

Vergleich zweier Navigierungstypen für Nutzerumfragen auf mobilen Endgeräten

Alexander Barbie
Christian-Albrechts-Universität
zu Kiel
Institut für Informatik
24098 Kiel, Deutschland
abar@informatik.uni-
kiel.de

Tilman Beck
Christian-Albrechts-Universität
zu Kiel
Institut für Informatik
24098 Kiel, Deutschland
stu127568@informatik.uni-
kiel.de

Fabian Fröhlich
Christian-Albrechts-Universität
zu Kiel
Institut für Informatik
24098 Kiel, Deutschland
faf@informatik.uni-kiel.de

ABSTRACT

Diese Arbeit präsentiert eine Nutzerstudie zu unterschiedlichen Navigationsarten bei Umfrage-Apps auf mobilen Endgeräten. Ziel ist es eine möglichst optimale Navigationsart für eine Umfrage-Applikation zu finden. Dafür wird zwischen zwei Darstellungsarten der Fragen innerhalb der Applikation unterschieden. Erstere stellt alle Fragen untereinander auf einer ununterbrochenen Seite dar und der Nutzer navigiert durch Scrollen. In der zweiten Darstellungsart wird jede Frage separat auf einer Seite dargestellt und Nutzer navigieren über Buttons. Insgesamt haben 42 Probanden an der Nutzerstudie teilgenommen. Alle Probanden wurden zufällig auf 4 unterschiedliche Szenarien verteilt. Für die unterschiedlichen Navigationsarten konnte kein signifikanter Unterschied festgestellt werden.

Keywords

mobile Nutzerstudien, Webapplikation, Scroll-Navigation, Button-Navigation, Design von Nutzerstudien, Bearbeitungszeit, Nutzerzufriedenheit

1. EINLEITUNG

Nutzerstudien sind ein bewährtes Mittel um qualitative und quantitative Daten zu einem Untersuchungsgegenstand zu ermitteln. Diese Studien werden oft mit Stift und Papier, aber auch zunehmend an Computern durchgeführt. Mit dem unaufhaltsamen Einzug mobiler Endgeräte in den privaten Alltag bietet sich die Möglichkeit solche Nutzerstudien auch auf diesen Geräten durchzuführen. Dadurch müssen aber viele neue Faktoren (z.B. unterschiedliche Eingabegestiken, verschiedene Displaygrößen, wechselnde Umgebungen, etc.) in Betracht gezogen werden. Es stellt sich die Frage, inwiefern das Ergebnis einer Studie durch die Unterschiede im Umgang mit den jeweiligen Geräten beeinflusst wird. Ein Problem hierbei ist, dass es nicht eindeutig ist wie die Inhalte einer Studie auf einer oder mehreren Seite eines mo-

bilen Gerätes dargestellt werden müssen, um eine schnelle und übersichtliche Bearbeitung für die Benutzer zu ermöglichen. Ziel dieser Arbeit ist es einen Aspekt diesbezüglich zu prüfen und gegebenenfalls eine Designempfehlung für die Entwicklung von mobilen Nutzerstudien zu geben.

Innerhalb dieser Nutzerstudie werden zwei Darstellungsarten miteinander verglichen. Die Erste stellt alle Fragen untereinander auf einer ununterbrochenen Seite dar und der Nutzer navigiert durch Scrollen. In der zweiten Darstellungsart wird jede Frage separat auf einer Seite dargestellt und der Nutzer navigiert über Buttons. Es wurde untersucht, ob es einen Unterschied zwischen der Bearbeitungszeit des Fragebogens in den beiden Ausprägungen gibt.

Die Ergebnisse der Evaluation zeigen keine signifikanten Unterschiede bezüglich der Durchführungszeit zwischen den beiden Designs der Nutzerstudie. Auch hinsichtlich der Effizienz konnten keine deutlichen Unterschiede beider Designs gefunden werden. Die Daten der Nutzerzufriedenheit lassen offen, ob es ein 'besseres' Design gibt, da beide Probandengruppen im Durchschnitt gut zurecht gekommen sind.

In Kapitel 2 wird der Designaufbau der Nutzerstudie vorgestellt. Darauf aufbauend wird in Kapitel 3 die Methodik erläutert, womit die in Kapitel 4 präsentierten Ergebnisse untersucht werden. Diese Ergebnisse werden in Kapitel 5 diskutiert und in Kapitel 6 mit anderen Arbeiten zu diesem Thema in Verbindung gebracht. Kapitel 7 fasst die Ergebnisse und Diskussion dieser Arbeit zu einem Fazit zusammen.

2. HINTERGRUND UND SZENARIO

Diese Nutzerstudie untersucht eine möglichst optimale Navigationsart für eine Umfrage-App. Für die Nutzerstudie wurde eine Webapplikation entwickelt, die in Kapitel 2.2 vorgestellt wird. Jeder Proband mit Smartphone kann entweder selbst an der Studie teilnehmen oder eines der bereitgestellten Geräte nutzen.

Um den Probanden das Gefühl einer alltäglichen Nutzerumfrage zu vermitteln, wurden zwei unterschiedlich lange Fragebögen entworfen. Der kürzere hat als Thema "Fast Food", der längere stellt Fragen zum Thema "Politik". Die unterschiedlichen Szenarien werden im folgenden Kapitel 2.1 näher erklärt.

2.1 Szenarien

Die Webapplikation unterscheidet zwischen zwei Darstellungen von Fragebögen. Der erste Typ von Fragebögen besteht aus genau einer Seite, welche alle Fragen untereinander

der in aufsteigender Reihenfolge enthält. (Single Page, Multi-Questions - im Folgenden *MQ* genannt). Folglich muss der Proband scrollen, um den ganzen Fragebogen zu beantworten. Bei dem zweiten Typ von Fragebögen wird pro Frage eigene Seite angezeigt und der Proband kann über *Weiter* bzw. *Zurück*-Buttons durch den Fragebogen navigieren (Single Page, Single Question - im Folgenden *SQ* genannt). Die Nutzerstudie besteht aus vier Szenarien. In jedem dieser Szenarien füllen die Probanden zwei Fragebögen aus. Die beiden Fragebögen unterscheiden sich sowohl im Thema als auch in der Länge. Fragebogen 1 behandelt dabei das Thema "Fast-Food" und umfasst 20 Fragen und Fragebogen 2 behandelt in 40 Fragen das Thema "Politik". Die Szenarien unterscheiden sich also in der Darstellungsart (*SQ/MQ*) und ob die Probanden zuerst Fragenkatalog 1 oder Fragenkatalog 2 ausfüllen.

1. Szenario: *MQ* mit Fragenkatalog 1 zu Anfang
2. Szenario: *MQ* mit Fragenkatalog 2 zu Anfang
3. Szenario: *SQ* mit Fragenkatalog 1 zu Anfang
4. Szenario: *SQ* mit Fragenkatalog 2 zu Anfang

2.2 Die Applikation

Für die Nutzerstudie wurde eine Client-basierte Webapplikation mithilfe des Polymer-Frameworks¹ von Google entwickelt. Ziel war es eine möglichst plattformunabhängige Webapplikation zu nutzen, um viele Probanden erreichen zu können und eine unkomplizierte Durchführung an Ort und Stelle zu ermöglichen.

Für die Nutzung innerhalb der Studie wurde die Anwendung auf einem Webserver installiert und öffentlich zugänglich gemacht. Somit konnte jeder Proband mit einem Smartphone und Internetzugang über den Webbrowser auf die Anwendung zugreifen. Abhängig von einer ID, welche den Probanden zu Beginn mitgeteilt wird, wird eine der verschiedenen Szenarien gewählt.

Mit dem Abschluss der Nutzerumfrage auf dem Smartphone werden die gesammelten Daten auf einen Webserver übertragen. Dieser führt eine kurze Validierung durch und speichert die Daten in einem Format ab, welches anschließend von einer Statistik-Software (z.B. SPSS² oder R³) gelesen werden kann. Es wird angemerkt, dass eine bestehende Internetverbindung zur Übertragung des Datensatzes notwendig ist.

2.3 Methodologie

Um unterschiedliche Daten messen zu können und Unterschiede festzustellen, bearbeitet jeder Nutzer zwei Fragebögen zu unterschiedlichen Themenbereichen. Die Antworten der Probanden spielen hier eine untergeordnete Rolle, vielmehr geht es darum die Probanden in die Situation einer realen Umfrage zu versetzen. Ausschlaggebend sind die Antwortzeiten und ob Fragen unbewusst ausgelassen worden sind. Die Fragen sollen möglichst keine Emotionen auslösen, sondern stets den gleichen Gemütszustand bei den Probanden hervorrufen. Beide Fragebögen werden direkt hintereinander ausgeführt. Die Reihenfolge wird durch die Szenarien festgelegt.

¹<https://www.polymer-project.org/1.0/>

²<http://www.ibm.com/analytics/us/en/technology/spss/>

³<https://www.r-project.org/>

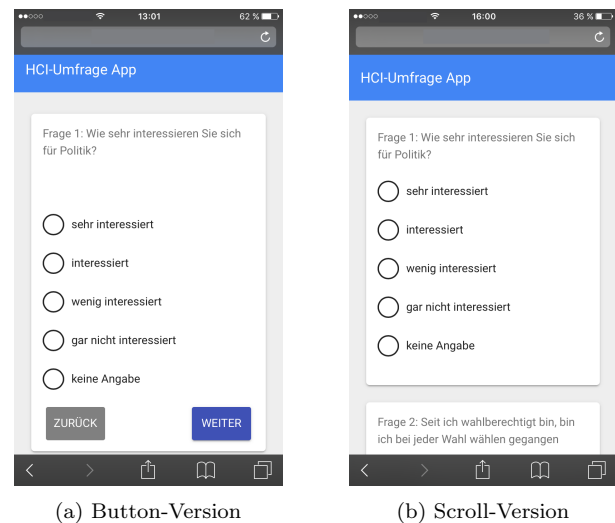


Abbildung 1: Die Darstellung der Webapplikation auf einem iPhone 6S.

Eine Gleichverteilung an weiblichen und männlichen Probanden für die verschiedenen Szenarien wird angestrebt. Vor Beginn der Studie werden die Probanden den Szenarien zugeteilt. Statt eines Würfels werden IDs von 1000-1999 gezogen und *modulo 4* gerechnet. Die IDs werden nacheinander in aufsteigender Reihenfolge vergeben. Am Ende der Nutzerstudie sollen alle Szenarien in gleicher Häufigkeit vorhanden sein.

Gemessen werden die Daten in der Webapplikation. Jede Interaktion mit den Fragebögen speichert einen Zeitstempel. Dieser Zeitstempel wird zusammen mit der Antwort des Probanden gespeichert. Sollte eine Frage nicht beantwortet werden, wird auch das gespeichert und später ausgewertet. Bevor die Nutzerstudie startet, wird eine Pilotstudie durchgeführt. Dafür werden drei Personen im Alter der Zielgruppe gesucht. Diese sollen vorab die Webapplikation testen.

2.4 Hypothesen

In dieser Nutzerstudie werden vier Aspekte untersucht, die für die Nutzung der Webapplikation interessant erscheinen. Untersucht wird der Unterschied der Gesamtdurchführungsdauer der Fragebögen im Stil *SQ* und *MQ*. Weiterhin werden die beiden kurzen und längeren Fragebögen im Stil *SQ* und *MQ* auf die Durchführungszeit miteinander verglichen. Als Maß für der Effizienz wird die Anzahl der ausgelassenen Fragen zwischen den beiden Designs verglichen. Zuletzt wird mithilfe eines Papierfragebogens die wahrgenommene Zufriedenheit der Probanden überprüft. Hierbei ist von Interesse, ob die Probanden sich im Umgang mit der Navigation wohlfühlt haben oder mit einer alternativen Navigation den Fragebogen schneller bearbeiten hätten können. Im Folgenden sind die zu untersuchenden Hypothesen aufgelistet:

- H_{10} : es gibt keinen Unterschied zwischen der Bearbeitungszeit des Fragebogens im Stil *SQ* und des Fragebogens im Stil *MQ*.
- H_{11} : die Bearbeitungszeit des Fragebogens im Stil *SQ* unterscheidet sich zur Bearbeitungszeit des Fragebogens im Stil *MQ*.

- H_{20} : eine größere Bearbeitungszeit des Fragebogens wirkt sich nicht auf den in H_1 beschriebenen Unterschied aus.
- H_{21} : je größer die Bearbeitungszeit des Fragebogens, desto deutlicher wird der in H_1 beschriebene Unterschied.
- H_{30} : die Wahrscheinlichkeit, dass Fragen im Stil MQ nicht beantwortet/übersehen werden ist nicht höher/gleich als im Stil SQ.
- H_{31} : die Wahrscheinlichkeit, dass Fragen im Stil MQ nicht beantwortet/übersehen werden ist höher als im Stil SQ.
- H_{40} : die wahrgenommene Zufriedenheit unterscheidet sich nicht zwischen SQ und MQ.
- H_{41} : die wahrgenommene Zufriedenheit unterscheidet sich zwischen SQ und MQ.

2.5 Pilotstudie

Es wurde eine Pilotstudie mit drei Personen durchgeführt. Die Personen gaben sowohl zur technischen Umsetzung als auch zum Inhalt der Umfrage Rückmeldungen.

Bei der Pilotstudie zeigte sich, dass die Applikation auf einigen Geräten nicht richtig dargestellt wurde oder Fehler aufgetaucht sind. Eine Person berichtete, dass im nativen WindowsPhone-Browser Eingabefelder so angezeigt wurden, dass keine Eingabe möglich war. Auf dem mobilen Opera-Browser wurde die Webapplikation gar nicht angezeigt. In einer anschließenden Untersuchung wurde festgestellt, dass die Applikation nur auf neueren Smartphones (Veröffentlichung nicht älter als drei Jahre) und den letzten beiden aktuellen Versionen der Browser (Safari 7+, IE 11+, Chrome, Firefox, Edge) fehlerfrei funktioniert. In der Darstellung gab es später auch Überlappungen bei Text- und Eingabefeldern. Dies ist der Displaygröße der Geräte geschuldet.

Auf Grund dieser Probleme, wurden Probanden gesucht, deren mobile Geräte diese Bedingungen erfüllten (aktuelles Smartphone mit ausreichender Displaygröße und aktueller Browserversion). Alternativ haben wir Probanden ohne passendes Smartphone ein Gerät gegeben.

Die Pilotprobanden gaben zudem an, dass einige Fragen uneindeutig gestellt sind oder die Antwortmöglichkeiten nicht zur Fragestellung passen. Die Fragen wurden daraufhin überarbeitet.

3. EVALUIERUNGSMETHODIK

Das folgende Kapitel beschreibt die Methodik wie die Daten gemessen werden, um sie anschließend auswerten zu können.

3.1 Verfahren

Das Experiment ist als Between-Group-Design entworfen worden. Es werden zwei unabhängige Variablen, die Navigation per Scrollen und die Navigation per Button, untersucht.

3.2 Probanden

Die Durchführung der Nutzerumfrage wurde in der Mensa I der Universität Kiel durchgeführt. Es wurden beliebig Gruppen oder einzelne Studenten gebeten, an der Studie teilzunehmen. Nachdem das Thema und der Ablauf erläutert

wurden, haben die Probanden eine Einverständniserklärung unterzeichnet. Anschließend bekamen die Probanden eine ID zur Durchführung der Studie mitgeteilt und falls benötigt ein Smartphone für die Dauer der Teilnahme. Nach dem Abschluss der Umfrage auf dem Gerät, wurde den Probanden ein schriftlicher Umfragebogen gegeben, welcher Fragen zu persönlichen Daten (Alter, Geschlecht, Studiengang, etc.), zur Erfahrung mit dem Smartphone und zur Benutzerzufriedenheit stellte. Nach dem Ausfüllen durften sich die Probanden als Dankeschön eine Süßigkeit oder Obst aussuchen. Durch die Trennung der Einverständniserklärung und des zweiten schriftlichen Fragebogens wurde die Anonymität der Datensätze sichergestellt. Dennoch konnten die Probanden später der Verwendung ihrer Daten widersprechen, indem sie ihre ID mit einer entsprechenden Nachricht an die Kontaktadresse schicken.

Die Zielgruppe für Umfragen auf mobilen Endgeräten sind Personen, die ein Smartphone besitzen oder Erfahrung im Umgang damit haben und regelmäßig an Nutzerstudien teilnehmen. Die Universitätsstudenten stellen unter anderem die Zielgruppe für universitäts-interne Umfragen und Nutzerstudien dar, so dass es Sinn ergibt, sich Probanden aus dieser Zielgruppe zu suchen.

Laut einer Umfrage aus dem Jahr 2014 liegt der Anteil der Smartphone-Nutzer bei den 14 bis 29-Jährigen bei 78%⁴. Da das Durchschnittsalter der Studenten [6] innerhalb dieser Altersspanne liegt, kann davon ausgegangen werden, dass die Probanden für diese Nutzerstudie durch die regelmäßige Benutzung Erfahrung im Umgang mit einem Smartphone haben. Somit kann die Wahrscheinlichkeit einer Verfälschung der Daten durch Probleme mit der Bedienung verringert werden.

Insgesamt wurden im Rahmen der Studie 42 Probanden befragt ($N = 42$), davon 23 Frauen und 19 Männer. Das Durchschnittsalter der Probanden liegt bei 24,17 Jahren ($SD = 3,07$). 39 der 42 Probanden haben angegeben im Besitz eines Smartphones zu sein. Zudem wurde von den Probanden die Selbsteinschätzung zur Erfahrung mit dem Smartphone erfragt (Frage 1: *"Ich bin ein erfahrener Smartphone-Nutzer"*, Frage 2: *"Ich fühle mich sicher im Umgang mit dem Smartphone"*). Als Antwortmöglichkeiten gab es Werte innerhalb einer 5-Likert-Skala von 1 (*"trifft zu"*) bis 5 (*"trifft nicht zu"*) oder *"keine Angabe"* wählen. Hierbei haben die Probanden im Besitz eines Smartphones ($N = 39$) eine hohe Erfahrung (Frage 1: $M = 1,72$ und $SD = 0,92$; Frage 2: $M = 1,64$ und $SD = 0,90$), während die Probanden ohne Smartphone weniger bis durchschnittlich viel Erfahrung (Frage 1: $M = 5,00$ und $SD = 0,00$; Frage 2: $M = 3,00$ und $SD = 0,00$) angaben.

In Abbildung 2 ist die Verteilung der mobilen Endgeräte über die Probanden zu sehen.

3.3 Aufgaben

Alle Probanden mussten jeweils zwei Fragebögen direkt nacheinander ausfüllen. In welcher Reihenfolge diese Fragebögen bearbeitet werden sollten, hat die gezogene ID festgelegt.

Vor Beginn der Umfrage wurden die Probanden über ihre Aufgaben aufgeklärt, jedoch ohne auf das Untersuchungsziel dieser Nutzerstudie explizit hinzuweisen. Wichtig war an der Stelle, dass den Probanden gesagt wurde, dass sie

⁴<https://www.bitkom.org/Presse/Presseinformation/Smartphones-staerker-verbreitet-als-normale-Handys.html>

	gesamt	Probanden männlich	weiblich
1. Szenario	10	5	5
2. Szenario	10	4	6
3. Szenario	11	6	5
4. Szenario	11	4	7

Tabelle 1: Anzahl der Probanden pro Szenario

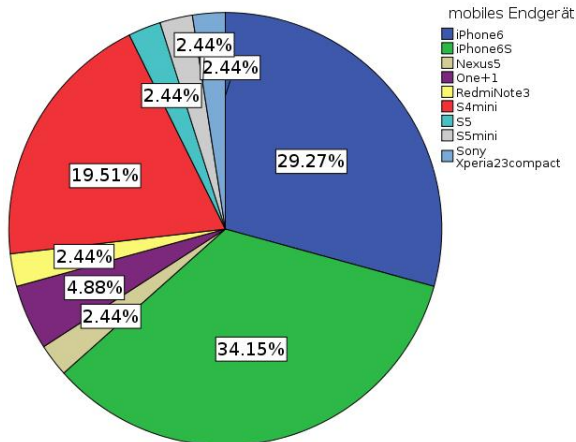


Abbildung 2: Verteilung der mobilen Endgeräte auf die Probanden

möglichst alle Fragen beantworten sollen. Im Anschluss an die Umfrage, wurden die Probanden noch mittels Papierfragebogen mit sechs Fragen zu ihrer Zufriedenheit mit der Webapplikation befragt.

3.4 Vorbereitung

Es wurde darauf geachtet, dass jedes Experiment gleich abläuft. Dazu wurde folgende Checkliste erstellt, welche den Ablauf beschreibt:

- Begrüßung und Bitte um Zeit für eine kurze Studienteilnahme
- kurze Einführung in das Thema
- Erklärung des Ablaufs (Einverständniserklärung unterschreiben, ID eingeben, Fragebogen vollständig und gewissenhaft ausfüllen, Benutzerzufriedenheit und persönliche Daten auf Papierfragebogen ausfüllen, Geschenk erhalten)
- offene Fragen klären
- Smartphone des Probanden auf Tauglichkeit testen (alternativ: eigenes Gerät übergeben)
- ID mitteilen
- Durchführung der Umfrage auf dem Gerät
- Benutzerzufriedenheitsbogen ausfüllen lassen
- bedanken und Geschenk überreichen

3.5 Messwerte

Die Daten wurden auf zwei unterschiedlichen Wegen erhoben. Zum einen werden die Daten zur Durchführungszeit und zur Anzahl der ausgelassenen Fragen über die Webapplikation gemessen. Die persönlichen Daten und die Daten zur Zufriedenheit werden über einen separaten, zweiseitigen Fragebogen auf Papier erfasst. Jeder Proband bekommt einen Papierbogen mit seiner ID, so dass später beide Datensätze über die ID miteinander verknüpft werden können. Die beiden Varianten der Datenerfassung werden im Folgenden detaillierter erläutert. Vor Beginn des Experiments haben die Probanden den Bedingungen der Einverständniserklärung schriftlich zugestimmt.

3.5.1 Webapplikation

Um Daten für die Auswertung der Effektivität und Effizienz zu erhalten, zeichnet die Webapplikation während des Experiments mehrere Daten auf. Zu Beginn wird die ID des Probanden gespeichert und basierend darauf eines der vier in Kapitel 2.1 beschriebenen Szenarios ausgewählt. Anschließend wird beim Klicken der *Start-Buttons* (*StartSurvey*) als auch der *Abschließen-Buttons* (*EndSurvey*) ein Zeitstempel gespeichert. Während eines Fragebogens wird jede Auswahl einer Antwort als Zeitstempel zu der entsprechenden Frage gespeichert. Zu jeder Frage wird auch der Typ der Frage gespeichert (Mehrfachauswahl oder einfache Wahl). Für den Design-Typ *SQ* wird zusätzlich noch jeder *Weiter-Button* bzw. *Zurück-Button*-Klick aufgezeichnet. Für den Design-Typ *MQ* können zur Navigation zwischen den Fragen keine sinnvollen Daten abgespeichert werden. Gespeichert wird eine Liste, wobei jeder Eintrag einer Frage entspricht. Dieser Eintrag speichert die Art der Antwortmöglichkeiten und die abgegebenen Antworten. Falls ein Eintrag in dieser Liste leer ist, wurde die Frage ausgelassen. Alle Daten werden als JSON-String an den Server geschickt und dort gespeichert. Ein PHP-Script transformiert die Datensätze in Comma-Separated-Values (CSV), um diese später leichter in statistische Auswertungsprogramme importieren zu können.

Auch wenn die Webapplikation Messpunkte bei jeder Interaktion enthält, beschränkt sich die Auswertung der Daten auf zwei entscheidende Messpunkte. Im folgenden werden diese Messpunkte *StartSurvey* und *EndSurvey* genannt. Wie der Name schon zeigt, speichert *StartSurvey* einen Zeitstempel direkt bei Beginn jedes Fragebogens. *EndSurvey* speichert den Zeitstempel nach dem Beenden jedes Fragebogens. Die Gesamtdauer eines Fragebogens ist also gegeben durch

$$durQuestionair = EndSurvey - StartSurvey$$

Die Dauer der gesamten Ausführungszeit ist dann die Addition beider Ausführungszeiten der einzelnen Fragebögen, also

$$durTotal = durShort + durLong$$

Die Anzahl der ausgelassenen Fragen je Fragebogen und die Summe beider werden gespeichert,

$$NullGes = Null1 + Null2$$

3.5.2 Fragebogen

Der zweiseitige Fragebogen erfasst persönliche Daten und Daten zur Zufriedenheit mit der Webapplikation. Die erste Seite liefert Daten zum Alter, Geschlecht, Studiengang

und ob der Proband im Besitz eines Smartphones ist. Zudem werden zwei Angaben zur Selbsteinschätzung hinsichtlich der Erfahrung mit dem Smartphone mithilfe einer 5-Punkt-Likert-Skala von 1 ("trifft zu") bis 5 ("trifft nicht zu") oder alternativ "keine Angabe" erhoben. Außerdem wurde notiert mit welchem Gerät der Proband die Studie durchgeführt hat.

Die zweite Seite erhebt Daten zur Nutzerzufriedenheit. Die Items orientieren sich an einem Iso-Metrics Fragebogen des Fachbereichs Psychologie der Universität Osnabrück [3] und gewährleisten damit die Validität des Fragebogens. Die Items wurden teilweise den für diese Nutzerstudie geforderten Konditionen entsprechend angepasst.

1. zufr1: Die Umfrage-App zwingt mich, überflüssige Eingaben durchzuführen
2. zufr2: Es müssen zu viele Eingaben für die Beantwortung der Fragen durchgeführt werden
3. zufr3: Die Button/Scroll-Navigation der Umfrage-App bietet mir gute Bedienmöglichkeiten, um mich im Fragekatalog zu bewegen.
4. zufr4: Ich bin mit der Umfrage-App insgesamt zufrieden
5. zufr5: Eine andere Navigationsart (z.B. Swipen, Scrollen/Buttons) würde mich beim Ausfüllen des Fragebogens besser unterstützen.
6. zufr6: Ich hatte beim Ausfüllen des Fragebogens technische Probleme (z.B. Ruckler, falsche Erkennung von Gesten auf dem Touchscreen, lange Ladezeiten)

Dieser Datensatz wird manuell in ein CSV-Format übertragen und an den Server gesendet. Ein weiteres PHP-Skript verbindet beide Datensätze zu einem großen Datensatz der alle Daten enthält.

Die Anpassung an das Treatment erfolgt in Items *zufr3* und *zufr5*. Die Items *zufr3* und *zufr4* sind umgepolt gestellt. Die Items *Zufr1* bis *Zufr6* (*zufrSum*) bilden einen Summenscore für die Zufriedenheit. Zusätzlich werden die Items *Zufr1*, *Zufr2*, *Zufr4-pol* (*zufrSum124*) zum Summenscore aufaddiert.

Die Objektivität des Messergebnisses wird dadurch gewährleistet, dass jeder Versuchsleiter eine Checkliste (vgl. Kapitel 3.4) abarbeitet, nach welcher die Probanden akquiriert und durch den Versuch begleitet werden.

Die Reliabilität des Fragebogens wird durch die gleichbleibende Umgebung der Probanden gewährleistet. Diese werden am Nachmittag innerhalb von zwei Wochen in der Mensa I der Universität Kiel befragt.

All diese Messgrößen werden in Kapitel 4 untersucht. Tabelle 2 gibt einen Überblick über die eingeführten Kennzahlen.

4. ERGEBNISSE DER EVALUIERUNG

In diesem Kapitel werden die gemessenen Daten ausgewertet. Welche Daten erhoben wurden und wie, wird nochmal detailliert in Kapitel 3 beschrieben. Die zu untersuchenden Hypothesen sind in Kapitel 2.4 dargestellt. Im Folgenden wird jede der Hypothesen überprüft. Anschließend werden die Ergebnisse im darauffolgenden Kapitel 5 diskutiert.

4.1 Vorbemerkungen

Bei der Überprüfung der Messergebnisse ist aufgefallen, dass für einen Probanden (ID = 1715) besonders hohe Werte für die Durchführungszeit (438,46 für *durTotal*) des Fragebogens gemessen wurden. Während der Durchführung mit dem Probanden wurde notiert, dass besonders viele Fragen gestellt wurden. Daher wurde dieser Datensatz von der weiteren Betrachtung ausgeschlossen. Für die Evaluation werden von nun an 41 Probanden ($N = 41$, für $n_{SQ} = 21$, für $n_{MQ} = 20$) betrachtet.

Des Weiteren wurden im Benutzerzufriedenheitsbogen eine Aussage bezüglich technischer Probleme überprüft ("Ich hatte beim Ausfüllen des Fragebogens technische Probleme. (z.B. Ruckler, falsche Erkennung von Gesten auf dem Touchscreen, lange Ladezeiten)"). Bei der Evaluierung ist aufgefallen, dass acht Probanden technische Probleme festgestellt haben. Insgesamt ergab sich für diese Frage ein Mittelwert von 4,48 ($SD = 1,23$) über alle Probanden. Diese Frage wurde nicht in die Evaluierung der Zufriedenheit mit einbezogen, da aus der Antwort keine direkten Rückschlüsse auf die Zufriedenheit der Probanden mit der Navigation gezogen werden können. In der Diskussion 5 wird nochmal näher auf dieses Problem eingegangen.

4.2 Effektivität

Eine der Hauptuntersuchungsgegenstände dieser Nutzerstudie ist die Effektivität der Webapplikation, also welche Navigation sich positiv auf die Bearbeitungsdauer auswirkt.

Ergebnisse für beide Fragebögen.

Um einen *t*-Test für unabhängige Stichproben durchführen zu können, sind folgende Voraussetzungen notwendig:

- das untersuchte Merkmal ist in den Grundgesamtheiten der beiden Gruppen normalverteilt
- Homogenität der Varianzen: Die Gruppen kommen aus Grundgesamtheiten mit annähernd identischer Varianz

Zunächst werden die beiden Gruppen auf Normalverteilung geprüft. Da die Anzahl der Probanden für beide Gruppen kleiner 50 ($n_{SQ} = 20$, $n_{MQ} = 21$) ist, wird der Shapiro-Wilk-Test zur Überprüfung der Normalverteilung angewendet. Für die Gesamtdauer errechnet der Shapiro-Wilk Test (Abbildung 3) die Signifikanzlevel $p_{mq} = ,39$ und $p_{sq} = ,86$ und die statistischen Werte $T_{sq} = 0,98$ und $T_{mq} = 0,95$. Die *p*-Werte sind sowohl für *MQ* als auch für *SQ* nicht signifikant und damit kann für beide Gruppen die Nullhypothese des Shapiro-Wilk Tests, also die Normalverteilung, angenommen werden. Um einen Fehler zweiter Art an dieser Stelle möglichst auszuschließen, werden beide Werte mit dem kritischen Wert des Shapiro-Wilk-Tests für eine Stichprobe mit 21 bzw. 20 Probanden verglichen. Diese Werte können Abbildung [4] entnommen werden. Für 20 Probanden ist der kritische Werte $\alpha_{krit,20} = 0,905$, für 21 Probanden gilt $\alpha_{krit,21} = 0,908$. Ein direkter Vergleich zeigt $T_{sq} > \alpha_{krit,20}$ und $T_{mq} > \alpha_{krit,21}$. Visualisiert werden die beiden Verteilung in dem Histogramm in Abbildung 4.

Damit wird eine Normalverteilung angenommen. Zur Überprüfung der Homogenität der Varianzen wird ein Levene-Test durchgeführt. Dieser ergibt ein Signifikanzlevel $p = ,395$ und eine *F*-Verteilung $F = 0,74$. Für das *f*-Quantil der Freiheitsgrade $df_1 = 20$ und $df_2 = 21$ gilt $F_{krit} = 2,10$. Der

Abhängige Variable	Kennzahl	Messung	Nr.	Skalierung
Effektivität	Gesamtdurchführungszeit (durTotal)	Webapplikation -		ratioskaliert
	Durchführungszeit des kürzeren Fragebogens (durShort)	Webapplikation -		ratioskaliert
	Durchführungszeit des längeren Fragebogens (durLong)	Webapplikation -		ratioskaliert
Effizienz	Anzahl der insgesamt ausgelassenen Fragen (NullGes)	Webapplikation -		ratioskaliert
Zufriedenheit	aufsummierter Score aller Fragen (zfrSum)	Fragebogen	1-6	ordinal
	aufsummierter Score der Fragen 1,2,4 (zfrSum124)	Fragebogen	1,2,4	ordinal

Tabelle 2: Anzahl der Probanden pro Szenario

treatment		Kolmogorov-Smirnoŕ			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
durTotal	mq	,112	21	,200*	,953	21	,394
	sq	,122	20	,200*	,975	20	,860

Abbildung 3: Shapiro-Wilk-Test für Normalverteilung bezüglich der Variable *durTotal*

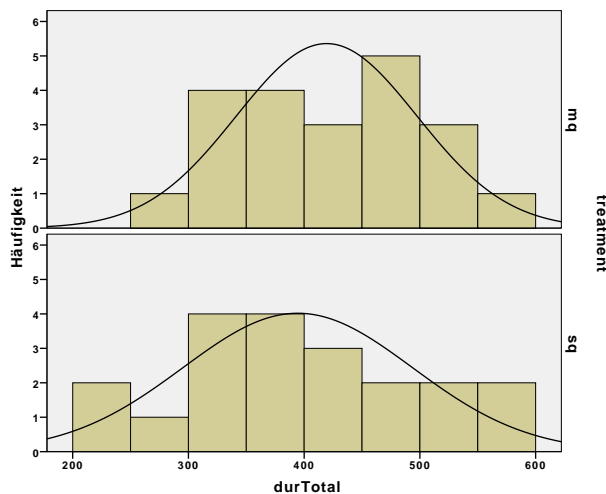


Abbildung 4: Histogramm mit Normalverteilungskurve der Gesamtdauer über alle Stichproben

direkte Vergleich zeigt $F < F_{krit}$. Damit wird die Nullhypothese des Levene-Test akzeptiert, es werden annähernd identische Varianzen für die beiden Gruppen angenommen. Der

		T-Test für die Mittelwertgleichheit						
		T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
							Untere	Obere
durTotal	Varianzen sind gleich	-.898	39	.375	-24.979	27.831	-81.272	31.315

Abbildung 5: T-Test für die Variable *durTotal*.

T-Test (Abbildung 5) zeigt keinen signifikanten Unterschied für die beiden Gruppen *SQ* ($M = 394,45$, $SD = 99,26$) und *MQ* ($M = 419,43$, $SD = 78,18$), ($p = ,375$, $t(39) = -0,898$, 2-seitiger Test, $d \approx -0,287$). Daher wird die in Kapitel 2.4 aufgestellte H_{10} -Hypothese akzeptiert. Da keine signifikanten Unterschiede festgestellt worden sind, wird hier auf die Durchführung einer Power-Analyse verzichtet.

Ergebnisse für den kleineren Fragebogen.

Für den kleineren Fragebogen zum Thema Fast-Food wird die Variable *durShort* untersucht. Dazu wird erneut ein Shapiro-Wilk-Test auf die unabhängigen Variablen *SQ* und *MQ* durchgeführt.

treatment		Kolmogorov-Smirnoŕ			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
durShort	mq	,123	21	,200*	,935	21	,173
	sq	,140	20	,200*	,945	20	,303

Abbildung 6: Shapiro-Wilk-Test für Normalverteilung bezüglich der Variable *durShort*

Das Ergebnis des Shapiro-Wilk Tests in Abbildung 6 nimmt für beide Gruppen mit jeweiligen Signifikanzleveln $p_{MQ} = 0,173$ und $p_{SQ} = 0,30$ die Normalverteilung an. Der Vergleich der statistischen Werte $T_{sq} = 0,945$ und $T_{mq} = 0,935$ mit den kritischen Werten $T_{mq} > \alpha_{krit,21}$, $T_{sq} > \alpha_{krit,20}$, bestätigt die Annahme der Normalverteilung. Der Levene-Test nimmt eine annähernd identische Varianzen für die beiden Gruppen an ($p = 0,845$, $F = 0,039$).

		T-Test für die Mittelwertgleichheit						
		T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
							Untere	Obere
durShort	Varianzen sind gleich	-.760	39	.452	-10,298	13,54382	-37,6925	17,09744

Abbildung 7: T-Test für die Variable *durShort*.

Der T-Test (Abbildung 7) zeigt keinen signifikanten Unterschied für die beiden Gruppen *SQ* ($M \approx 149,57$, $SD \approx 44,87$) und *MQ* ($M \approx 159,87$, $SD \approx 41,86$), ($p = ,452$, $t(39) = -0,76$, 2-seitiger Test, $d \approx -0,243$).

Ergebnisse für den größeren Fragebogen.

Für den größeren Fragebogen zum Thema Politik wird die Variable *durLong* untersucht. Dazu wird ein Shapiro-Wilk Test durchgeführt, um die Stichproben auf Normalverteilung zu überprüfen. Das Ergebnis des Shapiro-Wilk Tests

treatment		Kolmogorov-Smirnoŕ			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
durLong	mq	,161	21	,162	,909	21	,052
	sq	,117	20	,200*	,980	20	,940

Abbildung 8: Shapiro-Wilk Test für Normalverteilung bezüglich der Variable *durLong*

in Abbildung 8 nimmt für beide Gruppen mit jeweiligen Signifikanzleveln $p_{MQ} = ,052$ und $p_{SQ} = ,940$ die Normalverteilung an. Der Vergleich der statistischen Werte $T_{sq} = 0,980$ und $T_{mq} = 0,909$ mit den kritischen Werten $T_{mq} > \alpha_{krit,21}, T_{sq} > \alpha_{krit,20}$, deutet auch auf eine Normalverteilung. Eine Betrachtung der Histogramme beider Verteilungen (Abbildung 9) zeigt vor allem für die Gruppe MQ keine eindeutige Normalverteilung. Daher kann man die Voraus-

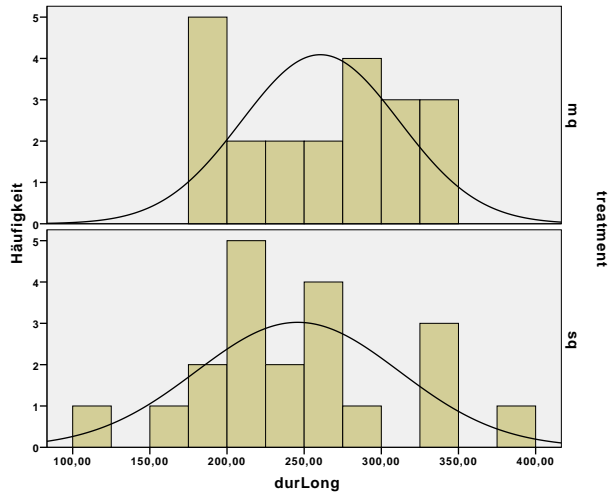


Abbildung 9: Histogramm mit Normalverteilungskurve der Dauer des Fragebogens zu Politik über alle Stichproben.

setzungen für einen T-Test nicht als erfüllt ansehen. Daher wird ein nicht parametrischer Test durchgeführt. Der Mann-Whitney U Test für zwei unabhängige Variablen zeigt für die Gruppen SQ ($Mdn = 379,50$) und MQ ($Mdn = 429,00$) keinen signifikanten Unterschied in der Bearbeitungszeit des längeren Fragebogens, $U = 179,00$, ns , $r = \frac{Z}{\sqrt{N}} = \frac{-0,809}{\sqrt{41}} \approx -0,13$.

	durLong
Mann-Whitney-U	179,000
Wilcoxon-W	389,000
Z	-,809
Asymptotische Signifikanz (2-seitig)	,419

Abbildung 10: Man-Whitney-U Test für die Variable *durLong*.

Unterschied zwischen den Szenarien.

Auch die T-Tests für die Szenarien 1-4 (vgl. Kapitel 2.1) bezüglich der Variablen *durShort*, *durLong* und *durTotal* zeigen keine signifikanten Unterschiede in der Bearbeitungszeit. Einzig für die Variable *durShort* zeigt der T-Test (Abbildung 11) signifikante Unterschiede zwischen dem Szenario 1 ($M \approx 140,86$, $SD \approx 26,46$) und Szenario 2 ($M \approx 180,78$, $SD \approx 46,77$), ($p = ,025$, $t(19) = -2,44$, 2-seitiger Test, $d \approx -1,12$) Zwischen Szenario 3 und Szenario 4 gibt es keine signifikanten Unterschiede für die Variable *durShort*. Daher wird die H_{20} -Hypothese akzeptiert.

T-Test für die Mittelwertgleichheit							
		T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz
durShort	Varianzen sind gleich	-2,44	19	,025	-39,9181	16,37644	Untere: -74,19438 Obere: -5,64180

Abbildung 11: T-Test für die Variable *durShort* zwischen Szenario 1 und Szenario 2.

4.3 Effizienz

Als Maß der Effizienz der Probanden wird die Anzahl der ausgelassenen Fragen (siehe Kapitel 2.4), *NullGes* herangezogen. Insgesamt gab es neun Probanden, die eine oder mehrere Fragen ausgelassen haben (siehe Abbildung 12). Es wurden insgesamt 14 Fragen ausgelassen. Wie in Kapitel 4.2 werden auch hier die Voraussetzungen für den t-Test bei unabhängigen Stichproben getestet. Ein Shapiro-Wilk-Test zur Normalität mit $p < ,001$ für jeweils beide Konditionen zeigt (*sq* und *mq*), dass keine Normalverteilung bezüglich der Anzahl der ausgelassenen Fragen in den beiden Gruppen *SQ* und *MQ* vorliegt.

		NullGes				Gesamt
		0	1	2	4	
treatment	mq	17	2	2	0	21
	sq	15	4	0	1	20
Gesamt		32	6	2	1	41

Abbildung 12: Kreuztabelle für Werte *nullGes* und die Konditionen

Daher wird der nichtparametrische Mann-Whitney-U-Test für die weitere Berechnung verwendet. Die Probanden der Gruppe *SQ* ($Mdn = 0$, *mittlererRang* = 21,53) und der Gruppe *MQ* ($Mdn = 0$, *mittlererRang* = 20,50) haben sich hinsichtlich der ausgelassenen Fragen nicht signifikant unterschieden, $U = 199,50$, $z = -0,378$, $p = ,71$, und es wurde keine Beziehung zwischen beiden Gruppen festgestellt ($r = -0,06$).

Folgerichtig wird die in Kapitel 2.4 aufgestellte Hypothese H_{30} akzeptiert. Da keine signifikanten Unterschiede festgestellt worden sind, wird hier auf die Durchführung einer Power-Analyse verzichtet.

4.4 Zufriedenheit

Die Häufigkeitsverteilung in Abbildung 13 zeigt eine Antworttendenz über alle Fragen zu "4: trifft eher nicht zu" und "5: trifft nicht zu". Eine Normalverteilung liegt nicht vor. Die nach Gruppen aufgeschlüsselten Mittelwerte (vgl. Ab-

treatment		zufr1	zufr2	zufr3_pol	zufr4_pol	zufr5	zufr6
mq	Mittelwert	3,95	4,14	4,38	4,10	3,57	4,24
	N	21	21	21	21	21	21
	SD	1,244	1,236	,805	,944	1,207	1,411
sq	Mittelwert	4,10	4,50	4,30	3,95	3,84	4,63
	N	20	20	20	20	19	19
	SD	1,410	,761	,657	1,146	1,500	1,012
Ges.	Mittelwert	4,02	4,32	4,34	4,02	3,70	4,43
	N	41	41	41	41	40	40
	SD	1,313	1,035	,728	1,037	1,344	1,238

Abbildung 13: Häufigkeitsverteilung der Nutzerzufriedenheit

Abbildung 14) zeigen höhere Werte für *MQ* in den Variablen

$zufr1$, $zufr2$, $zufr5$ und $zufr6$. In der Gruppe SQ sind die Werte für $zufr3_{pol}$ und $zufr4_{pol}$ höher. Im Folgenden werden diese Werte auf Signifikanz geprüft.

treatment		zufr1	zufr2	zufr3_pol	zufr4_pol	zufr5	zufr6
mq	Mittelwert	3,95	4,14	4,38	4,10	3,57	4,24
	H	21	21	21	21	21	21
	SD	1,244	1,236	,805	,944	1,207	1,411
sq	Mittelwert	4,10	4,50	4,30	3,95	3,84	4,63
	H	20	20	20	20	19	19
	SD	1,410	,761	,657	1,146	1,500	1,012
Ges.	Mittelwert	4,02	4,32	4,34	4,02	3,70	4,43
	H	41	41	41	41	40	40
	SD	1,313	1,035	,728	1,037	1,344	1,238

Abbildung 14: Gruppenweise Mittelwerte der Nutzerzufriedenheit

Zum Testen auf eine Manipulation der Nutzerzufriedenheit durch die Treatments SQ und MQ wird ein Mann-Whitney-U-Test durchgeführt. Dieser zeigt keinen signifikanten Effekt vom Treatment auf die Nutzerzufriedenheit ($zufrSum$), $U = 209,00$, $p = ,98$, $r = -0,04$ (mittlerer Rang für $SQ = 21,05$, für $MQ = 20,95$). Dies lässt auf ein Problem mit dem Design des Fragebogens schließen. Ein Chrombachs-Alpha von ,685 und eine Item-to-Item Korrelation mit dem Summenscore bestätigen diese Annahme. Problematisch erscheinen dabei die Items $Zufr3$, $Zufr5$, $Zufr6$. Sie haben die geringste Item-to-Item Korrelation ($p < ,05$) mit dem Summenscore ($Zufr3$: 0,472, $Zufr5$: 0,567, $Zufr6$: 0,588). Ein Ausschließen dieser Items verbessert Chrombachs-Alpha auf 0,812. Der Mann-Whitney-U-Test zwischen dem Summenscore aus $Zufr1$, $Zufr2$, $Zufr4_{pol}$ ($zufrSum124$) und dem Treatment verbessert sich auf $U = 188,00$, $p = 0,56$, $r = -0,09$ (mittlerer Rang für $SQ = 22,10$, für $MQ = 19,95$). Eine Signifikanz ist somit nicht nachweisbar und $H4_1$ kann nicht angenommen werden.

5. DISKUSSION

Nachdem Kapitel 4 die Ergebnisse der Nutzerstudie vorstellt, werden diese im folgenden Kapitel diskutiert. Vor allem im Hinblick auf ihre Aussagekraft bezüglich der Hypothesen in Kapitel 2.4.

5.1 Effektivität

Die Ergebnisse in Kapitel 4.2 deuten auf keinen signifikanten Unterschied in der Bearbeitungszeit zwischen den untersuchten Darstellungen SQ und MQ . Sowohl in der Gesamtzeit ($durTotal$) als auch in den beiden Zeiten der einzelnen Fragebögen ($durShort$, $durTotal$) zeigen die Tests, dass es keine Unterschiede gibt ($H1_0$ bestätigt).

treatment			Werte
durShort	mq	Mittelwert	159,8695
		Median	153,9700
	sq	Mittelwert	149,5720
		Median	141,1850
durLong	mq	Mittelwert	260,4900
		Median	260,7100
	sq	Mittelwert	245,8505
		Median	235,9900
durTotal	mq	Mittelwert	419,43
		Median	429,00
	sq	Mittelwert	394,45
		Median	379,50

Abbildung 15: Mittelwerte und Mediane der Gruppen SQ und MQ bzgl der Variablen

Ein Vergleich der Mittelwerte der Stichproben bezüglich der Gruppen und der Variablen (Abbildung 15) zeigt, dass der Mittelwert und der Median der Gruppe SQ für jede untersuchte Variable immer kleiner sind, als die entsprechenden Werte der Gruppe MQ . Dieser Zeitunterschied kann durch verschiedene Einflüsse entstehen. Probanden, die Fragebögen per Scroll-Navigation ausfüllen, müssen jede Frage auf der Seite zu erst richtig anordnen, während die Probanden mit Button-Navigation lediglich auf den *Weiter*-Button klicken. Weiterhin ist es für die Probanden der Gruppe MQ leichter nochmal zu einer Frage zurückzukehren oder die nächste bereits zu lesen, was sich auf die Gesamtbearbeitungszeit auswirken kann. Dieser Einfluss wurde in dieser Nutzerstudie jedoch nicht berücksichtigt.

Auch der Vergleich zwischen den Szenarien (vgl. Kapitel 2.1) zeigt keine signifikanten Unterschiede. Lediglich zwischen Szenario 1 und Szenario 2 ist ein signifikanter Unterschied für die Variable $durShort$, also den kurzen Fragebogen, festzustellen. Die Effektgröße $d = -1,12$ zeigt, dass Probanden, die den längeren Fragebogen zuerst ausfüllten (Szenario 1), deutlich schneller sind, als die Probanden mit dem kürzeren Fragebogen zuerst (Szenario 2). Dies kann auf eine Ermüdung während des Tests zurückgeführt werden, aber auch auf die Rückmeldungen der Probanden (Kapitel 5.4). Für viele Probanden war der Fragebogen zu dem Thema Politik zu lang und wer diesen zu erst bearbeitet hat, ist für den zweiten Fragebogen vielleicht nicht mehr so konzentriert und antwortet schneller als bei dem ersten Fragebogen. Da die Länge des Fragebogens zu Politik sehr oft kritisiert wurde, ist der signifikante Unterschied eher auf die mangelnde Bereitschaft, erneut einen Fragebogen auszufüllen, zurückzuführen. Auf Grund von fehlenden Unterschieden zwischen den anderen Szenarien wird auch hier die Nullhypothese $H2_0$ angenommen.

Der Einfluss der Leistung des jeweiligen genutzten mobilen Endgerätes wurde an dieser Stelle vernachlässigt. Die Geräte, die in der Pilotstudie als tauglich erkannt wurden, wurden bei fast allen Probanden eingesetzt. Latenzen beim Aufbau der Seite und beim Scrollen sind eher auf die Webapplikation zurückzuführen als auf die Endgeräte.

5.2 Effizienz

Der Blick auf die Daten (Abbildung 12) zeigt, dass ein Proband (ID 1712) innerhalb der Gruppe SQ vier Fragen ausgelassen hat. Es ist hierbei nicht auszuschließen, dass der Proband Fragen mit Absicht ignoriert oder ausgelassen hat. Da die Durchführungszeiten dieses Probanden ($T_{short} = 87,16$ und $T_{long} = 184,59$) sehr nahe an den Minimum-Werten liegen, könnte es auch sein, dass der Proband Fragen durch unabsichtliches Doppelklicken der Buttons übersprun-

gen hat ohne es zu merken. Da die Webapplikation nicht auf ausgelassene Fragen hinweist, ist dies vorstellbar. Die Probanden mit ausgelassenen Fragen sind auf beide Gruppen gleichmäßig verteilt. 9 der 14 Fragen wurden innerhalb des Fast-Food-Fragebogens ausgelassen. Dies könnte im Zusammenhang damit stehen, dass es Begriffsunklarheiten während der Durchführung bei einzelnen Probanden gab (siehe Kapitel 5.4) und diese deshalb die Fragen ausgelassen haben. An dieser Stelle fehlt eine Frage in dem Papierfragenbogen, ob der Probanden Fragen bewusst ausgelassen hat oder nicht.

Als Ergebnis können den beiden Designs keine Vor- oder Nachteile bezüglich der Effizienz zugesprochen werden, da die Gründe für die Verteilung der Daten nicht eindeutig hergeleitet werden können. Daher wird die Hypothese $H3_0$ akzeptiert.

5.3 Zufriedenheit

In der hier gezeigten Nutzerstudie konnten keine Signifikanten Zusammenhänge zwischen dem Treatment und der Nutzerzufriedenheit nachgewiesen werden. Dieses Ergebnis muss jedoch nicht ausschließlich auf zwei gleichwertige Navigationsarten zurückzuführen sein. Rückblickend konnten Schwachstellen im Fragebogen-Design aufgedeckt werden, welche innerhalb der Pilotstudie nicht sichtbar waren. Hierbei stechen die Items *zufr3*, *zufr5* und *zufr6* negativ heraus und wurden für die Auswertung von $H4$ ausgeschlossen.

1. *zufr3*: Dieses Item weist die geringste Korrelation mit dem Summenscore auf. Gleichzeitig ist die Varianz mit 0,728 die geringste unter den Items für die Zufriedenheit. Dies wirft die Frage auf, warum die Antworten der Probanden im Verhältnis so wenig streuen. Dies kann durch die mangelnde Fähigkeit begründet werden, sich in eine nicht angewandte Navigationsart hineinzusetzen, nachdem man auf eine andere Navigationsart konditioniert worden ist. Dies ist ein generelles Problem in Between-Group-Designs. Es hätte gedämpft werden können, indem zwischen der App-Nutzung und der Bearbeitung des Nutzerzufriedenheitsfragebogens ein Informations-Blatt zu den verschiedenen Navigationsarten gezeigt worden wäre.
2. *zufr5*: Dieses Item weist die zweit niedrigste Korrelation mit dem Summenscore auf. Das Problem der mangelnden Fähigkeit sich in eine andere als der angewandten Navigationsart hineinzusetzen wirkt hier noch stärker, da explizit nach anderen Navigationsarten gefragt wird. Des weiteren wurde die Frage nicht von allen Probanden verstanden. Der Begriff "*Wischgeste*" anstatt "*Swipen*" könnte diese Verwirrung reduzieren.
3. *zufr6*: Dieses Item fragt nach technischen Problemen während der Nutzung der Webapplikation. Dabei wird nicht eindimensional gefragt, da verschiedene Probleme (Ruckler, falsche Erkennung von Gesten auf dem Touchscreen, lange Ladezeiten) beispielhaft angegeben wurden. Es ist daher nicht nachvollziehbar, welche technischen Probleme bei dem jeweiligen Probanden aufgetreten sind. Es wurde im Verlauf des Experimentes beispielsweise angemerkt, dass das visuelle Feedback beim Auswählen von Radio-Buttons irreführend sei.

Dies konnte auf älteren Geräten reproduziert werden. Ein ungenaues Berühren des Radio-Buttons löst dabei einen Schattierungs-Effekt aus, welcher den Button nachhaltig umgibt. Erst ein erneutes und genaues Auswählen des Radio-Buttons aktiviert ihn.

Außerdem kam es auf älteren Geräten zu spürbaren Ladezeiten bzw. Rucklern von unter einer Sekunde, welche nicht in die Zeitmessung mit eingeflossen sind. Diese könnten das Antwortverhalten beeinflusst haben. Die Effekte dieser Problemzustände auf die Nutzerzufriedenheit können nicht als äquivalent zueinander angenommen werden.

Die übrigen Items bilden in der Auswertung den Summenscore für die Messung der Nutzerzufriedenheit (vgl. 4.4). Sie können die Unterschiede der Mittelwerte zwischen den einzelnen Treatments MQ und SQ besser beschreiben als der ursprünglich angesetzte Summenscore, erreichen dabei aber keine Signifikanz. Dies kann damit begründet werden, dass die übrigen Items nicht genügend Möglichkeiten zum differenzieren bieten, um Unterschiede messbar zu machen. Allerdings muss auch in Erwägung gezogen werden, dass die Grundgesamtheit unserer Zielgruppe mit beiden Navigationsarten gleich zufrieden ist.

5.4 Rückmeldungen der Probanden

Während und nach der Bearbeitung der Fragebögen, gaben einige Probanden Feedback zu der Nutzerstudie. Einige Probanden haben zusätzlich zur sechsten Frage des Nutzerzufriedenheitsbogens die Art ihrer technischen Probleme mitgeteilt. Es wurden stehende Schatten beim Anklicken der Radiobuttons erwähnt. Andere sprachen von leichten "Display-Hängern" beim Scrollen.

Einzelne Probanden haben Nachfragen zu Fragen der Fragebögen gestellt, da diese teilweise unklar formuliert erschienen. Zusätzlich wurde auch angemerkt, dass auf dem Papierfragebogen ein Feld mit "*Ich habe bewusst Fragen ausgelassen*" fehlt. Ein weiterer Kritikpunkt war häufig die Länge der beiden Fragebögen. Insbesondere der Politik-Fragebogen wurde auf Grund seiner Länge oft kritisiert.

Im Gespräch mit Probanden wurden auch verschiedene Vor- und Nachteile der beiden Designs angesprochen. Ein Proband (ID: 1721) gab an per Scroll-Navigation schneller den Fragebogen bearbeiten zu können, wohingegen ein anderer Proband (ID: 1722) angegeben hat sich Dank der Button-Navigation besser auf die einzelnen Fragen konzentrieren zu können und keine Zeit auf die richtige Positionierung des Displays durch Scrollen verwenden musste.

6. VERWANDTE ARBEITEN

Seitdem Umfragen nicht mehr nur mit Stift und Papier durchgeführt werden, werden optimale Designs für Umfragen auf Computern untersucht. Mit der Einführung von Smartphones und mobilem Internet verlagern sich die Untersuchungen von den Unterschieden in der Darstellungen auf Desktop PCs zu den Darstellungen auf mobilen Endgeräten. Eine Zusammenfassung zu verschiedenen Studien in diesem Bereich liefert [5].

Ein Kapitel beschäftigt sich speziell mit der Button-Navigation. Untersucht wird der Einfluss eines Zurück-Buttons. [2] untersucht, wie Webapplikationen für online Umfragen aussehen sollten.

Eine Arbeit die sehr ähnliche Untersuchungsgegenstände zu

dieser Nutzerstudie untersucht wird in [1] vorgestellt. Diese Arbeit prüft, ob es zeitliche Unterschiede in der Bearbeitung zwischen Umfragen auf Desktop PCs und mobilen Endgeräten gibt. Ein zentrales Ergebnis ist, dass Smartphone Nutzer 17% mehr Zeit gegenüber PC-Nutzern zum Bearbeiten von Fragebögen benötigen.

7. FAZIT

Wie in Kapitel 4 erläutert, können für die Hypothese $H1$ keine signifikanten Unterschiede festgestellt werden. Es wurden zwar Zeitunterschiede gemessen, diese können jedoch verschiedenen Einflüssen geschuldet sein. Auch für die Hypothese $H2_1$ konnten keine Signifikanzen festgestellt werden. Trotz einer gemessenen Signifikanz zwischen Szenario 1 und 2 für den kurzen Fragebogen, kann kein eindeutiger Unterschied festgestellt werden. Grund sind die Rückmeldungen der Probanden, diese schließen eher auf ein absichtliche schnellere Beantwortung der Fragen des zuletzt bearbeiteten Fragenkatalogs hin. Somit wird die Hypothese $H2_1$ abgelehnt. Die Effektivität unterscheidet sich dementsprechend nicht zwischen den beiden Navigationsarten.

Es kann auch zur Effizienz keine Empfehlung gegeben werden. Keine der beiden Designs führte bei den Probanden zu mehr unabsichtlich ausgelassenen Fragen, so dass auch Hypothese $H3_1$ abgelehnt wird.

Die statistische Auswertung der wahrgenommenen Zufriedenheit konnte ebenfalls keine signifikanten Unterschiede. Die Hypothese $H4_1$ wird abgelehnt. Die Ergebnisse dieser Studie lassen keine klare Designempfehlung zu. Für Nutzerstudien können also je nach Kontext beide Navigationsarten eingesetzt werden. Für zukünftige Arbeiten könnte ein Vergleich mit der *Swipe*-Navigation von Interesse sein und neue Erkenntnisse bringen.

Literatur

- [1] I. Andreadis. Comparison of response times between desktop and smartphone users. *Mobile Research Methods*, page 63, 2015.
- [2] I. Andreadis. Web surveys optimized for smartphones: Are there differences between computer and smartphone users? *The Collection of Survey Data using Mixed Devices*, page 213, 2015.
- [3] K.-C. Hamborg. Isometricsl - fragebogen zur evaluation von graphischen benutzungsschnittstellen. 1997.
- [4] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
- [5] R. Tourangeau, F. G. Conrad, and M. P. Couper. *The science of web surveys*. Oxford University Press, 2013.
- [6] Wissenschaftsrat. Basisdaten zu Hochschulen und Forschungseinrichtungen in Deutschland. 2016. <http://www.wissenschaftsrat.de/download/archiv/Basisdaten.pdf>; zuletzt gesehen: 15.07.2016.