wie akzeptabel waren die Wartezeiten zwischen den Operationen?

Grundsätzlich scheinen alle Frauen der Meinung zu sein, die Wartezeiten sind akzeptabel. Frauen mit der Tonalität **höflich** hatten die höchste Akzeptanz gegenüber Wartezeiten. Hingegen hatten Frauen mit der Tonalität **humorvoll** die schlechteste Akzeptanz.

Der Unterschied zwischen höflich und humorvoll ist **signifikant** (p-value=0.01403).

D.h. die Tonalität hat einen Einfluss darauf ob die Wartezeit mehr oder weniger akzeptiert wird.

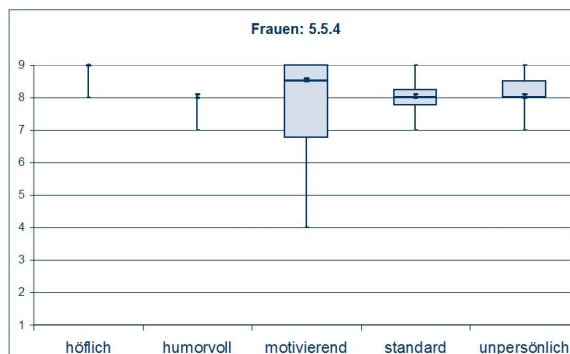


Abbildung 65: Frauen: Frage 5.5.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	8.0000	9.0000	8.2500	8.5000
50%	9.0000	8.0000	8.5000	8.0000	8.0000
25%	9.0000	8.0000	6.7500	7.7500	8.0000

Kruskal-Wallis-Test: p-value = 0.2575

Wilcoxon-Test: höflich-humorvoll: p-value = 0.01403



Frage 5.6: Wie hilfreich waren die Fehlermeldungen?

Gemäss Paarvergleichstest haben die Tonalitäten unpersönlich und humorvoll den signifikantesten Unterschied von allen. D.h. **unpersönliche** Fehlermeldungen wurden als die hilfreichsten deklariert und die **humorvollen** als die am wenigsten hilfreich.

Gemäss Grafik scheint die Tonalität standard besser zu sein als unpersönlich, jedoch basierend auf der Anzahl Probanden pro Tonalität, steht unpersönlich mit 7 Probanden gegenüber standard mit 4 Probanden besser da.

Der Unterschied zwischen unpersönlich und humorvoll ist nicht signifikant (p-value=0.1936).

Es kann durchaus sein, dass Ausdrücke mit denen man immer wieder konfrontiert wird, hilfreicher erscheinen, da man evtl. bereits einmal Fehler auf diese Weise erfolgreich beheben konnte. Neue Formulierungen erfordern mehr Denkleistung vom Benutzer. Dieser Aspekt könnte mit einer spezifischen Untersuchung in diesem Bereich besser erforscht werden.

Frage 5.6.1: Wie oft erklärten die Fehlermeldungen das Problem?

Motivierende Fehlermeldungen scheinen das Problem besser zu erklären als **humorvolle**. Paarvergleichstests haben ergeben, dass bei diesen zwei Tonalitäten der grösste Unterschied besteht.

Der Unterschied zwischen humorvoll und motivierend ist **signifikant** (p-value=0.03011)

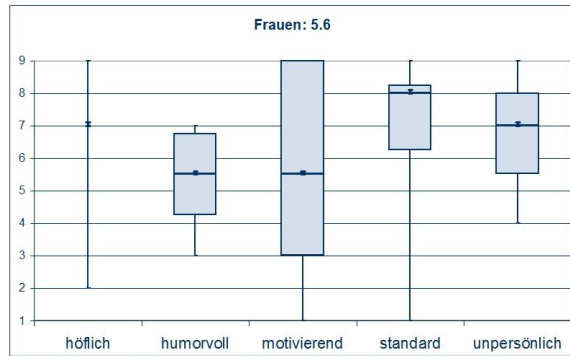


Abbildung 66: Frauen: Frage 5.6

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	6.7500	9.0000	8.2500	8.0000
50%	7.0000	5.5000	5.5000	8.0000	7.0000
25%	7.0000	4.2500	3.0000	6.2500	5.5000

Kruskal-Wallis-Test: p-value = 0.7359

Wilcoxon-Test: unpersönlich-humorvoll: p-value = 0.1936

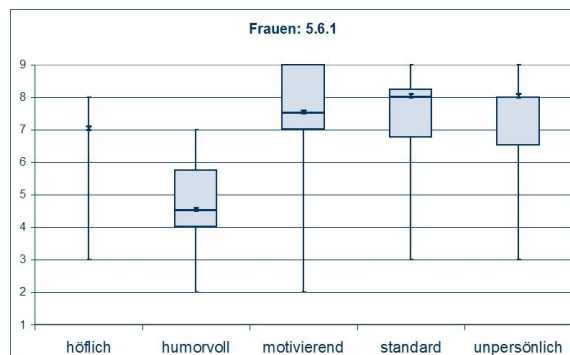


Abbildung 67: Frauen: Frage 5.6.1

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	5.7500	9.0000	8.2500	8.0000
50%	7.0000	4.5000	7.5000	8.0000	8.0000
25%	7.0000	4.0000	7.0000	6.7500	6.5000

Kruskal-Wallis-Test: p-value = 0.1365



Frage 5.6.2: Wie angenehm waren die Formulierung der Fehlermeldungen?

Die angenehmste Formulierung der Fehlermeldung ist gemäss Aussage der Frauen die **unpersönliche**. Hingegen die unangenehmste Formulierung ist die **humorvolle** Art und Weise.

Der Unterschied zwischen humorvoll und unpersönlich ist **nicht signifikant** (p-value=0.1942).

Wilcoxon-Test: motivierend -humorvoll: p-value = 0.03011

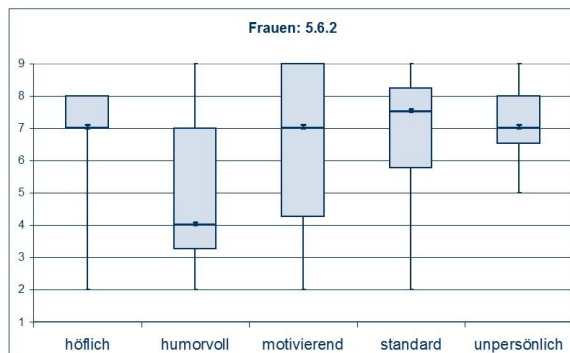


Abbildung 68: Frauen: Frage 5.6.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	9.0000	8.2500	8.0000
50%	7.0000	4.0000	7.0000	7.5000	7.0000
25%	7.0000	3.2500	4.2500	5.7500	6.5000

Kruskal-Wallis-Test: p-value = 0.8608

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.1942



Auswertung über alle Männer

- Datenmenge n=80
- höflich: 17 Männer
- humorvoll 16 Männer
- motivierend 14 Männer
- standard 18 Männer
- unpersönlich 15 Männer

Zum Thema Terminologie und Systeminformationen wurde bei den Männern *motivierend* als die beste Tonalität und *unpersönlich* als die schlechteste Tonalität festgestellt.

Der Paarvergleichstest hat den grössten Unterschied zwischen diesen zwei Tonalitäten festgestellt. Der Unterschied ist jedoch *nicht signifikant* (p-value=0.09698). Aber der Wert liegt noch immer *unter dem 2. Signifikanzniveau von 0.1*, und kann demzufolge als Tendenz durchaus erwähnt werden.

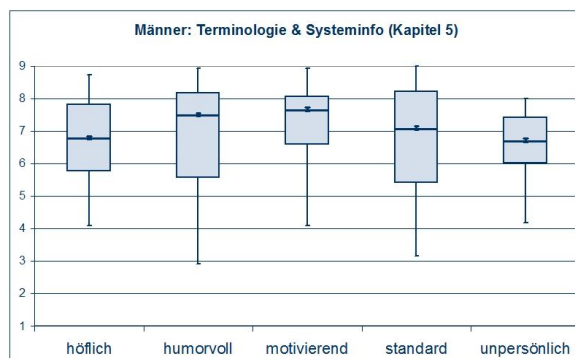


Abbildung 69: Männer: Kapitel 5

	höflich	humorv.	motivier.	standard	unpers.
75%	7.8182	8.1667	8.0625	8.2083	7.4167
50%	6.7500	7.4583	7.6250	7.0417	6.6667
25%	5.7500	5.5625	6.5708	5.3958	6.0000

Kruskal-Wallis-Test: p-value = 0.6068

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.09698



Frage 5.5.4: Wie akzeptabel waren die Wartezeiten zwischen den Operationen?

Für Männer mit der Tonalität *motivierend* war die Wartezeit viel akzeptabler als für Männer mit der Tonalität *unpersönlich*.

Der Unterschied zwischen diesen beiden Tonalitäten ist *signifikant* (p-value=0.02647).

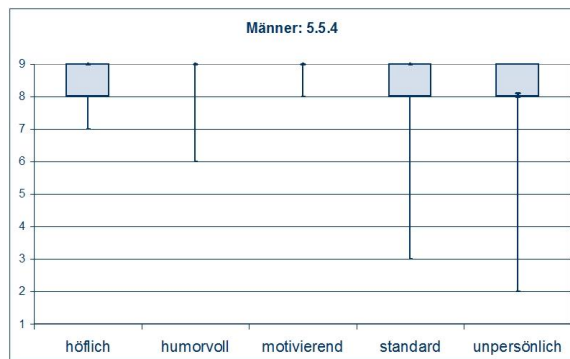


Abbildung 70: Männer: Frage 5.5.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	9.0000
50%	9.0000	9.0000	9.0000	9.0000	8.0000
25%	8.0000	9.0000	9.0000	8.0000	8.0000

Kruskal-Wallis-Test: p-value = 0.1444

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.02647



Auswertung über Personen 35 jährig und jünger

- Datenmenge n=42
- höflich 7 Personen
- humorvoll 13 Personen
- motivierend 7 Personen
- standard 8 Personen
- unpersönlich 7 Personen

Bei den jüngeren Teilnehmern ist bezüglich dem Thema Terminologie und Systeminformationen die favorisierte Tonalität **motivierend**. Hingegen ist **standard** die Tonalität welche die schlechteste Wertung erzielte.

Der Unterschied zwischen den beiden Tonalitäten ist in diesem Fall jedoch **nicht signifikant** (p-value=0.2972).

Die Frage ist: vermag die Untersuchung nach einzelnen Fragen aus diesem Kapitel eine signifikante Aussage zu machen?

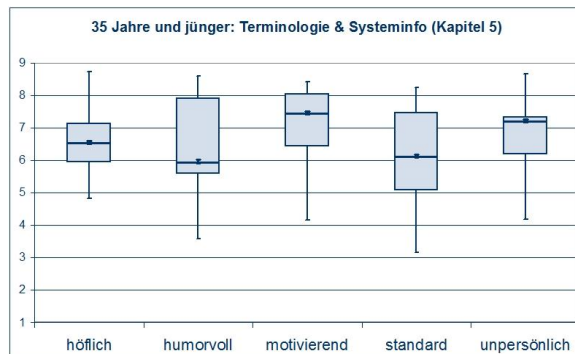


Abbildung 71: 35 Jahre und jünger: Kapitel 5

	höflich	humorv.	motivier.	standard	unpers.
75%	7.1250	7.9091	8.0417	7.4583	7.3333
50%	6.5000	5.9167	7.4167	6.0833	7.1667
25%	5.9417	5.5833	6.4167	5.0625	6.1667

Kruskal-Wallis-Test: p-value = 0.7813

Wilcoxon-Test: motivierend-standard: p-value = 0.2972

Frage 5.1: Wie konsistent war die Ausdrucksweise im Formular?

Bei den Jüngeren fördert eine **unpersönliche** Fehlermeldung den Eindruck einer konsistenten Ausdrucksweise im Formular. Hingegen leistet die **humorvolle** Tonalität von allen Tonalitäten am wenigsten für eine konsistente Ausdrucksweise.

Dieser Unterschied ist **signifikant** (p-value=0.02935).

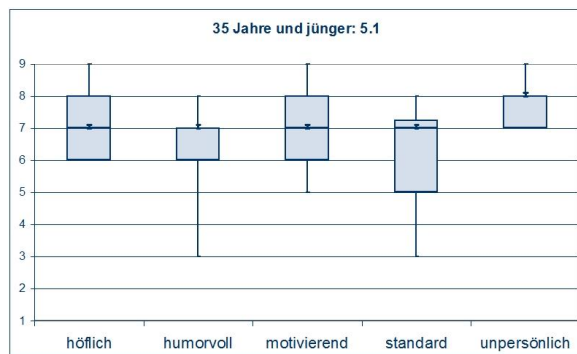


Abbildung 72: 35 Jahre und jünger: Frage 5.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.0000	7.2500	8.0000
50%	7.0000	7.0000	7.0000	7.0000	8.0000
25%	6.0000	6.0000	6.0000	5.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.2572

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.02935

Frage 5.1.2: Wie konsistent war die Aufgabenbezogene Ausdrucksweise?

Bei den Jüngeren erzeugt die **unpersönliche** Fehlermeldung eine stärkere aufgabenbezogene Ausdrucksweise als die **humorvolle** Tonalität.

Dieser Unterschied ist **hoch signifikant** (p-value=0.005387).

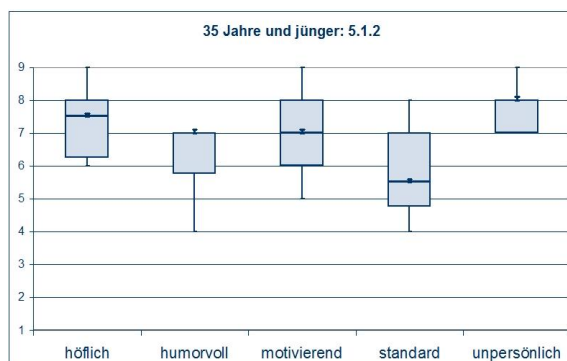


Abbildung 73: 35 Jahre und jünger: Frage 5.1.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.0000	7.0000	8.0000
50%	7.5000	7.0000	7.0000	5.5000	8.0000
25%	6.2500	5.7500	6.0000	4.7500	7.0000

Kruskal-Wallis-Test: p-value = 0.03769

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.005387

Frage 5.2: Wie oft passte die Ausdrucksweise zur auszuführenden Arbeit?

Bei den Jüngeren hat bezüglich zur ausführender Arbeit passender Ausdrucksweise die **unpersönliche** Tonalität den positivsten Einfluss gegenüber allen anderen Tonalitäten. Hingegen trägt eine **humorvolle** Tonalität am wenigsten zur passenden Ausdrucksweise bei.

Dieser Unterschied zwischen den beiden Tonalitäten ist **nicht signifikant**, jedoch **unterhalb des 2. Signifikanzniveau von 0.1**. Demzufolge sollte diese Aussage tendenziell beachtet werden.

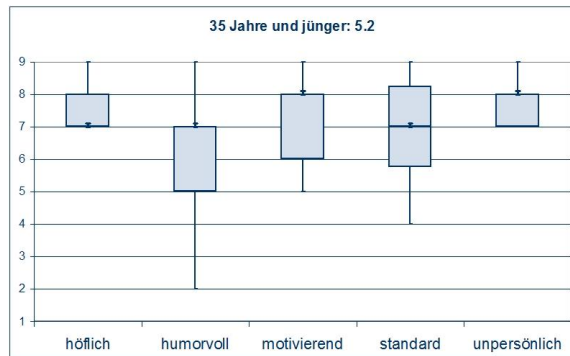


Abbildung 74: 35 Jahre und jünger: Frage 5.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	7.0000	8.0000	8.2500	8.0000
50%	7.0000	7.0000	8.0000	7.0000	8.0000
25%	7.0000	5.0000	6.0000	5.7500	7.0000

Kruskal-Wallis-Test: p-value = 0.3645

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.06918

Frage 5.3: Wie konsistent waren die Fehlermeldungen?

Die Fehlermeldungen bei welchen die Konsistenz am höchsten eingestuft wurde, ist die Tonalität **standard**. Hingegen die niedrigste Konsistenz wurde gemäss Bewertung der Tonalität **unpersönlich** attestiert.

Der Unterschied zwischen standard und unpersönlich **signifikant** (p-value=0.0366).

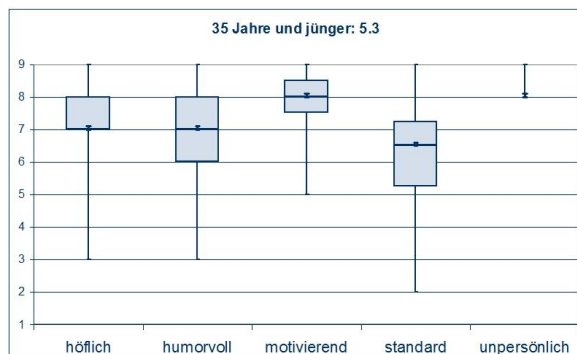


Abbildung 75: 35 Jahre und jünger: Frage 5.3

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.0000	8.5000	7.2500	8.0000
50%	7.0000	7.0000	8.0000	6.5000	8.0000
25%	7.0000	6.0000	7.5000	5.2500	8.0000

Kruskal-Wallis-Test: p-value = 0.1596

Wilcoxon-Test: unpersönlich -standard: p-value = 0.0366

Frage 5.5: Wie oft informiert der Computer über die laufende Aktion?

Junge Leute mit **motivierender** Tonalität fühlten sich massiv besser über laufende Aktionen informiert als junge Leute mit **höflicher** Tonalität.

Der Unterschied zwischen diesen beiden Tonalitäten ist **signifikant** (p-value=0.03025).

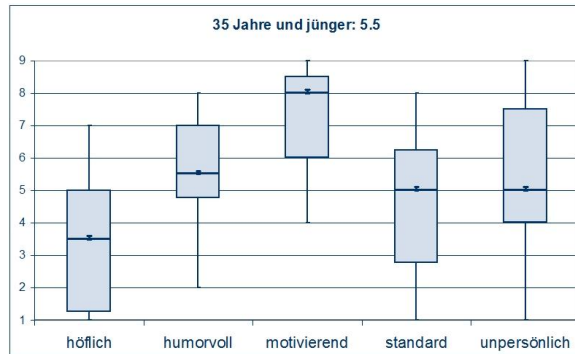


Abbildung 76: 35 Jahre und jünger: Frage 5.5

	höflich	humorv.	motivier.	standard	unpers.
75%	5.0000	7.0000	8.5000	6.2500	7.5000
50%	3.5000	5.5000	8.0000	5.0000	5.0000
25%	1.2500	4.7500	6.0000	2.7500	4.0000

Kruskal-Wallis-Test: p-value = 0.12

Wilcoxon-Test: motivierend -höflich: p-value = 0.03025

Frage 5.5.2: Wie oft führte die Ausführung einer Operation zu einem vorhersehbaren Ergebnis?

Ein vorhersehbares Ergebnis wurde bei den Jüngeren durch die Tonalität **motivierend** öfters attestiert als durch die Tonalität **höflich**.

Der Unterschied zwischen motivierend und höflich ist **signifikant** (p-value=0.03547).

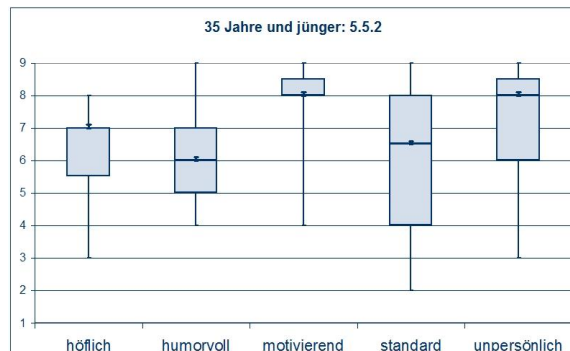


Abbildung 77: 35 Jahre und jünger: Frage 5.5.2

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	7.0000	8.5000	8.0000	8.5000
50%	7.0000	6.0000	8.0000	6.5000	8.0000
25%	5.5000	5.0000	8.0000	4.0000	6.0000

Kruskal-Wallis-Test: p-value = 0.314

Wilcoxon-Test: motivierend -höflich: p-value = 0.03547

Auswertung über Personen 45 jährig und älter

- Datenmenge n=38
- höflich 9 Personen
- humorvoll 4 Personen
- motivierend 8 Personen
- standard 10 Personen
- unpersönlich 7 Personen

Bei den älteren Benutzer (45 Jahre und älter) ist **motivierend** die favorisierte Tonalität. Im Gegensatz dazu ist die Tonalität **unpersönlich** am negativsten aufgefallen. D.h. die Terminologie und Systeminformationen bzw. die Effektivität der Rückmeldungen werden mit der Tonalität motivierend positiv beeinflusst. Hingegen die Tonalität unpersönlich beeinflusst die Effektivität der Rückmeldungen eher negativ.

Der Unterschied dieser beiden Tonalitäten ist **nicht signifikant** (p-value=0.07259). Der Wert ist jedoch *unter dem 2*. *Signifikanzniveau von 0.1* und sollte demzufolge tendenziell beachtet werden.

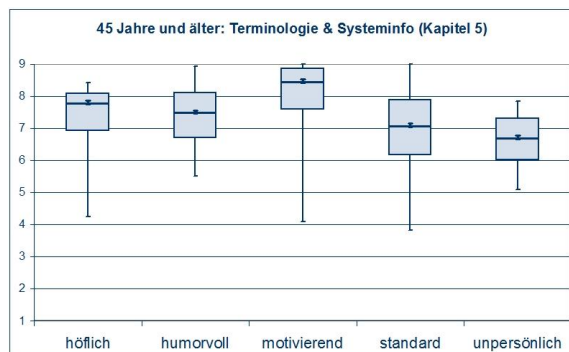


Abbildung 78: 45 Jahre und älter: Terminologie & Systeminformation (Kapitel 5)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0833	8.1042	8.8542	7.8807	7.3056
50%	7.7500	7.4583	8.4242	7.0417	6.6667
25%	6.9167	6.6875	7.5625	6.1458	6.0000

Kruskal-Wallis-Test: p-value = 0.2924

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.07259

Frage 5.1: Wie konsistent war die Ausdrucksweise im Formular?

Bei älteren Benutzer beeinflusst die **humorvolle** Tonalität auf positive Art und Weise. Hingegen wirkt sich die **unpersönliche** Tonalität, im Vergleich zu den anderen am schlechtesten auf die Konsistenz der Ausdrucksweise im Formular aus.

Der Unterschied zwischen diesen beiden Tonalitäten ist **nicht signifikant** (p-value=0.158).

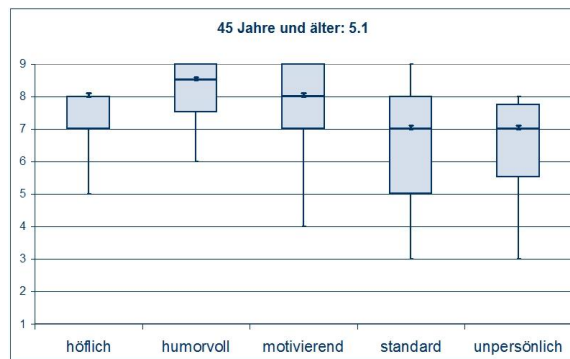


Abbildung 79: 45 Jahre und älter: Frage 5.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.0000	7.7500
50%	8.0000	8.5000	8.0000	7.0000	7.0000
25%	7.0000	7.5000	7.0000	5.0000	5.5000

Kruskal-Wallis-Test: p-value = 0.4035

Wilcoxon-Test: humorvoll-unpersönlich: p-value = 0.158

Frage 5.1.2: Wie konsistent war die Aufgabenbezogene Ausdrucksweise?

Die Konsistenz der Aufgabenbezogene Ausdrucksweise wird am stärksten durch die **motivierende** Tonalität positiv beeinflusst. Auf der anderen Seite wirkt sich die **unpersönliche** Tonalität eher negativ auf die Konsistenz aus.

Der Unterschied zwischen den beiden Tonalitäten ist **nicht signifikant** (p-value=0.08608). Aber dadurch dass der Wert *unterhalb des 2. Signifikanzniveau von 0.1* befindet, muss die Stellung der beiden Tonalitäten tendenziell beachtet werden.

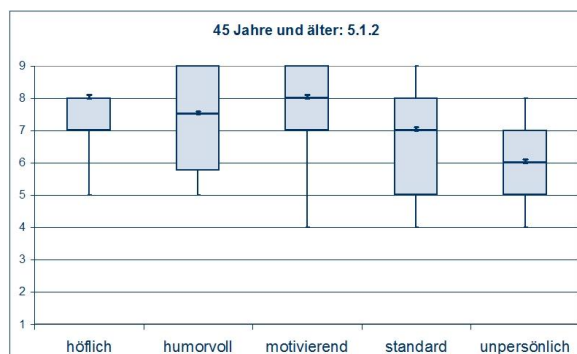


Abbildung 80: 45 Jahre und älter: Frage 5.1.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.0000	7.0000
50%	8.0000	7.5000	8.0000	7.0000	6.0000
25%	7.0000	5.7500	7.0000	5.0000	5.0000

Kruskal-Wallis-Test: p-value = 0.2849

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.08608

Frage 5.4: Wie klar und eindeutig waren die Fehlermeldungen?

Bei der Frage nach Klarheit bei Fehlermeldungen ist die **motivierende** Tonalität diejenige welche die Fehlermeldungen am wenigsten Verwirrung verursachte. Hingegen war von allen anderen Tonalitäten **unpersönlich** diejenige welche bezüglich Klarheit am schlechtesten bewertet wurde.

Der Unterschied zwischen den beiden Tonalitäten ist **nicht signifikant** (p-value=0.1056).

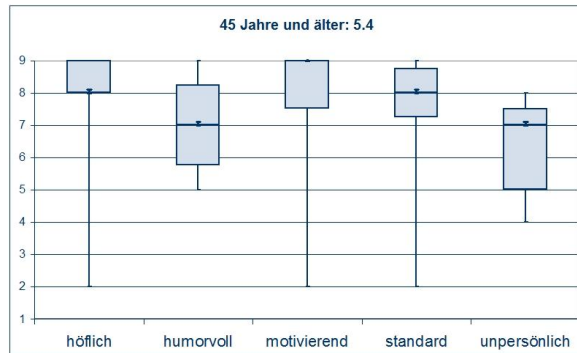


Abbildung 81: 45 Jahre und älter: Frage 5.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	8.2500	9.0000	8.7500	7.5000
50%	8.0000	7.0000	9.0000	8.0000	7.0000
25%	8.0000	5.7500	7.5000	7.2500	5.0000

Kruskal-Wallis-Test: p-value = 0.3188

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.1056

Frage 5.5: Wie oft informiert der Computer über die laufende Aktion?

Die Tonalität **motivierend** hat den stärksten Einfluss darauf dass der Benutzer sich über laufende Aktionen der Anwendung öfters informiert fühlt. Alle anderen Tonalitäten wirken sich nicht so stark darauf aus. Als letzte Tonalität wirkt sich **unpersönlich** auf den Informationsgrad aus.

Der Unterschied zwischen motivierend und unpersönlich ist **nicht signifikant** (p-value=0.06512). Der Wert ist jedoch unterhalb des 2. Signifikanzniveau und sorgt dafür dass die Aussage tendenziell trotzdem relevant ist.

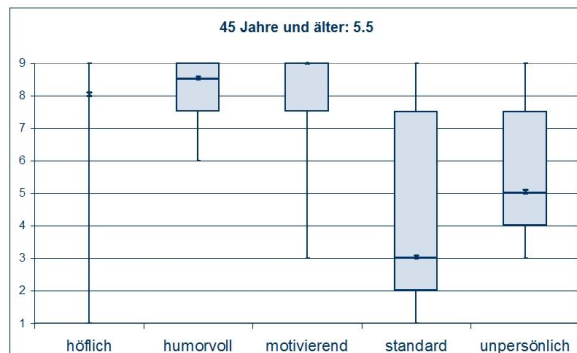


Abbildung 82: 45 Jahre und älter: Frage 5.5

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	7.5000	7.5000
50%	8.0000	8.5000	9.0000	3.0000	5.0000
25%	8.0000	7.5000	7.5000	2.0000	4.0000

Kruskal-Wallis-Test: p-value = 0.1184

Wilcoxon-Test: motivierend-standard: p-value = 0.06512

Frage 5.5.2: Wie oft führte die Ausführung einer Operation zu einer vorhersehbarem Ergebnis?

Die Tonalität **höflich** beeinflusst die Vorhersehbarkeit des Ergebnisses am stärksten auf positive Art und Weise.

Motivierend hat ebenfalls sehr gute Bewertung erhalten, aber statistisch gesehen (Freiheitsgrade und Verteilung werden ebenfalls berücksichtigt) ist die Tonalität **höflich** welche die grösste Distanz zu **unpersönlich** hat.

Der Unterschied von **höflich** und **unpersönlich** ist **nicht signifikant**, (p-value=0.06101). Der Wert ist jedoch **unterhalb des 2. Signifikanzniveau von 0.1** und demzufolge ist die Aussage tendenziell zu beachten.

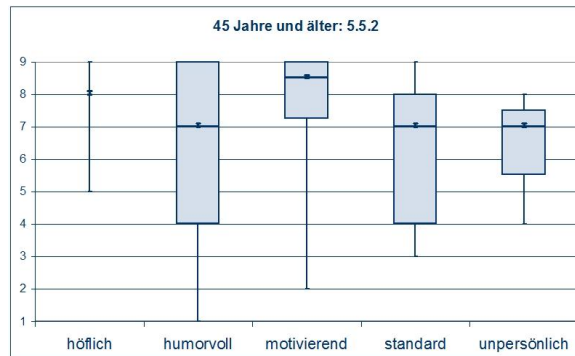


Abbildung 83: 45 Jahre und älter: Frage 5.5.2

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.0000	7.5000
50%	8.0000	7.0000	8.5000	7.0000	7.0000
25%	8.0000	4.0000	7.2500	4.0000	5.5000

Kruskal-Wallis-Test: p-value = 0.3414

Wilcoxon-Test:

höflich-unpersönlich: p-value = 0.06101

motivierend-unpersönlich: p-value = 0.1248

Frage 5.5.4: Wie akzeptabel waren die Wartezeiten zwischen den Operationen?

Die Wartezeiten waren bei allen Tonalitäten grundsätzlich akzeptabel. Die Tonalität welche die Anwendung bezüglich Wartezeiten am positivsten zu beeinflussen vermag ist **motivierend**. Die Tonalität **unpersönlich** senkt die Akzeptanz gegenüber Wartezeiten.

Der Unterschied zwischen **motivierend** und **unpersönlich** ist **hoch signifikant** (p-value=0.001384).

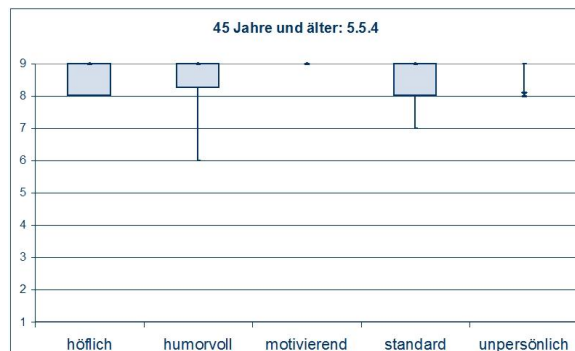


Abbildung 84: 45 Jahre und älter: Frage 5.5.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	8.0000
50%	9.0000	9.0000	9.0000	9.0000	8.0000
25%	8.0000	8.2500	9.0000	8.0000	8.0000

Kruskal-Wallis-Test: p-value = 0.0407

Wilcoxon-Test:

motivierend -höflich: p-value = 0.04271

motivierend-unpersönlich: p-value = 0.001384

Frage 5.6.1: Wie oft erklärten die Fehlermeldungen das Problem?

Bei der Tonalität **motivierend** ist der positive Einfluss am stärksten, dass die Fehlermeldung das Problem öfters erklärt. Bei allen anderen Tonalitäten ist die Wertung nicht so stark. Die Tonalität **unpersönlich** bildet dabei das Schlusslicht. Diese hat den kleinsten Einfluss darauf, wie oft eine Fehlermeldung das Problem erklärt.

Der Unterschied zwischen motivierend und unpersönlich ist **signifikant** (p-value=0.03818)

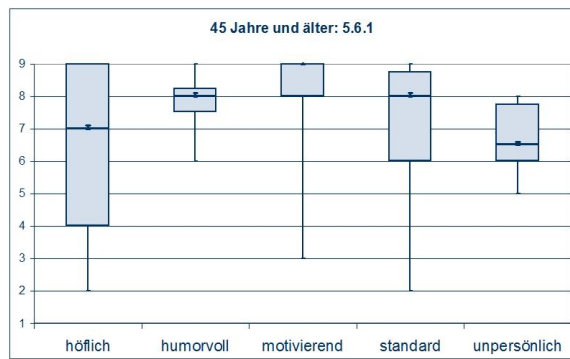


Abbildung 85: 45 Jahre und älter: Frage 5.6.1

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	8.2500	9.0000	8.7500	7.7500
50%	7.0000	8.0000	9.0000	8.0000	6.5000
25%	4.0000	7.5000	8.0000	6.0000	6.0000

Kruskal-Wallis-Test: p-value = 0.3015

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.03818

10.5.4 KAPITEL 6: LERNFORTSCHRITT (UNTERSTÜTZUNG BEI PROBLEMLÖSUNG)

Datenmenge: n=110

Höflich 22 Personen

Humorvoll 22 Personen

Motivierend 22 Personen

Standard 22 Personen

Unpersönlich 22 Personen

Beim Thema Lernfortschritt geht es um die Frage: wie gut wird der Benutzer durch die Anwendung unterstützt. Dabei fühlten sich Benutzer welche die Tonalität **humorvoll** hatten, am besten unterstützt. Am schlechtesten unterstützt fühlten sich die Benutzer mit der Tonalität **höflich**.

Der Unterschied zwischen höflich und humorvoll ist **nicht signifikant** (p-value=0.09137). Aber dennoch *unterhalb des 2. Signifikanzniveau von 0.1*. Demzufolge sollte der Aussage tendenziell Beachtung geschenkt werden.

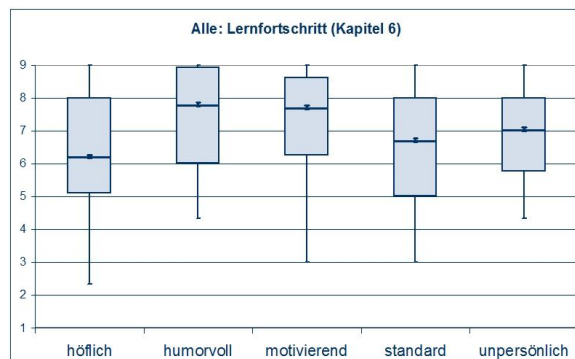


Abbildung 86: Alle: Lernfortschritt (Kapitel 6)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.9167	8.6250	8.0000	8.0000
50%	6.1667	7.7500	7.6667	6.6667	7.0000
25%	5.0833	6.0000	6.2500	5.0000	5.7500

Kruskal-Wallis-Test: p-value = 0.2384

Wilcoxon-Test: humorvoll -höflich: p-value = 0.09137

Frage 6.4: Wie einfach konnte die Aufgabe gelöst werden?

Die Tonalität **motivierend** sorgte am stärksten dafür, dass die Aufgabe einfach gelöst werden konnte. Auf der anderen Seite hat die Tonalität **höflich** am wenigsten dazu beigetragen.

Der Unterschied zwischen motivierend und höflich ist **hoch signifikant** (p-value=0.004397)

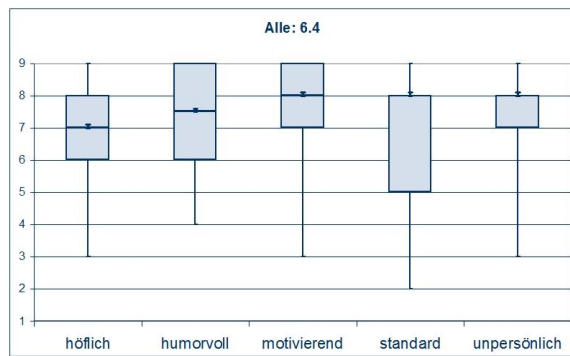


Abbildung 87: Alle: Frage 6.4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.0000	8.0000
50%	7.0000	7.5000	8.0000	8.0000	8.0000
25%	6.0000	6.0000	7.0000	5.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.07523

Wilcoxon-Test: motivierend -höflich: p-value = 0.004397

Auswertung über Gruppe Frauen

- Datenmenge n=30
- höflich: 5 Personen
- humorvoll: 6 Personen
- motivierend: 8 Personen
- standard: 4 Personen
- unpersönlich: 7 Personen

Grundsätzlich wird beim Thema Lernfortschritte die Unterstützung bei der Problemlösung behandelt.

Grundsätzlich unterstützt die Tonalität **höflich** die Problemlösung am besten. Hingegen bietet die Tonalität **standard** die schlechteste Unterstützung zur Problemlösung.

Der Unterschied zwischen den Tonalitäten höflich und standard ist **nicht signifikant** (p-value=0.1905).

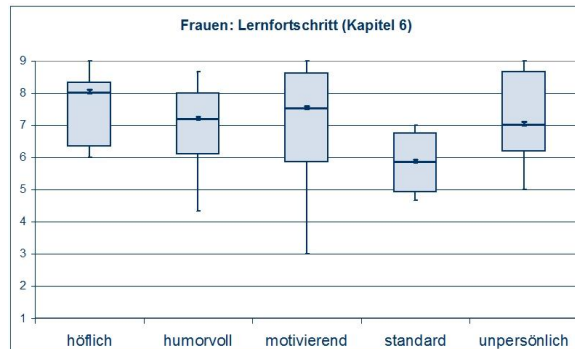


Abbildung 88: Frauen: Lernfortschritt (Kapitel 6)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3333	8.0000	8.6250	6.7500	8.6667
50%	8.0000	7.1667	7.5000	5.8333	7.0000
25%	6.3333	6.0833	5.8333	4.9167	6.1667

Kruskal-Wallis-Test: p-value = 0.5924

Wilcoxon-Test: höflich-standard: p-value = 0.1905



Auswertung über Gruppe Männer

- Datenmenge n=80
- höflich: 17 Personen
- humorvoll: 16 Personen
- motivierend: 14 Personen
- standard: 18 Personen
- unpersönlich: 15 Personen

Bei den Männern ist die Unterstützung bei Problemlösung bzw. Lernfortschritt am stärksten ausgeprägt bei der Tonalität **humorvoll**. Männer mit der Tonalität fanden die Unterstützung der Anwendung besser als z.B. Männer mit der Tonalität **höflich**. Die Tonalität höflich hat in dieser Beziehung am schlechtesten abgeschlossen. Der Unterschied zwischen humorvoll und höflich ist **signifikant** (p-value=0.02826).

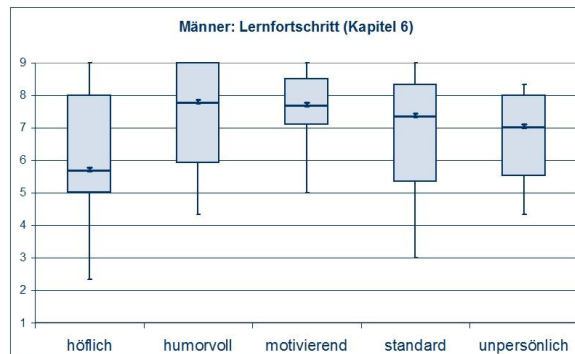


Abbildung 89: Männer: Lernfortschritt (Kapitel 6)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	8.5000	8.3333	8.0000
50%	5.6667	7.7500	7.6667	7.3333	7.0000
25%	5.0000	5.9167	7.0833	5.3333	5.5000

Kruskal-Wallis-Test: p-value = 0.1103

Wilcoxon-Test: humorvoll -höflich: p-value = 0.02826



Frage 6.4: Wie einfach konnte die Aufgabe gelöst werden?

Wie einfach die Aufgabe gelöst werden konnte, haben die Männer mit der Tonalität **motivierend** am besten bewertet. Am schlechteste Bewertung kam von den Männern mit der Tonalität **höflich**.

Der Unterschied zwischen motivierend und höflich ist **hoch signifikant** (p-value=0.001427).

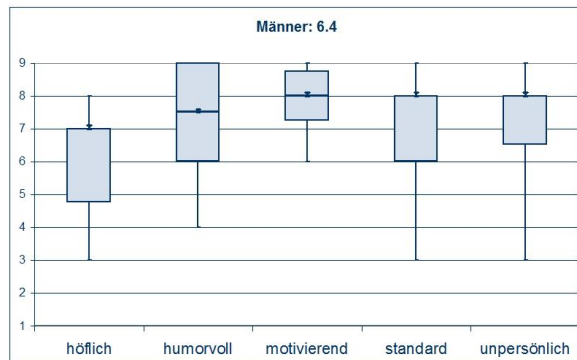


Abbildung 90: Männer: Frage 6.4

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	9.0000	8.7500	8.0000	8.0000
50%	7.0000	7.5000	8.0000	8.0000	8.0000
25%	4.7500	6.0000	7.2500	6.0000	6.5000

Kruskal-Wallis-Test: p-value = 0.0447

Wilcoxon-Test:

motivierend -höflich: p-value = 0.001427

humorvoll -höflich: p-value = 0.04575



Auswertung über Personen 35 jährig und jünger

- Datenmenge n=42
- höflich 7 Personen
- humorvoll 13 Personen
- motivierend 7 Personen
- standard 8 Personen
- unpersönlich 7 Personen

Bei den Jüngeren Probanden wird der Lernfortschritt bzw. Unterstützung bei der Problemlösung durch die Tonalität **humorvoll** am stärksten gefördert. Am schlechtesten schneidet hier die Tonalität **standard** ab. Der Unterschied zwischen humorvoll und standard ist **signifikant** (p-value=0.02649).

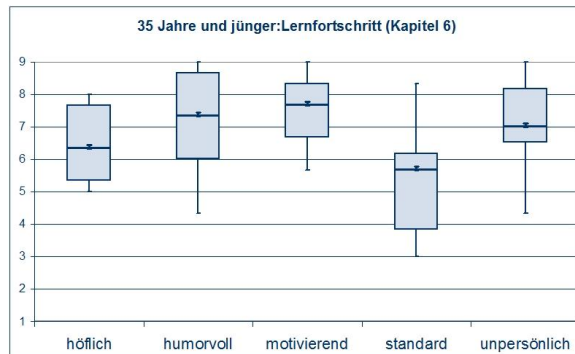


Abbildung 91: 35 Jahre und jünger: Lernfortschritt (Kapitel 6)

	höflich	humorv.	Motivier.	Standard	unpers.
75%	7.6667	8.6667	8.3333	6.1667	8.1667
50%	6.3333	7.3333	7.6667	5.6667	7.0000
25%	5.3333	6.0000	6.6667	3.8333	6.5000

Kruskal-Wallis-Test: p-value = 0.1056

Wilcoxon-Test:

humorvoll-standard: p-value = 0.02649

motivierend-standard: p-value = 0.0421

Frage 6.4.3: Herrscht Klarheit darüber, welche Schritte bereits erledigt sind?

Die Tonalität humorvoll sorgt bei den Jüngeren dafür, dass es Klarheit über bereits erfolgten Schritte herrscht.

Der Rangsummentest ergibt, dass **humorvoll** die beste Bewertung darstellt. Hingegen bildet die Tonalität **standard** hier das Schlusslicht.

Der Unterschied zwischen humorvoll und standard ist **signifikant** (p-value=0.02149).

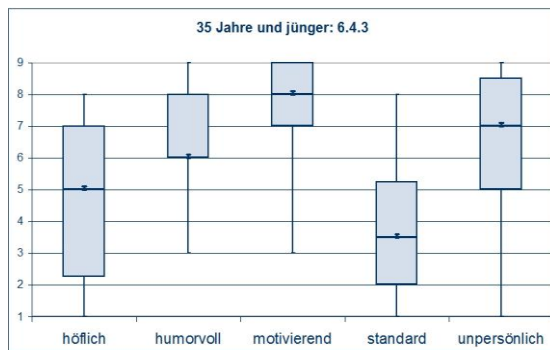


Abbildung 92: 35 Jahre und jünger: Frage 6.4.3

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	8.0000	9.0000	5.2500	8.5000
50%	5.0000	6.0000	8.0000	3.5000	7.0000
25%	2.2500	6.0000	7.0000	2.0000	5.0000

Kruskal-Wallis-Test: p-value = 0.07546

Wilcoxon-Test: Humorvoll-standard: p-value = 0.02149

Bei den Jüngeren ist prinzipiell die Tonalität humorvoll diejenige welche beim Thema Lernfortschritt die besseren Karten hat. Hingegen kommen Fehlermeldungen mit der Tonalität standard bezüglich Lernfortschritt bzw. Unterstützung bei Problemlösung nicht gut an.

Auswertung über Personen 45 jährig und älter

- Datenmenge n=38
- höflich 9 Personen
- humorvoll 4 Personen
- motivierend 8 Personen
- standard 10 Personen
- unpersönlich 7 Personen

Bei den Älteren ist bezüglich Lernfortschritt die Tonalität **motivierend** diejenige welche den besten Einfluss hat. Diese Tonalität vermag den Eindruck: die Anwendung liefert genügend Unterstützung zur Problemlösung zu stärken. Hingegen hinterlässt die Tonalität **unpersönlich** die schlechteste Bewertung.

Der Unterschied zwischen motivierend und unpersönlich ist **nicht signifikant** (p-value=0.1773).

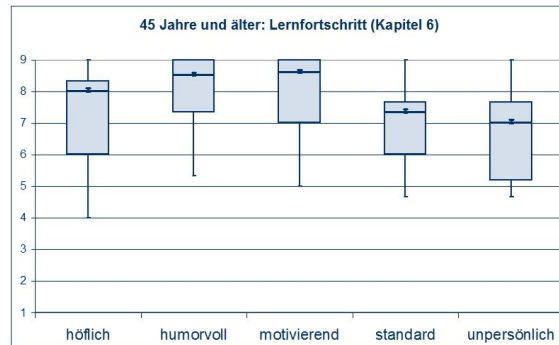


Abbildung 93: 45 Jahre und älter: Lernfortschritt (Kapitel 6)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3333	9.0000	9.0000	7.6667	7.6667
50%	8.0000	8.5000	8.5833	7.3333	7.0000
25%	6.0000	7.3333	7.0000	6.0000	5.1667

Kruskal-Wallis-Test: p-value = 0.3735

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.1773

Bei den Älteren Probanden kann keine Aussage zum Thema Lernfortschritt gemacht werden, da die Unterschiede zwischen den Tonalitäten nicht signifikant sind.

10.5.5 KAPITEL 7: SYSTEMEIGENSCHAFTEN (KOMPLEXITÄT DES PROBLEMLÖSUNGSPROZESSES)

Datenmenge: n=110

Höflich 22 Personen

Humorvoll 22 Personen

Motivierend 22 Personen

Standard 22 Personen

Unpersönlich 22 Personen

Beim Thema Systemeigenschaften geht es grundsätzlich um die Frage: wie einfach ist die Problemlösung. Dabei waren die Benutzer mit der Tonalität *motivierend* der Meinung die Anwendung sei am wenigsten komplex, bzw. am einfachsten zu bedienen. Die höchste Komplexität empfanden hier die Benutzer mit der Tonalität *höflich*.

Der Unterschied zwischen höflich und humorvoll ist jedoch *nicht signifikant* (p-value=2634).

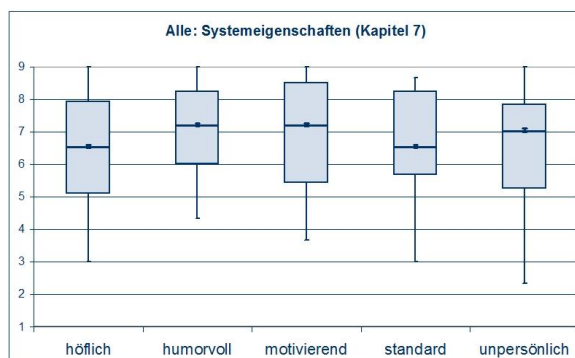


Abbildung 94: Alle: Systemeigenschaften (Kapitel 7)

	höflich	humorv.	motivier.	standard	unpers.
75%	7.9167	8.2500	8.5000	8.2500	7.8333
50%	6.5000	7.1667	7.1667	6.5000	7.0000
25%	5.0833	6.0000	5.4167	5.6667	5.2500

Kruskal-Wallis-Test: p-value = 0.7261

Wilcoxon-Test: humorvoll -höflich: p-value = 0.2634

Das Thema Systemeigenschaften beschäftigt sich grundsätzlich mit der Frage: wie einfach oder komplex ist die Problemlösung mit dieser Anwendung? Aufgrund fehlender Signifikante konnte hier keine Aussage gemacht werden.

Auswertung über Gruppe Frauen

- Datenmenge n=30
- höflich: 5 Frauen
- humorvoll: 6 Frauen
- motivierend: 8 Frauen
- standard: 4 Frauen
- unpersönlich: 7 Frauen

Auch bei den Frauen konnte keine Eindeutige Aussage bezüglich Systemeigenschaften gemacht werden. Grundsätzlich hat bei den Frauen die Tonalität *unpersönlich* die höchste Bewertung erhalten, und *humorvoll* die schlechteste.

Der Unterschied zwischen unpersönlich und humorvoll ist *nicht signifikant* (p-value=0.1967).

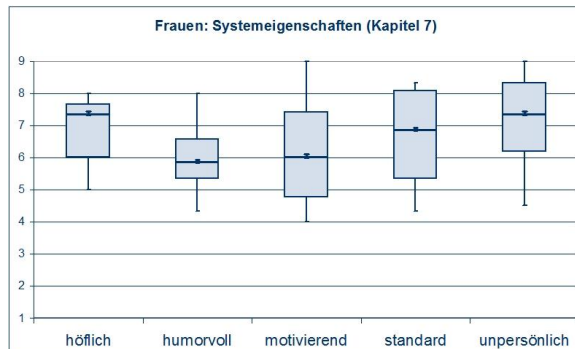


Abbildung 95: Frauen: Systemeigenschaften (Kapitel 7)

	höflich	humorv.	motivier.	standard	unpers.
75%	7.6667	6.5833	7.4167	8.0833	8.3333
50%	7.3333	5.8333	6.0000	6.8333	7.3333
25%	6.0000	5.3333	4.7500	5.3333	6.1667

Kruskal-Wallis-Test: p-value = 0.6339

Wilcoxon-Test: unpersönlich - humorvoll: p-value = 0.1967



Auswertung über Gruppe Männer

- Datenmenge n=80
- höflich: 17 Männer
- humorvoll: 16 Männer
- motivierend: 14 Männer
- standard: 18 Männer
- unpersönlich: 15 Männer

Bei den Männern ist die Tonalität **humorvoll** diejenige welche das Thema Systemeigenschaften im besten Licht erscheinen lässt. Die schlechteste Bewertung hat **unpersönlich** erhalten. Interessant ist, dass es genau umgekehrt wie bei den Frauen ist.

Der Unterschied bei den Männern zwischen humorvoll und unpersönlich ist **nicht signifikant** (p-value=0.07048). Der Wert liegt jedoch *unterhalb des 2. Signifikanzniveau von 0.1*. Tendenziell stimmt die Aussage bei den Männern.

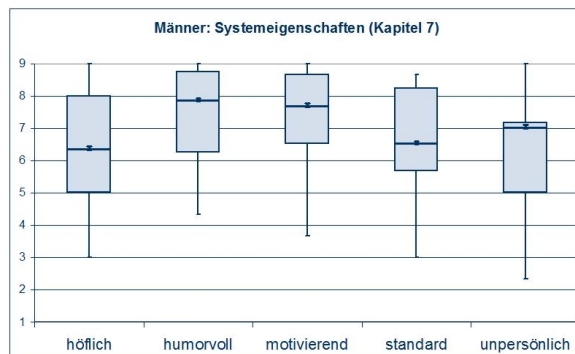


Abbildung 96: Männer: Systemeigenschaften (Kapitel 7)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.7500	8.6667	8.2500	7.1667
50%	6.3333	7.8333	7.6667	6.5000	7.0000
25%	5.0000	6.2500	6.5000	5.6667	5.0000

Kruskal-Wallis-Test: p-value = 0.1565

Wilcoxon-Test: humorvoll-unpersönlich: p-value = 0.07048

Leider haben auch hier bei den Männern keine weiteren Fragen signifikante Resultate geliefert. Erwähnenswert ist die komplett unterschiedliche Auslegung zwischen Mann und Frau.



Auswertung über Personen 35 jährig und jünger

- Datenmenge n=42
- höflich 7 Personen
- humorvoll 13 Personen
- motivierend 7 Personen
- standard 8 Personen
- unpersönlich 7 Personen

Bei den jüngeren Probanden (35 Jahre und jünger) steigert die Tonalität **motivierend** die Einfachheit des Problemlösungsprozesses. Im Gegensatz dazu erscheint der Problemlösungsprozess mit Tonalität **standard** komplexer zu sein. Der Unterschied zwischen motivierend und standard ist **nicht signifikant** ($p\text{-value}=0.1809$).

Bei den einzelnen Fragen gibt es keine weiteren Erkenntnisse welche signifikantere Unterschiede zwischen den Tonalitäten hätte zu Tage führen können.

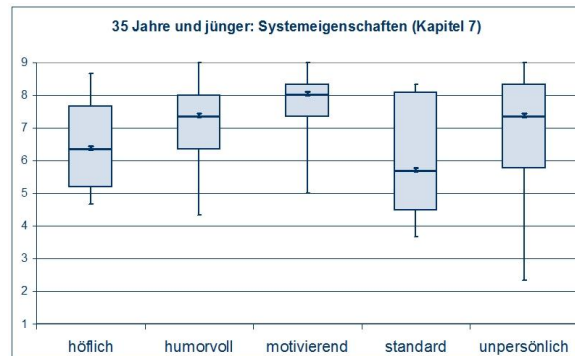


Abbildung 97: 35 Jahre und jünger: Systemeigenschaften (Kapitel 7)

	höflich	humorv.	motivier.	standard	unpers.
75%	7.6667	8.0000	8.3333	8.0833	8.3333
50%	6.3333	7.3333	8.0000	5.6667	7.3333
25%	5.1667	6.3333	7.3333	4.4583	5.7500

Kruskal-Wallis-Test: $p\text{-value} = 0.5212$

Wilcoxon-Test: motivierend-standard: $p\text{-value} = 0.1809$

Auswertung über Personen 45 jährig und älter

- Datenmenge n=38
- höflich 9 Personen
- humorvoll 4 Personen
- motivierend 8 Personen
- standard 10 Personen
- unpersönlich 7 Personen

Bei den älteren Probanden scheint die Tonalität *humorvoll* den Problemlösungsprozess leichter zu gestalten. Hingegen lässt die Tonalität *unpersönlich* diesen Prozess komplexer erscheinen.

Der Unterschied zwischen humorvoll und unpersönlich ist *nicht signifikant* (p-value=0.1042).

Evtl. können die einzelnen Fragen etwas mehr Licht in diese Angelegenheit bringen.

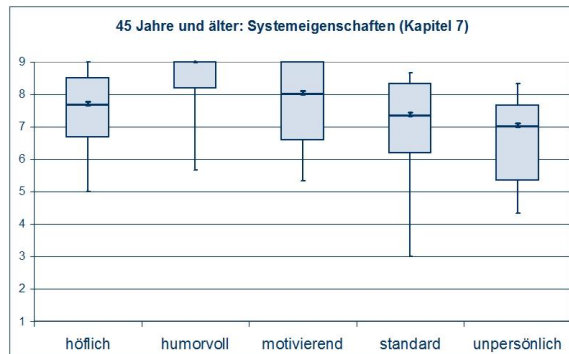


Abbildung 98: 45 Jahre und älter: Systemeigenschaften (Kapitel 7)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.5000	9.0000	9.0000	8.3333	7.6667
50%	7.6667	9.0000	8.0000	7.3333	7.0000
25%	6.6667	8.1667	6.5833	6.1667	5.3333

Kruskal-Wallis-Test: p-value = 0.2988

Wilcoxon-Test: humorvoll-unpersönlich: p-value = 0.1042

Frage 7.4: Wie leicht war das Beheben von Fehler?

Bei den Älteren scheint die Tonalität **motivierend** die Fehler der Anwendung leichter behebbar zu machen. Hingegen lässt die Tonalität **unpersönlich** den Eindruck erwecken, die Fehler seien schwieriger zu beheben.

Der Unterschied zwischen motivierend und unpersönlich ist **signifikant** (p-value=0.04414).

D.h. also die Tonalität motivierend senkt scheinbar die Komplexität der Fehlerbehebung, und unpersönlich lässt diese Komplexität ansteigen.

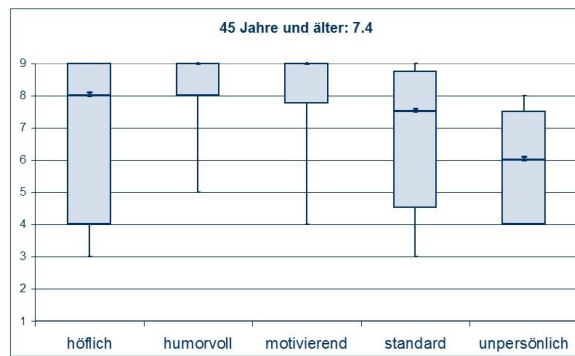


Abbildung 99: 45 Jahre und älter: Frage 7.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	8.7500	7.5000
50%	8.0000	9.0000	9.0000	7.5000	6.0000
25%	4.0000	8.0000	7.7500	4.5000	4.0000

Kruskal-Wallis-Test: p-value = 0.1713

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.04414

Frage 7.4.1: Wie einfach war das Korrigieren von Tippfehler?

Auch in diesem Bereich ist es erstaunlich, dass die Tonalität überhaupt einen Einfluss in diesen sehr mechanischen Prozess des Tippfehlers korrigieren. Grundsätzlich ist es in jeder Tonalität der selbe Aufwand Zeichen aus einem Eingabefeld zu löschen und wieder schreiben. Trotzdem scheint es mit der Tonalität **motivierend** deutlich einfacher zu gelingen als mit der Tonalität **unpersönlich**.

Der Unterschied zwischen der Tonalität motivierend und unpersönlich ist **signifikant** (p-value=0.04867).

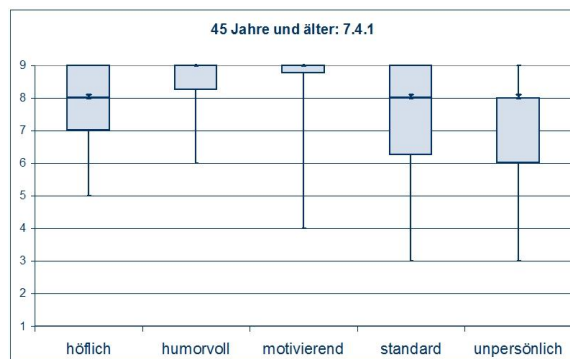


Abbildung 100: 45 Jahre und älter: Frage 7.4.1

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	8.0000
50%	8.0000	9.0000	9.0000	8.0000	8.0000
25%	7.0000	8.2500	8.7500	6.2500	6.0000

Kruskal-Wallis-Test: p-value = 0.2794

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.04867

10.5.6 KAPITEL 9: ONLINE-ANLEITUNG (EFFEKTIVITÄT DER HILFESTELLUNG)

Auswertung über alle Probanden

Datenmenge: n=110

Höflich 22 Personen

Humorvoll 22 Personen

Motivierend 22 Personen

Standard 22 Personen

Unpersönlich 22 Personen

Den Benutzern der Tonalität **motivierend** ist bezüglich Hilfestellung der Anwendung und Leichtigkeit der Bedienung, die Anwendung positiver aufgefallen, als allen anderen Benutzern mit anderen Tonalitäten. Am schlechtesten ist es den Benutzern der Tonalität **höflich** aufgefallen.

Der Unterschied zwischen höflich und motivierend ist **signifikant** (p-value=0.04864).

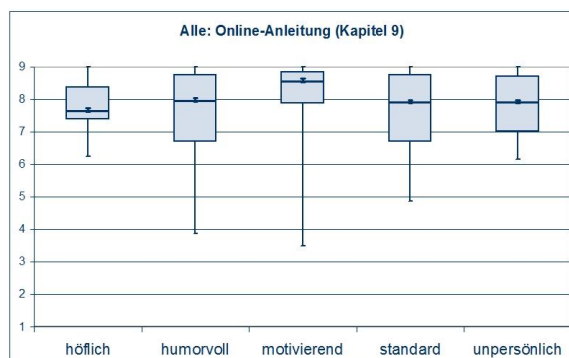


Abbildung 101: Alle: Online-Anleitung (Kapitel 9)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3750	8.7545	8.8438	8.7545	8.7143
50%	7.6250	7.9375	8.5357	7.8750	7.8750
25%	7.3750	6.6875	7.8616	6.6920	7.0000

Kruskal-Wallis-Test: p-value = 0.4624

Wilcoxon-Test: motivierend -höflich: p-value = 0.04864

Frage 9.3.1: Wie komplett und informativ waren Informationen des Systems für spezielle Aspekte?

Mit **motivierender** Tonalität scheint für die Probanden die Anwendung bezüglich Vollständigkeit und Information die Anwendung besser zu sein, als für Probanden mit der Tonalität **höflich**. Der Unterschied zwischen motivierend und höflich ist **signifikant** (p-value=0.04881).

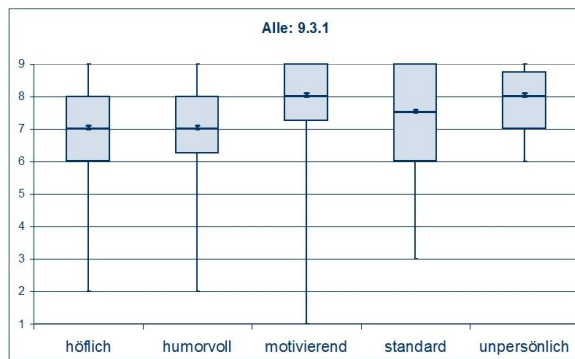


Abbildung 102: Alle: Frage 9.3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.0000	9.0000	9.0000	8.7500
50%	7.0000	7.0000	8.0000	7.5000	8.0000
25%	6.0000	6.2500	7.2500	6.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.2242

Wilcoxon-Test: motivierend -höflich: p-value = 0.04881

Frage 9.4: Wie leicht war die Lösung der Aufgabe?

Motivierende Tonalität ist hauptsächlich verantwortlich dass die Lösung der Aufgabe leicht erscheint. Hingegen scheint die Tonalität **standard** die Lösung der Aufgabe schwieriger erscheinen.

Der Unterschied zwischen motivierend und standard ist **hoch signifikant** (p-value=0.004546).

Auch der Unterschied zwischen motivierend und höflich ist **hoch signifikant** (p-value=0.006341).

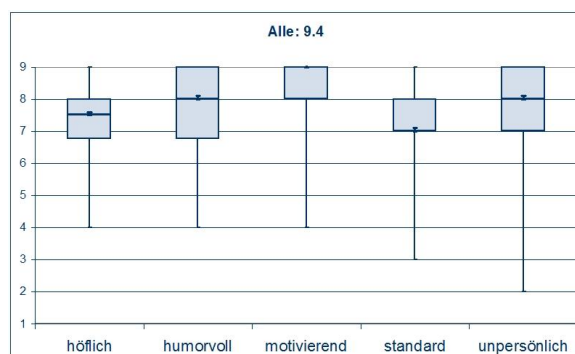


Abbildung 103: Alle: Frage 9.4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.0000	9.0000
50%	7.5000	8.0000	9.0000	7.0000	8.0000
25%	6.7500	6.7500	8.0000	7.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.02139

Wilcoxon-Test:

motivierend- höflich: p-value = 0.006341

motivierend-standard: p-value = 0.004546

Auswertung über Gruppe Frauen

- Datenmenge n=30
- höflich: 5 Frauen
- humorvoll: 6 Frauen
- motivierend: 8 Frauen
- standard: 4 Frauen
- unpersönlich: 7 Frauen

Bei den Frauen steht bezüglich Hilfestellung die Tonalität **unpersönlich** hoch im Kurs. Hingegen lässt die Tonalität **humorvoll** die Hilfsbereitschaft der Anwendung schlechter aussehen.

Der Unterschied zwischen unpersönlich und humorvoll ist **nicht signifikant** (p-value=0.1156).

Evtl. ist die Analyse der einzelnen Fragen aufschlussreicher.

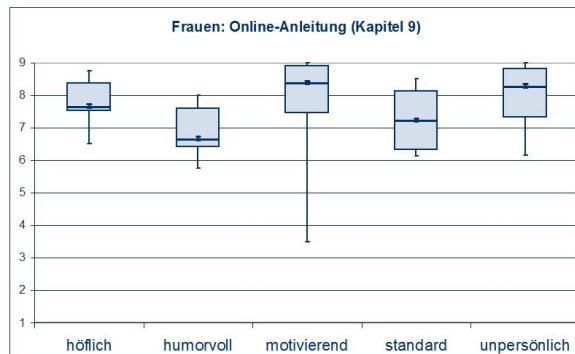


Abbildung 104: Frauen: Online-Anleitung (Kapitel 9)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3750	7.5938	8.9063	8.1250	8.8125
50%	7.6250	6.6250	8.3482	7.1875	8.2500
25%	7.5000	6.4063	7.4375	6.3125	7.3036

Kruskal-Wallis-Test: p-value = 0.3518

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.1156

Frage 9.1: Wie hilfreich war die Online-Anleitung?

Auf die Frage hin, wie hilfreich die Online Anleitung war, haben die Probandinnen mit der Tonalität **unpersönlich** die beste Bewertung abgegeben. Die schlechteste Bewertung kam von den Probandinnen mit der Tonalität **humorvoll**. Der Unterschied zwischen unpersönlich und humorvoll ist **signifikant** (p-value=0.04282).

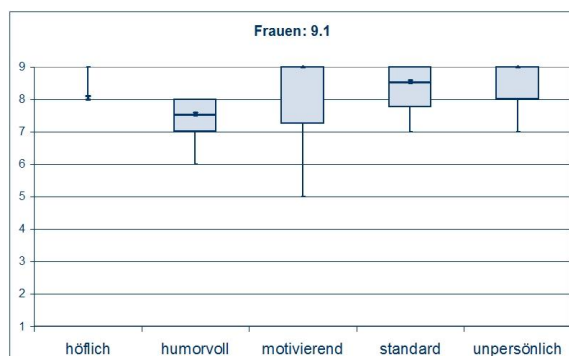


Abbildung 105: Frauen: Frage 9.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.0000	9.0000	9.0000	9.0000
50%	8.0000	7.5000	9.0000	8.5000	9.0000
25%	8.0000	7.0000	7.2500	7.7500	8.0000

Kruskal-Wallis-Test: p-value = 0.2576

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.04282



Auswertung über Gruppe Männer

- Datenmenge n=80
- höflich: 17 Männer
- humorvoll: 16 Männer
- motivierend: 14 Männer
- standard: 18 Männer
- unpersönlich: 15 Männer

Bei den Männern ist ebenfalls die Tonalität **motivierend** diejenige welche bezüglich Hilfestellung der Anwendung den besten Einfluss hatte. Hingegen lässt die Tonalität **höflich** die Anwendung bezüglich Hilfestellung eher schlechter aussehen.

Der Unterschied zwischen motivierend und höflich ist **nicht signifikant** (p-value=0.0609). Der Wert ist jedoch *unterhalb des 2. Signifikanzniveau von 0.1*. Demzufolge sollte die Aussage tendenziell beachtet werden.

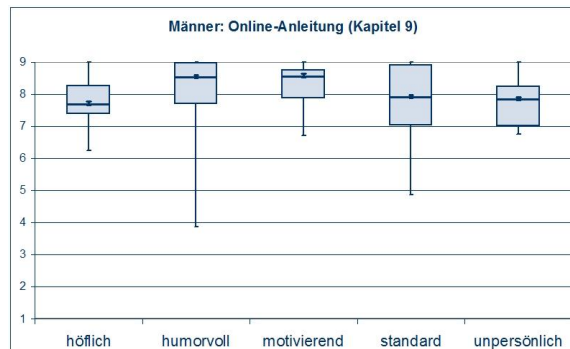


Abbildung 106: Männer: Online-Anleitung (Kapitel 9)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.2571	8.9688	8.7500	8.9063	8.2500
50%	7.6696	8.5000	8.5357	7.8750	7.8125
25%	7.3646	7.6771	7.8616	7.0223	7.0000

Kruskal-Wallis-Test: p-value = 0.3073

Wilcoxon-Test: motivierend -höflich: p-value = 0.0609



Frage 9.3.1: Wie komplett und informativ waren Informationen des Systems für spezielle Aspekte?

Bei Männern ist die Tonalität **motivierend** massgeblich dafür verantwortlich dem System eine Aura der Vollständigkeit und Informationsgehalt zu vermitteln. Die Tonalität **höflich** wirkt sich hingegen gegenteilig aus.

Der Unterschied zwischen motivierend und höflich ist **signifikant** (p-value=0.011816)

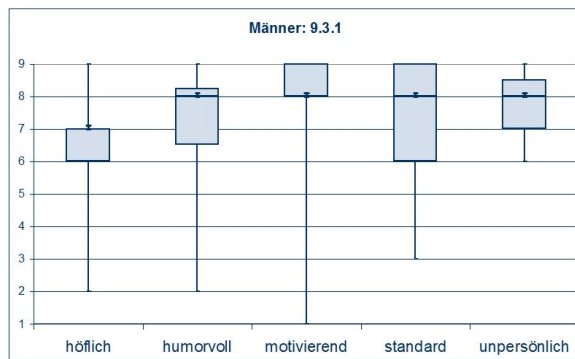


Abbildung 107: Männer: Frage 9.3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	8.2500	9.0000	9.0000	8.5000
50%	7.0000	8.0000	8.0000	8.0000	8.0000
25%	6.0000	6.5000	8.0000	6.0000	7.0000

Kruskal-Wallis-Test:

Wilcoxon-Test: motivierend -höflich: p-value = 0.01816

Frage 9.4: Wie leicht war die Lösung der Aufgabe?

Die Tonalität **motivierend** lässt die Lösung der Aufgabe leichter erscheinen. Hingegen erschwert die Tonalität **höflich** die Lösung der Aufgabe deutlich.

Der Unterschied zwischen motivierend und höflich ist **hoch signifikant** (p-value=0.008443).

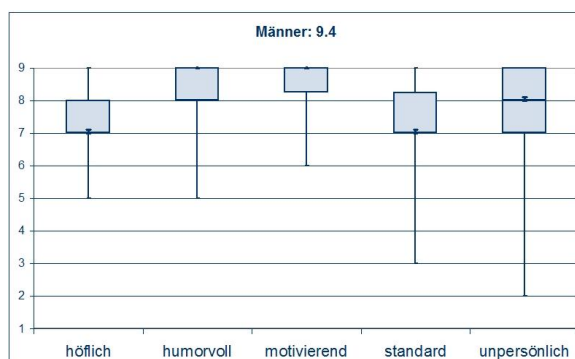


Abbildung 108: Männer: Frage 9.4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	9.0000	9.0000	8.2500	9.0000
50%	7.0000	9.0000	9.0000	7.0000	8.0000
25%	7.0000	8.0000	8.2500	7.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.03959

Wilcoxon-Test:

motivierend -höflich: p-value = 0.008443

motivierend-standard: p-value = 0.01285



Auswertung über Personen 35 jährig und jünger

- Datenmenge n=42
- höflich 7 Personen
- humorvoll 13 Personen
- motivierend 7 Personen
- standard 8 Personen
- unpersönlich 7 Personen

Bei den Jüngeren ist wird die Hilfestellung der Anwendung durch die Tonalität **unpersönlich** verstärkt. Die Hilfsbereitschaft lässt unter der Tonalität **standard** etwas nach.

Der Unterschied zwischen unpersönlich und standard ist **nicht signifikant** (p-value=0.1307).

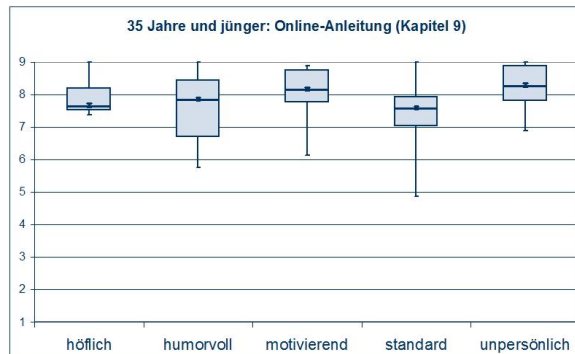


Abbildung 109: 35 Jahre und jünger: Online-Anleitu (Kapitel 9)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.1875	8.4375	8.7500	7.9375	8.8750
50%	7.6250	7.8125	8.1250	7.5500	8.2500
25%	7.5000	6.6875	7.7500	7.0223	7.8036

Kruskal-Wallis-Test: p-value = 0.4002

Wilcoxon-Test: unpersönlich -standard: p-value = 0.1307

Frage 9.3.1: Wie komplett und informativ waren Informationen des Systems für spezielle Aspekte?

Jüngere Probanden erhalten vor allem bei der Tonalität **unpersönlich** den Eindruck der Vollständigkeit. Im Gegensatz dazu erscheint unter der Tonalität **humorvoll** eher das Gegenteil davon.

Der Unterschied von unpersönlich und humorvoll ist **signifikant** (p-value=0.02705).

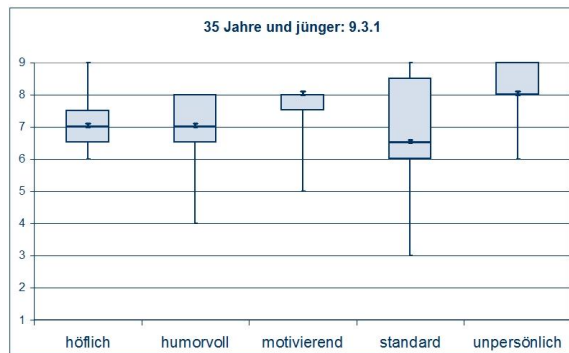


Abbildung 110: 35 Jahre und jünger: Frage 9.3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	7.5000	8.0000	8.0000	8.5000	9.0000
50%	7.0000	7.0000	8.0000	6.5000	8.0000
25%	6.5000	6.5000	7.5000	6.0000	8.0000

Kruskal-Wallis-Test: p-value = 0.2249

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.02705

Auswertung über Personen 45 jährig und älter

- Datenmenge n=38
- höflich 9 Personen
- humorvoll 4 Personen
- motivierend 8 Personen
- standard 10 Personen
- unpersönlich 7 Personen

Bei älteren Probanden sieht es ein wenig anders aus. Hier wird die Hilfestellung des Systems besser beurteilt durch die Tonalität **motivierend**. Im Gegensatz dazu steht die Tonalität **unpersönlich**, welches die Hilfsbereitschaft des Systems etwas schmälert.

Der Unterschied zwischen motivierend und unpersönlich ist **signifikant** (p-value=0.02597).

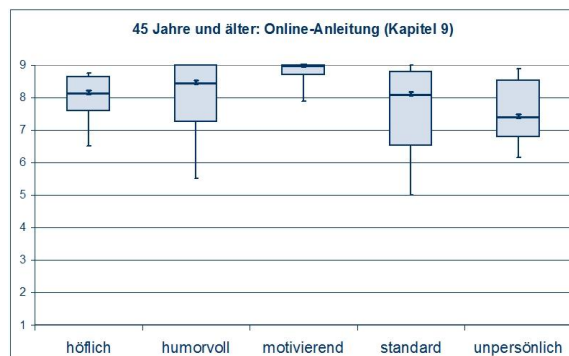


Abbildung 111: 45 Jahre und älter: Online-Anleitung (Kapitel 9)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.6375	9.0000	9.0000	8.7857	8.5313
50%	8.1161	8.4167	8.9375	8.0625	7.3750
25%	7.5625	7.2500	8.6875	6.5000	6.7813

Kruskal-Wallis-Test: p-value = 0.1362

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.02597

Frage 9.1.1: Wie leicht war der Zugang zur Online-Anleitung?

Bei den älteren Probanden wird der Zugang zu der Online-Anleitung durch die Tonalität **motivierend** vereinfacht. Hingegen wird derselbe Zugang durch die Tonalität **unpersönlich** massiv erschwert.

Der Unterschied zwischen motivierend und unpersönlich ist **signifikant** (p-value=0.0186).

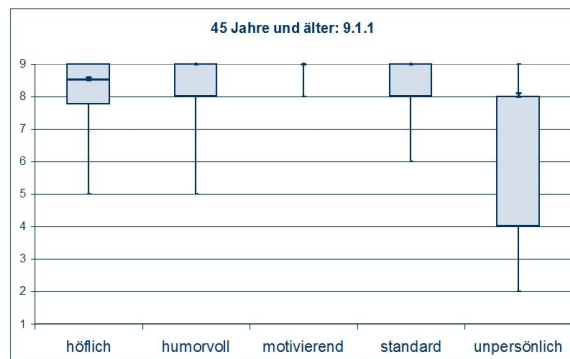


Abbildung 112: 45 Jahre und älter: Frage 9.1.1

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	9.0000	8.0000
50%	8.5000	9.0000	9.0000	9.0000	8.0000
25%	7.7500	8.0000	9.0000	8.0000	4.0000

Kruskal-Wallis-Test: p-value = 0.1268

Wilcoxon-Test: motivierend-unpersönlich: p-value = 0.0186

Frage 9.3.1: Wie komplett und informativ waren Informationen des Systems für spezielle Aspekte?

Die Tonalität **motivierend** vermittelt hier ein gutes Gefühl wenn es um Vollständigkeit und Informationsgehalt geht. Anders sieht es bei der Tonalität **höflich** aus. Hier scheint das Gefühl der Vollständigkeit nicht mehr so ausgeprägt zu sein.

Der Unterschied zwischen motivierend und höflich ist **signifikant** (p-value=0.04901).

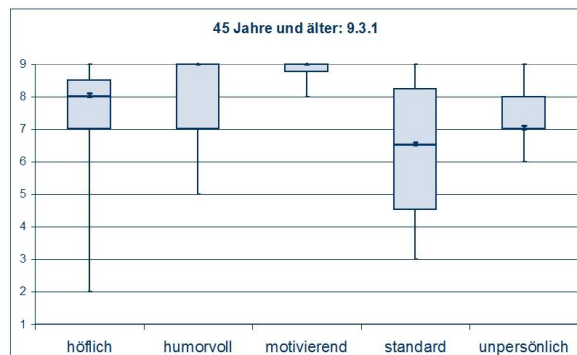


Abbildung 113: 45 Jahre und älter: Frage 9.3.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.5000	9.0000	9.0000	8.2500	8.0000
50%	8.0000	9.0000	9.0000	6.5000	7.0000
25%	7.0000	7.0000	8.7500	4.5000	7.0000

Kruskal-Wallis-Test: p-value = 0.1577

Wilcoxon-Test: motivierend -höflich: p-value = 0.04901

Frage 9.4: Wie leicht war die Lösung der Aufgabe?

Die Aufgabe scheint bei den Älteren durch die Tonalität **motivierend** vereinfacht zu werden. Anders sieht es mit der Tonalität **standard** aus. Hier scheint die Lösung der Aufgabe nicht mehr so einfach zu sein.

Der Unterschied zwischen den Tonalitäten motivierend und standard ist **signifikant** (p-value=0.02637).

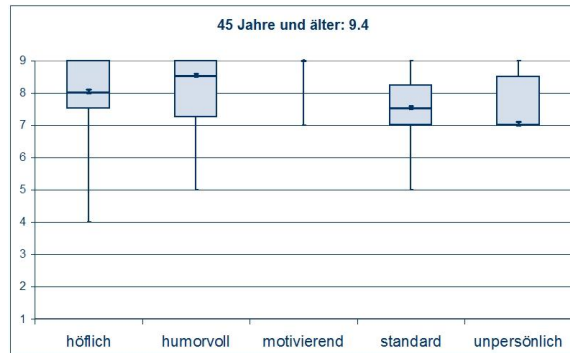


Abbildung 114: 45 Jahre und älter: Frage 9.4

	höflich	humorv.	motivier.	standard	unpers.
75%	9.0000	9.0000	9.0000	8.2500	8.5000
50%	8.0000	8.5000	9.0000	7.5000	7.0000
25%	7.5000	7.2500	9.0000	7.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.2008

Wilcoxon-Test: motivierend-standard: p-value = 0.02637

Frage 9.4.1: Wie klar waren die Anweisungen für die Lösung der Aufgabe?

Auch bei dieser Frage scheint die Tonalität **motivierend** bei den Älteren für mehr Klarheit zu verhelfen als durch andere Tonalitäten. Speziell steht die Tonalität **höflich** im Gegensatz zu motivierend. Durch die Tonalität höflich scheint die Anweisung für die Lösung der Aufgabe nicht mehr so klar zu sein.

Der Unterschied zwischen den Tonalitäten motivierend und höflich ist **signifikant** (p-value=0.03103).

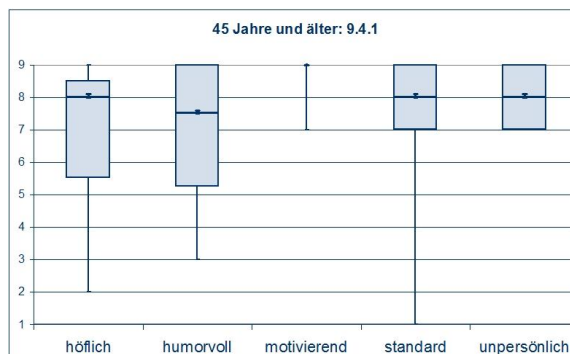


Abbildung 115: 45 Jahre und älter: Frage 9.4.1

	höflich	humorv.	motivier.	standard	unpers.
75%	8.5000	9.0000	9.0000	9.0000	9.0000
50%	8.0000	7.5000	9.0000	8.0000	8.0000
25%	5.5000	5.2500	9.0000	7.0000	7.0000

Kruskal-Wallis-Test: p-value = 0.2324

Wilcoxon-Test: motivierend -höflich: p-value = 0.03103

10.5.7 KAPITEL 10: MULTIMEDIA KOMponentEN (VISUELL-INFORMATIVE UNTERSTÜTZUNG)

Beim Thema Multimedia Komponenten geht es um die Frage wie unterstützend wirken Farben und Bilder bei dieser Anwendung?

Probanden mit der Tonalität **motivierend** fanden die visuellen Elemente hätten einen starken unterstützenden Wert. Hingegen haben Probanden mit der Tonalität **höflich** eine leicht schlechtere Bewertung abgegeben.

Der Unterschied zwischen motivierend und höflich ist **nicht signifikant** (p-value=0.5175).

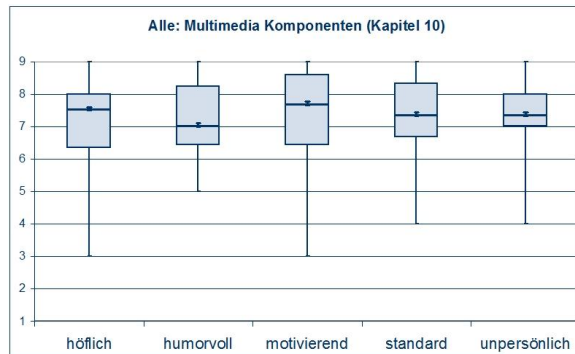


Abbildung 116: Alle: Multimedia Komponenten (Kapitel 10)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.2500	8.5833	8.3333	8.0000
50%	7.5000	7.0000	7.6667	7.3333	7.3333
25%	6.3333	6.4167	6.4167	6.6667	7.0000

Kruskal-Wallis-Test: p-value = 0.9693

Wilcoxon-Test: motivierend -höflich: p-value = 0.5175

Auswertung über Gruppe Frauen

- Datenmenge n=30
- höflich: 5 Frauen
- humorvoll: 6 Frauen
- motivierend: 8 Frauen
- standard: 4 Frauen
- unpersönlich: 7 Frauen

Bei den Frauen wird visuell-informative Unterstützung gestärkt durch die Tonalität **höflich**. Im Gegensatz dazu trägt die Tonalität **standard** am wenigsten zur Unterstützung bei.

Der Unterschied zwischen höflich und standard ist **nicht signifikant** (p-value=0.3873). Auch die Analyse von weiteren Fragen aus diesem Kapitel haben keine signifikanten Unterschiede gezeigt.

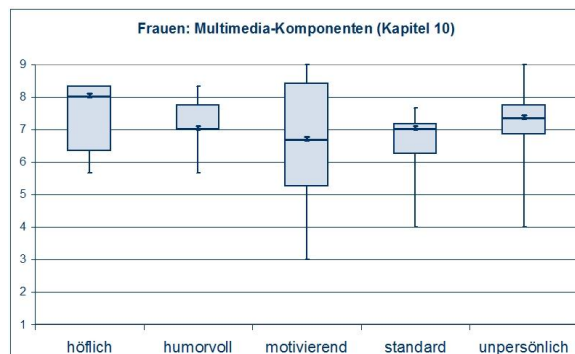


Abbildung 117: Frauen: Multimedia Komponenten (Kapitel 10)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3333	7.7500	8.4167	7.1667	7.7500
50%	8.0000	7.0000	6.6667	7.0000	7.3333
25%	6.3333	7.0000	5.2500	6.2500	6.8333

Kruskal-Wallis-Test: p-value = 0.9187

Wilcoxon-Test: höflich-standard: p-value = 0.3873



Auswertung über Gruppe Männer

- Datenmenge n=80
- höflich: 17 Männer
- humorvoll: 16 Männer
- motivierend: 14 Männer
- standard: 18 Männer
- unpersönlich: 15 Männer

Bei den Männern wirkt die Tonalität **motivierend** unterstützend, wenn es um das Visuelle geht. Im Gegenteil dazu wirkt bei den Männern die Tonalität **höflich** am wenigsten auf die visuelle Stimulanz.

Der Unterschied zwischen motivierend und höflich ist **nicht signifikant** (p-value=0.1855).

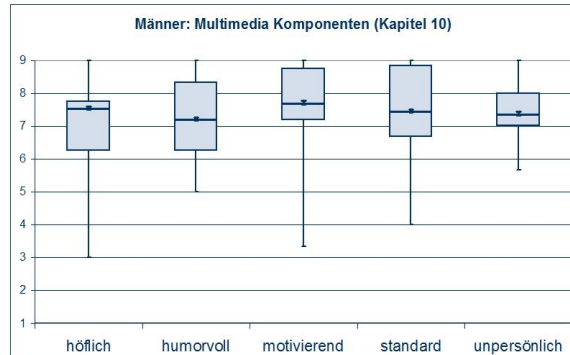


Abbildung 118: Männer: Multimedia Komponente (Kapitel 10)

	höflich	humorv.	motivier.	standard	unpers.
75%	7.7500	8.3333	8.7500	8.8333	8.0000
50%	7.5000	7.1667	7.6667	7.4167	7.3333
25%	6.2500	6.2500	7.1667	6.6667	7.0000

Kruskal-Wallis-Test: p-value = 0.7312

Wilcoxon-Test: motivierend -höflich: p-value = 0.1855

Frage 10.4: Wie natürlich war die Verwendung von Farben?

Bei der Frage nach der Natürlichkeit der verwendeten Farben ist **humorvoll** die Tonalität welche diese Empfindung am stärksten beeinflusst. Die Tonalität **unpersönlich** ist diejenige welche die Natürlichkeit der verwendeten Farben am schlechtesten bewerten lässt.

Der Unterschied zwischen humorvoll und unpersönlich ist **nicht signifikant** (p-value=0.09087). Der Wert ist jedoch **unterhalb des 2. Signifikanzniveau von 0.1**, und demzufolge als Tendenz zugelassen.

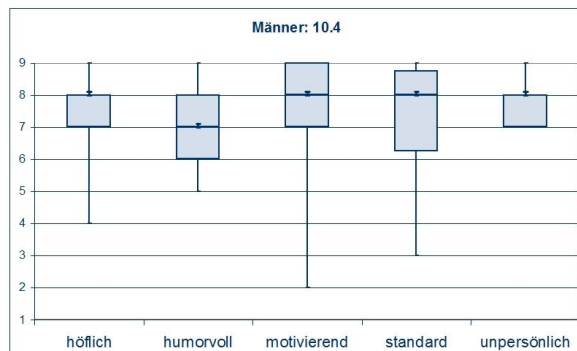


Abbildung 119: Männer: Frage 10.4

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	8.0000	9.0000	8.7500	8.0000
50%	8.0000	7.0000	8.0000	8.0000	8.0000
25%	7.0000	6.0000	7.0000	6.2500	7.0000

Kruskal-Wallis-Test: p-value = 0.4533

Wilcoxon-Test: unpersönlich -humorvoll: p-value = 0.09087



- Datenmenge n=42
- höflich 7 Personen
- humorvoll 13 Personen
- motivierend 7 Personen
- standard 8 Personen
- unpersönlich 7 Personen

Bei den Jüngeren wird die visuell-informative Unterstützung durch die Tonalität **motivierend** am positivsten beeinflusst. Hingegen wird durch die Tonalität **höflich** der negativste Einfluss aller Tonalitäten geltend gemacht.

Der Unterschied zwischen den Tonalitäten motivierend und höflich ist **nicht signifikant** (p-value=0.3035).

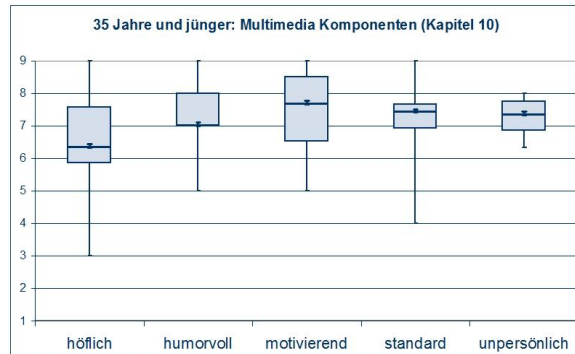


Abbildung 120: 35 Jahre und jünger: Multimedia Komponenten (Kapitel 10)

	höflich	humorv.	motivier.	standard	unpers.
75%	7.5833	8.0000	8.5000	7.6667	7.7500
50%	6.3333	7.0000	7.6667	7.4167	7.3333
25%	5.8333	7.0000	6.5000	6.9167	6.8333

Kruskal-Wallis-Test: p-value = 0.8108

Wilcoxon-Test: motivierend -höflich: p-value = 0.3035

Frage 10.1: Wie gut hat das Bildmaterial den Ablauf unterstützt?

Bei den Jüngeren besteht der Eindruck dass durch die Tonalität **humorvoll** das Bildmaterial den Ablauf besser unterstützt als durch die Tonalität **unpersönlich**.

Der Unterschied zwischen den Tonalitäten humorvoll und unpersönlich ist **nicht signifikant** (p-value=0.06925). Der Wert liegt jedoch *unterhalb des 2. Signifikantniveau von 0.1*, und demzufolge tendenziell zu beachten.

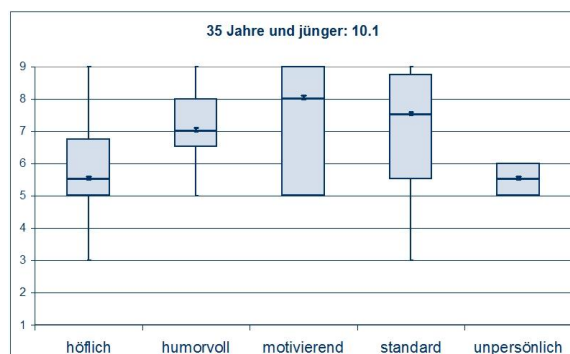


Abbildung 121: 35 Jahre und jünger: Frage 10.1

	höflich	humorv.	motivier.	standard	unpers.
75%	6.7500	8.0000	9.0000	8.7500	6.0000
50%	5.5000	7.0000	8.0000	7.5000	5.5000
25%	5.0000	6.5000	5.0000	5.5000	5.0000

Kruskal-Wallis-Test: p-value = 0.4478

Wilcoxon-Test: humorvoll-unpersönlich: p-value = 0.06925

Auswertung über Personen 45 jähig und älter

- Datenmenge n=38
- höflich 9 Personen
- humorvoll 4 Personen
- motivierend 8 Personen
- standard 10 Personen
- unpersönlich 7 Personen

Bei den Älteren bestehen die grössten Unterschiede zwischen den Tonalitäten höflich und unpersönlich. Die Tonalität **höflich** unterstützt den visuell-informativen Effekt am besten, wo hingegen die Tonalität **unpersönlich** am schwächsten wirkt.

Der Unterschied zwischen diesen Tonalitäten ist **nicht signifikant** (p-value=0.3379).

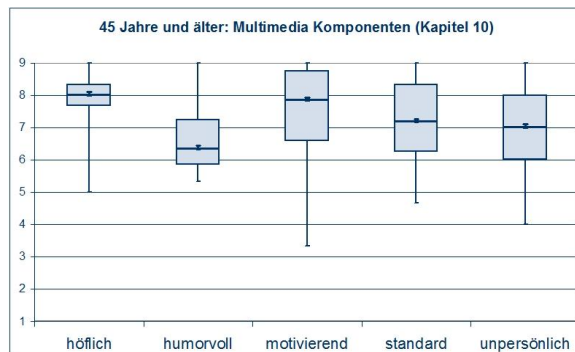


Abbildung 122: 45 Jahre und älter: Multimedia Komponenten (Kapitel 10)

	höflich	humorv.	motivier.	standard	unpers.
75%	8.3333	7.2500	8.7500	8.3333	8.0000
50%	8.0000	6.3333	7.8333	7.1667	7.0000
25%	7.6667	5.8333	6.5833	6.2500	6.0000

Kruskal-Wallis-Test: p-value = 0.8235

Wilcoxon-Test: höflich-unpersönlich: p-value = 0.3379

10.6 BEURTEILUNG DER TONALITÄTEN DURCH GRUPPEN

Beurteilung durch Männer

An erster Stelle schätzen Männer motivierende Sätze als präferierte Fehlermeldungen ein. Als letzte dann unpersönliche. Diese Einschätzung deckt sich grösstenteils mit den Ermittlungen durch den Rest des Fragebogens.

Dass humorvolle Sätze jedoch eher unangenehm sein sollten, das überrascht uns doch ein wenig. Denn die Tonalität humorvoll hat bei den Männer ein relativ hoher Stellenwert bezüglich User Experience. Hingegen sind höfliche Sätze besser bewertet als in Realität.

Als Vergleich zeigen wir die Auswertungsgrafik des Gesamteindruckes der Anwendung:

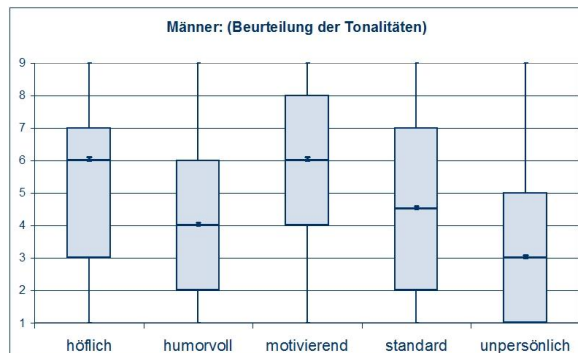
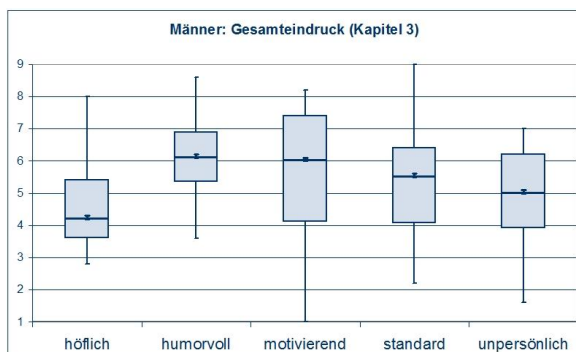


Abbildung 123: Beurteilung der Tonalität durch Männer

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	6.0000	8.0000	7.0000	5.0000
50%	6.0000	4.0000	6.0000	4.5000	3.0000
25%	3.0000	2.0000	4.0000	2.0000	1.0000



Beurteilung durch Frauen

Bei den Frauen sind es die motivierenden und höflichen Sätze welche stärker favorisiert wurden als z.B. unpersönlich. Diese Eigeneinschätzung verhält sich genau anders herum als die Bewertung der Anwendung unter Einfluss der Tonalität.

Als Vergleich zeigen wir die Auswertungsgrafik des Gesamteindruckes der Anwendung:

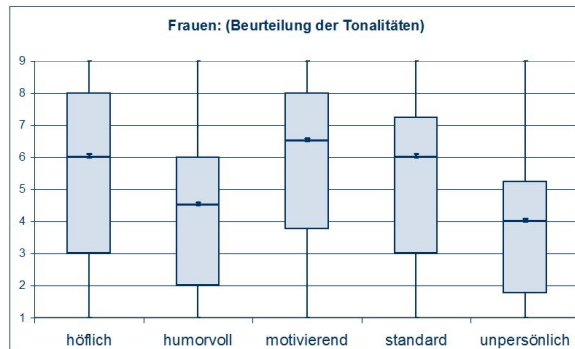
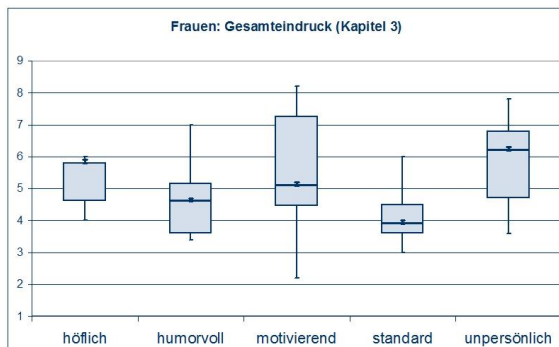


Abbildung 124: Beurteilung der Tonalität durch Frauen

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	6.0000	8.0000	7.2500	5.2500
50%	6.0000	4.5000	6.5000	6.0000	4.0000
25%	3.0000	2.0000	3.7500	3.0000	1.7500



Beurteilung durch 35 Jahre und jünger

Auch bei den Jüngeren schiesst die Selbsteinschätzung an der Realität vorbei. Die Tonalitäten humorvoll und unpersönlich werden durch die explizite Beurteilung als eher schlecht gewertet, jedoch beim Test haben Jüngere mit der Tonalität humorvoll und unpersönlich viel bessere Beurteilung der User Experience bekundet. Hingegen sind die Tonalitäten höflich und motivierend viel höher eingeschätzt.

Als Vergleich zeigen wir die Auswertungsgrafik des Gesamteindruckes der Anwendung:

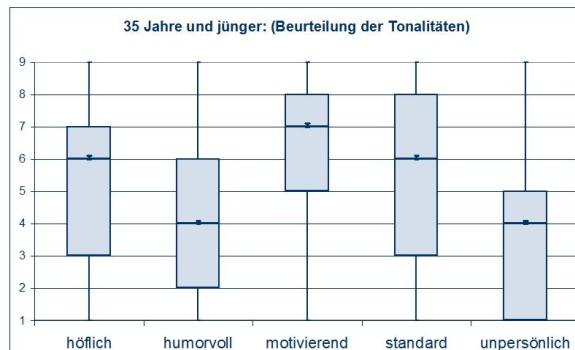
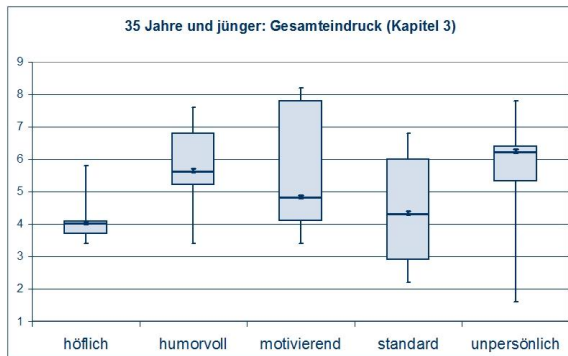


Abbildung 125: Beurteilung der Tonalität durch 35 Jahre und jünger

	höflich	humorv.	motivier.	standard	unpers.
75%	7.0000	6.0000	8.0000	8.0000	5.0000
50%	6.0000	4.0000	7.0000	6.0000	4.0000
25%	3.0000	2.0000	5.0000	3.0000	1.0000

Ergebnisse



Beurteilung durch 45 Jahre und älter

Bei den Älteren ist die Selbsteinschätzung sehr gut ausgefallen. Diese deckt sich mehr oder weniger mit den Beobachtungen aus dem Test. Z.B. sind höflich und motivierend eher hoch eingeschätzt, das

Als Vergleich zeigen wir die Auswertungsgrafik des Gesamteindruckes der Anwendung:

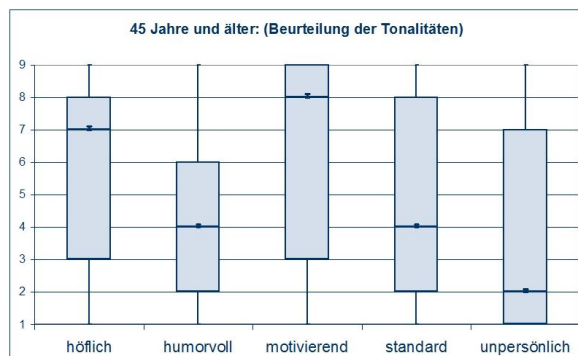
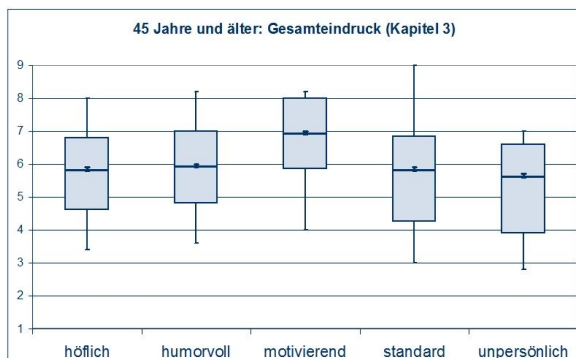


Abbildung 126: Beurteilung der Tonalität durch 45 Jahre und älter

	höflich	humorv.	motivier.	standard	unpers.
75%	8.0000	6.0000	9.0000	8.0000	7.0000
50%	7.0000	4.0000	8.0000	4.0000	2.0000
25%	3.0000	2.0000	3.0000	2.0000	1.0000

11 REFLEXION

Theodor Huber

Der Beginn unserer Masterarbeit war für mich von Anfang an problemlos. Das ist nicht selbstverständlich, da ich meine Kollegen nicht im zuvor absolvierten Lehrgang kennen lernen konnte. Meine MAS HCID-Ausbildung begann mit dem CAS Lehrgang «Interaction Design» und danach erst folgte «Requirements Engineering». Bei der Masterarbeit lernte ich somit eine dritte Weiterbildungsgruppe kennen. Ich habe mit Dorota und Manuel ein gutes Team gefunden. Die Zusammenarbeit war äusserst interessant und kollegial trotz der z.T. hohen Komplexität unseres gewählten Themas. Wir haben kritische Situationen, z.B. bei der statistischen Auswertung unserer Daten, gut überstanden und dies führte schlussendlich zu einem spannenden Forschungsergebnis. Unsere gewählten Methoden haben sich bewährt und es hat Spass gemacht, mit dem eigens im Team entwickelten Test zu sehen, wie wir Schritt für Schritt unserem Ziel näher kamen. Ein besonderes Highlight für mich war, bei den ersten qualitativen Interviews zu beobachten, wie die erhobenen Daten aus dem Gespräch nicht immer mit denjenigen aus den konkreten Tätigkeiten des Probanden während des Tests übereinstimmten.

Die statistische Auswertung haben wir m.E. aufgrund unseres mangelnden Vorwissens in diesem Gebiet vom Aufwand her unterschätzt. Schlussendlich haben wir aber in ein neues Wissensgebiet vordringen können und dadurch einen guten Überblick über die verschiedenen Möglichkeiten zur Messung der User Experience erhalten.

Manuel Sanchez

Wir waren uns am Anfang nicht einig ob wir wirklich eine Arbeit dieser Art machen sollten. Ich wusste eigentlich nicht so recht auf was wir uns da eingelassen hatten. Ich war jedoch hoch motiviert und war bereit einiges auf mich zu nehmen. Das Team war nicht von Beginn weg so aufgestellt. Zusammen mit Dorota hatte ich bereits zwei Abschlussarbeiten (Requirement und Interaction Design) realisiert und wusste wie es war mit ihr zusammen zu arbeiten. Teddy kam danach zu uns. Ich kannte Teddy bis zu diesem Zeitpunkt gar nicht. Seine ruhige und zuvorkommende Art hat uns sehr gut getan. Ich finde, ich durfte mit einem spitze Team die Arbeit machen, und es hat sehr viel Spass gemacht.

Das von uns gewählte Thema hat mir persönlich sehr viel gebracht. Ich durfte z.B. mich in die Script-sprache PHP einarbeiten. Zusammen haben wir dann den Prototypen so weit gebracht, dass er genau das tat, was wir uns vorgenommen hatten. Vor allem war es nicht das Werk eines einzelnen, sondern ein Zusammenspiel von verschiedenartigem Knowhow, das jeder von uns mitbrachte. Als wäre dieses Team das Resultat einer Rekrutierung. Jeder von uns hatte seine Fähigkeiten die er oder sie spezifisch einsetzen konnte. Darüber hinaus haben wir sehr viel über Tonalität gelernt. Eine grossartige Erfahrung war für mich der Tag im Flughafen, als wir die qualitativen Tests durchgeführt haben. Ich war mir nicht sicher bis zu diesem Zeitpunkt, ob ich fähig wäre auf Leute zu zugehen und sie um ein Interview zu bitten. Wir haben sehr viel Feedback bekommen an diesem Tag. Es wäre schade gewesen, wenn wir am Online-Test hätten erfahren müssen, welche Fehler noch vorhanden sind.

Was mir ganz speziell Sorge bereitet hat, war die Tatsache dass wir lange Zeit nicht wussten wie wir die Auswertung statistisch korrekt durchführen sollten. Selbst als wir dank einem

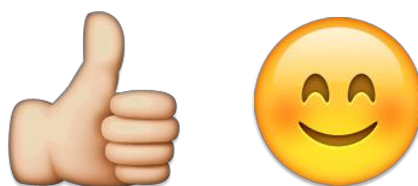
Kollegen von Teddy, der Mathematik studiert hatte, etwas Licht ins Dunkle bringen konnten, war trotzdem noch vieles unklar. Was wir besser hätten machen können, ist ganz klar: Wir hätten uns viel früher mit der Thematik Datenauswertung und Statistik befassen müssen. Nun sind wir aber trotz allem fertig, und ich bin froh darüber, dass wir uns entschlossen haben diese Arbeit zu machen.

Dorota Przygienda

Den Zusammenhang zwischen Fehlermeldungen und Emotionen der Benutzer zu untersuchen, fand ich von Anfang an sehr spannend. Erstens, weil es noch so wenig erforscht war und zweitens wegen meinen Job und Erfahrungen als Software Engineer, wo ich mit den beiden Themen oft konfrontiert bin.

Das Neuland zu betreten, wo wir mit wissenschaftlichen Studien, Datenauswertungen und Statistik zu tun bekommen hatten, war ein mutiger Schritt. Der Pioniergeist unseres Projektes zeigte sich nicht nur in dem Thema aber auch bei der Vorgehensweise einer wissenschaftlicher Studie. Mit ist das erst ganz bewusst geworden, als Markus Stolze uns mal sagte, dass wir die Ersten an der HSR sind, die das wagten ...

Dank guter Zusammenarbeit und Kommunikation im Team, haben wir in dieser Zeit viel gelernt und interessante Erfahrungen gesammelt. Eine der wichtigsten Erfahrungen für mich persönlich waren unsere qualitativen UX Tests, sowohl in den Büros von UBS, AdNovum, CSS als auch am Flughafen Zürich. Besonders motivierend waren das Interesse und die Unterstützung bei der Arbeit, die uns in unserem privaten Umfeld entgegen kam, sowohl von den freiwilligen Studienteilnehmern, als auch von den Profis in der Statistik, Psychologie und Werbung. Das Team hat sehr gut funktioniert und wir konnten uns gut gegenseitig inspirieren, motivieren und unterstützen. In dieser Arbeit ging es um die User Experience aber für mich war sie auch eine tolle und spannende Life Experience. Danke Euch, Manuel und Teddy!



12 AUSBLICK

Unser Test ist stark kontextabhängig (Hotelbranche). Die User Experience wird stark beeinflusst durch die Grundstimmung des Benutzers. Diese Erkenntnis haben wir aus der Beurteilung der Tonalitäten erhalten. Benutzer welche ihren Test in einer bestimmten Grundtonalität ausgeführt haben, machten grundsätzlich eine andere Beurteilung dieser Tonalität. Basierend auf dieser Erkenntnis wäre es interessant folgendes weiter zu untersuchen:

- Wie verhält sich die Grundstimmung bei Fehlermeldungen, wenn die Tonalität über die gesamte Anwendung angewandt wird?
- Ist die Akzeptanz der Tonalität höher wenn diese sich dem Kontext anpasst und nicht dem Benutzer?
- Ergeben sich dieselben Resultate wenn die Tonalität auf mobile devices angewendet wird, wenn es weniger visuellen Kontext gibt? Oder gelten generell andere Regeln für mobile devices?

Wie wir nun wissen, hat Tonalität z.T. einen starken Einfluss auf die User Experience. Interessant wäre auch regionale und kulturelle Unterschiede zu untersuchen.

Zu beachten wäre die demographische Verteilung und die Grösse der Testpopulation, damit sie der Bevölkerungsstruktur in der Schweiz entspricht. Dadurch wäre eine repräsentative Studie möglich.

13 GLOSSAR

Affinity Diagramm	Affinity Diagramm («Wall») ist ein Werkzeug um grosse Mengen von Daten in logische Gruppen aufzuteilen und diese zu benennen. Das Affinity Diagramm ist eine «bottom-up» Methode. Konsolidierte Daten können auch verwendet werden, um eine Reihe von Personas zu beschreiben. Siehe auch Beyer, H. & Holtzblatt, K. (1998) Contextual Design
Anforderung	Eine geforderte Eigenschaft, die vom neuen System erfüllt werden soll (engl. requirement).
Artefakt	Dinge, die Personen erstellen oder benutzen, um ihre Arbeit zu erledigen. Beispiele sind Listen, Formulare, Rapporte etc.
Box-Whisker-Plot	Der Boxplot (auch Box-Whisker-Plot oder deutsch Kastengrafik) ist ein Diagramm, das zur grafischen Darstellung der Verteilung kardinalskalierten Daten verwendet wird. Es fasst dabei verschiedene robuste Streuungs- und Lagemasse in einer Darstellung zusammen. Ein Boxplot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen. Deshalb werden alle Werte der sogenannten Fünf-Punkte-Zusammenfassung, also der Median, die zwei Quartile und die beiden Extremwerte, dargestellt.
CSS	Cascading Style Sheets - steht für hintereinander geschachtelte Gestaltungsvorlagen, auch CSS genannt. CSS ist eine deklarative Sprache für Stilvorlagen (engl. stylesheets) von strukturierten Dokumenten. Sie wird zusammen mit HTML und XML (zum Beispiel bei SVG) eingesetzt.
Datenbank View	Ein Datenbank Objekt, welches eine bestimmte logische Sicht auf die Daten und deren Relationen darstellt. Eine View wird mittels einer SQL Abfrage definiert und kann gleich wie eine Tabelle benutzt werden.
Design Patterns	Entwurfsmuster welches eine wiederverwendbare Vorlage zur Problemlösung darstellt, die in einem bestimmten Zusammenhang einsetzbar ist. Im Software Design können die z.B. Formular-Elemente mit bestimmten Verhalten, Funktionalität und visuellem Design sein.
Expert Review	Usability Analyse Methode zum überprüfen der Usability eines Produkts durch mehrere, voneinander unabhängige Spezialisten.
GUI	Graphical User Interface: grafische Benutzeroberfläche ist eine Software Komponente, die den Benutzern Interaktion mit einer Software durch grafische Elemente ermöglicht.
HCID	Human Computer Interaction Design (HCID) besteht einerseits aus der Wissenschaft zur Erforschung der Kommunikation zwischen Mensch und Computer und im weiteren aus der Gestaltung der Interaktion zwischen Benutzer und System, dem Interaktionsdesign.

HTML	Hypertext Markup Language, ist eine einfache Notation zur Beschreibung von Form und Inhalt eines Textes durch einfache Anmerkungen oder «Tags» auf den Text. HTML wird als grundlegendes Format für die Präsentation von Dokumenten im Web verwendet.
Javascript	JavaScript (JS) ist eine Skriptsprache die Möglichkeiten von HTML und CSS erweitert, in dem sie Benutzerinteraktionen auswertet, Inhalte verändert oder nachlädt oder generiert.
Kruskal-Wallis-Test	Der Kruskal-Wallis-Test (nach William Kruskal und Wilson Allen Wallis; auch H-Test) ist ein parameterfreier statistischer Test, mit dem im Rahmen einer Varianzanalyse getestet wird, ob unabhängige Stichproben (Gruppen oder Messreihen) hinsichtlich einer ordinalskalierten Variable einer gemeinsamen Population entstammen.[1] Er ähnelt einem Mann-Whitney-U-Test und basiert wie dieser auf Rangplatzsummen, mit dem Unterschied, dass er für den Vergleich von mehr als zwei Gruppen angewendet werden kann.
Lautes Denken	Technik die u.a. in Usability Tests eingesetzt wird. Benutzer werden gebeten während dem Ausführen einer Tätigkeit ihre Gedanken laut auszusprechen.
Mann Whitney Test	Der Wilcoxon-Mann-Whitney-Test (auch: „Mann-Whitney-U-Test“, „U-Test“, „Wilcoxon-Rangsummentest“) ist ein parameterfreier statistischer Test. Der U-Test ist ein Homogenitätstest. Er dient zur Überprüfung der Signifikanz der Übereinstimmung zweier Verteilungen, also ob zwei unabhängige Verteilungen A und B (zum Beispiel eine unbeeinflusste und eine beeinflusste) zu derselben Grundgesamtheit gehören.
Median	Der Median oder Zentralwert ist ein Mittelwert für Verteilungen in der Statistik. Der Median einer Auflistung von Zahlenwerten ist derjenige Wert, welcher an der mittleren Stelle steht, wenn man die Werte der Größe nach sortiert. Zum Beispiel für die Werte 4, 1, 37, 2, 1 ist die Zahl 2 der Median, nämlich die mittlere Zahl in 1, 1, 2, 4, 37.
MySQL	MySQL ist ein weltweit verbreitetes relationales Datenbankverwaltungssystem. Es ist gratis für verschiedene Betriebssysteme verfügbar und bildet die Grundlage für viele dynamische Webauftritte. MySQL wurde 1994 in Schweden entwickelt welches dann 1994 von Sun gekauft wurde. Inzwischen hat Oracle Sun gekauft und damit auch MySQL.
Persona	Eine Persona ist ein fiktiver Charakter, der die Bedürfnisse einer ganzen Reihe echter Anwender repräsentieren kann. Zweck von Personas ist, dass die Produkt-Designer ihre Anwender während der Gestaltung nie aus den Augen lassen. Alan Cooper wurde vor allem mit seinem Konzept der Personas bekannt.
PHP	Rasmus Lerdorf entwickelte 1995 die Skriptsprache die bis heute hauptsächlich zur Erstellung dynamischer Webseiten oder Webanwendungen verwendet wird. PHP zeichnet sich u.A. durch breite Datenbankunterstützung aus und als freie Software unter der PHP-Lizenz verbreitet.

PSPP	Freie Software zur Analyse von statistischen Daten, kompatibel mit SPSS
Quartil	Quartile (lateinisch „Viertelwerte“) sind die Quantile, die auch als Q1 („unteres Quartil“), Q2 („mittleres Quartil“) und Q3 („oberes Quartil“) bezeichnet werden. Sie sind die in der Statistik mit am häufigsten verwendete Form der Quantile. Quantile erlauben einem ganz praktische Aussagen im Stile von „25% aller Frauen sind kleiner als 1,62 m“ – wobei 1,62 m hier das 25%-Quantil ist.
QUIS	"The Questionnaire For User Interaction Satisfaction" ist ein standardisierter Fragebogen um die subjektive Zufriedenheit von spezifischen Aspekten der Schnittstelle in Mensch-Computer-Systemen zu ermitteln und wurde von 1987 von Ben Shneidermann entwickelt.
Responsive Design	Ist ein gestalterischer und technischen Ansatz zur Erstellung von Websites, welcher auf die Eigenschaften des jeweils benutzten Endgeräts reagieren kann. Beispiel: Die klickbaren Flächen einer Website auf Mobile Devices müssen ausreichend gross sein, um sie mit der Fingerspitze bedienen zu können. Der gleiche Button muss auf einer Website für Desktop PCs wesentlich kleiner dargestellt werden, da er dort mit der Maus oder der Tastatur bedient wird.
RStudio	Ein freies Software Packet zur statistischen Datenanalyse und Darstellung, es basiert auf Skriptsprache R
SPSS	IBM SPSS Statistics: ein modular aufgebautes Software Packet zur statistischen Datenanalyse
Signifikanz	Als signifikant werden in der Statistik die Unterschiede zwischen Messgrößen oder Variablen bezeichnet, wenn die Wahrscheinlichkeit, dass sie durch Zufall derart zustande kommen würden, nicht über einer gewissen Schwelle liegt. Der Ausdruck signifikant bedeutet daher in der induktiven Statistik, anders als in der Alltagssprache, nicht bedeutsam, wesentlich, wichtig oder groß, sondern allein, dass ein überzufälliger Zusammenhang angenommen wird, bezogen auf eine zuvor festgelegte Schwelle
SQL	Structured Query Language, eine Datenbanksprache die für Abfrage, Definition und Bearbeitung von Daten benutzt wird.
Szenario	Ein Szenario ist eine kurze, einfache Erzählung die beschreibt, wie eine Persona versucht ihre Nutzerziele zu erreichen. Durch das Erstellen von Szenarios resultieren Anforderungen an das System, damit die Personas ihre Ziele erreichen können.
T-Test	Der t-Test ist ein Begriff aus der mathematischen Statistik, er bezeichnet eine Gruppe von Hypothesentests mit t-verteilter Testprüfgröße. Oft ist jedoch mit dem t-Test der Einstichproben- bzw. Zweistichproben-t-Test gemeint.
UCD	User Centered Design besteht aus Vorgehensmodellen, welche die späteren Benutzer systematisch in die Entwicklung von Systemen und Produkten einbeziehen.
Usability	Ein Mass für die Benutzbarkeit eines technischen Systems (auch Gebrauchstauglichkeit oder Benutzerfreundlichkeit genannt).

Use Case Ein Use Case, auch Anwendungsfall genannt, beschreibt eine Abfolge von Interaktionen zwischen Akteur und System, die ein Ergebnis von Wert für den Akteur liefert.

User Experience (UX) Wahrnehmungen und Reaktionen einer Person, die sich bei der Benutzung oder der erwarteten Verwendung eines Produktes ergeben.

14 ABBILDUNGSVERZEICHNIS

Abbildung 1: User Experience vs. Usability	5
Abbildung 2: Facetten der UX nach Hassenzahl und Tractinsky (2006, S. 95)	6
Abbildung 3: Einflussfaktoren der Usability (Shackel, 2001)	20
Abbildung 4: Projektplan	22
Abbildung 5: Projekt Tätigkeiten im Bezug auf ISO Standard Prozess	23
Abbildung 6: Rollen und Aufgaben im Team	28
Abbildung 7: Anwendung der Tonalität bei Fehlermeldungen und Komponenten des Mensch-Maschine-Systems beim Online-UX-Test	32
Abbildung 8: Erste Ideen zum Konzept der Tonalität wurden an der Wandtafel festgehalten und diskutiert	33
Abbildung 9: Definition der Tonalitäten bei Fehlermeldungen	34
Abbildung 10: Fehlermeldungs-Beispiele in jeder Tonalität	36
Abbildung 11: Einstiegs-Seite für den Online-UX-Test	37
Abbildung 12: Personalien Erfassung im Online-UX-Test	37
Abbildung 13: Formular und Hilfestellung (rechts) des Online-UX-Tests	38
Abbildung 14: Komponenten und Aufbau des Online-UX-Tests	39
Abbildung 15: Datenmodell für Online-UX-Test	43
Abbildung 16: Veröffentlichung des Online-UX-Tests auf Google+	47
Abbildung 17: Tabelle mit Quartilen pro Tonalität	51
Abbildung 18: Gesamtwertung über alle Fragen	55
Abbildung 19: Beurteilung der Tonalität durch allen Probanden	56
Abbildung 20: Beurteilung der Tonalität durch Probanden der Tonalität humorvoll	56
Abbildung 21: Beurteilung der Tonalität durch Probanden der Tonalität motivierend	57
Abbildung 22: Beurteilung der Tonalität durch Probanden der Tonalität standard	57
Abbildung 23: Beurteilung der Tonalität durch Probanden der Tonalität unpersönlich	57
Abbildung 24: Verteilungskurve der Erfahrungsgruppen	58
Abbildung 25: 4x pro Jahr online buchen	59
Abbildung 26: 1x pro Jahr online buchen	60
Abbildung 27: seltener als alle 3 Jahre einmal online buchen	60
Abbildung 28: Beurteilung der Tonalität durch sehr erfahrene Benutzer	61
Abbildung 29: Beurteilung der Tonalität durch unerfahrene Benutzer	62
Abbildung 30: Gesamtwertung über alle Fragen	62

Abbildung 31: Beurteilung der User Experience aufgrund des Kapitel 3 von QUIS.....	63
Abbildung 32: Alle: Kapitel 3	64
Abbildung 33: Frauen: Kapitel 3	65
Abbildung 34: Frauen: Frage 3.1	65
Abbildung 35: Frauen: Frage 3.3	66
Abbildung 36: Frauen: Frage 3.4	67
Abbildung 37: Frauen: Frage 3.6	68
Abbildung 38: Männer: Kapitel 3	69
Abbildung 39: Männer: Frage 3.1.....	70
Abbildung 40: Männer: Frage 3.3.....	70
Abbildung 41: 35 Jahre und jünger: Kapitel 3	71
Abbildung 42: 35 Jahre und jünger: Frage 3.1	72
Abbildung 43: 35 Jahre und jünger: Frage 3.2	72
Abbildung 44: 35 Jahre und jünger: Frage 3.3	73
Abbildung 45: 35 Jahre und jünger: Frage 3.6	73
Abbildung 46: 45 Jahre und älter: Kapitel 3	74
Abbildung 47: 45 Jahre und älter: Frage 3.1	75
Abbildung 48: 45 Jahre und älter: Frage 3.3	75
Abbildung 49: 45 Jahre und älter: Frage 3.6	76
Abbildung 50: Alle: Kapitel 4	77
Abbildung 51: Frauen: Kapitel 4	78
Abbildung 52: Männer: Kapitel 4	79
Abbildung 53: Männer: Frage 4.1.....	80
Abbildung 54: 35 Jahre und jünger: Kapitel 4	81
Abbildung 55: 45 Jahre und älter: Kapitel 4	82
Abbildung 56: 45 Jahre und älter: Frage 4.1	83
Abbildung 57: 45 Jahre und älter: Frage 4.3	83
Abbildung 58: Alle: Kapitel 5	84
Abbildung 59: Frauen: Kapitel 5	85
Abbildung 60: Frauen: Frage 5.1	86
Abbildung 61: Frauen: Frage 5.1.2	86
Abbildung 62: Frauen: Frage 5.2	87
Abbildung 63: Frauen: Frage 5.3	87
Abbildung 64: Frauen: Frage 5.5.2	88

Abbildung 65: Frauen: Frage 5.5.4	88
Abbildung 66: Frauen: Frage 5.6	89
Abbildung 67: Frauen: Frage 5.6.1	89
Abbildung 68: Frauen: Frage 5.6.2	90
Abbildung 69: Männer: Kapitel 5	91
Abbildung 70: Männer: Frage 5.5.4.....	92
Abbildung 71: 35 Jahre und jünger: Kapitel 5	93
Abbildung 72: 35 Jahre und jünger: Frage 5.1	94
Abbildung 73: 35 Jahre und jünger: Frage 5.1.2	94
Abbildung 74: 35 Jahre und jünger: Frage 5.2	95
Abbildung 75: 35 Jahre und jünger: Frage 5.3	95
Abbildung 76: 35 Jahre und jünger: Frage 5.5	96
Abbildung 77: 35 Jahre und jünger: Frage 5.5.2	96
Abbildung 78: 45 Jahre und älter: Terminologie & Systeminformation (Kapitel 5).....	97
Abbildung 79: 45 Jahre und älter: Frage 5.1	98
Abbildung 80: 45 Jahre und älter: Frage 5.1.2	98
Abbildung 81: 45 Jahre und älter: Frage 5.4	99
Abbildung 82: 45 Jahre und älter: Frage 5.5	99
Abbildung 83: 45 Jahre und älter: Frage 5.5.2	100
Abbildung 84: 45 Jahre und älter: Frage 5.5.4	100
Abbildung 85: 45 Jahre und älter: Frage 5.6.1	101
Abbildung 86: Alle: Lernfortschritt (Kapitel 6)	102
Abbildung 87: Alle: Frage 6.4	103
Abbildung 88: Frauen: Lernfortschritt (Kapitel 6)	104
Abbildung 89: Männer: Lernfortschritt (Kapitel 6).....	105
Abbildung 90: Männer: Frage 6.4.....	106
Abbildung 91: 35 Jahre und jünger: Lernfortschritt (Kapitel 6)	107
Abbildung 92: 35 Jahre und jünger: Frage 6.4.3	108
Abbildung 93: 45 Jahre und älter: Lernfortschritt (Kapitel 6)	109
Abbildung 94: Alle: Systemeigenschaften (Kapitel 7)	110
Abbildung 95: Frauen: Systemeigenschaften (Kapitel 7)	111
Abbildung 96: Männer: Systemeigenschaften (Kapitel 7).....	112
Abbildung 97: 35 Jahre und jünger: Systemeigenschaften (Kapitel 7)	113
Abbildung 98: 45 Jahre und älter: Systemeigenschaften (Kapitel 7)	114

Abbildung 99: 45 Jahre und älter: Frage 7.4	115
Abbildung 100: 45 Jahre und älter: Frage 7.4.1	115
Abbildung 101: Alle: Online-Anleitung (Kapitel 9)	116
Abbildung 102: Alle: Frage 9.3.1	117
Abbildung 103: Alle: Frage 9.4	117
Abbildung 104: Frauen: Online-Anleitung (Kapitel 9)	118
Abbildung 105: Frauen: Frage 9.1	118
Abbildung 106: Männer: Online-Anleitung (Kapitel 9)	119
Abbildung 107: Männer: Frage 9.3.1.....	120
Abbildung 108: Männer: Frage 9.4.....	120
Abbildung 109: 35 Jahre und jünger: Online-Anleitung (Kapitel 9)	121
Abbildung 110: 35 Jahre und jünger: Frage 9.3.1	122
Abbildung 111: 45 Jahre und älter: Online-Anleitung (Kapitel 9)	122
Abbildung 112: 45 Jahre und älter: Frage 9.1.1	123
Abbildung 113: 45 Jahre und älter: Frage 9.3.1	123
Abbildung 114: 45 Jahre und älter: Frage 9.4	124
Abbildung 115: 45 Jahre und älter: Frage 9.4.1	124
Abbildung 116: Alle: Multimedia Komponenten (Kapitel 10).....	125
Abbildung 117: Frauen: Multimedia Komponenten (Kapitel 10).....	125
Abbildung 118: Männer: Multimedia Komponenten (Kapitel 10)	126
Abbildung 119: Männer: Frage 10.4.....	126
Abbildung 120: 35 Jahre und jünger: Multimedia Komponenten (Kapitel 10).....	127
Abbildung 121: 35 Jahre und jünger: Frage 10.1	127
Abbildung 122: 45 Jahre und älter: Multimedia Komponenten (Kapitel 10).....	128
Abbildung 123: Beurteilung der Tonalität durch Männer.....	129
Abbildung 124: Beurteilung der Tonalität durch Frauen	130
Abbildung 125: Beurteilung der Tonalität durch 35 Jahre und jünger	130
Abbildung 126: Beurteilung der Tonalität durch 45 Jahre und älter	131

15 LITERATURVERZEICHNIS

Bücher

Cooper Alan, Robert Reimann, David Cronin (2010): *About Face*, 1 Auflage, mitp Verlag Gruppe Hüthig Jehle Rehm GmbH, ISBN 978-3-8266-5888-4

Wroblewski Luke (2008): *Web Form Design: Filling in the Blanks*, Rosenfeld Media, New York, USA

Wroblewski Luke (2009): *Mobile First*, A Book Apart, New York, USA

Norman Donald A. (2002): *The design of everyday things*, Basic Books, New York, USA

Johnson Jeff (2010): *Designing with the Mind in Mind, Simple Guide to Understanding User Interface Design Rules*, ELSEVIER, Burlington, USA

Garrett Jesse James (2012): *Die Elemente der User Experience*, 2 Auflage, Addison-Wesley Verlag, München, Deutschland

Sauro Jeff, James R. Lewis (2012), *Quantifying the user experience*, Morgan Kaufmann, Massachusetts, USA

Richter Michael, Markus Flückiger (2010), *Usability Engineering kompakt*, Spektrum Akademischer Verlag, Heidelberg, Deutschland

Heinecke M. Andreas, Prof. Dr. (2004) *Mensch-Computer-Interaktion*

Sarodnick Florian, Hennig Brau (2011), *Methoden der Usability Evaluation, Wissenschaftliche Grundlagen und praktische Anwendung*, 2 Auflage, Verlag Hans Huber, Hogrefe AG, Bern, Schweiz

Links

Wikipedia, [Fehlermeldung](http://de.wikipedia.org/wiki/Fehlermeldung), <http://de.wikipedia.org/wiki/Fehlermeldung> (Zugriff: 02.11.2013)

Duden, [Tonalität](http://www.duden.de/rechtschreibung/Tonalitaet), <http://www.duden.de/rechtschreibung/Tonalitaet> (Zugriff: 23.1.2014)

Wikipedia, [Tonalität](http://de.wikipedia.org/wiki/Tonalit%C3%A4t_%28Musik%29), http://de.wikipedia.org/wiki/Tonalit%C3%A4t_%28Musik%29 (Zugriff: 23.1.2014)

Wikipedia, [Tonality](http://de.wikipedia.org/wiki/Tonality), <http://de.wikipedia.org/wiki/Tonality> (Zugriff: 23.1.2014)

Fischer Mike B. (2009): [Improved error handling](http://completeusability.com/improved-error-handling-part-2-writing-effective-error-messages), part 2: writing effective error messages <http://completeusability.com/improved-error-handling-part-2-writing-effective-error-messages> (Zugriff: 02.11.2013)

Travis Dr. David, [Userfocus](http://www.userfocus.co.uk/resources/helpchecklist.html) (2013): 37 usability guidelines for help, feedback and error tolerance, www.userfocus.co.uk/resources/helpchecklist.html (Zugriff: 11.11.2013)

Wroblewski Luke (2009): [The Apple Store's Checkout Form Redesign](http://www.lukew.com/ff/entry.asp?968), www.lukew.com/ff/entry.asp?968 (Zugriff: 10.11.2013)

Wroblewski Luke (2009): [Web Form Design: Adobe's Error Messages](http://www.lukew.com/ff/entry.asp?949), www.lukew.com/ff/entry.asp?949 (Zugriff: 10.11.2013)

Nielsen Jakob (1995): [10 Usability Heuristics for User Interface Design](http://www.nngroup.com/articles/ten-usability-heuristics), www.nngroup.com/articles/ten-usability-heuristics (Zugriff: 14.11.2013)

Nielsen Jakob (2001): [Error Message Guidelines](http://www.nngroup.com/articles/error-message-guidelines), www.nngroup.com/articles/error-message-guidelines (Zugriff: 9.11.2013)

Norman Donald A. (2013) [Google's non-error Messages](http://www.jnd.org/dn.mss/googles_non-error_m.htm), www.jnd.org/dn.mss/googles_non-error_m.htm (Zugriff: 19.11.2013)

Wilska Emily (2013): [Non-Fatal Errors: Creating Usable, Effective Error Messages](http://www.writersua.com/articles/message) www.writersua.com/articles/message (Zugriff: 18.11.2013)

Bargas-Avila J.A., O. Brenzikofer, S.P. Roth, A.N. Tuch, S. Orsini and K. Opwis (2010). [Simple but Crucial User Interfaces in the World Wide Web: Introducing 20 Guidelines for Usable Web Form Design](http://cdn.intechopen.com/pdfs/10814/InTech-Simple%20but%20Crucial%20User%20Interfaces%20in%20the%20World%20Wide%20Web%20Introducing%2020%20Guidelines%20for%20Usable%20Web%20Form%20Design.pdf), [http://cdn.intechopen.com/pdfs/10814/InTech-](http://cdn.intechopen.com/pdfs/10814/InTech-Simple%20but%20Crucial%20User%20Interfaces%20in%20the%20World%20Wide%20Web%20Introducing%2020%20Guidelines%20for%20Usable%20Web%20Form%20Design.pdf)

Simple_but_crucial_user_interfaces_in_the_world_wide_web_introducing_20_guidelines_for_usable_web_form_design.pdf (Zugriff: 18.11.2013)

Hassenzahl, M. & Tractinsky, N. (2006): *User Experience – a Research Agenda*. Behaviour and Information Technology, Vol. 25, No. 2, March–April 2006, pp. 91-97, <https://ccrma.stanford.edu/~sleitman/UserExperienceAResearchAgenda.pdf> (Zugriff: 11.1.2013)

Schneider Wolfgang (2013): *Grundsatz Fehlertoleranz*, www.ergo-online.de/site.aspx?url=html/software/ergonomische_dialoggestaltung/fehlertoleranz.htm (Zugriff: 19.11.2013)

Kienpointner Manfred, Prof. (2005): *Über den höflichen Umgang mit Menschen*, www.dieuniversitaet-online.at/dossiers/beitrag/news/uber-den-hoflichen-umgang-mit-menschen/367.html (Zugriff: 19.11.2013)

Hassenzahl Marc, (2004): *Funology: From Usability to Enjoyment, Chapter 3: The Thing and I: Understanding the relationship between user and product* <http://books.google.ch/books?id=Y8kBx3nZa9QC&lpg=PA30&dq=marc%20hassenzahl&hl=de&pg=PA30#v=onepage&q=marc%20hassenzahl&f=false> (Zugriff: 1.1.2014)

Universität Zürich, Methoden Beratung, *Datenanalyse*, <http://www.methodenberatung.uzh.ch/datenanalyse.html> (Zugriff: 3.1.2014)

Grote, Gudela, Prof., *Empirische Methoden für betriebliche Untersuchungen*, Skript zur Vorlesung ETH Zürich www.oat.ethz.ch/education/winter_05_06/material_empirical_methods/Skript_VL_Empirische_Methoden.pdf (Zugriff: 10.11.2013)

Studien

Pauwels Stefan L., Christian Hübscher, Stefan Leuthold, Javier A. Bargas-Avila, Klaus Opwis (2009): *Error prevention in online forms*: Use color instead of asterisks to mark required-fields, University of Basel, Faculty of Psychology, Department of Cognitive Psychology and Methodology, Basel, Switzerland. In *Interacting with Computers*, Band 21 Ausgabe 4, August 2009, S. 257-262, Elsevier Science Inc., New York, USA

Bargas-Avila Javier A., Glenn Oberholzer, Peter Schmutz, Marco de Vito, Klaus Opwis (2007): *Usable error message presentation in the World Wide Web: Do not show errors right away*, Department of Psychology, University of Basel, Switzerland. In *Interacting with Computers*, Band 19, Ausgabe 3, Mai 2007, S. 330-341, Elsevier Science Inc. New York, USA

Bargas-Avila Javier A., Sébastien Orsini, Hannah Piosczyk, Dominic Urwyler, Klaus Opwis (2011): *Enhancing online forms: Use format specifications for fields with format restrictions to help respondents*, University of Basel, Faculty of Psychology, Department of Cognitive Psychology and Methodology, Basel, Switzerland. In *Interacting with Computers*, Band 23, 2011, S. 33-39, Elsevier Science Inc., New York, USA

Seckler Mirjam, Alexandre N. Tuch, Klaus Opwis, Javier A. Bargas-Avila (2012): *User-friendly Locations of Error Messages in Web Forms: The right place is on the right side of the erroneous input field*, University of Basel, Faculty of Psychology, Department of Cognitive Psychology and Methodology, Basel, Switzerland. In *Interacting with Computers*, Band 24, 2012, S. 107-118, Elsevier Science Inc., New York, USA

Niculescu Andreea (2002): *Mensch-Maschine Kommunikation*, Ruhr-Universität Bochum, Fakultät für Elektrotechnik und Informationstechnik, Institut für Kommunikationsakustik, Fakultät für Philosophie und Publizistik, Sektion für Publizistik und Kommunikationswissenschaften, Bochum, Deutschland

Normen

Internationale Organisation für Normung (2010): *Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme, Deutsche Fassung* DIN EN ISO 9241-210:2010, Beuth, Berlin

Internationale Organisation für Normung (2010): *Grundsätze der Dialoggestaltung DIN EN ISO 9241-10: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten*, http://wiki.infowiss.net/Grundsätze_der_Dialoggestaltung (Zugriff: 19.11.2013)

16 ANHANG

A001 – QUIS Fragebogen

Nr.	Frage / Kapitel	von	bis
1	Systemerfahrung		
1.2	Wie oft verwenden Sie Online-Buchungssysteme für Hotels?		
3	Gesamteindruck		
3.1	Wie ist ihr Gesamteindruck	unangenehm	angenehm
3.2	Wie ist ihr Gesamteindruck	frustrierend	zufriedenstellend
3.3	Wie ist ihr Gesamteindruck	langweilig	anregend
3.4	Wie ist ihr Gesamteindruck	schwierig	leicht
3.6	Wie ist ihr Gesamteindruck	starr	flexibel
4	Bildschirm		
4.1	Die Zeichen auf dem Bildschirm waren	schwer erkennbar	leicht erkennbar
4.2	Die Hervorhebungen auf dem Bildschirm waren	nicht hilfreich	hilfreich
4.3	Die Bildschirmlayouts waren hilfreich	niemals	immer
4.3.1	Die Menge der auf dem Bildschirm dargestellten Informationen war	nicht ausreichend	ausreichend
4.3.2	Die Anordnung der Informationen auf dem Bildschirm war	unlogisch	logisch
5	Terminologie und System-Informationen		
5.1	Die einheitliche Ausdrucksweise im Formular war	inkonsistent	konsistent
5.1.2	Die aufgabenbezogene Ausdrucksweise war	inkonsistent	konsistent
5.2	Die Ausdrucksweise passte gut zu der auszuführenden Arbeit	niemals	immer
5.3	Die Fehlermeldungen waren	inkonsistent	konsistent
5.4	Die Fehlermeldungen waren	verwirrend	klar
5.4.2	Die Anweisungen zur Fehlerbehebung waren	verwirrend	klar
5.5	Der Computer informiert Sie über die laufende Aktion	niemals	immer
5.5.2	Das Ausführen einer Operation führte zu einem vorhersehbaren Ergebnis	niemals	immer
5.5.4	Die Wartezeiten zwischen den Operationen waren	inakzeptabel	akzeptabel
5.6	Die Fehlermeldungen waren	nicht hilfreich	hilfreich
5.6.1	Die Fehlermeldungen erklärten das Problem	niemals	immer
5.6.2	Die Formulierungen der Fehlermeldungen waren	unangenehm	angenehm
6	Lernfortschritt		
6.4	Die Aufgabe konnte auf einfache Weise gelöst werden	niemals	immer
6.4.2	Die Ausführungsschritte zum Lösen einer Aufgabe folgten einer logischen Reihenfolge	niemals	immer
6.4.3	Bei einer Reihenfolge abzuarbeitender Schritte gab das System Auskunft darüber, welche Schritte bereits abgearbeitet worden waren	unklar	klar

7	Systemeigenschaften		
7.4	Das Beheben von Fehler war	schwierig	leicht
7.4.1	Das Korrigieren von Tippfehlern war	komplex	einfach
7.5	Stimmt diese Aussage: je mehr Erfahrung Sie besitzen, desto leichter können Sie mit der Hotelbuchungssystem umgehen?	niemals	immer
9	Online Anleitung (Reiter rechter Rand "Hilfe zur Umfrage")		
9.1	Die Online Anleitung war	nicht hilfreich	hilfreich
9.1.1	Der Zugang zur Onlineanleitung war	schwierig	leicht
9.3	Der Inhalt der Anleitung war	nicht hilfreich	hilfreich
9.3.1	Informationen für spezielle Aspekte des Systems waren komplett und informativ.	niemals	immer
9.3.2	Die Informationen waren kurz und prägnant	niemals	immer
9.4	Die Lösung der Aufgabe war	schwierig	leicht
9.4.1	Die für die Aufgabenlösung gegebenen Anweisungen waren	verwirrend	klar
9.4.2	Die Zeit zum Lösen der Aufgabe war	nicht angemessen	angemessen
10	Multimedia Komponenten		
10.1	Der Einsatz von Bildmaterial hat den Ablauf	gestört	unterstützt
10.4	Die Farbverwendung war	unnatürlich	natürlich
10.4.1	Die Menge der Verwendbaren Farben war	nicht angemessen	angemessen
11	Welche Fehlermeldung spricht Sie persönlich am besten an?		
11.1	Damit wir Sie bei uns gebührend willkommen heissen können, benötigen wir Ihre Ankunftszeit	schlecht	gut
11.2	Sagen Sie uns doch, wann Sie anreisen, damit wir den Willkommens-Apéro kühl stellen können. Santé!	schlecht	gut
11.3	Bitte nennen Sie uns Ihre Ankunftszeit, damit wir Sie noch besser bedienen können.	schlecht	gut
11.4	Eingabe erforderlich	schlecht	gut
11.5	Bitte geben Sie die Ankunftszeit des Gastes an.	schlecht	gut

A002 - Fehlermeldungen Texte

Tonalität höflich

Gast, Vorname, Nachname

Wir möchten Sie gerne persönlich ansprechen. Bitte geben Sie Ihren Vornamen und den Nachnamen ein.

Ankunftszeit

- Damit wir Sie bei uns gebührend willkommen heißen können, benötigen wir Ihre Ankunftszeit.
- Leider können wir Ihre Ankunftszeit nicht richtig entziffern. Bitte geben Sie sie in folgendem Format ein: hh:mm

Anrede

Ob Doktor, Professor oder einfach Herr oder Frau: Es ist uns wichtig, Sie korrekt anzusprechen.

Vorname

Wir würden gerne Ihren Vornamen wissen. Bitte teilen Sie ihn uns mit.

Nachname

Ihr Nachname ist für die Reservierung sehr wichtig. Bitte teilen Sie uns mit wie wir Sie ansprechen dürfen.

Geburtstag

- Damit wir unseren Service bestmöglichst gestalten und Sie eventuell angenehm überraschen können, benötigen wir Ihr Geburtsdatum.
- Leider können wir Ihr Geburtsdatum nicht richtig entziffern. Bitte geben Sie es in folgendem Format ein: TT.MM.JJJJ

PLZ

An welche Postleitzahl dürfen wir Ihnen die Reservationsbestätigung senden?
Leider können wir Ihre Postleitzahl nicht richtig zuordnen. Die Formatierung zur Eingabe lautet: 9999

Ort

Wir würden gerne wissen, woher Sie sind. Bitte teilen Sie uns Ihren Wohnort mit.

Land

Aufgrund Ihrer Angabe können wir Ihr Herkunftsland leider nicht ermitteln. Bitte wählen Sie Ihr Land aus.

Telefon

- Damit wir Sie schnell und kurzfristig kontaktieren können, benötigen wir Ihre Telefonnummer.
- Bitte überprüfen Sie noch einmal Ihre Eingabe. Die Formatierung lautet: +41 44 1234567

Mobil

Bitte überprüfen Sie noch einmal Ihre Eingabe. Die Formatierung lautet: +41 79 1234567

Email Adresse

- Bitte teilen Sie uns Ihre Email Adresse mit, damit wir Sie schriftlich kontaktieren können.
- Bitte überprüfen Sie noch einmal Ihre Eingabe. Die Formatierung lautet: info@myhotel.ch

Email bestätigen

Die beiden Email Adressen sind nicht identisch. Würden Sie bitte Ihre Eingabe noch einmal überprüfen?

Kreditkarte

- Bitte teilen Sie uns mit, mit welcher Kreditkarte Sie bezahlen möchten.
- Der Kreditkartenservice Ihrer Karte ist zur Zeit nicht erreichbar. Bitte versuchen Sie es später noch einmal oder wählen Sie einen anderen Karten-Typ.

Kartennummer

Für die einfache Abrechnung benötigen wir Ihre Kreditkartennummer. Bitte teilen Sie uns diese mit.

Karteninhaber

Bitte teilen Sie uns mit, auf welchen Namen (Vor- und Nachname) die Kreditkarte lautet.

Gültig bis

Bitte teilen Sie uns das Ablaufdatum der Kreditkarte mit.

AGB akzeptieren

Um die Transaktion zu starten, bitten wir Sie die AGBs zu bestätigen.

Tonalität humorvoll

Gast, Name, Vorname

Auch wenn Sie inkognito unterwegs sind, wir bräuchten Ihren Namen und versprechen, die Angaben diskret zu behandeln.

Ankunftszeit

- Sagen Sie uns doch, wann Sie anreisen, damit wir den Willkommens-Apéro kühl stellen können. Santé!
- Da hat jemand an der Uhr gedreht. Bitte geben Sie Ihre Ankunftszeit in folgendem Format ein: hh:mm

Anrede

Herr? Frau? Vor dem ersten Prosecco können wir Sie unmöglich duzen. Bitte sagen Sie uns wie wir Sie anreden sollen.

Vorname

Vielleicht duzen wir uns nach dem Prosecco? Bitte sagen Sie uns doch schon mal Ihren Vornamen.

Nachname

Damit wir Sie anständig begrüßen können, benötigt die Dame an der Reception Ihren Nachnamen. Danach wird's lockerer.

Geburtsdatum

- An welchem Datum hatte Ihr Geburtstagskuchen noch keine Kerze drauf?
- An welchem Datum hatte Ihr Geburtstagskuchen noch keine Kerze drauf? Bitte in folgendem Format: TT.MM.JJJJ

Strasse

Von welcher Adresse aus fliehen Sie, um bei uns eine schöne Zeit zu verbringen?

PLZ

An welcher Postleitzahl sind Sie zu Hause, wenn Sie sich nicht bei uns zu Hause fühlen?

Ort

An welchem Ort sind Sie zu Hause, wenn Sie sich nicht bei uns zu Hause fühlen?

Land

Wir freuen uns auf Gäste aus allen Herren Länder. Aus welchem sind Sie?

Telefon

- Der heiße Draht zu Ihnen scheint kalt zu sein. Bitte überprüfen Sie Ihre Eingabe noch einmal.
- Sagen Sie uns doch den heißen Draht zu Ihnen in folgendem Format: +41 44 1234567

Mobil

Sorry, kein Anschluss unter dieser Nummer. Bitte geben Sie uns Ihre Handy-Nummer in folgendem Format: +41 79 1234567

Email

- Ohne E-Mail-Adresse geht heute gar nichts mehr. Sagen Sie uns Ihre?
- An diese Adresse geht keine Post. Geben Sie Ihre E-Mail-Adresse in folgendem Format ein: info@myhotel.ch

Email bestätigen

- Haben Sie mehrere Adressen? Bitte bestätigen Sie die zuerst eingegebene E-Mail-Adresse.
- Haben Sie mehrere Adressen? Bitte bestätigen Sie Ihre E-Mail-Adresse in folgendem Format: info@myhotel.ch

Kreditkarte

- Mit welcher Kreditkarte bezahlen Sie Ihr Vergnügen?
- Ihr Kreditkarten-Service ist momentan in den Ferien. Versuchen Sie es später nochmal oder wählen Sie einen anderen Karten-Typ.

Kartennummer

Auf welche Kreditkartennummer dürfen wir Ihr Vergnügen verrechnen?

Karteninhaber

Sagen Sie uns bitte wer das Vergnügen bezahlt.

Gültig bis

Bis wann ist Ihre Kreditkarte gültig?

AGB

Erst die AGB, dann das Vergnügen.

Tonalität motivierend

Vorname, Name

Bitte sagen Sie uns Ihren Namen, damit wir Sie persönlich begrüßen können.

Ankunftszeit

- Bitte nennen Sie uns Ihre Ankunftszeit, damit wir Sie noch besser bedienen können.
- Damit es schneller weiter geht, geben Sie Ihre Ankunftszeit bitte in diesem Format ein: hh:mm

Anrede

Wenn wir Ihre Anrede wissen, erhalten Sie von uns korrekte Korrespondenz.

Vorname

Sobald wir Ihren Vornamen haben, geht die Reservierung schnell voran.

Nachname

Für die persönliche Reservation des Zimmers brauchen wir Ihren Nachnamen.

Geburtstag

- Wenn wir Ihren Geburtstag wissen, können wir Sie überraschen.
- Damit wir Sie überraschen können, geben Sie Ihr Geburtsdatum bitte in folgendem Format ein: TT.MM.JJJJ

Strasse

Eine korrekte Strassenangabe ist wichtig für eine schnelle Abwicklung der Reservation.

PLZ

- Für die schnelle Abwicklung der Reservation brauchen wir Ihre Postleitzahl.
- Damit alles korrekt ist, hier das Format der PLZ: 9999

Ort

Für eine sichere Reservierung brauchen wir Ihren Wohnort.

Land

Aus welchem Land sind Sie? Das ist wichtig für eine sichere Reservierung.

Telefon

- Ihre Telefonnummer ist uns wichtig. Nur für alle Fälle.
- Damit es schnell voran geht, benutzen Sie bitte folgendes Format für die Telefonnummer: +41 44 1234567

Mobil

Damit es schnell voran geht, benutzen Sie bitte folgendes Format für Ihre Mobil-Nummer:
+41 79 1234567

Email

- Mit der Angabe Ihrer E-Mail-Adresse erhalten Sie schneller eine Bestätigung.
- Damit es schnell voran geht, benutzen Sie bitte folgendes Format bei Ihrer E-Mail-Adresse: info@myhotel.ch

Email bestätigen

- Beide E-Mail-Adressen müssen identisch sein. Dann geht's schnell weiter.
- Damit es schnell voran geht, bestätigen Sie bitte Ihre E-Mail-Adresse in folgendem Format: info@myhotel.ch

Kreditkarte

- Nur noch die Kreditkarte eingeben, und die Reservation ist schon fast gesichert.
- Ihr Kreditkarten-Service ist momentan nicht erreichbar. Um gleich jetzt zu reservieren, geben Sie bitte eine andere Kreditkarte ein.

Kartennummer

Nur noch die Kartennummer eingeben, und wir heissen Sie schon bald willkommen.

Karteninhaber

Bitte geben Sie den Namen des Karteninhabers ein. Ihre Reservation ist damit schon fast abgeschlossen.

Gültig bis

Jetzt fehlt nur noch das Gültigkeitsdatum Ihrer Kreditkarte, und Ihrem Aufenthalt bei uns steht nichts mehr im Weg.

AGB

Bitte bestätigen Sie noch unsere AGBs. Nur der guten Ordnung halber.

Tonalität standard und unpersönlich

Siehe Kapitel 8.3 FEHLERMELDUNGEN – DEFINITION UND VALIDIERUNG. Eine detaillierte Übersicht ist im Daten-Anhang auf dem Datenträger ersichtlich.

A003 – Projektplan

MAS HCID "Einfluss Fehlermeldungen auf UX"

29.01.2014

Tasks

2

Vorgang	Anfang	Ende	Fortschritt
Kickoff HSR	27.04.13	27.04.13	100
Coach Meeting1	29.05.13	29.05.13	100
Coach Meeting2	05.07.13	05.07.13	100
Zwischenreview	31.08.13	31.08.13	100
CoachMeeting3	16.10.13	16.10.13	100
CoachMeeting4	06.11.13	06.11.13	100
CoachMeeting5	27.11.13	27.11.13	100
Ferien Dorota	27.06.13	07.07.13	100
Ferien Teddy	17.08.13	31.08.13	100
Ferien Manuel	20.07.13	31.07.13	100
Weihnachtsferien	20.12.13	29.12.13	100
Research und Analyse Phase	01.05.13	13.06.13	100
Literatursuche und -studie	01.05.13	26.05.13	100
Hypothesen formulieren	27.05.13	03.06.13	100
Wissenschaftliche Studie	04.06.13	13.06.13	100
Vorgehens & Methodik	20.05.13	21.05.13	100
UX Fragenbogen	20.05.13	21.05.13	100
Design Phase	22.05.13	20.08.13	100
Test Szenario definieren	22.05.13	28.05.13	100
Ablauf von UX Test definieren	29.05.13	12.06.13	100
Wettbewerb & Empfehlungen Konzept	12.06.13	14.06.13	100
Test Inhalt ausarbeiten	18.06.13	31.07.13	100
Expert Review - Texter	01.08.13	20.08.13	100
Prototyping Phase	29.05.13	29.08.13	100
Tools & Frameworks evaluieren	29.05.13	10.06.13	100
Test Szenario aufbauen	11.06.13	13.08.13	100
Wettbewerb & Empfehlungen	14.06.13	19.06.13	100
Datenbank Design und Aufsetzung	19.06.13	28.06.13	100
Prototyp testen	16.08.13	22.08.13	100
Test Varianten erstellen	23.08.13	29.08.13	100
Durchführung Phase	28.08.13	28.01.14	100
Qualitative Test vorbereiten	28.08.13	31.08.13	100
Qualitative Test durchführen	12.09.13	02.11.13	100
UX Test1 @AdNovum	12.09.13	12.09.13	100
UX Test2 @AdNovum	24.09.13	24.09.13	100
UX Test3 @CSS	10.10.13	10.10.13	100
UX Test4 @UBS	14.10.13	14.10.13	100
UX Tests am Flughafen	02.11.13	02.11.13	100
Fazit: Auswertung und Korrekturen	03.11.13	07.11.13	100
Online Test - Email Versand	10.11.13	10.11.13	100

MAS HCID "Einfluss Fehlermeldungen auf UX"

29.01.2014

Tasks

3

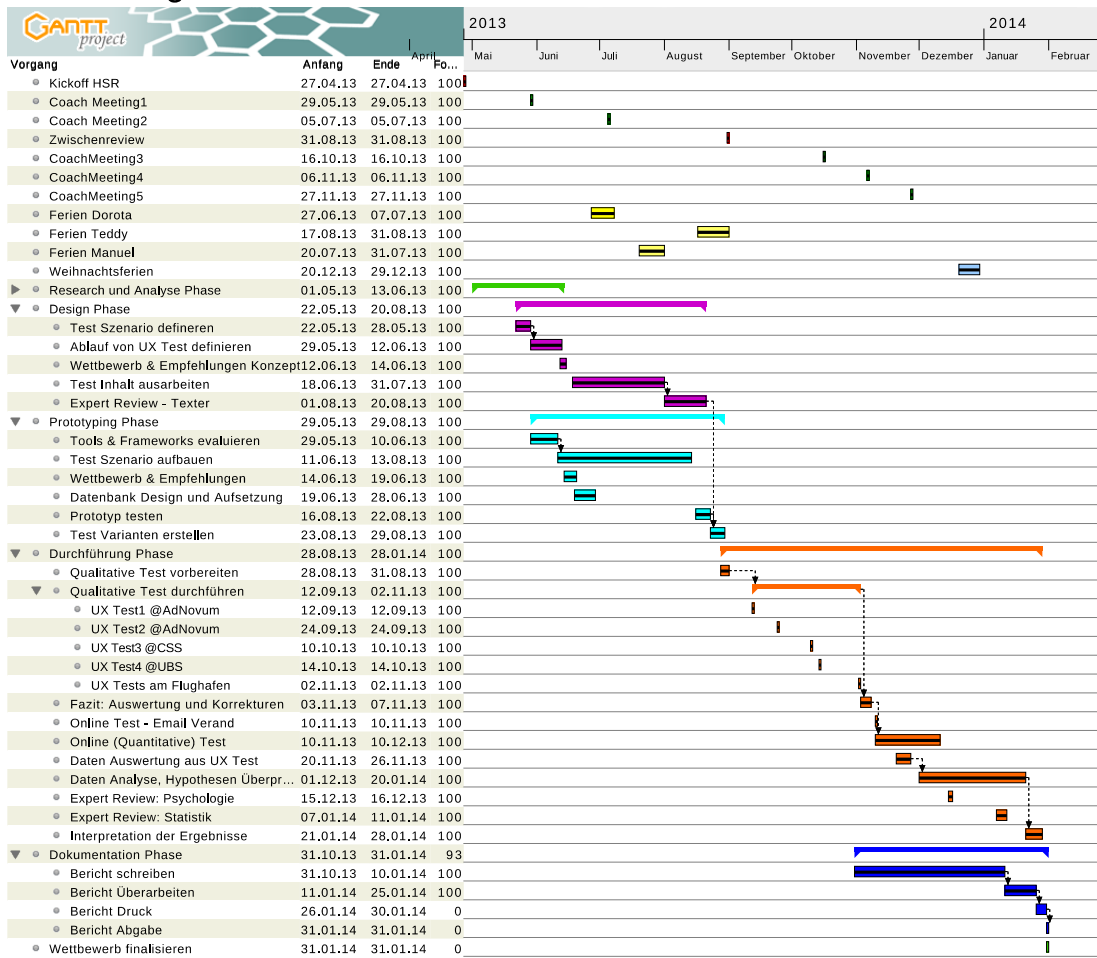
Vorgang	Anfang	Ende	Fortschritt
Online (Quantitative) Test	10.11.13	10.12.13	100
Daten Auswertung aus UX Test	20.11.13	26.11.13	100
Daten Analyse, Hypothesen Überprüfung	01.12.13	20.01.14	100
Expert Review: Psychologie	15.12.13	16.12.13	100
Expert Review: Statistik	07.01.14	11.01.14	100
Interpretation der Ergebnisse	21.01.14	28.01.14	100
Dokumentation Phase	31.10.13	31.01.14	93
Bericht schreiben	31.10.13	10.01.14	100
Bericht Überarbeiten	11.01.14	25.01.14	100
Bericht Druck	26.01.14	30.01.14	0
Bericht Abgabe	31.01.14	31.01.14	0
Wettbewerb finalisieren	31.01.14	31.01.14	0

MAS HCID "Einfluss Fehlermeldungen auf UX"

29.01.2014

Gantt-Diagramm

5



A003 – Risiko Liste

Risiko	Beschreibung	Auswirkung bzw. Schaden	Schaden klein=1gross=6 (A)	Wahrscheinlichkeit klein=1, gross=6(B)	Risiko Impact (AxB) = Prio	Massnahmen	Verantwortung	Erfasst am	Letzte Überprüfung
Know-How Aufbau	Recherche und Einarbeitung in die relevante Themengebiete: UX, Statistik, Fehlermeldungen brauchen viel Zeit	Fehlendes Wissen kann zu falschen Annahmen und Ergebnissen Führen	5	5	25	Selbststudium, Recherchen, Analyse von Beispielen und Best Practices, Expert Reviews	alle	01.05.2013	20.01.2014
Online Umfrage dauert zu lange	Fragebogen ist lang und einige Fragen sehr ähnlich oder für die Studie irrelevant	Hohe Abbruchrate und zu wenig valide Daten	6	5	30	Bei Peer Review MAS HCID die Erfahrungen und Meinungen in der Klasse abholen.. QUIS Fragebogen kurzen.	Manuel, Dorota	20.06.2013	16.08.2013
Zu wenig Testpersonen nehmen an der online Umfrage teil	Rekrutierung von Testpersonen findet im privaten Umfeld, dadurch ist die Menge beschränkt und eventuell nicht genügend für die Studie	Zu wenig Daten für plausible Auswertungen	6	6	36	Wettbewerb Konzept als Motivation und Belohnung, Empfehlungen, Datenwachstum während der Laufzeit der Studie überwachen, Link zur Umfrage auf Socialmedia Plattformen publizieren. Laufzeit der Studie verlängern	alle	20.05.2013	20.11.2013
Daten Verlust	Datenmanipulationen oder Programmfehler können Daten verfälschen, löschen	Studie muss abgebrochen werden	6	6	36	DB Kopieren erstellen	Dorota	30.06.2013	25.12.2013
Code Verlust (DB,GUI)	Eine technische Panne bei dem Service Provider MetaNet Plesk	Die Entwicklung Umgebung muss neu aufgebaut werden. Extra Aufwand und Zeitverlust	5	3	15	Backup von DB Skripten und Frontend Code erstellen. Es soll möglich sein DB Strukturen und GUI Code auf anderen Server zu installieren	Dorota, Manuel	15.05.2013	01.11.2013
Verlust von Ergebnissen, Dokumenten	Die verlorene Files müssen rekonstruiert werden	Extra Aufwand und Zeitverlust	5	5	25	Online arbeiten, lokale Kopien von Google-Docs erstellen, DropBox Backup nutzen	alle	15.06.2013	29.01.2013
Statistische Methoden bei Daten Analyse ergeben keine signifikante Unterschiede über gesamte Population	Die Suche nach signifikanten Unterschieden kann zu lange dauern.	Extra Aufwand und Zeitverlust bei Auswertungen	3	5	15	Population anders Gruppieren und nach signifikanten Unterschieden zwischen den kleinen Gruppen suchen. Auf die Signifikanz-Prüfung und statistischen Methoden verzichten.	Manuel Dorota	10.11.2013	20.01.2013
Statistische Methoden bei Daten Auswertung passen nicht	Auf Grund von speziellen Daten Konstellationen andere, passende Methoden angewenden	Mehr Zeitaufwand bei der Suche nach passenden statistischen Methoden	3	5	15	Expert Review, Statistik Beratung (ETH) Statistische Auswertungen unterlassen.	Manuel, Dorota	30.11.2013	10.01.2014