



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DEPARTMENT INFORMATION

Masterarbeit

Gegenüberstellung dreier Verfahren zur Durchführung von Usability-Tests –
Empirischer Vergleich von klassischem Labortest, toolbasiertem asynchronen
Remote-Test und Remote-Videotest

vorgelegt von

Sascha Küchler

Studiengang Informationswissenschaft und -management

erster Prüfer: Prof. Dr. Dirk Lewandowski
zweite Prüferin: Prof. Dr. Ulrike Spree

Hamburg, August 2012

Abstract (dt.)

Usability-Tests dienen der Identifikation von Benutzungsproblemen, die Nutzer mit einer Anwendung haben. Neben dem Verfahren des klassischen Usability-Tests im Labor wurden neue Verfahren entwickelt, die ein asynchrones Testen einer Anwendung durch den Nutzer von zu Hause aus ermöglichen. Neben dem schon länger verbreiteten Verfahren des asynchronen Remote-Usability-Tests kommt das neue Verfahren eines asynchronen Videotests auf, bei dem die Nutzer ihren Bildschirm und ihre gesprochenen Kommentare während der Nutzung aufzeichnen. Um dieses neue Verfahren einzustufen, wurde es mit den beiden etablierten Verfahren an einem Fallbeispiel verglichen. In Bezug auf die gefundenen Usability-Probleme schnitt es sehr gut ab und lieferte auch relativ tiefgründige Verbesserungsvorschläge für die Webanwendung. In der Ausführlichkeit und der Klarheit von Verbesserungsvorschlägen kann der Usability-Test im Labor noch deutlich hervorstechen, da der interaktive Austausch mit dem Testleiter deutliches Potenzial für bereits sehr durchdachte Verbesserungsvorschläge bietet.

Abstract (engl.)

Usability tests aim to identify problems which users have with an application. Beside the classical usability test in the lab new methods have been developed which make it possible for the user to test an application at home. Apart from the common method of the asynchronous remote usability test there is a new method coming up which is best described as asynchronous video test. With this method the users record their screen and their spoken comments during the use of the application. To rate this new method it was compared with the two established procedures on an example. Regarding the amount of identified usability problems it performed very good and brought reasonably profound suggestions for improvement of the web application to surface. Taking into consideration the detailedness and clarity of the suggestions for improvement, the usability test in the lab can stand out because the interactive exchange with the person who looks after the test person includes clear potential for suggestions for improvement which are already well thought-out.

Schlagwörter

Usability/Benutzerfreundlichkeit | Methodik | Methodenvergleich | Asynchroner Usability-Test

Inhaltsverzeichnis

1	Einleitung	7
2	Usability-Testing	9
3	Notwendigkeit von Usability-Tests	11
4	Stand der Forschung	14
5	Untersuchungsobjekt	20
6	Grundlagen für die Erhebung	22
6.1	Hypothesen	22
6.2	Probandenauswahl	23
6.3	Methoden und Studiendesign	25
6.3.1	Der Usability-Test im Labor	27
6.3.2	Der asynchrone Remote-Usability-Test	29
6.3.3	Der Remote-Videotest	30
7	Ergebnisse der Erhebungen	32
7.1	Usability-Test im Labor	34
7.2	Asynchroner Remote-Usability-Test	35
7.3	Remote-Videotest	39
7.4	Schwerwiegende Probleme	41
7.5	Anzahl der Testpersonen und Auswirkung auf die Ergebnisse	42
7.6	Aufwände der einzelnen Methoden im Vergleich	44
8	Prüfung der Hypothesen	50
9	Fazit	59
10	Ausblick	67
11	Literatur	72
12	Anhang	76

Abbildungsverzeichnis

Abbildung 1: Startseitenausschnitt des Onlineshops www.mein-schnulli.de	20
Abbildung 2: Sitzplatz des Probanden in einem typischen Usability-Labor (Foto der eResult GmbH).	28
Abbildung 3: Beispielhafte Clickstream-Visualisierung des Anbieters Loop 11 (LOOP11 2012).....	29
Abbildung 4: Vorliegende Daten zur Auswertung einer Remote-Studie von RapidUserTests (RAPIDUSERTESTS 2012).	31
Abbildung 5: Ausschnitt aus der Gesamtproblemliste aller drei Methoden.	32
Abbildung 6: Diagramm zur Gegenüberstellung der gefundenen Probleme der drei Methoden.	34
Abbildung 7: Ablauf des aRUT, wie er als Abfolge im Loop11-Tool visualisiert wird (Screenshot aus der Anwendung).	36
Abbildung 8: Aufgabenabarbeitung eines einzelnen Probanden bei Loop11 (Screenshot aus der Anwendung).	37
Abbildung 9: Auswertungsansicht des Videotests von RapidUserTests (Screenshot aus der Anwendung).	39
Abbildung 10: Gefundene Probleme der Methoden im Diagramm bei herunter gerechneten Stichproben für $n=12$	43
Abbildung 11: Kosten und Manntage für die drei Verfahren im Vergleich.	46

Tabellenverzeichnis

Tabelle 1: Zielgruppenverteilung anhand der drei Merkmale Geschlecht (m/w), Zielgruppenpriorität und Erfahrungshorizont für die drei Testverfahren.	25
Tabelle 2: Anzahl und Priorisierung der mit den drei Verfahren aufgedeckten Usability-Probleme....	33
Tabelle 3: Auf n=12 umgerechnete Problemanzahl für die Methoden Videotest und asynchroner Remote-Test.	42
Tabelle 4: Bei n=12 ermittelte Probleme von Labor- und Videotest und Problemverhältnis.	44
Tabelle 5: Gegenüberstellung der Aufwände und Kosten für die drei verglichenen Verfahren.	45
Tabelle 6: Abschließende Gegenüberstellung von Vor- und Nachteilen der drei Methoden.....	64

Abkürzungsverzeichnis

aRUT	asynchroner Remote-Usability-Test
URL	Uniform Resource Locator (eindeutige Identifizierung einer Webadresse)

1 Einleitung

Websites und insbesondere Onlineshops müssen vom Anwender gut nutzbar sein, damit sie den gewünschten Erfolg für die Betreiber bewirken. Bereits seit mehreren Jahren hat sich der Begriff der Web Usability, dessen zunehmende Verbreitung u. a. von RICHTER/FLÜCKIGER (2010, S. 124 f.) dargestellt wird, als Beschreibung der Gebrauchstauglichkeit eines Webangebots etabliert. Er weist daraufhin, dass bei der großen Anzahl an Webangeboten, die stets noch weiter zunimmt, der Nutzer bei Problemen in der Bedienung problemlos auf einen alternativen Anbieter zurückgreifen kann und so kein Umsatz beim erstbesuchten Shop entsteht. Daher spielt die Web Usability eine zunehmende Rolle und hält durch Anwendung wissenschaftlicher Methoden zu diesem Themenkomplex Einzug in die Praxis und wirkt auf diesem Weg den Problemen eines Online-Angebots entgegen.

Untersuchungen zur Gebrauchstauglichkeit von Internetangeboten haben sich als ein Verfahren etabliert, das die Schwachpunkte von Sites aufdeckt. Deren Behebung sorgt z. B. für gesteigerte Umsätze, im Allgemeinen zunächst für eine gesteigerte Zufriedenheit der Nutzer. Anhand der üblichen und weit verbreiteten Auswertung von sog. Analytics-Daten (vgl. EBERHARD-YOM 2010, S. 140 f.) können schwierige Stellen in Onlineangeboten identifiziert werden, und zwar etwa solche, wo Nutzer Bestellprozesse abrechnen oder bestimmte Aktionen, auf die das Angebot abzielt, nicht durchführen. Um die Intention der Nutzer in Erfahrung zu bringen und von diesen im realen Nutzungskontext zu erfahren, was ihnen an dem Angebot Probleme bereitet, werden diese Untersuchungen durchgeführt. Denn diese gehen über die reine Datenanalyse der Analytics-Daten hinaus, da sie zusätzlich die Nutzermeinung berücksichtigen und so für ein tiefergehendes Verständnis des Nutzerverhaltens sorgen. Ausgegangen wird hierbei stets vom nutzerzentrierten Ansatz (vgl. a. a. O., S. 121 ff.), der dem Nutzer niemals Fehlverhalten unterstellt. Vielmehr geht es hierbei darum, jedem Nutzer unabhängig von seinen technischen und kognitiven Fähigkeiten eine für ihn optimal nutzbare Seite anzubieten. Schließlich sollen alle zur Zielgruppe gehörenden Nutzer unabhängig von Alter und Internetaffinität die einfache Möglichkeit zum Kauf oder zum Auffinden der gewünschten Information haben.

Neben bestimmten Konventionen, die sich hierzu im Lauf der Jahre herausgebildet haben (vgl. u. a. PUSCHER 2009, S. 47 ff.), muss i. d. R. für viele Angebote die Bedienung separat evaluiert werden, um eine konkrete Site möglichst erfolgversprechend aufzustellen. Die Konventionen dienen dann als Basis. Die weitergehende Optimierung des spezifischen Angebots ist dann noch als individuelle Verbesserung des Angebots zu sehen.

Für diese Evaluierung bietet sich dann der Usability-Test an. Als qualitative Methode angelegt, führt er mit der subjektiven Bewertung des Angebots durch eine begrenzte Zahl an Testnutzern (Probanden) in aller Regel zum Aufdecken eines Großteils der Usability-Probleme (vgl. EBERHARD-YOM 2010, S. 136 f.). Dieser Test kann in mehreren Variationen durchgeführt werden, die sich

teilweise bereits als eigenständige (Teil-)Methode etabliert haben. Wichtig ist hierbei, das Potenzial der jeweiligen Methode zu kennen und entsprechend die adäquate Methode für die eigene Forschungsfrage in Abhängigkeit von Budget und zur Verfügung stehender Zeit auszuwählen, um möglichst gute Ergebnisse zu erhalten. Einen solchen Vergleich hat die vorliegende Arbeit zum Thema, denn sie vergleicht drei Verfahren für Usability-Tests, und zwar den klassischen Usability-Test im Labor, den asynchronen toolbasierten Remote-Usability-Test und den Remote-Videotest. Das letztgenannte Verfahren ist noch relativ neu am Markt und auch der Name ist bisher nur eine Umschreibung, die sich noch nicht durchgesetzt hat und so nicht als eindeutige Benennung dienen kann. Um diese neue Methode einordnen zu können, soll sie mit den beiden etablierten Verfahren verglichen werden. Hierbei ist zu beachten, dass von vornherein nicht angezweifelt wird, dass die Methoden starke Unterschiede aufweisen und nicht als gleichwertige Substitute eingesetzt werden können. Vergleiche zwischen dem klassischen Labortest und dem asynchronen Remote-Usability-Test hat es auch bereits gegeben (s. Kapitel 4). Diese werden für die Arbeit berücksichtigt. Um jedoch den ganzheitlichen Kontext zu gewährleisten und die Untersuchung ausreichend abzusichern, werden alle Methoden für diese Arbeit nach gleichem Muster und am selben Untersuchungsobjekt angewandt, um verlässliche und vergleichbare Ergebnisse zu erhalten. Nichtsdestotrotz müssen die Spezifika der einzelnen Methoden beachtet werden, damit die Tests überhaupt durchführbar sind. So kann im Anschluss eine Einordnung der drei Verfahren erfolgen und insbesondere die Eignung und Ergiebigkeit des neuen Video-Testverfahrens herausgearbeitet werden.

Die Arbeit ist unterteilt in folgende Abschnitte: Im Kapitel 2 wird noch ein kurzer Überblick zum Usability-Testing gegeben, woraufhin im Kapitel 3 der Kontext für die Notwendigkeit von Usability-Testing aufgrund der Marktsituation dargestellt wird. Der Stand der Forschung wird in Kapitel 4 aufgearbeitet. Das Untersuchungsobjekt wird in Kapitel 5 erläutert, um deutlich zu machen, worauf die Methoden angewandt wurden. Kapitel 6 erläutert die Rahmenbedingungen der Erhebung mit den Hypothesen, der Auswahl der Probanden, dem Studienkonzept und den drei verwendeten Methoden. Im darauf folgenden Kapitel 7 werden die Ergebnisse dargelegt, woraufhin im Kapitel 8 die Hypothesen geprüft werden. Die Arbeit schließt mit dem Fazit (Kapitel 9), das die Ergebnisse noch einmal kurz diskutiert, und gibt in Kapitel 10 einen Ausblick auf mögliche Folgeforschung.

Als Grundlage und Intention für die Anfertigung der Arbeit dienten folgende Forschungsfragen, die im Vorfeld formuliert wurden. In Kapitel 6.1 werden diese in eine größere Anzahl an Hypothesen, die sich konkreten Einzelaspekten widmen, transformiert.

- Wo liegen die Vor- und Nachteile der einzelnen Verfahren? Für welches Untersuchungsziel ist welche Methode geeignet?
- Welches ist die kosteneffizienteste Methode?
- Wie können die einzelnen Methoden zielführend eingesetzt werden? Ist eine Ergänzung um weitere Methoden bei ausgewählten Verfahren unumgänglich?

2 Usability-Testing

Usability-Tests sind eine am Markt relativ stark verbreitete Methode, denn sie zeigen direkt am Anwendungsfall, an welchen Stellen konkrete Nutzer Probleme mit einem Online-Angebot oder einer Software haben. Vom Grundsatz her als qualitative Methode ausgerichtet, wollen sie mit verhältnismäßig wenigen Probanden die zentralen Probleme einer Anwendung aufdecken. Die Validität der Methode wird bisweilen diskutiert. So weisen etwa MOLICH ET AL. (2004) auf die Abhängigkeit der Testergebnisse vom Studiendesign, der Testleitung und den Probanden hin. Ihres Erachtens sei die Methode demnach umstritten, da ein durchgeführter Test häufig einmalige Ergebnisse zutage fördert und dieser mit anderen Testpersonen oder einer anderen Testleitung zu abweichenden Ergebnissen führen würde. Bei dem von ihnen hierzu durchgeführten Vergleich ließen sie mehrere Expertengruppen Usability-Tests derselben Site durchführen und erhielten größtenteils individuelle Ergebnisse mit nur geringen Überschneidungen bei den Ergebnissen. Fraglos muss dieser Aspekt stets im Hinterkopf behalten werden, wenn ein Test durchgeführt wird. Durch die qualitative Ausrichtung der Methode, die begrenzte Zeit und den begrenzten finanziellen Rahmen, der für kommerzielle Tests i. d. R. zur Verfügung steht, muss dieser Umstand jedoch hingenommen werden und kann schwer aufgelöst werden. Schließlich ist der nutzerzentrierte Ansatz nach wie vor die entscheidende Grundlage. Da dieser bereits das einmalige Auftreten eines Problems als hinreichenden Nachweis ansieht, betrifft dieser Umstand dann nicht aufgefundene Probleme (wenn diese nicht einmal einmalig aufgetreten sind) und das Verhalten des Testleiters. Eine andere Testperson wäre eventuell noch auf andere Probleme gestoßen. Eine abgewandelte Anleitung durch den Testleiter hätte ein abweichendes Verhalten des Probanden verursachen können. So kann hier keine abschließende Sicherheit gewährleistet werden. Im Rahmen dieser Arbeit wird aber durch personelle Kontinuität und die spezifische Anpassung des Studienkonzepts an die einzelnen Methoden der Versuch unternommen, dieses Problem zumindest so gut wie möglich zu vermeiden.

Der klassische Usability-Test im Labor ist mit einem relativ hohen Aufwand verbunden: Testperson und Testleitung müssen zur selben Zeit am selben Ort sein, um den Test überhaupt durchführen zu können. Daneben ist ein Mindestmaß an technischer und räumlicher Ausstattung nötig (vgl. SARODNICK/BRAU 2011, S. 168 f.). An den Testleiter werden hohe Ansprüche gestellt, da dieser einen entscheidenden Einfluss auf den Verlauf des Tests hat und schon in der Konzeptionsphase Einfluss auf den Testverlauf und den potenziellen Erkenntnisgewinn nehmen kann. Aufgrund der qualitativen Ausrichtung der Methode wird versucht, mit wenigen Testpersonen ein möglichst umfassendes Ergebnis zu erhalten. Auch deswegen ist der Usability-Test als solches nicht unumstritten. So kann u. a. ein nicht ausreichend elaboriertes Studienkonzept dafür sorgen, dass die Ergebnisse nicht die wichtigsten Usability-Probleme zutage fördern. Auch hängt der Erfolg von den Probanden ab, die individuell innerhalb der Testsitzungen agieren und so direkten Einfluss auf den Testverlauf und Testerfolg haben. Auch andere Testpersonen können dafür sorgen, dass andere

Ergebnisse erzielt werden. Auf die möglichen Probleme (vgl. MOLICH ET AL. 2004) wurde bereits hingewiesen (s. o.). Diese müssen also bei jeder Erhebung im Hinterkopf bleiben und zeigen auf, dass die Ergebnisse stets unter dieser Prämisse zu betrachten wären. Für die Erhebung im Rahmen dieser Arbeit wird daher zumindest auf Seiten der Projektleitung auf personelle Kontinuität geachtet, d. h. der Verfasser der Arbeit wird gemeinsam mit einem fest zugewiesenen Mitarbeiter die Erhebungen planen und durchführen. Auf Seiten der Probanden ist dies kaum zu gewährleisten. Eine doppelte Teilnahme der Probanden an mehreren Methoden ist nicht praktikabel, da diese sich dann aufgrund der Bekanntheit der Oberfläche beim folgenden Test anders verhalten würden. Es muss daher mit drei unterschiedlichen Gruppen an Probanden gearbeitet werden (zur näheren Methodik s. Kapitel 6.3). Deren individuelle Verhaltensweisen können kaum beeinflusst werden. Lediglich durch eine bestimmte Vorauswahl können ähnliche Zielgruppen akquiriert werden, die z. B. eine vergleichbare Interneterfahrung vorweisen, die gleiche Geschlechterverteilung haben und gleiche Altersspannen abdecken. Auch das Interesse für das Untersuchungsobjekt sollte über alle Probanden in etwa ähnlich verteilt sein, um einen angemessenen Bezug zum Produkt zu gewährleisten.

Dennoch lässt sich insbesondere beim Labortest mit der geringsten Anzahl an Probanden eine Verzerrung aufgrund des individuellen Verhaltens der teilnehmenden Personen nicht gänzlich ausschließen. Unter Berücksichtigung dieser und der anderen Grenzen der anzuwendenden Methoden lassen sich mit jeder der Methoden bei entsprechendem Problembewusstsein dennoch gute und v. a. praxisrelevante Ergebnisse erzielen, die wie viele wissenschaftliche Methoden spezifischen Grenzen ausgesetzt sind. So soll der Usability-Test auch als eine Teilmethode von verschiedenen möglichen Instrumenten eingesetzt werden und zeigt seine Stärken meist erst durch das Zusammenspiel mit anderen Methoden, wenn etwa die Analytics-Daten einer Website (vgl. EBERHARD-YOM 2010, S. 140 f.) Fragestellungen liefern, die im Usability-Test dann aufgegriffen und überprüft werden können.

3 Notwendigkeit von Usability-Tests

Die Motivation für Usability-Forschung lässt sich v. a. auf die Marktverhältnisse zurückführen bzw. erhält dadurch besondere Relevanz für die Praxis, für die die vermeintlichen methodischen Defizite (vgl. Kapitel 2) nur eine untergeordnete Rolle spielen.

Der deutsche Markt für Onlineshops ist sehr fragmentiert und war im Jahr 2010 auf ca. 150.000 Onlineshops aufgeteilt (vgl. GROSS 2010).

An erster Stelle muss ein Onlineshop also gewährleisten, dass er überhaupt besucht wird. Dies kann er über ein Maßnahmenpaket erreichen, das unter anderem aus Suchmaschinenoptimierung besteht und Online-Werbung beinhalten kann (vgl. EBERHARD-YOM 2010, S. 9 f.). So kann der Shop seine Bekanntheit steigern und Besuche generieren. Aufgrund der genannten Zahlen über die Gesamtanzahl an Anbietern ist die Wahrscheinlichkeit sehr hoch, dass es für den Anbieter eines sehr speziellen Shops auch einen Mitbewerber mit einem identischen oder sehr ähnlichen Sortiment gibt.

Wird nun ein Onlineshop von einem Nutzer aufgesucht, den dieser beispielsweise auf der Suche nach einem speziellen Produkt über eine Suchmaschine aufgerufen hat, entsteht beim Nutzer sofort ein Eindruck in der ersten Nutzungsphase des neuen Anbieters. Denn die Nutzer haben sehr spezifische Erwartungen an einen Shop, die vor allem aus ihrer Erfahrung stammen. Ein Shop soll für sie erwartungskonform sein, d. h. sie haben sehr konkrete Vorstellungen aus ihrem bisherigen Erfahrungsschatz, die sie dann auf einen ihnen bislang unbekannten Shop ebenso anwenden (vgl. MOHS/NAUMANN/KINDSMÜLLER 2007, S. 3). Kann der bislang unbekannte Shop diese Erwartungen nicht in ausreichendem Maße erfüllen oder kommt der Nutzer an einer bestimmten Stelle einfach nicht weiter, da ihm nicht klar ist, wie er eine Aktion ausführen kann, entsteht schnell ein Frustrationserlebnis, das bis zum Abbruch der Sitebenutzung führen kann. In einem solchen Fall waren die vorherigen Bemühungen und Ausgaben, um den Nutzer zum Besuch der Site zu bewegen (s. o.) nicht zielführend. Hat der Nutzer nun das Gefühl, an den oder die gewünschten Artikel auch in einem anderen Shop gelangen zu können, genügt die kurze Suche nach einem Alternativangebot und dieser Nutzer trägt nicht mehr zum Umsatz des Shops bei.

Aufgrund der oben angeführten Fragmentierung des deutschen Marktes für Onlineshops sind daher die Erwartungskonformität und das Nutzererlebnis nicht als rein kosmetische Faktoren zu betrachten, die es dem Nutzer ermöglichen, schneller zum Ziel zu kommen oder Zufriedenheit bei ihm zu erzeugen. Dies betrifft allenfalls Intranets oder ähnliche firmeninterne Anwendungen, die sowieso benutzt werden müssen und mit einer Steigerung der Benutzerfreundlichkeit zu schnelleren Bearbeitungszeiten und damit auch geringeren Kosten führen können.

Im Onlineshopping wird damit ein Kaufabschluss verhindert, der aber das Geschäftsziel des Shops ist. Aufgrund der hohen Anzahl an Alternativen kann ein Shop zwar mit einem besonderen

Sortiment hervorstechen. Da die Wahrscheinlichkeit sehr hoch ist, dass es den entsprechenden Mitbewerber mit einem vergleichbaren Sortiment gibt, muss die Usability auf dem Shop so gut umgesetzt sein, dass der Erstbenutzer sich sofort zurechtfindet. Usability lässt demnach nicht viel Spiel für Freiraum auf Seiten des Anbieters. Es geht darum, dem Nutzer einerseits durch die Gestaltung einen Anreiz für die Nutzung zu bieten. Durch Erwartungskonformität kann auf der anderen Seite erreicht werden, dass der Nutzer sich schnell zurechtfindet und das Ziel im Shop auf zufriedenstellende Weise erreicht. Solche Maßgaben können bereits in der Phase der Sitegestaltung berücksichtigt werden, also bevor die eigentlich Webpräsenz online geht. Zu diesem Zweck gibt es eine Reihe an sog. Guidelines, die die wichtigsten Punkte beinhalten, wenn man mit der Erstellung von Interaktionsprozessen startet (vgl. u. a. SARODNICK/BRAU 2011, S. 126 ff.). Die so entstehende Grundlage sollte schon zum Großteil erwartungskonform sein und dürfte daher relativ gut benutzbar sein.

Da aber die Komplexität eines Onlineangebots insbesondere im Bereich des Onlineshoppings sehr hoch ist, empfiehlt es sich, Nutzer einzubeziehen. Dies kann und sollte bereits im Entwicklungsprozess erfolgen, da so verhindert werden kann, dass bestimmte Probleme erst zu spät festgestellt werden und dann nicht mehr oder nur noch mit sehr hohem Aufwand behoben werden können (vgl. KRUG 2006, S. 134 f.). Daher ist trotz der Einhaltung von Guidelines und der vermeintlich durchdachten Gestaltung der Website in jedem Fall das Einbeziehen der Nutzer als wichtiges Instrument zu betrachten. Schließlich wird dadurch verhindert, dass es später zu Kaufabbrüchen kommt und so der Gewinn des Shops verringert wird.

Denn der Wettbewerb ist groß und es gibt einen Mitbewerber, der die Kundenbedürfnisse besser befriedigen kann und daher den Vorzug erhalten kann. Erst die ganzheitliche Kombination von Maßnahmen, die aus der richtigen Maßnahmenkombination zur Besuchergewinnung, der Beachtung von Guidelines bei der Sitegestaltung und der Erhebung tatsächlicher Nutzermeinungen besteht, ermöglicht den erfolgreichen Betrieb eines Onlineshops. Experimente in der Gestaltung und Bedienung des Shops sind daher nur selten zielführend, da die Nutzer schnell einen Kauf abschließen wollen und bei ungewohnten Erlebnissen auf einem Shop eher Abneigung als Interesse an Neuem verspüren. Um im Wettbewerb zu bestehen, muss daher diese Kombination an Maßnahmen angewandt werden, in der die Einbindung der Nutzer eine zentrale Rolle spielt, da erst dadurch qualitative Äußerungen gewonnen werden, die tatsächliche inhaltliche Erklärungen für Frustration beim Nutzer erlauben. Dass daneben etwa auch der ständige Nutzerstrom überwacht und über die Webanalyse-Daten ausgewertet werden sollte, ist offenbar. Denn gerade im Dauerbetrieb sind diese Kennzahlen die wichtigste Datenquelle. Wurde etwa ein Teil am Shop verändert oder ist ohne Veränderung am Shop ein deutlicher Besucher- oder Kaufrückgang festzustellen, müssen die vorangegangenen Maßnahmen hinterfragt werden oder es muss festgestellt werden, ob es eine Veränderung am Markt gab. Liegen externe Gründe vor, kann auch ein Usability-Test helfen. Dies ist z. B. dann der Fall, wenn ein

Mitbewerber Veränderungen an seinem Shop vorgenommen hat und dadurch einen Teil der Besucher und Käufer vom Shop abgeworben hat. Dann kann neben der Konkurrenzbeobachtung eine Untersuchung des eigenen Shops Möglichkeiten zutage fördern, wie das Potenzial desselben gesteigert werden kann.

Da Usability-Tests im Labor relativ teuer sind und ein entsprechender Vorlauf für ein derartiges Projekt einzuplanen ist, erscheint die Betrachtung von alternativen Testmöglichkeiten, wie er in dieser Arbeit erfolgt, sinnvoll zu sein. So soll festgestellt werden, inwieweit das Potenzial von Usability-Tests auch mit neueren Alternativen und ggf. verringerten Kosten und Aufwänden realisiert werden kann.

4 Stand der Forschung

Der konkrete Vergleich der drei Methoden wurde in dieser Form bislang nicht wissenschaftlich durchgeführt bzw. nicht dokumentiert. Daher wird hier weitestgehend auf Untersuchungen verwiesen, die zwar auch den Vergleich von Usability-Methoden zum Thema haben. Diese vergleichen dann jedoch meist andere Untersuchungstypen, können aber dennoch hilfreiche Anhaltspunkte liefern. Dieser Überblick macht auch die Beweggründe deutlich, die hinter der Anfertigung dieser Studie stehen.

Auf einen Text mit veröffentlichten Ergebnissen aus einem Vergleich von einem Videotest und dem klassischen Usability-Test sei hier verwiesen. So schreibt STENFELDT (2012) über einen in der Praxis durchgeführten Vergleich, bei dem der Videotest gleich viele Probleme aufdeckte wie der klassische Usability-Test. Empfohlen wird hier die Formulierung klarer Aufgabenstellungen, um den Videotest möglichst ergiebig zu machen.

BOSENICK (2009, S. 280 f.) gibt einen Kurzüberblick über Remote-Testing. Beim synchronen Test wird das Potenzial ähnlich hoch eingeschätzt wie im Labor, da dort die Betreuung durch den Testleiter ebenso vorhanden ist. Den Remote-Verfahren räumt er aufgrund der natürlicheren Testumgebung per se ein höheres „Involvement der Probanden“ (a. a. O., S. 281) ein. Dies ist eine im Rahmen der Erhebungen in jedem Fall zu prüfende Aussage, denn sie erscheint durchaus nachvollziehbar, könnte aber genauso auch gegenteilig der Fall sein, wenn die Probanden in der vertrauten Umgebung und ohne unmittelbare Anleitung durch den Testleiter eher an einer raschen Durchführung zur Erlangung der Aufwandsentschädigung als an einer konzentrierten Abarbeitung der gestellten Aufgaben interessiert sind.

BOSENICK (2009, S. 284) weist dennoch abschließend daraufhin, dass für reine Usability-Fragestellungen dem Labortest aufgrund der Möglichkeit der Interaktion mit dem Nutzer eher der Vorzug zu gewähren ist. Der Remote-Test ist seines Erachtens daher eher geeignet, um z. B. die Wege zu dokumentieren, die der Nutzer zur Erledigung einer bestimmten Aufgabe gegangen ist, um so die „Customer Experience“ (ebd.) zu erfassen und dies ggf. sogar mit einer größeren Zielgruppe zu tun, um die Ergebnisse durch größere Häufigkeiten absichern zu können.

Eine grundlegende Studie für den Vergleich von Usability-Methoden lieferte ECKERT (2008), die die drei Verfahren Labortest, synchronen Remote-Test und den asynchronen Remote-Test miteinander vergleicht. Den Untersuchungsgegenstand legte sie auf das Web 2.0, die Ergebnisse sind dennoch eine wichtige Hilfestellung für diese Studie. Anhand von Thesen wurde der Vergleich unterstützt. So wies ECKERT (2008, S. 80 ff.) u. a. nach, dass die Kommentare im asynchronen Remote-Test deutlich weniger ausführlich sind als die gesprochenen Kommentare in den beiden anderen Testverfahren. Entgegen ihrer thesenhaft geäußerten Vermutung über eine kritischere Nutzerschaft in den

ortsunabhängig durchgeführten Testverfahren wurden im Labortest deutlich mehr kritische Kommentare geäußert als im asynchronen Remote-Test. Der synchrone Remote-Test erreichte einen ähnlichen Wert wie der Labortest. Die soziale Dynamik des angeleiteten bzw. begleiteten Tests führte also wider Erwarten eher zu kritischeren Äußerungen als zur Zurückhaltung der Probanden. Dies zeigt auch deutlich, dass die Erwartungen an die verschiedenen Testverfahren von verschiedenen Autoren im Vorfeld durchaus kontrovers gewertet werden (s. o.), sodass i. d. R. erst ein Test mit den tatsächlichen Probanden aufzeigt, wie es um das Involvement und die Menge an kritischen Äußerungen tatsächlich bestellt ist.

Nicht bestätigt werden konnte die Vermutung, die Probanden würden sich im asynchronen Remote-Test tendenziell eher zu kritischen als zu positiven Gesichtspunkten äußern. Dies wird jedoch auch auf das Studienkonzept zurückgeführt, das eine freie Nutzung nur sehr eingeschränkt ermöglichte. Zu dieser hätten sonst eher positive Kommentare erwartet werden können. Im asynchronen Verfahren ist ohnehin die Auswertung einer freien Nutzungsphase schwierig, da in aller Regel die aktuelle Intention des Nutzers nicht bekannt ist und so der Klickpfad zwar angeschaut, aber nicht nachvollzogen werden kann, weshalb bestimmte Schritte durchgeführt wurden.

Als zentrale These wurde die Anzahl der zu identifizierenden Usability-Probleme genannt, die von einer gleichen Anzahl an aufgedeckten Problemen im Labortest und im synchronen Remote-Test ausging, wohingegen im asynchronen Remote-Test deutlich weniger erkannte Probleme erwartet wurden. Dies wurde in Teilen bestätigt, der asynchrone Remote-Test machte insgesamt weniger Probleme ausfindig. Aber auch der synchrone Remote-Test blieb hinter den Quantitäten des Labortests etwas zurück. In jedem Fall sollte im asynchronen Remote-Test mit einer geringeren Anzahl an aufgedeckten Usability-Problemen gerechnet werden. Im Labortest und im synchronen Remote-Test können bestimmte Probleme durch eine expertenbasierte Vorevaluation ggf. identifiziert werden oder live im Verlauf eines Tests in Ansätzen gefunden werden, sodass daraufhin individuelle Nachfragen möglich werden, die im asynchronen Remote-Test allenfalls durch eine problemorientierte Studienkonzeption in Ansätzen zu realisieren sind, jedoch nur mit einer geringeren Wahrscheinlichkeit tatsächlich identifiziert werden dürften.

Die möglichen Vorteile einer Interaktion mit den Probanden wurden in Abhängigkeit von den erkannten Usability-Problemen evaluiert. Hierbei zeigte sich zwar, dass alle drei Verfahren die schwersten Usability-Probleme aufdeckten. Bei weniger schweren Problemen zeigt sich indes, dass die gelenkten Verfahren – teilweise auch nur aufgrund des diesen Testmethoden inhärenten lauten Denkens – durch die Interaktion des Testleiters und die Möglichkeit, gezielte Nachfragen zu stellen, bewusst Aussagen zu weniger gravierenden Problemen provozieren können. Dies ist als Vorteil dieser Methode zu sehen, denn es ist z. B. für die ganzheitliche Usability-Optimierung wichtig, einem Kunden neben den gravierenden Usability-Problemen auch weitere kosmetische bzw. weniger dringende Verbesserungsmöglichkeiten aufzeigen zu können. ECKERT (2008, S. 88) stellt jedoch noch

heraus, dass z. B. ein erfahrener Studienersteller ggf. durch bereits im Vorfeld aufgestellte Nachfragen auch mit dem asynchronen Remote-Test ein gutes Ergebnis im Hinblick auf weniger gravierende Usability-Probleme erzielen kann, wenngleich der Verzicht auf die direkte Interaktion und Möglichkeit zu individuellen Nachfragen hier Grenzen setzt.

Die von ECKERT (2008) dokumentierten Ergebnisse dienen als eine Konzeptionsgrundlage für die Erhebung im Rahmen dieser Arbeit, da sie bereits einen ausführlichen Vergleich durchführte, wenn auch an einem spezifischen Fallbeispiel. Schließlich soll für diese Untersuchung auch ein Fallbeispiel verwendet werden, das dann auch mit drei Methoden untersucht wird, wobei entsprechend auf den synchronen Remote-Test verzichtet wird, da dieser bereits in der genannten und weiteren Studien eine ausreichende Würdigung erfuhr. Im Gegenzug wird dann der Videotest betrachtet, da bei diesem ein großes Erkenntnisinteresse besteht.

Bereits seit etwa zehn Jahren gibt es weitere, veröffentlichte Studien, die auf Unterschiede bei den verschiedenen Erhebungsverfahren eingehen.

TULLIS ET AL. (2002) verglichen den Labortest mit ortsunabhängigen Verfahren und stellten ein ähnliches Verhalten der Nutzer in beiden Testumgebungen fest. Im ortsunabhängig durchgeführten Test erhielten die Studienleiter ausführliche Kommentare, die vom Gehalt her mit denen aus dem Labortest vergleichbar sind, obgleich diese durch Daten der Verhaltensbeobachtung noch ergänzt wurden. Von Vorteil ist ihrer Ansicht nach auch die höhere Stichprobengröße, die der asynchrone Test ermöglicht. Weiterhin stellen sie die geringe Stichprobengröße eines Labortests als Nachteil heraus, da dessen Verlässlichkeit dadurch limitiert wird. Hier bevorzugen sie die größeren Stichproben im asynchronen Testing, da diese mehr unterschiedlichere Testpersonen abdecken können. Dennoch weisen sie auch darauf hin, dass bestimmte Probleme nur im Labor zu identifizieren sind, was als Vorteil des Labortests zu werten ist. Geht es dem Auftraggeber nur um das Ausfindigmachen der schwerwiegendsten Probleme, kann er dies auch mit einem (evtl. preisgünstigeren) asynchronen Remote-Test bewerkstelligen.

WATERSON ET AL. (2002) stellen ein Tool zur Visualisierung von Clickstreams vor. Clickstreams werden zur Auswertung von asynchronen Usability-Tests verwendet, da sie die Darstellung des Weges ermöglichen, den der Nutzer gegangen ist. So kann festgestellt werden, ob der Nutzer den vorgesehenen Weg zur Erfüllung einer Aufgabe gegangen ist oder einen alternativen Weg eingeschlagen hat. Auch weisen sie auf weitere Funktionen des Tools hin wie beispielsweise die Anzeige der Aufenthaltsdauer auf einer Seite, was ein Hinweis auf Probleme mit bestimmten Seitenbereichen sein kann. Eine solche Visualisierung ist der Schlüssel zum raschen Verständnis des Nutzerverhaltens in asynchronen Remote-Tests, da dieses Verhalten – von einem Nutzer oder aggregiert von mehreren Nutzern – übersichtlich dargestellt am ehesten schnell verstanden werden

kann. Dies vereinfacht so die Auswertung. In Abhängigkeit von einer zielgerichteten und adäquaten Studienkonzeption kann so eine zeitsparende und ergiebige Auswertung begünstigt werden.

BRUSH/AMES/DAVIS 2004 dokumentieren eine Studie, in der sie den synchronen Remote-Test mit dem Labortest verglichen, und zwar bei der Betrachtung einer Nutzeroberfläche für Experten. Die Menge an gefundenen Usability-Problemen ist für beide Methoden identisch. Einige Probanden, die an Tests mit beiden Methoden teilnahmen, empfanden die Remote-Variante als praktischer. Ein Viertel dieser Teilnehmer fühlte sich allerdings im Labortest wohler dabei, mit dem Testleiter zu sprechen als über das Telefon, das hier als Tonverbindung für den Remote-Test Verwendung fand. Die Ergebnisse werden dahingehend problematisiert, dass die getestete Software allen Teilnehmern geläufig war. Eventuell ist ein anderes Verhalten, das weniger Usability-Probleme zutage fördert, zu erwarten, wenn eine für die Probanden neue Software oder Internetseite getestet wird.

Eine Reihe weiterer Arbeiten beschäftigte sich mit dem ortsunabhängigen Usability-Test und befasste sich mit speziellen Teilthemen. KODIYALAM (2003) entwickelte etwa ein eigenes Tool für die Aufzeichnung von Remote-Usability-Tests. Dies lässt darauf schließen, dass das Remote-Testing schon länger aktuell ist und hier bereits der Versuch unternommen wurde, die User Experience einer schwer zu erreichenden Zielgruppe bei der Arbeit mit einer Website oder einem Programm beobachten zu können.

In seiner Abschlussarbeit dokumentiert SELVARAJ (2004) eine vergleichende Studie von klassischem Labortest und synchronem Remote-Usability-Test. Die beiden Verfahren sind in Bezug auf die Ergebnisse insgesamt ähnlich. Allerdings stellte sich in der Erhebung heraus, dass im Labortest mehr schwerwiegende Usability-Probleme ausfindig gemacht werden konnten.

ANDREASEN ET AL. (2007) stellten dem Labortest drei unterschiedliche Varianten des ortsunabhängigen Tests gegenüber, und zwar den synchronen Remote-Test, den asynchronen Remote-Test mit einem potenziellen Endnutzer und den asynchronen, durch einen Experten durchgeführten Remote-Test. Ihr Untersuchungsobjekt war ein E-Mail-Programm. Für die asynchrone Datensammlung standen hier ausschließlich freie Texteingaben der Probanden zur Verfügung, die die Schwere der Usability-Probleme nach eigenem Ermessen festlegen durften. Auch hier wird festgestellt, dass der synchrone Remote-Test beinahe austauschbar ist mit einem laborgebundenen Test, was die aufgefundenen Probleme angeht. Die asynchronen Verfahren fördern wiederum weniger Usability-Probleme zutage und sind i. d. R. aufwändiger für die Teilnehmer, die hier stets nach Erledigung einer Aufgabe zu einer freien Texteingabe aufgefordert werden. Beim lauten Denken kann dies direkt und unbewusst im Laufe der Bearbeitung erfolgen, was für den Probanden praktischer ist. Dennoch lassen sich mit den getesteten asynchronen Verfahren auch wichtige Erkenntnisse erzielen, die dann entsprechend Berücksichtigung finden. Mit den mittlerweile üblichen Clickstream-Analysen bei asynchronen Remote-Tests ließen sich diese Daten heute noch sehr sinnvoll ergänzen. Weiterhin

ist die mögliche höhere Anzahl an Probanden auch ein wichtiger Punkt, da dies beim Labortest und dem synchronen Remote-Test aufgrund der deutlich höheren Datenmenge schwerer zu realisieren wäre.

SCHULZ (2008) verglich auch drei verschiedene Verfahren, allerdings die heuristische Evaluation, das Eyetracking und den Labortest, den er v. a. durch seine Spezifikation des lauten Denkens charakterisiert. Er stelle die allgemeinen Probleme der Verfahren dar wie etwa das nicht ausblende Expertenwissen, wenn die Experten in die Rolle des Nutzers schlüpfen (vgl. a. a. O., S. 8), oder die Beeinflussung eines Tests durch die künstliche Laborumgebung (vgl. ebd.). Für diese Untersuchung bleibt insbesondere festzuhalten, dass der Autor dem lauten Denken den Vorzug einräumt. So kann die Testleitung die verbalen Äußerungen der Probanden unmittelbar mit dem zeitlich synchronen Verhalten in Verbindung bringen und so eine stark erleichterte Auswertung von Problemen durchführen (vgl. a. a. O., S. 11). Weiterhin stellt SCHULZ (2008, S. 18) heraus, dass er bereits mit acht Teilnehmern beim Labortest alle schwerwiegenden Probleme der untersuchten Anwendung aufgedeckt hatte und so diese Teilnehmeranzahl als ausreichend einstufte.

Daneben gibt es noch eine Reihe weiterer Studien, die für die Vorbereitung der empirischen Erhebung der Arbeit gesichtet wurden. GERTH/KRACH/PILOTTEK (2008) entwickelten etwa ein Software-Tool für asynchrone Remote-Tests, das deren Potenzial erweiterte und den Versuch unternahm, diese noch mehr an den Labortest anzunähern. So war es ihr Bestreben, beim asynchronen Test durch einen erweiterten Funktionsumfang und einen sog. virtuellen Moderator ergiebiger Funde zu erhalten. Der Funktionsumfang ermöglichte u. a. das Abspielen eines Videos mit den Nutzeraktionen, was als besondere Toolfunktionalität zu sehen ist, die über den üblichen Funktionsumfang von Remote-Tools wie das Aufzeichnen von Clickstreams noch hinausgeht und damit die Auswertung erleichtert. Für diese Arbeit lag kein derart mächtiges Tool vor, sodass keine vergleichbar große Datenmenge erhoben werden konnte. Die genannte Ausarbeitung zeigt aber auf, dass das Potenzial des asynchronen Remote-Testings durchaus mit entsprechend gut ausgestatteter Software angehoben werden kann. Dies wirkt sich dann jedoch auch deutlich auf die benötigte Auswertungszeit aus, die so stark ansteigt, aber auch in Teilen eine Vermeidung der ansonsten vorhandenen Nachteile des asynchronen Remote-Verfahrens bewirken kann. Dennoch können auch mit einem elaborierten Tool gewisse konzeptuelle Nachteile nicht verhindert werden. Diese liegen in der un gelenkten Durchführung des Tests, die keine Kontrolle über die Testperson bietet und keine Möglichkeit zum Stellen gezielter Nachfragen an den Probanden bietet.

Der Überblick über die betrachteten und ausgewerteten Forschungsarbeiten zeigt auf, dass die Möglichkeiten und Grenzen einzelner Verfahren schon von vielen Forschern bewertet wurden. Auch wenn sich diese vorwiegend an Fallbeispielen orientierten, zeigen die Ergebnisse dennoch, dass es verschiedene Usability-Test-Verfahren gibt, die eingesetzt werden können. Ist man sich über die Stärken und Schwächen der ausgesuchten Methode im Klaren, kann durch die problematisierte

Betrachtung der Ergebnisse eine zielführende Würdigung für den praktischen Kontext oder die Anwendung im Forschungsbereich erfolgen.

5 Untersuchungsobjekt

Die Erhebungen für die Arbeit wurden an einem konkreten Untersuchungsobjekt durchgeführt, und zwar an einem deutschsprachigen Onlineshop.

Der Shop ist ein Special-Interest-Shop, der unter der URL www.mein-schnulli.de erreichbar ist und u. a. Schnuller mit eingraviertem Namen anbietet (vgl. Abbildung 1). Das Sortiment ist relativ überschaubar und wird unterteilt in vier Hauptrubriken angeboten. Eine Suchfunktion wird dort ebenso wenig angeboten wie die Filterung der Artikel, da je nach Rubrik maximal 27 Produkte vorhanden sind, die so nach kurzem Scrollen bereits komplett erfasst werden können. Hauptmerkmal des Shops ist die Individualisierung der Produkte. Im Vorfeld wurde der Shop analysiert, um Untersuchungsansätze zu finden. Auch fand ein Austausch mit der Betreiberin des Shops statt, die auf mögliche Probleme hinwies, die sich anhand der Webanalyse-Daten des Shops bereits erahnen ließen. Diese wurden dann in der Studienkonzeption berücksichtigt, um dort durch die Erhebung qualitativer Daten zu diesen Problemen weitere Erkenntnisse liefern zu können.

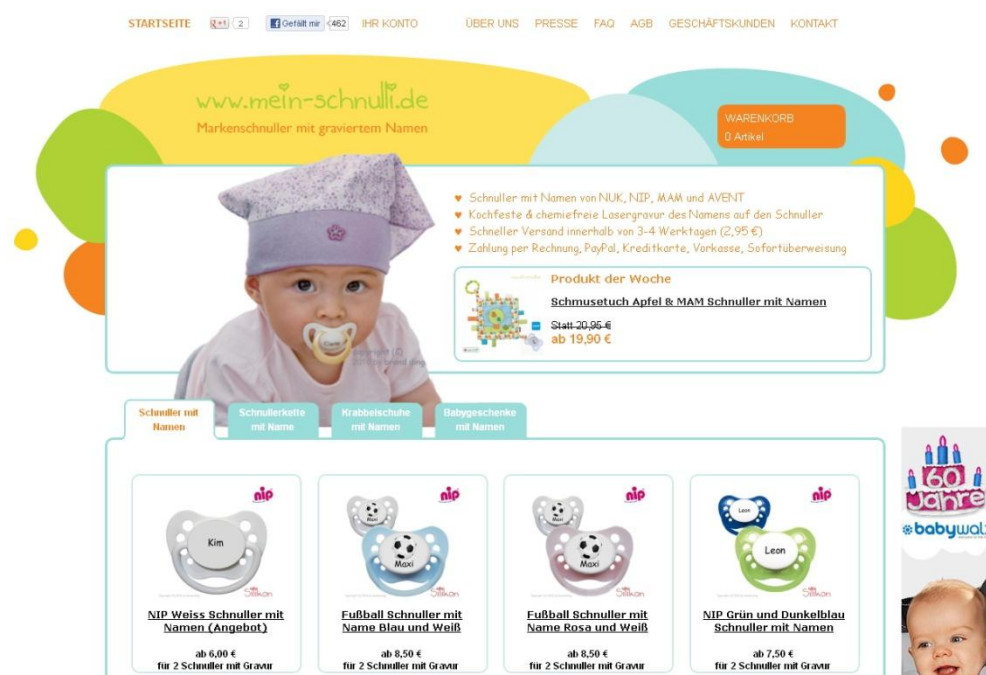


Abbildung 1: Startseitenausschnitt des Onlineshops www.mein-schnulli.de.

So wurden von der Betreiberin mögliche Probleme dargelegt, die mit der Bestellung individualisierter Artikel zusammenhängen können. Auch wurden Probleme im Hinblick auf den Bezahlprozess erläutert, wenn die Probanden etwa eine von der Rechnungsadresse abweichende Versandadresse wählen sollen. Ein solcher „Kick-Off-Workshop“ (EBERHARD-YOM 2010, S. 137) mit einem Shopbetreiber bildet eine wichtige Grundlage, um die Ergiebigkeit eines Tests zu erhöhen und eine praxisnahe Zusammenstellung der Aufgabenstellungen zu erreichen, da Aufgabenstellungen, die den Nutzern keine Probleme bereiten, nicht unbedingt in einem aufwändigen Test geprüft werden sollten. Stattdessen sollte eher darauf hingearbeitet werden, bei problematischen Stellen möglichst viele Rückmeldungen der Nutzer zu diesem Seitenbereich zu erhalten, um ein tatsächliches Optimierungspotenzial auszeigen zu können. Eine verbreitete Nutzerzufriedenheit muss in einem Usability-Test nicht nachgewiesen werden, da dies nicht dessen Zielsetzung ist.

Kurz vor dem Test gab es am Shop noch eine größere Umstrukturierung, und zwar wurden die Versandkosten verringert und dafür Zahlungsgebühren bei einigen Zahlungsarten eingeführt bzw. erhöht. Insbesondere hierzu kann ein qualitativer Austausch zu diesen Rahmenbedingungen zu wichtigen Erkenntnissen führen, die deutlich mehr Informationsgewinn bieten als die vermehrte Feststellung von Bestellabbrüchen, wie sie die Analysedaten aufzeigen können. Dies wäre zwar kein klassisches Usability-Problem, muss jedoch als zentrales Problem, das die Nutzer mit dem Shop haben, im Rahmen eines solchen Tests ebenso festgehalten werden. So kann sich bei einem solchen Aspekt auch zeigen, dass dieser Umstand nicht ausreichen kommuniziert wird und die Nutzer nur unzufrieden damit sind, so spät informiert worden zu sein. Auch um die Ursache einer solchen Frustration aufzuzeigen, kann der Usability-Test auf den Webanalyse-Daten aufbauen und die Lücken für nicht erklärbare Tatsachen auffüllen.

6 Grundlagen für die Erhebung

Dieser Abschnitt beschreibt die drei untersuchten Methoden für Usability-Tests und grenzt sie durch die jeweiligen Beschreibungen voneinander ab. Als Vorannahme für alle drei Verfahren gilt, dass für alle Testarten im Vorfeld Studienkonzepte zu erstellen sind, die den Ablauf eines Tests mit Szenarien vorgeben, um sicherzugehen, dass alle Probanden identische Aufgaben abarbeiten, um so einen Überblick über das Nutzerverhalten verschiedener Probanden zu erhalten und die Probanden nicht nur ungelenkt durch die Anwendung/Website zu schicken, da ohne vorformuliertes Erkenntnisinteresse die Ableitung von Empfehlungen schwierig ist. Auch erleichtert diese „szenariobasierte Nutzung“ (EBERHARD-YOM 2010, S. 136) den Vergleich der verschiedenen Testverfahren, da schließlich alle Nutzer einen identischen Ablauf absolviert haben. In diesem Rahmen ist es aber sinnvoll, dem Nutzer dennoch etwas Freiraum zu lassen, um einen Einkauf nach seinem Interesse durchzuführen. So kann der Kauf eines Wunschartikels authentischer für den Nutzer sein als der Kauf eines vorgegebenen Artikels.

Um neben den Erkenntniszielen für den Shop auch die Erkenntnisziele für den Methodenvergleich als Vorgabe für alle Tests zu haben, wurde neben dem Aufstellen der im folgenden Abschnitt folgenden Hypothesen die Auseinandersetzung mit den einzelnen Methoden anhand bisheriger Erfahrungen im Unternehmen und ausgewählter Literatur durchgeführt, um sicherzustellen, dass die Methoden korrekt angewandt werden und es keine vermeidbaren Fehler bei der Planung und Durchführung der Erhebungen gibt, die die angestrebte Allgemeingültigkeit der Ergebnisse verhindern würden.

6.1 Hypothesen

Um den Methodenvergleich systematisch auswerten zu können, wurden Hypothesen aufgestellt, anhand derer ein guter Vergleich der untersuchten Methoden möglich ist. Es wurden acht Hypothesen aufgestellt, deren Beantwortung im Anschluss an die Auswertung der drei erhobenen Methoden möglich gemacht werden soll (vgl. Kapitel 8). Diese wurden einerseits aus Vermutungen und bisherigen Erfahrungen mit Usability-Tests hergeleitet und auch durch Vorschläge von den Kooperationspartnern bei den Usability-Verfahren ergänzt und dann in den Hypothesen zusammengeführt. Diese lauteten:

1. Der Labortest ist durch seine Interaktivität ergiebiger als die beiden asynchronen Verfahren und deckt so mehr Usability-Probleme auf als diese. Der Videotest deckt indes mehr Probleme auf als der asynchrone Remote-Usability-Test.

2. Im asynchronen Videotest werden genauso viele schwere Usability-Probleme aufgedeckt wie im Labortest.
3. Im Labortest und im asynchronen Videotest werden in der gleichen Erhebungszeit gleich viele Usability-Probleme aufgedeckt.
4. In den testleiterunabhängigen Verfahren ist ein natürlicheres Verhalten der Teilnehmer festzustellen als im Labortest, wo Nutzer etwa Kommentare beeinflusst durch soziale Erwünschtheit abgeben oder unnatürlich später aufgeben, um unbedingt zur Lösung zu gelangen, wenn sie zeitgleich beobachtet werden.
5. Der Videotest und der asynchrone Remote-Usability-Test können bei jeder Testperson zu individuelleren Abläufen und Ergebnissen führen, da die Nutzer sich nur an den Aufgabenstellungen orientieren und mit diesen nicht interagiert wird. Im Labortest kann der Nutzer hingegen während des gesamten Testverlaufs gezielt gelenkt werden und z. B. gewollt auf einzelne Aspekte gestoßen werden.
6. Der Videotest und der asynchrone Remote-Usability-Test können vielfältige Probleme aufdecken, da jeder Testrechner eine individuelle Konfiguration hat (z. B. unterschiedliche Bildschirmgrößen). In der Aufzeichnung des Videotests können diese Auswirkungen sogar direkt gesehen werden und so bei der Auswertung besser berücksichtigt werden.
7. Der Videotest liefert in kurzer Zeit mit relativ geringem Aufwand für den Auftragnehmer direkt verwertbare Ergebnisse, aus denen dieser auch selbsttätig relativ gut die wichtigsten Erkenntnisse ableiten kann.
8. Die Ergebnisse sind von der verwendeten Testmethode weitgehend unabhängig, da diese eher von anderen Rahmenbedingungen wie dem Studienkonzept und der Interpretation der Ergebnisse im Nachgang abhängig sind.

6.2 Probandenauswahl

Um zielführende Ergebnisse zu erhalten und auch die weitere Verwertbarkeit der Ergebnisse für den Shopbetreiber gewährleisten zu können, war die Wahl der Probanden von großer Wichtigkeit, da

der Shop ein begrenztes Artikelspektrum für eine klar abgesteckte Zielgruppe anbietet und nur von tatsächlich zum Sortiment in Bezug stehenden Personen wertvolle Erkenntnisse zu erwarten sind.

Hierfür wurden vom Shop-Management Eckdaten zur Verfügung gestellt, die bei der Rekrutierung der Testpersonen Berücksichtigung fanden, um die Zielgruppe des Shops möglichst genau abzubilden. Der Kundenstamm des Shops besteht zu 80 Prozent aus Frauen. Als Hauptzielgruppe (Priorität A) gelten Mütter und Väter, die Kinder im Alter von null bis drei Jahren haben, sowie Schwangere bzw. Paare, die ein Kind erwarten. Diese Gruppe sollte also den größeren Anteil der Probanden ausmachen. Als Nebenzielgruppe (Priorität B) gelten Personen, die die Shopartikel an jemanden verschenken möchten, der ein Kind bekommen hat. Hiermit sind z. B. Großeltern oder Freunde von jungen Eltern gemeint. In beiden Zielgruppen waren außerdem zwei Erfahrungshorizonte abzudecken, und zwar die eher unerfahrenen und die erfahrenen Kunden. Als unerfahren können z. B. Schwangere gelten, die das erste Kind erwarten, oder Schenkende, die selbst keine Kinder haben. Erfahren sind dagegen beispielsweise Eltern, die das zweite Kind erwarten, oder Großeltern, die als Schenkende auf die entsprechende Erfahrung aufgrund ihrer eigenen Kinder zurückgreifen können. Tabelle 1 zeigt die Verteilung der Testpersonen für die drei Verfahren, wie sie idealtypisch geplant wurde. Sie konnte in dieser Form fast exakt umgesetzt werden. Die Rekrutierung der Probanden für den Labortest und den asynchronen Remote-Usability-Test wurde selbsttätig mit Unterstützung des Unternehmens durchgeführt, wohingegen die Akquise der Probanden für den asynchronen Videotest durch den Kooperationspartner RapidUserstests erfolgte, da dieses Unternehmen ein eigenes Panel für seine Tests betreibt. Alle Probanden erhielten für die Teilnahme an der Studie eine Aufwandsentschädigung.

Durch die systematische Auswahl der Probanden konnte sichergestellt werden, dass die Ergebnisse dem untersuchten Shop dienlich sind. Auch kann hierdurch eine bessere Beteiligung der Probanden erreicht werden, wenn diese nämlich auch persönlich einen Bezug zum Sortiment haben. Mit der Zielgruppe der Priorität B wurde hingegen gerade nach Probanden gesucht die nur einen geringen Bezug zu den Produkten haben und auch kaum Erfahrung damit. Da diese nur eine geringe Rolle spielen, wurden von diesen auch deutlich weniger rekrutiert. Dennoch ist gerade eine solche Aufspaltung der Zielgruppe dann sinnvoll, wenn dadurch die Nutzerschaft entsprechend abgebildet wird (vgl. SARODNICK/BRAU 2011, S. 167). Dadurch kann schließlich sichergestellt werden, dass alle potenziellen Nutzer auch bei der Studie berücksichtigt wurden, denn es ergibt wenig Sinn, ausschließlich mit den Nutzern zu testen, die ohnehin Stammkunden sind. Diese sind zwar deutlich einfacher zu rekrutieren, da sie dem Shop eventuell sowieso positiv gegenüber eingestellt sind. Da aber Neukunden ebenso von großem Interesse für jeden Shop sind, sollten geeignete Maßnahmen in Angriff genommen werden, um auch bisherige Nicht-Kunden zu gewinnen. Voraussetzung dafür ist allerdings, dass die mögliche Reichweite des Shops vorab festgestellt wird, um auch tatsächlich nur in Frage kommende Nutzer zu testen. Das Testen mit nicht adäquaten Probanden würde keinen Mehrwert

bieten, sondern nur das Testen mit der tatsächlichen Zielgruppe oder potenziellen Neukunden, die als zukünftige Nutzer in Frage kämen, da so deren Gründe für die bisherige Nichtnutzung ggf. im Test direkt mit erhoben werden können.

Tabelle 1: Zielgruppenverteilung anhand der drei Merkmale Geschlecht (m/w), Zielgruppenpriorität und Erfahrungshorizont für die drei Testverfahren.

Labortest				
n=12	Priorität A		Priorität B	
	unerfahren	erfahren	unerfahren	erfahren
	4	4	2	2
m/w	1/3	0/4	1/1	0/2

Videotest				
n=20	Priorität A		Priorität B	
	unerfahren	erfahren	unerfahren	erfahren
	6	6	4	4
m/w	2/4	0/6	2/2	0/4

aRUT				
n=40	Priorität A		Priorität B	
	unerfahren	erfahren	unerfahren	erfahren
	12	12	8	8
m/w	2/10	2/10	2/6	2/6

6.3 Methoden und Studiendesign

Dieser Abschnitt stellt getrennt voneinander die drei eingesetzten Methoden vor, und zwar einmal die Methode selbst mit ihren allgemeinen Spezifikationen und im Anschluss die Umsetzung bzw. Adaption der Methode für die im Rahmen der Erhebung durchgeführten Studie, sodass nachvollzogen werden kann, was sich hinter den einzelnen Methoden verbirgt und wie diese ggf. angepasst werden mussten, um für die Erhebung angewandt werden zu können.

Als Grundlage für die Untersuchung diente eine expertenbasierte Vorevaluation des Shops, die ergänzt um die Hinweise des Shopmanagements (s. u.) die Grundlage für das weitere Vorgehen bildete. Aufbauend darauf wurden wichtige Anwendungsfälle auf dem Shop mein-schnulli.de zusammengetragen, die als Untersuchungsansätze für die szenariobasierten Tests Verwendung fanden. Hier sei u. a. der Kauf des gleichen Schnullers mit unterschiedlichen Namen genannt oder die Änderung der Lieferadresse, die im Vorfeld vom Shopmanagement als derzeit suboptimal umgesetzt dargestellt wurde und daher einer Verbesserung bedurfte. So wurde für alle drei Verfahren ein typischer Ablauf skizziert, der für das jeweilige Testverfahren noch entsprechend angepasst wurde und ggf. um retrospektive Fragen im Labortest ergänzt wurde, die im Videotest konzeptbedingt

weggelassen wurden, um den selbst gelenkten Testablauf nicht zu stören. Im asynchronen Remote-Usability-Test wurden einige Fragen über Skalen beantwortbar gemacht, um die quantitativen Vorzüge der Methode auszunutzen und eine Analyse über die 40 Teilnehmer dieser Teilstudie zu vereinfachen.

Die Tests orientierten sich demnach an folgendem Ablauf:

- Erfassung der Erwartungen der Probanden, wenn sie an einen Shop hinter der URL www.mein-schnulli.de denken und im Gegenzug an einen Shop hinter der URL www.schnuller.de
- Freie Nutzung des Shops (nur im Labortest und Videotest möglich, da im asynchronen Remote-Test keine Hintergründe zum Erkenntnisinteresse des Nutzers vorliegen und so keine Interpretation der Klickstrecken möglich ist)
- Einführung in das Testszenario: „Eine gute Freundin von Ihnen hat gerade eineiige Zwillinge bekommen. Sie möchten ihr Schnuller der Marke NUK mit den Namen der Kinder schenken. Da es eineiige Zwillinge sind, sollen die Jungen auch die gleichen Schnuller erhalten.“
- Kauf des ersten Schnullers für das Kind Philipp. Hier war dann auch der Zusatzkauf eines empfohlenen Produkts von der Warenkorb-Zwischenseite gewünscht.
- Namensänderung für die Schnullerbeschriftung
- Kauf des gleichen Schnullers für den Bruder Tom
- Kauf einer Schnullerkette mit Namen für das zweite Kind
- Durchlaufen des Checkouts inklusive einer Änderung der Lieferadresse und der Auswahl der bevorzugten Zahlungsart
- Die Bestellungen wurden nicht abgeschickt, sondern im letzten Schritt abgebrochen.
- Zum Abschluss wurden Fragen zum Gesamteindruck der Nutzer gestellt.

Im Labortest waren außerdem entsprechende Rückfragen hinter die einzelnen Teilaufgaben angestellt. In den asynchronen Testverfahren wurden teilweise Aufgaben zusammengefasst, um eine zügige Erledigung der Tests zu ermöglichen, da hierfür kürzere Zeitspannen als für den Labortest vorgesehen waren, da bei den Verfahren, die nicht in Echtzeit durch einen Testleiter betreut werden, kürzere Aufmerksamkeitsspannen der Teilnehmer angenommen werden. Auch ist das Konzept der Erledigung einer Aufgabe in Schritten hier nicht analog zum Labortest umsetzbar, wo der Testleiter während einer Aufgabe mit mehreren Schritten auch zwischendurch immer wieder kleine Teilanweisungen geben kann. Auf der anderen Seite können diese mit kürzeren Testdauern potenziell die gleiche Menge an Aufgaben bewältigen, da es keine oder deutlich weniger Rückfragen dazu gibt.

Ob und inwieweit diese Daten dann dennoch vergleichbar sind, war u. a. die Themenstellung der Thesis und wird daher in den späteren Kapiteln ausführlich erläutert werden.

6.3.1 Der Usability-Test im Labor

Das klassische Verfahren des qualitativen Website-Testings ist der Usability-Test im Labor (vgl. EBERHARD-YOM 2010, S. 136 f.). Dieser wird mit einer geringen Anzahl an Probanden durchgeführt, wobei die Testpersonenanzahl nicht allzu hoch sein muss, um viele Usabilityprobleme ausfindig machen zu können (vgl. NIELSEN 2000). Hierfür muss eine Infrastruktur für den Test vorgehalten werden, die aus einem Usability-Labor mit umfangreicher technischer Ausstattung (vgl. Abbildung 2) und einem Beobachtungsraum besteht. Es werden mindestens zwei Personen benötigt, da eine den Probanden direkt betreut und die andere den Test beobachtet, wichtige Notizen anfertigt und dem Studienleiter ggf. abweichende Anweisungen oder Tipps zum Nachfragen geben kann. Häufig ist der Kunde des Tests vertreten und befindet sich im Raum des Protokollanten und kann dort auch bestimmte Nachfragen oder Erweiterungen des Tests über diesen in den Testraum weiterleiten. Für diese Tests ist neben den Räumlichkeiten die technische Ausstattung bereitzustellen, die aus zwei miteinander verbundenen Computern, Möglichkeiten zur Tonaufzeichnung und Tonübertragung und der Möglichkeit zum Aufzeichnen des Bildschirms der Testperson und ihres Gesichtsausdrucks besteht. Zwar sind hieran angelehnte Tests auch mit weniger Aufwand und Technik realisierbar (vgl. KRUG 2006, S. 142 f.). Im Rahmen der Arbeit soll dennoch der klassische, professionelle Labortest Beachtung finden. Das Verfahren ist zum Auffinden von Usabilityproblemen gut geeignet, aber auch verhältnismäßig teuer, zeit- und personalintensiv.



Abbildung 2: Sitzplatz des Probanden in einem typischen Usability-Labor (Foto der eResult GmbH).

Der Usability-Test im Labor ist häufig die bevorzugte Methode für Nutzerstudien, da er zwar aufwändig, aber potenziell am ergiebigsten und flexibelsten ist, wenn z. B. während des Tests noch Änderungen berücksichtigt werden sollen, die die Ergänzung um eine Frage meinen oder eine konkrete Rückfrage an einen Probanden, wenn dem Testleiter spontan etwas an dessen Verhalten auffällt, das dann direkt geklärt werden kann. Auch kann der Großteil der Eindrücke direkt verwertet werden, da Testleitung und Protokollant dem Test direkt in Echtzeit beiwohnen und dies so schon die erste Sichtung darstellt, während bei den asynchronen Verfahren die Ergebnisse erst hinterher vorliegen. Der Kunde kann ebenso direkt dabei sein und seine Eindrücke aus den Tests mitnehmen. Dies erscheint sinnvoll, da selbst Erlebtes i. d. R. einen prägenderen Eindruck hat als das Sichten der dokumentierten Ergebnisse, ohne dem Test selbst beigewohnt zu haben. Ein weiterer konzeptioneller Vorteil ist die Möglichkeit, die Tests aufeinander aufbauen zu lassen und so etwa Bereiche, deren Probleme bereits ausreichend dokumentiert wurden, in den nächsten Tests wegzulassen und dafür eventuell ergänzend einen neuen Bereich zusätzlich zu testen, sofern es sich im Studienkonzept so berücksichtigen lässt.

Im Rahmen der Arbeit wurde der Usability-Test im Labor mit folgenden Rahmenbedingungen durchgeführt: Die Erhebung fand am 20. und 21.06.2012 in Göttingen statt. Teilgenommen haben 12 Testpersonen. Die Tests dauerten jeweils 30 Minuten und wurden im fest eingerichteten Usability-Labor der eResult GmbH durchgeführt und mit der Software Morae zur späteren Auswertung aufgezeichnet. Durch eine Funkverbindung vom Beobachtungsraum in den Testraum konnte der Protokollant dem Testleiter zwischendurch ergänzende Anweisungen geben, ohne dass die Testperson hiervon etwas mitbekommen konnte. Das für den Test angefertigte Studienkonzept orientierte sich im

groben Aufbau an den Vorgaben der eResult GmbH und wurde für den Test mit den Szenarien und den dazu passenden retrospektiven Interviewfragen versehen, um die Stärken dieser Methode mit dem Erkenntnisinteresse des kooperierenden Shops in Einklang zu bringen.

6.3.2 Der asynchrone Remote-Usability-Test

Dieses Verfahren kann örtlich und zeitlich unabhängig eingesetzt werden. Es wird teilautomatisiert durchgeführt und bedarf der vorherigen Strukturierung des Tests in Aufgabenteile, um eine Auswertung zu ermöglichen. Hat der Nutzer eine Teilaufgabe ausgeführt, kennzeichnet er diese als beendet. Hier können ihm dann noch Fragen zur gerade durchgeführten Aufgabe gestellt werden. Weitere erfasste Daten sind die Mausbewegungen und Mausklicks, die als sog. Clickstream (vgl. Abbildung 3) für die Auswertung verwendet werden können und die eingeschlagenen Wege der Nutzer auf der Site zeigen. Auch können zu jeder Aufgabe Zeitangaben für die Erledigung erhoben werden. Eine Aufzeichnung weiterer Daten erfolgt nicht, so wird kein Ton aufgezeichnet und auch der Proband ist nicht zu sehen. Für diese Methode kann mit einer sehr großen Stichprobe gearbeitet werden (vgl. SARODNICK/BRAU 2011, S. 181), da zu jeder Testperson eine begrenzte Menge an Daten vorliegt und diese je nach Fragearten auch automatisiert ausgewertet werden können (z. B. Fragen mit Radiobuttons oder Skalen). Die Methode ist demnach sowohl im quantitativen als auch im qualitativen Bereich angesiedelt und kann mit relativ geringem Aufwand viele Rückmeldungen zu einer Site generieren, die dann zu deren Optimierung eingesetzt werden können.

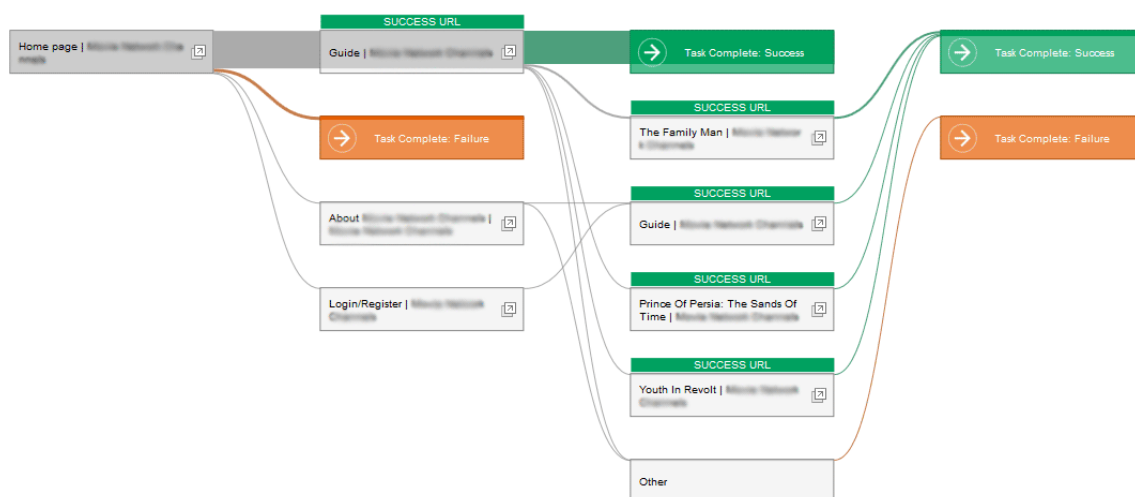


Abbildung 3: Beispielhafte Clickstream-Visualisierung des Anbieters Loop 11 (LOOP11 2012).

6.3.3 Der Remote-Videotest

Ein relativ neues Verfahren ermöglicht auch das orts- und zeitunabhängige Testen. Die Methode geht einen neuen Weg, denn sie versucht, eine Art klassischen Labortest ortsunabhängig und ohne betreuendes Personal durchzuführen. Einige Elemente des asynchronen Remote-Usability-Tests können hier auch verwendet werden. Das primäre Ergebnis des Tests ist ein Video, das alle Bildschirminteraktionen des Nutzers aufzeichnet. Der Ton ist ein Thinking-Aloud-Protokoll der Testperson. In dieser Hinsicht ähnelt das Verfahren am ehesten dem Labortest. Allerdings gibt es keinen Testleiter, sondern die Probanden führen den Test komplett eigenständig durch. Die Anweisungen erhalten Sie hierbei aus eingeblendeten Aufgaben. Das Abfragen von Meinungen zu einzelnen Aufgaben kann standardisiert erfolgen, sodass diese Ergebnisse dann in Datenform vorliegen und nicht aus den Videos entnommen werden müssen. Abbildung 4 zeigt die Bestandteile eines Testergebnisses des Anbieters RapidUserTests. Die Einrichtung des Tests (Aufgabenstellungen und Fragen) erfolgt weitestgehend eigenständig durch den Kunden. Die den Test anbietende Firma übernimmt i. d. R. nur die Durchführung der Tests mit ihren Ressourcen (Akquise der Probanden, Abspeichern der Videos).

Dieses Verfahren wird derzeit von einigen Firmen angeboten, z. B. von UI-CHECK (vgl. UI-CHECK 2012), die für eine Testperson i. d. R. einen Festpreis veranschlagen. Der Kunde kann wählen, wie viele Probanden am Test teilnehmen sollen und erhält innerhalb sehr kurzer Zeit (teilweise binnen 24 Stunden) die Videos als Testergebnisse.

Gegen Aufpreis wird Unterstützung bei der Auswertung angeboten, sonst wertet der Auftraggeber die Videos selbst aus. Anders als bei dem asynchronen Remote-Usability-Test besteht hier also eine Geschäftsbeziehung zwischen dem Kunden und dem Unternehmen, das die Untersuchung durchführt. Beim asynchronen Test kann die Software von jedem Lizenznehmer eigenständig eingesetzt werden, dieser muss dann aber auch den Test selbst organisieren, d. h. unter anderem die Probanden selbst akquirieren, was abhängig von den Vorgaben eine der schwierigsten Voraussetzungen sein kann. Bei diesem Verfahren wird indes das Panel an Testpersonen schon als eine wichtige Grundlage betrachtet, die v. a. für eine schnelle Ergebnislieferung sorgen kann.

Bei dem Verfahren der beiden genannten Anbieter wird daher auf eine All-in-one-Lösung gesetzt, die weitestgehend transparente Kosten mit schnellen Ergebnissen kombiniert. Das Angebot kann ohne langen Vorlauf in Anspruch genommen werden und es suggeriert jedem Interessenten eine rasche Durchführung und einen schnellen Ergebniserhalt, ohne dass ausgeprägte Kenntnisse im Bereich der Usability vorhanden sein müssen.

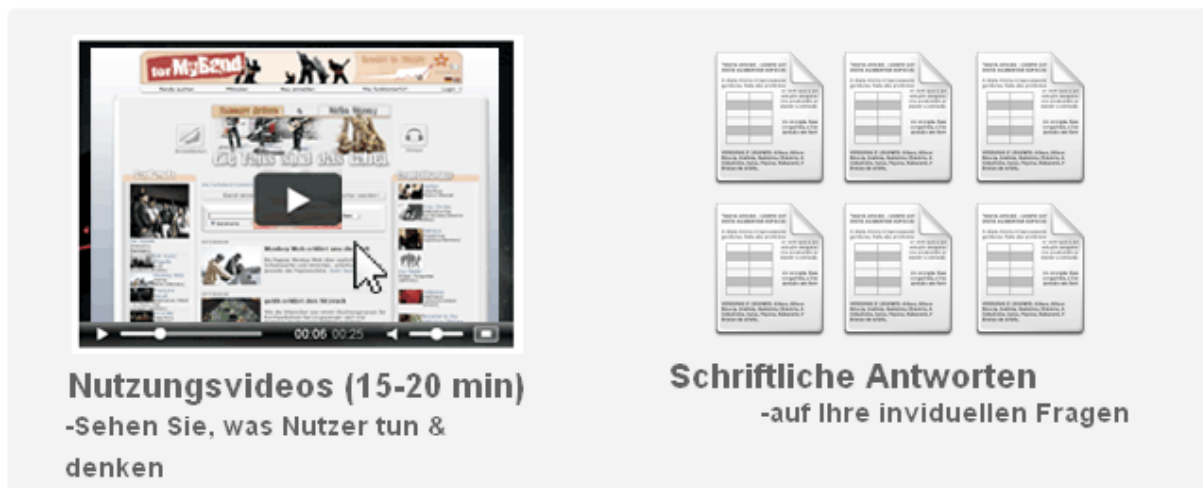


Abbildung 4: Vorliegende Daten zur Auswertung einer Remote-Studie von RapidUserTests (RAPIDUSERTESTS 2012).

Eine feststehende Benennung für dieses Verfahren hat sich noch nicht etabliert. Im Text dieser Arbeit werden sie zunächst als Videotests bzw. Remote-Videotests bezeichnet, wobei beachtet werden muss, dass sich dahinter das orts- und zeitunabhängige Verfahren des Usability-Tests verbirgt, den die Probanden mit lautem Denken ohne Betreuung selbst durchführen.

Ausgangspunkt für die Betrachtung dieser Methode ist deren Vorhandensein am Markt seit relativ kurzer Zeit. Eine wissenschaftliche Evaluation ist bislang nicht erfolgt und soll durch diesen Vergleich möglich werden. Als Thema sind sie in der Fachwelt bereits vorhanden (vgl. u. a. PUSCHER 2011), werden dort allerdings als sog. Online-Panel bezeichnet. Der Schwerpunkt liegt dort also eher auf der Akquise. Für diese Arbeit soll der Fokus auf den Stärken und Schwächen der Erhebung selbst liegen, wobei die Probandenakquise nur eine geringe Rolle spielt.

7 Ergebnisse der Erhebungen

Die drei Methoden wurden zunächst für sich ausgewertet und deren Erkenntnisse aufgelistet und interpretiert. Hierbei wurde ein gemeinsamer Rahmen für alle Methoden vorgegeben, sodass für den Shopbetreiber zunächst die gesammelten Erkenntnisse vorlagen. Außerdem dient diese Liste dem Abgleich der drei Methoden dahingehend, welche Probleme durch alle Verfahren und welche nur durch eins oder zwei der angewandten Verfahren aufgedeckt wurden. Abbildung 5 zeigt einen Ausschnitt aus der zusammengestellten Problemtabelle. Dabei ist links das Problem beschrieben, die Spalten in der Mitte sind für je eine Methode und werden beim Auftreten dieses Problems bei der Methode mit dem Buchstaben für den Schweregrad (s. u.) versehen. In der rechten Spalte befindet sich ein kurz dargestellter Kommentar zur Behebung dieses Problems.

Die Priorisierung der Probleme wurde durch Experteneinschätzung vorgenommen, um deutlich zu machen, welche Wichtigkeit den einzelnen Problemen zuzuweisen ist, um auch sichtbar zu machen, ob zumindest jedes Verfahren die schwerwiegendsten Probleme aufzudecken vermochte. Hierbei wurden die Kategorien „A“, „B“ und „C“ vergeben, wobei „A“ schwerwiegende Probleme beschreibt, die sich z. B. sehr negativ auf das Nutzererlebnis auswirken und ggf. sogar eine Aktion undurchführbar machen. Probleme der Kategorie „B“ sind weniger schlimm, aber auch störend, während Probleme der Kategorie „C“ entsprechend eher kosmetische Aspekte im Shop beschreiben.

Einige Artikelbenennungen wirken sperrig und sind ungünstig zusammengesetzt u. unstrukturiert (Bsp.: „Schnullerkette mit Name Blau mit Autos“) bzw. lassen keine ausreichende Unterscheidung zwischen den Artikeln einer Rubrik zu.	B	1	B	1	B	1	Die Zusammensetzungen könnten z. B. durch Kommas od. kursive Teiltelle besser aufbereitet werden. Z. B. Farbe immer in zweite Zeile u. Teil (Ente) nicht mehr in fett; ggf. Unterstreichung zur Steigerung der Lesbarkeit entfernen
Der Kontrast auf den Seiten ist insgesamt gering.	B	2	B	2	B	2	Die Farben sollten kontrastreicher gestaltet werden u. ausgewählte Texte (z. B. Fehlermeldungen) gefettet werden, um besser erkennbar zu sein. Pfeiler der Kategorien schlecht lesbar.
Die Mindeststückzahl von zwei Schnullern wird als zu viel empfunden. Viele Nutzer möchten einen einzelnen Schnuller kaufen.	C	3			C	5	Sollten das Anbieten einzelner Schnuller nicht möglich ist, sollte bei der Mengenauswahl eine Erklärung ergänzt werden, weshalb es die Mindestmenge gibt (z. B. technische Gründe)

Abbildung 5: Ausschnitt aus der Gesamtproblemliste aller drei Methoden.

Als Hauptprobleme im Shop zeigte sich dessen Aufmachung, die auf der Startseite nicht die gesamte Sortimentsbreite vermitteln konnte. Die Nutzer waren sehr fixiert auf die Produktkategorie der Schnuller, sodass ihnen andere Produkte entgingen. Unklar war vielen auch die Bestellung von Schnullern, die nur paarweise erhältlich sind und dann mit demselben Namen bedruckt werden müssen. Die Kommunikation der Zusatzgebühren von Zahlungsverfahren erfolgte erst spät im Checkout, sodass dieser Aspekte ebenso schlecht bewertet wurde. Die Änderung der Versandadresse bereitete einem Großteil der Nutzer in allen Verfahren Probleme, da u. a. das Wording der Adressen unklar blieb. Dies sind die Hauptprobleme, die sich in den Tests herauskristallisierten. Alle weiteren

Probleme sind der Problemtabelle im Anhang zu entnehmen. Von hieran werden die Probleme nicht mehr gesondert erwähnt, da der Fokus auf der Ergiebigkeit der Methoden im Hinblick auf die erkannte Problemanzahl lag und das Untersuchungsobjekt hierfür nur als exemplarische Plattform diente.

Anhand der ausgezählten Probleme, die die einzelnen Methoden aufdeckten, kann bereits ein erster Vergleich in Zahlen erfolgen, wobei beachtet werden muss, dass die Stichprobengrößen bei allen drei Tests unterschiedlich waren. Es zeigt sich, dass der Videotest eine relativ hohe Anzahl an Problemen aufgedeckt hat, während der asynchrone Remote-Usability-Test nur eine geringe Anzahl zutage förderte. Tabelle 2 zeigt die Anzahl der jeweils aufgedeckten Probleme der einzelnen Methoden. Der Labortest und der Videotest erkannten gleich viele schwerwiegende Probleme. Probleme mit den Prioritäten „B“ und „C“ wurden in größerer Zahl vom Videotest aufgedeckt. Daraus ergibt sich auch dessen höhere Problemanzahl über alle Prioritäten. Der asynchrone Remote-Usability-Test erkannte insgesamt eine deutlich geringere Problemanzahl, wenngleich er sechs der schwerwiegenden Probleme aufdeckte, was in Anbetracht des Aufwandes schon ein guter Wert ist. Anhand der Spalte „Alle Verfahren“ lässt sich ablesen, dass die Überschneidungen bei den gefundenen Problemen überschaubar ist, d. h. bestimmte Probleme nicht von jeder der drei Methoden identifiziert wurden. In Abbildung 6 ist der Vergleich als Diagramm visualisiert und zeigt so die Potenziale der Methoden im direkten Vergleich. Die folgenden Abschnitte befassen sich jeweils gesondert mit den Ergebnissen der Methoden und weisen dort auch auf weitere Erkenntnisse zu den Methoden hin, die über die reine Quantifizierung der Probleme hinausgehen.

Tabelle 2: Anzahl und Priorisierung der mit den drei Verfahren aufgedeckten Usability-Probleme.

<i>Methode/Anzahl der aufgedeckten Probleme</i>	Labortest (n=12)	Videotest (n=20)	Asynchroner Remote-Usability-Test (n=40)	Alle Verfahren
Priorität A	24	24	6	29
Priorität B	13	20	4	25
Priorität C	14	26	9	37
Gefundene Probleme insgesamt	51	70	19	91

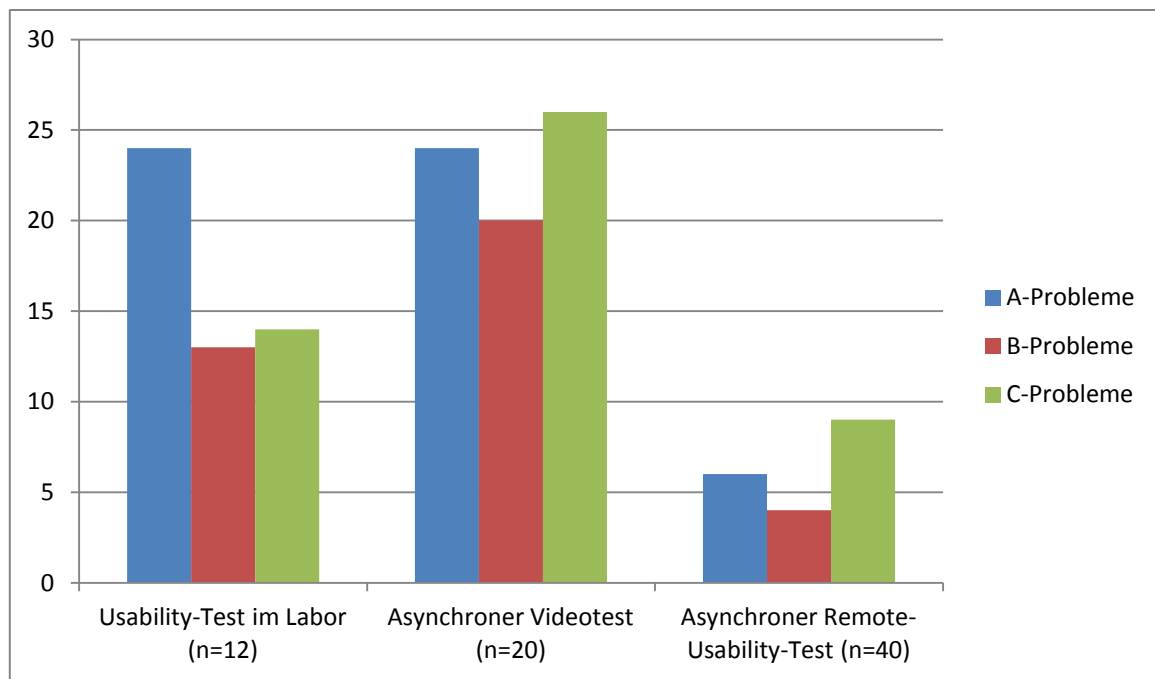


Abbildung 6: Diagramm zur Gegenüberstellung der gefundenen Probleme der drei Methoden.

7.1 Usability-Test im Labor

Der im Labor durchgeführte Usability-Test war das Verfahren, das der Studienleitung bereits bekannt war und daher relativ unkompliziert vorbereitet werden konnte. Diese Methode war aufgrund der bereits vorhandenen Erfahrung verhältnismäßig leicht durchführbar. Es ist daher nicht auszuschließen, dass die Erfahrung mit dieser Methode einen Vorteil gegenüber den anderen untersuchten Verfahren hat, da diese auf Seiten der Testleitung vorher unbekannt waren und daher zum ersten Mal Verwendung fanden.

Der wichtigste Vorzug des Labortests gegenüber nicht angeleiteten Testverfahren ist die stärkere Beteiligung (Involvement) der Probanden am Test. Diese fühlen sich durch die direkte Betreuung viel intensiver in das Testgeschehen eingebunden und handeln daher auch deutlich konzentrierter und zielgerichteter und nehmen sich ausreichend Zeit für die Durchführung der einzelnen Aufgaben. Daher sind die Rückmeldungen der Probanden i. d. R. auch qualitativ hochwertig, sodass relativ häufig auch direkt aus den Tests gut verwertbare Verbesserungsvorschläge zu bestimmten Siteelementen geäußert werden bzw. der Testleiter hierzu zielführende Nachfragen stellen kann, um die Wünsche der Probanden zu einzelnen Funktionen bereits mit einem hohen Detaillierungsgrad erheben zu können.

Anhand der Auszählung der gefundenen Usability-Probleme lässt sich feststellen, dass der Labortest mit 24 Problemen der Priorität „A“ genauso viele Probleme wie der Videotest zu dieser Priorität aufgefunden hat. Im Hinblick auf die schwerwiegendsten Probleme hat er also gut

abgeschnitten. Bei den Problemen der Prioritäten „B“ und „C“ wurde mit 13 und 14 Problemen auch eine relativ hohe Anzahl festgestellt. Diese Probleme sind für einen Sitebetreiber ebenso wichtig, wenngleich die „B“-Probleme eher behoben werden sollten als die „C“-Probleme, die weitestgehend kosmetische Aspekte behandeln. Insbesondere die „C“-Probleme werden in Abhängigkeit von der Aktivität und Offenheit der Probanden festgestellt. Je höher deren Affinität zur Websitegestaltung ist, desto eher darf mit spezifischen Kommentaren gerechnet werden, die z. B. die räumliche Aufteilung der Webseitenelemente betreffen. Je mehr Personen an einem Test teilnehmen, desto mehr (unterschiedliche) Äußerungen dürften ebenso zu solchen Aspekten erwartet werden. Hierbei muss jedoch auch immer berücksichtigt werden, inwieweit solche Aussagen überhaupt erwünscht sind bzw. Berücksichtigung finden können. Zudem ist die für einen Test vorgesehene Zeit hier weiterhin ein limitierender Faktor: Ist die freie Nutzungsphase relativ kurz, ist mit weniger freien Äußerungen zu rechnen, die keine konkrete Aufgabe betreffen. Je strenger sich der Testleiter am Studienkonzept orientiert, wenn er evtl. unbeabsichtigt den Probanden bereits zur nächsten Aufgabe drängt, desto weniger Rückmeldungen außer der Reihe wird es zu allgemeinen Aspekten geben. Die Abhängigkeit von der Aktivität des Probanden bleibt hier jedoch nach wie vor bestehen, wenngleich im nicht angeleiteten Test eine aktive Testperson nicht in ihre Schranken gewiesen wird, aber evtl. aus Gründen des geringeren Involvements ebenso wenig freie Äußerungen tätigt.

So lässt sich für den Usability-Test im Labor folgende Zusammenfassung geben: Es wurden alle schwerwiegenden Usability-Probleme erkannt. Es wurde auch eine Reihe von weniger schwerwiegenden Problemen erkannt, sodass mit diesen Ergebnissen der Shop bereits optimal verbessert werden kann. Insbesondere die ausführlichen Rückmeldungen zu einzelnen Teilaspekten machen diese Methode sehr wertvoll, da sie der Auswertung sehr viel Tiefe verleiht.

7.2 Asynchroner Remote-Usability-Test

Die Auswertung des asynchronen Remote-Usability-Teste erfolgte mit dem Online-Tool, das die Daten aller Testpersonen aggregierte und online aufrufbar machte. So ließen sich die Ergebnisse u. a. als Clickstream (vgl. Kapitel 6.3.2) darstellen, um die Wege aller Probanden sehen zu können.

Der bei der Konzeption dieser Umfrage vorgegebene Ablauf mit der Reihenfolge von Aufgaben und Fragen (vgl. Abbildung 7) kann im Tool auch bei der Auswertung als gute Orientierung genutzt werden.

Loop¹¹ Home Create New Project My Projects My Account Help Centre

1 User Test Details 2 **Tasks & Questions** 3 User Test Options 4 Invite Participants 5 Launch

Step 2: Tasks & Questions

- 1** Question Was erwarten Sie hinter der Adresse www.mein-schnulli.de für einen Shop? Was en [View](#) [Preview](#)
- 2** Question Wie erfahren sind Sie beim Kauf von Babyartikeln? [View](#) [Preview](#)
- 3** Question Bitte erläutern Sie kurz Ihre persönlichen Erfahrungen mit Babyartikeln. Kaufen Sie [View](#) [Preview](#)
- 4** Task Kauf erstes Schnullerpaar [View](#) [Preview](#)
- 5** Question Wie hat Ihnen dieser Schritt gefallen? [View](#) [Preview](#)
- 6** Question Gibt es noch etwas, das Sie zu diesem Schritt loswerden möchten (optional)? [View](#) [Preview](#)
- 7** Task Namensänderung [View](#) [Preview](#)
- 8** Question Wie hat Ihnen dieser Schritt gefallen? [View](#) [Preview](#)
- 9** Question Gibt es noch etwas, das Sie zu diesem Schritt loswerden möchten (optional)? [View](#) [Preview](#)
- 10** Task Zweites Schnullerpaar [View](#) [Preview](#)
- 11** Question Wie hat Ihnen dieser Schritt gefallen? [View](#) [Preview](#)
- 12** Question War Ihnen zu jeder Zeit klar, dass Sie an einem anderen Artikel arbeiten als mit der [View](#) [Preview](#)
- 13** Task Checkout [View](#) [Preview](#)
- 14** Question Wie hat Ihnen dieser Schritt gefallen? [View](#) [Preview](#)
- 15** Question Welche Zahlungsart(en) bevorzugen Sie normalerweise (mehrere Optionen wählbar [View](#) [Preview](#)
- 16** Question Sie haben nun das Ende des Tests erreicht. Wir haben noch ein paar Fragen zu Ihre [View](#) [Preview](#)
- 17** Question Bitte schildern Sie kurz Ihren Gesamteindruck. [View](#) [Preview](#)
- 18** Question Was hat Ihnen besonders gut gefallen? Was hat Ihnen gar nicht gefallen? Inwiefern [View](#) [Preview](#)
- 19** Question Gibt es noch etwas, das Sie den Betreibern des Shop mit auf den Weg geben möcht [View](#) [Preview](#)
- 20** Question Vielen Dank für Ihre Teilnahme. Um Ihnen die Bonopolis-Punkte für die Teilnahme z [View](#) [Preview](#)

Abbildung 7: Ablauf des aRUT, wie er als Abfolge im Loop11-Tool visualisiert wird (Screenshot aus der Anwendung).

Zur Auswertung war die Ansicht eines einzelnen Probanden möglich, sodass dessen Abwicklung einer Teilaufgabe im Online-Tool visualisiert werden konnte (vgl. Abbildung 8).

Kauf erstes Schnullerpaar

Participant 1 Path Analysis

[Next participant →](#)

Navigation path for **Participant 1** attempting task **Kauf erstes Schnullerpaar**

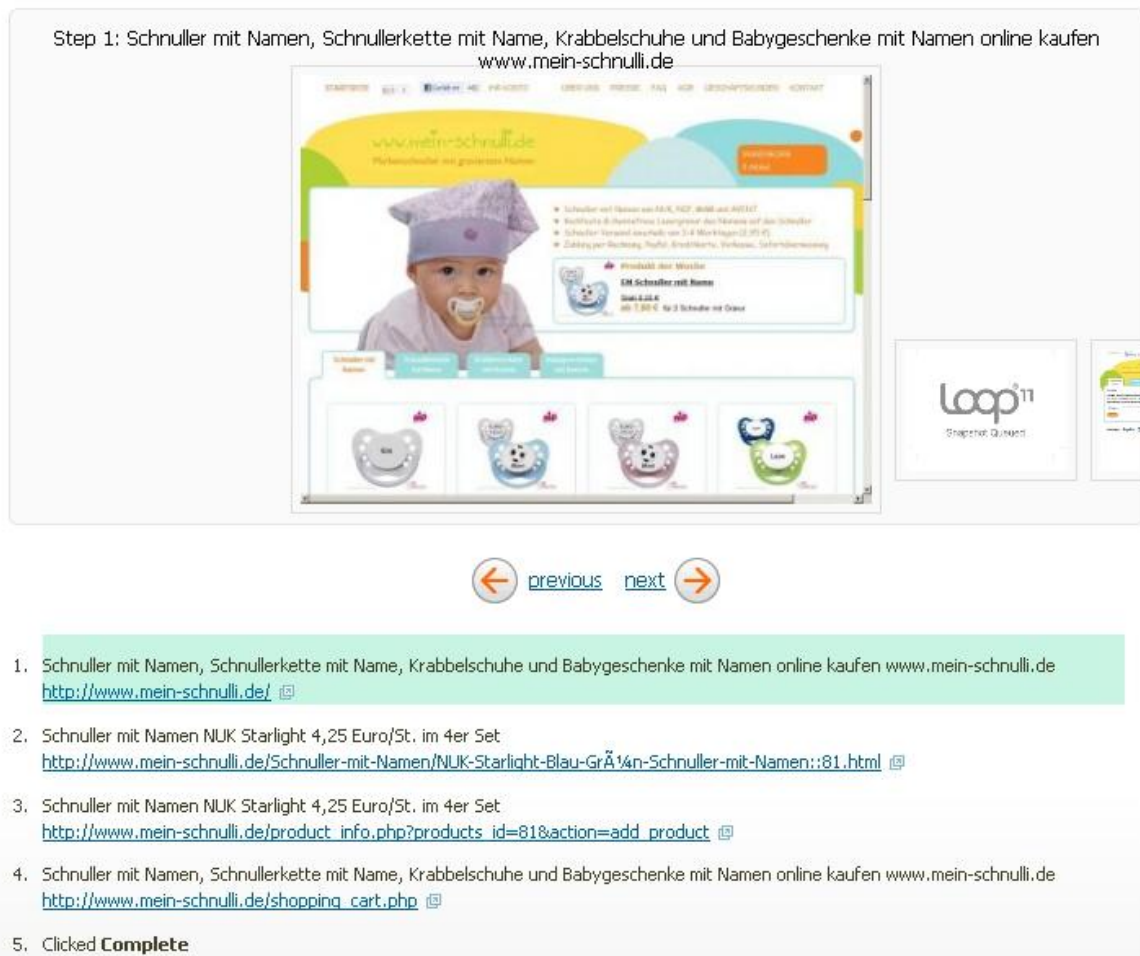


Abbildung 8: Aufgabenabarbeitung eines einzelnen Probanden bei Loop11 (Screenshot aus der Anwendung).

Der asynchrone Remote-Test, der toolbasiert mit Loop11 durchgeführt wurde, deckte insgesamt nur 19 Probleme auf. Hierunter waren sechs mit der Priorität „A“, vier der Priorität „B“ und neun mit der Priorität „C“.

Dieses Testverfahren hatte die geringste Einbindung der Probanden. Dies zeigte sich u. a. daran, dass viele Teilnehmer nicht alle Aufgaben vollständig lösten oder stellenweise verfrüht abbrachen, wenn sie nicht schnell zum Ziel fanden. Auch die erhobenen Daten begrenzten den Erkenntnisgewinn sehr stark: Die erzeugten Clickstreams sind erst nach Auflösen der kryptischen URLs

nachzuvollziehen und die vom System erzeugten Screenshots, die den Pfad darstellen, den ein einzelner Nutzer gegangen ist, werden stets live erzeugt, sodass ebenso wenig nachvollzogen werden kann, welches Produkt beim Nutzer im Warenkorb lag und welchen Weg er beim Checkout gegangen ist, da sich alle Seiten, die sich hinter dem Login verbergen, nicht mehr reproduzieren lassen. Genauso wenig lassen sich die von den Probanden gemachten Eingaben nachvollziehen, sodass nicht mehr festgestellt werden kann, ob deren Änderungen bei Versand- und Rechnungsadresse auch erfolgreich durchgeführt und in die korrekten Felder eingetragen worden sind. Der Großteil der Erkenntnisse kann daher nur in Ansätzen aus den Clickstreams gezogen werden, die zumindest durch die hohe Anzahl an Probanden eine quantitative Auswertung im Ansatz möglich machen. Ein weiterer Datenlieferant müssen daher die den Probanden gestellten Fragen sein, die teilweise als Skalen und teilweise als freie Texteingaben gestellt wurden, wobei hier auch festgestellt wurde, dass die Probanden eher zu sehr kurzen und wenig wertvollen Texteingaben tendieren, um den Test rasch abschließen zu können.

So lässt sich die geringe Anzahl an gefundenen Problemen erklären, wenngleich es durchaus sein kann, dass die Probanden während der Tests auf weitere Probleme gestoßen sind. Diese sind dann jedoch aufgrund der erfassten Daten und nicht vorgenommener Freitexteingaben zu diesem Teilproblem nicht dokumentiert und können nicht festgestellt werden. Dennoch wurden wichtige Kernprobleme erkannt, sodass der asynchrone Remote-Test in Anbetracht des Erhebungs- und Auswertungsaufwands durchaus vertretbare Ergebnisse erzielte. Dessen Stärke liegt daher auch eher in der Erhebung einer sehr großen Anzahl an Nutzern und ist daher als quantitativer Test auch stärker, je mehr Datensätze vorliegen. Die freien Äußerungen machen dann noch eher eine geringere Wichtigkeit aus. Dann bietet sich dieses Verfahren jedoch am ehesten für weniger interaktive Sites an, die z. B. keine individuellen Eingaben ermöglichen, sondern unter Vorgabe eines relativ festgelegten Weges durchlaufen werden sollen, sodass festgestellt werden kann, inwieweit die Nutzer vom vorgesehenen Weg abgekommen sind oder nicht. Insbesondere für eine sehr freie Nutzung ist der Test kaum geeignet, da die Aggregation der Daten dann kaum zu hilfreichen Ergebnissen führt, da es in dem Fall keinen idealen Weg gibt, den die Probanden hätten beschreiten können. Trotz der begrenzten Menge an Daten ließen sich einige Probanden feststellen, die die Aufgaben nicht hinreichend bearbeitet hatten und so direkt aussortiert werden mussten. Dieses geringe Involvement lässt sich ähnlich auch in Online-Befragungen feststellen, bei denen viele Probanden auch eher halbherzig teilnehmen, um die Aufwandsentschädigung zu erhalten. Hier ist offenbar die Hemmschwelle bei einer schlechten Teilnahme sehr gering, sodass ein Teil der Datensätze nicht ausgewertet werden kann. Bei einer verhältnismäßig hohen Aufwandsentschädigung ist dies natürlich ungünstig. Auf der anderen Seite kann dieses Testverfahren auch relativ unkompliziert z. B. auf einer Website eingebettet werden und allen Besuchern angezeigt werden. Als Aufwandsentschädigung könnte dann etwa eine Verlosung eingesetzt werden. Ein hoher Ausschuss von Daten wäre dann nicht schlimm und der Test kann auch über einen längeren Zeitraum online bleiben, bis die gewünschte Anzahl an Datensätzen gesammelt wurde. Hier muss dann lediglich bei einer bevorzugten Zielgruppe noch im Nachhinein eine Filterung

vorgenommen werden, falls dies gewünscht wird. In jedem Fall müssen dann diese Filterfragen in die Studie integriert sein, da ja die Rekrutierung nicht im Vorfeld stattfindet und es so kaum Einfluss auf die Eigenschaften der Teilnehmer gibt.

In jedem Fall muss zu dieser Teilmethode festgehalten werden, dass die Ergebnisse stark von dem verwendeten Tool für die Erhebung abhängig sind. Aus finanziellen und organisatorischen Gründen war diese Methode nur mit einem Tool durchführbar. Das Tool des Anbieters Loop11 diente daher als Anhaltspunkt für Funktionsumfang und Potenzial des asynchronen Remote-Tests. Dabei kann es durchaus sein, dass es Tools mit einem größeren Funktionsumfang gibt, die so mehr Ergebnisse zutage fördern. Diese Arbeit vermag daher nicht, eine allgemeingültige Aussage zum Potenzial des asynchronen Remote-Testings mit einem Tool zu treffen, sondern muss diesen Einzeltest als Maßstab annehmen.

7.3 Remote-Videotest

Die Videos des durch den Anbieter RapidUserTests durchgeführten Videotests können online aufgerufen und betrachtet werden. Notizen können direkt in der Auswertungsansicht (vgl. Abbildung 9) eingetragen werden und auch die Bewertung der Testperson ist auf einer Skala möglich, um dem Unternehmen zu zeigen, ob die Testperson für den Test hilfreich war.

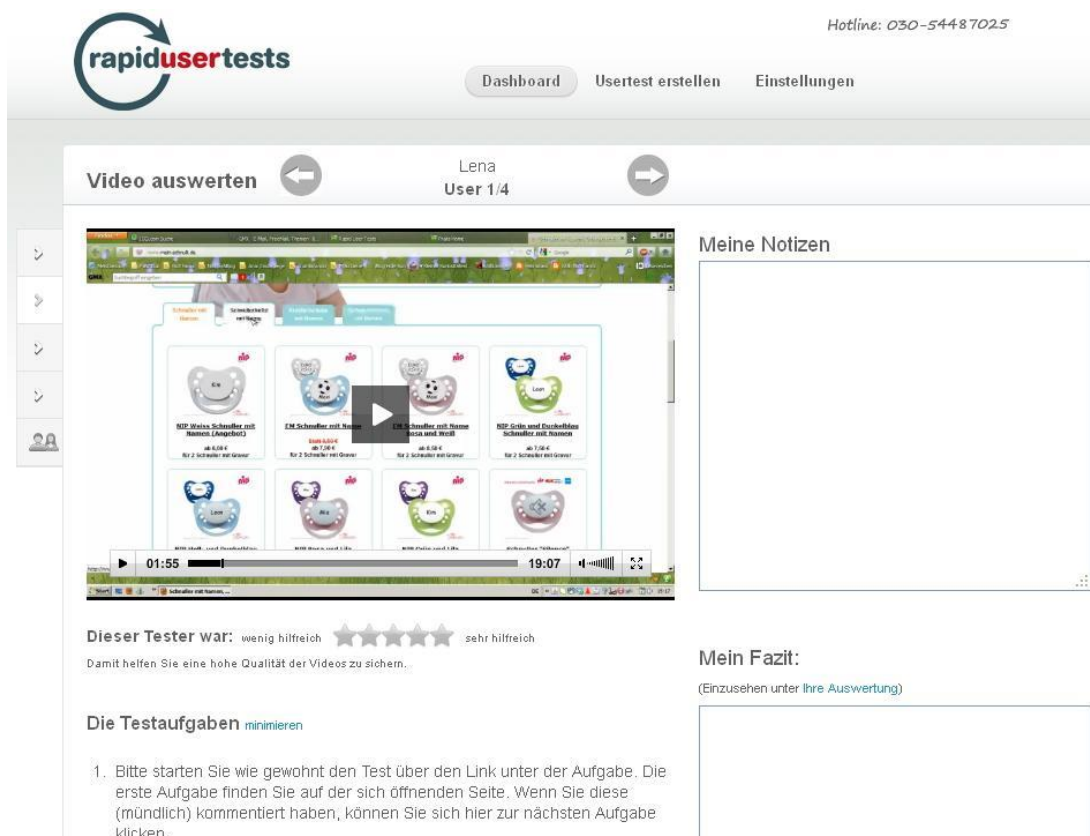


Abbildung 9: Auswertungsansicht des Videotests von RapidUserTests (Screenshot aus der Anwendung).

Damit wird der Ansatz verfolgt, einen Test unkompliziert und aus einer Hand auswerten und zentral dokumentieren zu können.

Das Verfahren erzielte von der Menge her die höchste Anzahl an gefundenen Usability-Problemen. Der erste Eindruck dieses Verfahrens ist subjektiv auch gut. Dem Verfahren wurde vorab nicht ein derartiges Potenzial zugeschrieben. Das Verfahren deckte genauso viele schwerwiegende Usability-Probleme auf wie der Labortest (24) und erkannte sogar noch mehr Probleme mit den Prioritäten „B“ (20) und „C“ (26). Hierbei muss beachtet werden, dass die Probleme mit den Prioritäten „B“ und „C“ zwar auch interessant sind, der Erkenntnisgewinn jedoch nicht immer allzu wichtig ist. Sinnvoll ist es dennoch, weitere Aussagen zum Shop zu erhalten. Beim Videotest, den die Probanden ja selbst anleiteten, ist die Anzahl an Rückmeldungen zur untersuchten Website wieder sehr stark von den Probanden und deren Aktivität abhängig. So gab es hier auch Probanden, die weitestgehend still durch den Test manövierten und so wenige Erkenntnisse erlangten, während andere sehr viel erzählten und dann auch die vorgesehene Testdauer überschritten, was für eine längere Auswertungszeit sorgte.

Insgesamt ist das Verfahren wertvoll und durch die Bündelung des Großteils der Dienstleistungen bei einem Anbieter kurzfristig durchführbar. So unterhalten die meisten Anbieter ein eigenes Panel, aus dem die Probanden rekrutiert werden. Innerhalb kürzester Zeit werden die Probanden dann eingeladen, sodass die Ergebnisse zeitnah vorliegen. Sind darunter einige aktive Probanden, erhält man bereits sehr wertvolle Erkenntnisse, ohne einen dem Labortest vergleichbaren Aufwand betreiben zu haben.

Lediglich in der Tiefe lässt der Videotest nach, da keine Nachfragen möglich sind. Einige Probanden machen Verbesserungsvorschläge. Wenn sie diese jedoch nicht hinreichend erklären, gehen hiervon wertvolle Informationen verloren oder diese können erst gar nicht verwertet werden.

Das Bestreben, den Test rasch abzuschließen, wurde hier nicht so stark festgestellt wie im asynchronen Remote-Test. Die Probanden sind i. d. R. verhältnismäßig aktiv. Vermutlich haben sie durch das selbsttätige Beschreiben ihrer Aktionen einen starken Bezug zu den ausgeführten Aktionen. Dieser Umstand ist in jedem Fall wertvoll. Dennoch fehlt stellenweise die Möglichkeit zum Eingriff durch den Testleiter. Kommen die Testpersonen z. B. an einer Stelle nicht weiter, können sie die Situation zwar umschreiben. Wenn sie von dort aus jedoch nicht weiter voranschreiten können, brechen sie diese Teilaufgabe dann gezwungenermaßen ab, wodurch dann Daten zu den Aufgaben fehlen können, die erst durch das Abschließen der vorangestellten Aufgabe durchführbar sind.

Als Potenzialsteigerung des Verfahrens wäre ggf. die Möglichkeit zur nachträglichen Kontaktaufnahme mit dem Probanden denkbar. Dies könnte z. B. telefonisch erfolgen, um etwa einen Verbesserungsvorschlag noch detaillierter erklärt zu bekommen. Diese Kontaktaufnahme müsste dann

jedoch zwingend sehr früh nach dem Test erfolgen, da ansonsten die Gefahr besteht, dass der Proband gar nicht mehr genau weiß, wie sein Verbesserungsvorschlag zu verstehen war.

Dennoch lässt sich so festhalten, dass der Videotest zur Identifikation der größeren Probleme einer Site ausreicht. Durch die lebensechte Visualisierung der Aktionen mit den dazu aufgezeichneten Kommentaren ist das Handeln des Nutzers jederzeit nachvollziehbar, sodass die Erkenntnisse deutlich praxisbezogener sind als im asynchronen Remote-Test. Unter dem Gesichtspunkt der schnellen und relativ preisgünstigen Durchführbarkeit stellt das Verfahren einen guten Ansatz für die Identifikation schwerwiegender Usability-Probleme dar und liefert weiterhin auch sinnvolle Anregungen, die teilweise direkt verwertet werden können. Auch Personen mit wenig oder gar keiner Erfahrung können binnen kurzer Zeit hiermit Erkenntnisse generieren, wenn sie etwa typische Anwendungsfälle in einem solchen Test abprüfen lassen, auf die sie z. B. durch die Analytics-Daten ihrer Website gestoßen sind. Gegebenenfalls kann so auf den Austausch mit einem Usability-Consultant verzichtet werden.

7.4 Schwerwiegende Probleme

Die Einstufung der Usability-Probleme erfolgte durch den Studienleiter und einen assistierenden Mitarbeiter, der ebenfalls die Laborerhebung unterstützte. So konnten zwei professionell geschulte Personen die Probleme rückblickend einstufen und die schwerwiegendsten Probleme mit der Priorität „A“ versehen, während die anderen Probleme mit „B“ od. „C“ gekennzeichnet wurden.

Hierbei zeigte sich, wie bereits in der Aufzählung der Probleme deutlich gemacht wurde, dass Labortest und Videotest gleich viele schwerwiegende Probleme identifizierten. So ist der Stellenwert des Videotests deutlich geworden, der für das Auffinden derartiger Probleme geeignet ist. Das Auffinden derartiger Probleme stellt den wichtigsten Erkenntnisgewinn eines Tests dar, da die so identifizierten Problemstellen in den meisten Fällen die Ursache für viele Kaufabbrüche sind. Dass der Videotest schon einen Großteil dieser Probleme identifizieren konnte, ist demnach schon als große Stärke dieses Verfahrens zu werten.

Fraglich bleibt beim asynchronen Videotest nur die Ableitung von Handlungsempfehlungen, die durch einen Austausch durch den Testleiter mit dem Probanden bereits angeregt im Rahmen des Tests erfolgen kann. Die nachträgliche Ableitung der Handlungsempfehlungen kann ggf. problematisch sein, wenn die Erklärungen nicht auf Anhieb verständlich genug waren. Aus dem asynchronen Remote-Test die richtigen Handlungsempfehlungen abzuleiten erscheint indes noch schwerer, da die Probanden bei Freitexteingaben meist nur kurze Eingaben vornehmen und schon eine ausreichende Expertise auf Seiten des Auswertenden vorhanden sein muss, damit dieser anhand der vorliegenden Daten aus dem Remote-Testing-Tool ableiten kann, wie ein Problem hätte vermieden werden können.

7.5 Anzahl der Testpersonen und Auswirkung auf die Ergebnisse

Die drei Methoden wurden jeweils mit unterschiedlich hohen Anzahlen an Probanden durchgeführt. Für den Labortest standen 12 Probanden, für den Videotest 20 Probanden und für den asynchronen Remote-Usability-Test 40 Probanden zur Verfügung. Dies erschwert zunächst den Vergleich, soll aber deutlich machen, dass die beiden asynchronen Verfahren bei ähnlichem oder sogar verringertem Aufwand deutlich höhere Probandenzahlen zulassen. Schließlich hat jedes der angewandten Verfahren etwas andere Charakteristika, sodass bei den asynchronen Verfahren eine erhöhte Anzahl an Probanden durchaus praktikabel sein kann und zu mehr Erkenntnissen führt.

Da bei den Vergleichsverfahren die Auswahl von 12 Probanden zum Vergleich mit dem Labortest schwierig ist, da deren einzelner Wert für den Erkenntnisgewinn schwer ermittelt werden kann und die Methoden zunächst vollständig ausgewertet wurden, erfolgt der Methodenvergleich hier rein rechnerisch über einen Dreisatz. Dies entspricht zwar nicht vollständig dem Vergleich bei der Anwendung qualitativer Methoden, soll aber als Anhaltspunkt für den Vergleich dienen. Da u. a. nicht ausgeschlossen werden kann, dass es aktivere und weniger aktive Probanden gibt, kann es sein, dass durch eine weitere Testperson die Problemanzahl deutlich hätte erhöht werden können. Daher wird an dieser Stelle die Annahme als Grundlage verwendet, dass jede Testperson potenziell gleich wertvoll ist und genauso viele Erkenntnisse generiert wie eine andere. Unter dieser Prämisse lässt die folgende Tabelle 3 einen ungefähren Rückschluss auf die Ergiebigkeit der Methoden zu. In Abbildung 10 findet sich die diagrammartige Aufbereitung der Ergebnisse, die in den Kategorien der B- und C-Probleme zeigt, dass der Videotest hier trotz umgerechneter Stichprobe beinahe gleichauf ist mit den gefundenen Problemen im Labortest.

Tabelle 3: Auf n=12 umgerechnete Problemanzahl für die Methoden Videotest und asynchroner Remote-Test.

<i>Methode/Anzahl der aufgedeckten Probleme</i>	Labortest	Videotest	Asynchroner Remote-Usability-Test
Priorität A	24	14	2
Priorität B	13	12	2
Priorität C	14	16	3
Gefundene Probleme insgesamt	51	42	7

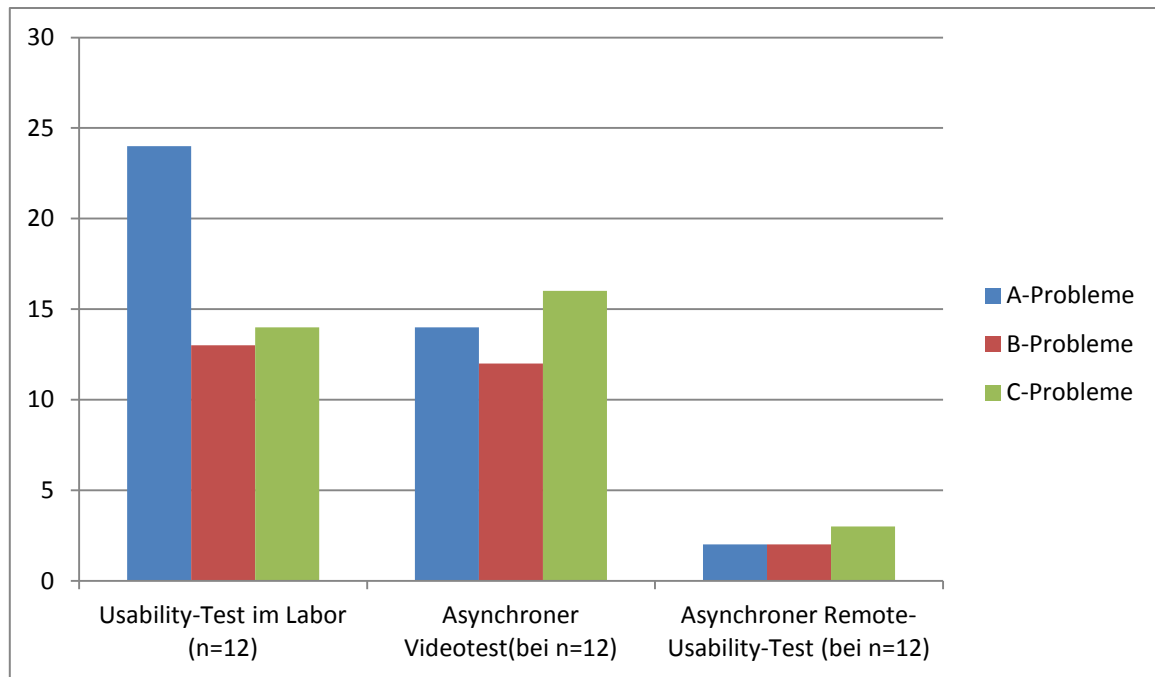


Abbildung 10: Gefundene Probleme der Methoden im Diagramm bei herunter gerechneten Stichproben für n=12.

Diese Vergleichsübersicht deutet an, dass das Potenzial der Vergleichsmethoden bei 12 Probanden deutlich geringer sein dürfte. Dies ist teilweise korrekt. Dabei muss jedoch beachtet werden, dass der asynchrone Videotest und der asynchrone Remote-Usability-Test ohne hohe Mehrkosten und Aufwände auch mit mehr Probanden durchgeführt werden können, um zu vermeiden, dass Erkenntnisse nicht erfasst werden können. Außerdem kann die Ergiebigkeit bereits durch eine weitere aktive Testperson deutlich gesteigert werden. Bei einem qualitativen Verfahren wie dem Videotest kann der Wert einer einzelnen Testperson immer erst nach dem Test festgestellt werden. Eine gute Panelpflege und entsprechend klare Anweisungen können dies schon gut unterstützen.

Bleibt man bei der rechnerischen Annahme der Gleichwertigkeit der Probanden kann bei Gegenüberstellung von Labortest und asynchronem Videotest ein Verhältnis errechnet werden, das zumindest einen groben Rückschluss auf die Verhältnismäßigkeit der Methoden zueinander zulässt. Diesen Versuch eines Vergleichs zeigt Tabelle 4. Es wird deutlich, dass der Videotest bei Problemen der Kategorie „B“ etwa gleichauf mit dem Labortest ist und bei Problemen der Kategorie „C“ sogar mehr Probleme erkennen kann. Bei den schwerwiegenden Problemen ist das Verhältnis dagegen wieder unausgeglich zugunsten des Labortests. Der asynchrone Remote-Usability-Test wird hier nicht gesondert berücksichtigt, da er aufgrund deutlich weniger identifizierter Probleme für diesen Verhältnisvergleich kaum Bedeutung hat.

Tabelle 4: Bei n=12 ermittelte Probleme von Labor- und Videotest und Problemverhältnis.

<i>Methode/Probleme</i>	Labortest	Videotest	Verhältnis Labor- /Videotest
Schwerwiegende Probleme (bei n=12)	24	14	1:0,6
Mittelschwere Probleme (bei n=12)	13	12	1:0,9
Leichte Probleme (bei n=12)	14	16	1:1,1
Anzahl Probleme (bei n=12)	51	42	1:0,8

Dennoch muss nach wie vor berücksichtigt werden, dass dieser Vergleich rechnerisch ist.

Auch wenn es nicht anhand der Daten belegt werden kann, ist der gefühlte Erkenntnisgewinn durch mehr getestete Probanden im Videotest zumindest gering. Dies kann aber auch damit zusammenhängen, dass zunächst der Labortest ausgewertet wurde und so bereits der Großteil der Probleme bekannt war und diese bei jeder weiteren Sichtung eines Auswertungsvideos bereits erwartet wurden. Dennoch bleibt festzuhalten, dass auch der Videotest mit weniger Probanden durchaus praktikabel und ergiebig sein kann, wenn diese von guter Qualität (Engagement und Involvement) sind. Denn die tatsächlich gezählten Probleme weichen aufgrund der sehr individuellen Verhaltensweisen der Probanden sehr stark von der rechnerischen Darstellung ab, die hier nur als Behelf für einen versuchten Vergleich diene.

7.6 Aufwände der einzelnen Methoden im Vergleich

Um nicht nur die Stärken und Schwächen der Methoden festzuhalten, sondern diese auch mit deren Rahmenbedingungen ins Verhältnis zu setzen, werden in diesem Teil die Aufwände für die Durchführung mit der Ergiebigkeit der Methoden verglichen, um eine Einschätzung dahingehend zu ermöglichen, welche Methode sich bei verringertem Budget dennoch lohnt oder was nur mit deutlich höheren Ausgaben überhaupt zielgerichtet herausgearbeitet werden kann. Hierfür müssen mehrere Faktoren berücksichtigt werden. Diese sind der personelle Aufwand, die Kosten und die zeitliche Komponente. Diese Faktoren zeigen dann den Umfang eines Projekts mit einer der drei Methoden. Gemeinsam mit den gesammelten Erkenntnissen zum Potenzial der Methoden kann dann eingeschätzt

werden, welche Methode sich lohnt oder welcher Kompromiss ggf. eingegangen werden muss und sich dennoch aufgrund der Ersparnis einer Methode lohnen kann. Hier werden die ursprünglichen Probandenzahlen wieder angesetzt, d. h. der Labortest wird mit 12 Probanden, der Videotest mit 20 und der asynchrone Remote-Test mit 40 Probanden angesetzt, da dies praktikable Werte sind und hier keine nicht adäquate Umrechnung erfolgen muss. Die angesetzten Werte wurden gemeinsam mit dem betreuenden Unternehmen ermittelt und stellen Anhaltspunkte zum Zeitpunkt der Erhebungen für die Masterarbeit dar. Manntage bezeichnen Arbeitstage à 8 Stunden und die angesetzten Beträge sind als ungefähre Richtwerte zu verstehen.

Hierbei gibt es noch einige Aspekte zu beachten. Die Erhebung des Labortests muss von zwei Mitarbeitern gleichzeitig durchgeführt werden, da einer die Testperson betreut und der andere als Protokollant fungiert. Die Erhebung des Videotests erfordert auf Seiten des Auftraggebers keine Aufwände, da diese durch das ausführende Unternehmen übernommen werden. Hier wird ein konzeptueller Unterschied nochmals deutlich: Der Videotest ist insgesamt als All-in-One-Lösung abgegrenzt von den beiden anderen Verfahren, da er außer bei der Bestimmung der Zielgruppe und der Konzeption der Studie keinen Arbeitsaufwand voraussetzt, der vor der Erhebung zu erbringen ist.

Tabelle 5: Gegenüberstellung der Aufwände und Kosten für die drei verglichenen Verfahren.

<i>Methode/Aufwände</i>	Usability-Test im Labor (n=12, 30 Minuten, 1,5 Erhebungstage)	Asynchroner Videotest (n=20, 20 Minuten)	Asynchroner Remote-Usability-Test (n=40, 20 Minuten)
Studienkonzept u. Tool-Setup	2 Manntage	1,5 Manntage	1,5 Manntage
Rekrutierung u. Incentivierung	pro Proband 50 €	Gesamtkosten v. 750 € (bei 20 Probanden)	pro Proband ca. 25 €
Toolkosten u. Ausrüstungsbedarf	Labor u. Ausrüstung pro Tag 300 €	in Gesamtkosten enthalten	400 €
Erhebungsaufwand	3 Manntage	keine Aufwände	0,5 Manntage
Auswertungsaufwand (schriftlicher Ergebnisband)	4 Manntage	5-6 Manntage	1 Manntag
Gesamtprojektdauer	~ 2,5 Wochen	~ 1,5 Wochen	~ 1-1,5 Wochen
Kosten	9 Manntage + 1.200 €	6,5-7,5 Manntage + 750 €	3 Manntage + 1.400 €

Wie sich in der Tabelle 5 zeigt, bedarf die Studienkonzeption für alle drei Verfahren in etwa gleich viel Aufwand. Der Großteil fällt hierbei auf die Konzeption der Aufgaben und damit des gesamten Tests, da dieser Teil ausreichend elaboriert zusammengestellt werden muss, um durchführbare Tests zu erhalten. Nicht eingeplant ist hierbei ein Pretest, der mit dem ersten Studienkonzept durchgeführt werden kann. Dieser könnte aber insbesondere bei den asynchronen Verfahren sinnvoll sein, da ein späteres Eingreifen während des Tests nicht mehr möglich ist. Dann würde sich dies allerdings deutlich auf die Erhebungsdauer auswirken, wenn erst ein Videotest mit einem Probanden erhoben und ausgewertet werden muss, da dies dann auch etwa einen Tag dauert und Aufwand für die Sichtung und eventuelle Überarbeitung entsteht. Auch würde dieser Pretest nur einen beispielhaften Test liefern, bei dem nicht klar ist, ob die anderen Probanden sich ähnlich oder deutlich abweichend verhalten werden. Dies ist zwar beim Pretest im Labor ähnlich. Dafür kann der Labortest aber jederzeit individualisiert werden, und zwar auch während des Tests.

Die Gegenüberstellung zeigt weiterhin, dass der Labortest den höchsten Aufwand und die höchsten Kosten hat. Dies hängt mit den räumlichen Bedürfnissen und der längeren Testdauer zusammen, die so potenziell mehr Probleme zutage fördern kann. Dafür bot er auch die ergiebigsten Testerkenntnisse (vgl. Kapitel 7.1). Abbildung 11 zeigt die angenommenen Gesamtkosten und die einzuplanenden Manntage für die drei Verfahren im Vergleich.

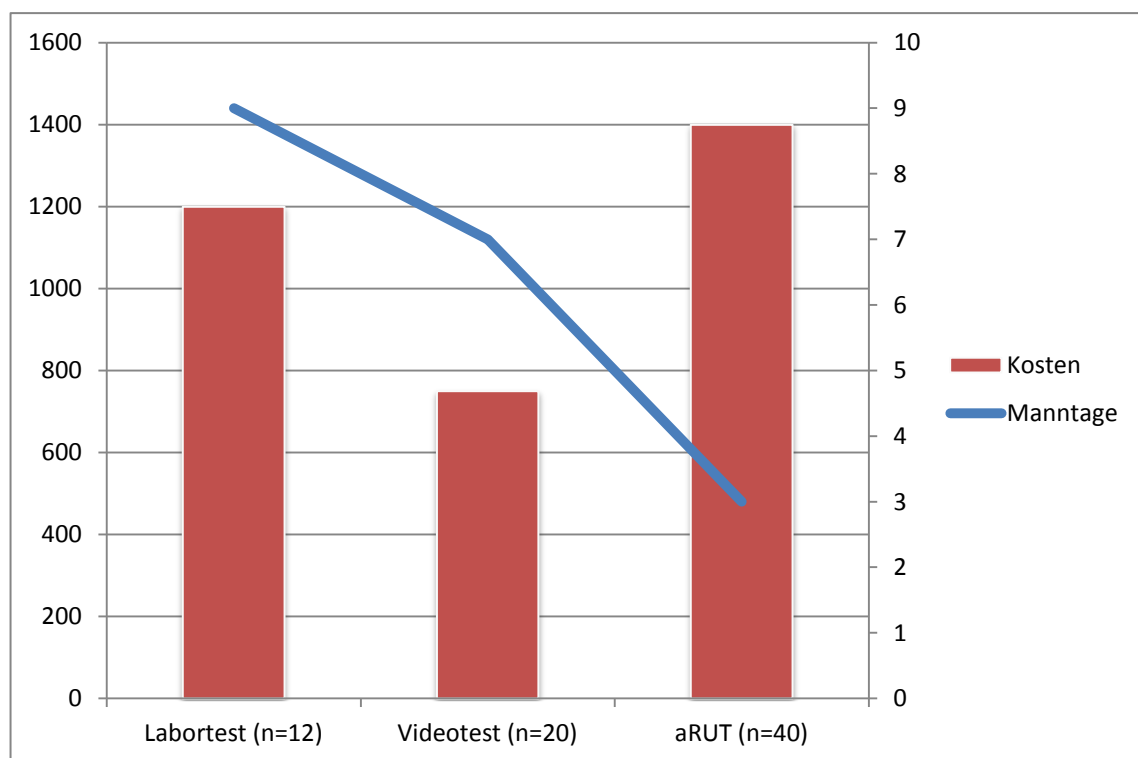


Abbildung 11: Kosten und Manntage für die drei Verfahren im Vergleich.

So kann festgestellt werden, dass sich die Testaufwände in etwa proportional zu dem zu erwartenden Erkenntnisgewinn verhalten: Der Labortest bietet mit hohem Aufwand die ganzheitliche Analyse mit qualitativ tiefgehenden Erkenntnissen und Verbesserungsvorschlägen, der Videotest bietet bei etwas verringertem Aufwand in etwa die gleiche Erkenntnismenge, wenngleich er nicht so viele elaborierte Verbesserungsvorschläge hervorzubringen vermag. Der asynchrone Remote-Usability-Test hat den geringsten Aufwand und liefert dafür auch die wenigsten Erkenntnisse. Wie oben erwähnt behält der Videotest hier eine Sonderstellung, da er für eine Usability-Agentur, die dieses Verfahren nicht selbst anbietet, mit einer immensen Auslagerung von Aufgaben verbunden ist, die beinahe vollständig extern erledigt werden. Bei den beiden anderen Verfahren müssen nur die Tools von extern zur Verfügung stehen. Bei einer Verfügbarkeit von einem entsprechenden Panel und der notwendigen Expertise für die Testvorbereitung und -auswertung kann dann der Großteil durch die Agentur selbst durchgeführt werden. Aus der Perspektive eines Kunden, in diesem Falle eines Anbieters einer Website od. Webanwendung betrachtet, dürften alle drei Verfahren als vollständig ausgelagert anzusehen sein, wodurch sich diese Sonderstellung teilweise relativiert. Dafür liegt dann beim Videotest in der Standardkonfiguration der meisten Anbieter noch die externe Konzeptionierung der Aufgabenstellungen, was sonst eine Agentur übernehmen kann. Aber auch dies bieten die Anbieter der Videotests bisweilen als zusätzliche Dienstleistung an.

Wenn also das Budget und die Zeit ausreichend sind, kann der Labortest gewählt werden, der sich weiterhin gut in ein größeres Projekt einbinden lässt und nebenbei auch die Möglichkeit bietet, den Auftraggeber live bei der Erhebung dabei sein zu lassen, wobei eine Vielzahl von Problemen und Verbesserungsvorschlägen bereits direkt während des Tests besprochen werden kann. In Teilen kann dadurch die Auswertung sogar ersetzt werden bzw. ergänzt werden und für den Auftraggeber werden die Erkenntnisse noch verständlicher, wenn diese tatsächlich durch eine im Nebenraum anwesende Testperson erbracht werden, deren Gesichtsausdruck auch zu sehen ist, sodass diese Informationen auch mit einem echten Nutzer verknüpft werden können. Dies ist zwar ein weicher Faktor, der sich aber als durchaus wertvoll erweisen kann.

Für den asynchronen Remote-Test spricht die mit relativ geringem Aufwand verbundene Auswertung, die aber im durchgeführten Methodenvergleich nur wenige Probleme aufdeckte (vgl. Kapitel 7.2). Hier kann aber unter Rückgriff auf ein mächtigeres Tool eine größere Datenmenge erhoben werden. Dies kann jedoch auch zu höheren Kosten führen. Auf der anderen Seite können Kosten etwa dadurch begrenzt werden, dass keine rekrutierten Probanden als Testpersonen genutzt werden, sondern Besucher der Site, die wie bei einer Onsite-Befragung um die Teilnahme an einem solchen Test gebeten werden können. Zwar sind hier vermutlich hohe Abbruchquoten zu erwarten. Dafür kann es aber auch in kurzer Zeit viele Site-Besucher geben, von denen dann auch eine gewisse Anzahl den Test vollständig durchführt. Die hohe Stichprobengröße würde dann Sicherheit bieten und

die Motivation zur Teilnahme bzw. zum Abschluss des gesamten Tests mit einer angekündigten Verlosung verbunden sein, was die Motivation steigern kann.

Es bleibt festhalten, dass es definitive Unterschiede bei den Methoden gibt, wenn man sich deren Aufwände anhand der Kosten und personellen Bedarfe anschaut und miteinander vergleicht. Dies kann auch als Orientierung für die Wahl der geeigneten Methode dienen. Hierbei kann die Entscheidung auf zwei Arten getroffen werden, und zwar aufwandsorientiert unter Eingang des Kompromisses, dass man ggf. nicht alle Usability-Probleme findet oder keine tiefgehenden Verbesserungsvorschläge generieren konnte. Oder man geht ergebnisorientiert vor und wählt so die passende Methode, deren Kosten ggf. höher sind, aber auch bessere Ergebnisse sicherstellen.

Es ist für das jeweilige Projekt im Vorfeld zu empfehlen, eine individuelle Gegenüberstellung der Kosten und Aufwände durchzuführen. So kann es z. B. sein, dass ein Team, das bislang keine Erfahrung mit asynchronen Remote-Testing-Tools hat, deutlich mehr Aufwand in die Testvorbereitung investieren muss, da zunächst das Tool und dessen Funktionsumfang gesichtet werden müssen, bevor dort die Aufgabenstellungen eingerichtet werden können. Auch kann es sein, dass für eine schwer zu erreichende Zielgruppe deutlich höhere Kosten für die Rekrutierung und Aufwandsentschädigung einkalkuliert werden müssen, die die Projektkosten in die Höhe schnellen lassen. Für derartige Projekte wird aber i. d. R. ein höheres Projektvolumen vorab bereitgestellt. In einem solchen Fall gelten die erhöhten Kosten dann auch für Remote-Verfahren. Die Probanden sind dann aber ggf. auch an verschiedenen Orten rekrutierbar und erhalten dennoch eine etwas geringere Aufwandsentschädigung. Die individuelle Betrachtung wird also zu abweichenden Ergebnissen führen. Die Verhältnisse werden sich aber stets im ähnlichen Rahmen bewegen, sodass eine identische Reihenfolge der Verfahren zu erwarten ist, bei der der Labortest am aufwändigsten ist, der asynchrone Videotest einen verringerten Aufwand hat und der asynchrone Remote-Usability-Test mit dem verhältnismäßig geringsten Aufwand zu kalkulieren sein wird.

Diese Werte sollen als Orientierung dienen. Für die vorgenommene Erhebung im Rahmen dieser Studie waren sie in jedem Fall korrekt und wurden im Nachhinein dokumentiert anhand der tatsächlich angefallenen Kosten. Insbesondere die Aufwände im Hinblick auf die einzuplanenden Manntage können aber deutlichen Schwankungen unterliegen, wenn es sich hierbei um ein umfangreiches Onlineangebot handelt oder der Mitarbeiter, der mit der Konzeption der Aufgaben betraut ist, noch nicht über eine ausreichende Sachkenntnis des Angebots verfügt. Dann erscheint es in jedem Fall sinnvoll, dafür mehr Zeit einzuplanen und es entstehen ggf. weitere Kosten für einen oder eventuell mehrere Abstimmungstermine, wenn es im Laufe des Projekts zu Problemen kommt. Dies zeigt noch einmal, dass der Fokus der beispielhaft durchgeführten Erhebung definitiv auf dem E-Commerce lag. Wenn andere Anwendungen getestet werden, ist mit anders gearteten Aufwänden zu rechnen. Handelt es sich dabei etwa um ein Intranet, das von den Nutzern auch von zu Hause aus genutzt werden kann, sind für den Test zwar meist keine höheren Aufwände einzukalkulieren. Dafür ist aber auf Seiten der

Studienkonzeption definitiv mit erhöhten Aufwänden in der Planung aufgrund der notwendigen Einarbeitung zu rechnen. Die Verhältnismäßigkeiten der Verfahren dürften sich nichtsdestotrotz annähern. Schließlich sind die erwähnten Mehraufwände für alle drei Verfahren notwendig, da sie sich an der Stelle noch nicht explizit auf das anzuwendende Verfahren beziehen, sondern erst die tatsächliche Bearbeitung des Projekts möglich machen sollen.

Eine vorherige konkrete Vergleichsbetrachtung der Methoden sollte daher die genannten Faktoren wie die bereits vorhandene Expertise der involvierten Mitarbeiter, die notwendige Einarbeitungszeit und die Rekrutierungsaufwände und Entschädigungszahlungen berücksichtigen, um im Vorfeld eine möglichst realistische Einschätzung der Aufwände zu ermöglichen und anhand dieser Gegenüberstellung die optimale Methode auswählen zu können.

8 Prüfung der Hypothesen

Als Basis für die Bestätigung oder Widerlegung der Hypothesen dienen die im Kapitel 7 zusammengetragenen Ergebnisse der Erhebungen. Aus diesen kann die Beantwortung der Hypothesen abgeleitet werden.

Für die übersichtliche Darstellung der Hypothesenauswertung werden diese hier noch einmal wiederholt und dann durch einen ergänzenden Text mit der jeweiligen Erklärung versehen, ob sie sich bewahrheitet haben oder widerlegt worden sind. Hierbei ist zu beachten, dass für alle drei Methoden jeweils die gesamte Anzahl an Teilnehmern berücksichtigt wurde und dies daher die Vergleichbarkeit etwas einschränkt. Da aber die Anzahl an Testpersonen für die jeweiligen Verfahren durchdacht festgelegt wurde und unter Annahme des Potenzials und der Eignung der Testmethode und des für solche Tests zur Verfügung stehenden Budgets berücksichtigt wurde, erscheint diese Betrachtung sinnvoll. Außerdem wurde im Kapitel 7.5 auf den Informationsgewinn durch die höheren Teilnehmerzahlen in Videotest und asynchronem Remote-Test eingegangen, sodass etwa die hohe Problemanzahl aus dem Videotest auch mit einer geringeren Anzahl an Probanden beinahe erreicht worden wäre.

1. Der Labortest ist durch seine Interaktivität ergiebiger als die beiden asynchronen Verfahren und deckt so mehr Usability-Probleme auf als diese. Der Videotest deckt indes mehr Probleme auf als der asynchrone Remote-Usability-Test.

Diese Hypothese konnte nicht vollständig bestätigt werden. Zwar ist der Testablauf im Labortest durch die Interaktivität stärker beeinflussbar und kann damit ein während des Tests plötzlich auftretendes neues Erkenntnisinteresse auf Seiten der Testleitung sofort befriedigen. Bei identischer Anzahl an Testpersonen dürfte der Labortest auch eine etwas größere Anzahl an Problemen aufdecken. Bei dem durchgeführten Vergleich ($n=12$ im Labor, $n=20$ im Videotest), zeigte sich ein quantitativer Vorsprung des Videotests, d. h. eine Testperson im Videotest steht dem informationellen Mehrwert einer Testperson im Labor nur geringfügig nach. Der asynchrone Remote-Usability-Test wurde zwar mit einer hohen Stichprobe ($n=40$) durchgeführt, blieb dennoch deutlich hinter den Erwartungen zurück und identifizierte nur sehr wenige Probleme der unterschiedlichen Schweregrade, sodass sein Erkenntnisgewinn als deutlich geringer einzuschätzen ist als die beiden qualitativen Verfahren, die beide deutlich kleinere Stichproben testeten.

Durch die Interaktivität sticht der Labortest aber dennoch heraus: So ist die Problemanzahl zwar nominell nur leicht erhöht bzw. in den Schweregraden „B“ und „C“ etwa gleichauf mit dem Videotest. Dafür fällt die Ableitung von Handlungsempfehlungen durch den interaktiven Charakter deutlich leichter. Schließlich kann der Testleiter sich mit dem Probanden intensiv über einen Verbesserungsvorschlag austauschen, sodass die

Handlungsempfehlung praktisch schon im Verlauf des Tests entstehen kann. Bei den anderen Verfahren ist die Wahrscheinlichkeit nicht ganz so hoch, einen ausformulierten und auf Anhieb verständlichen Verbesserungsvorschlag zu erhalten.

Dadurch kann mit dem Labortest auch dem Auftraggeber relativ einfach gezeigt werden, welche Bereiche optimierungsbedürftig sind. Ist dieser nämlich beim Test mit vor Ort und beobachtet den Probanden, kann der Protokollant im Beobachtungsraum weitere Erläuterungen liefern, sodass auch dem Auftraggeber klar wird, wie er seine Anwendung verbessern kann. Bei der Sichtung eines Videotests oder der Ergebnisse des asynchronen Remote-Tests ist die Abhängigkeit von den Ergebnissen größer, da auf diese nur nach einem Test Zugriff besteht und keine nachträglichen Erläuterungen vorgesehen sind, es sei denn sie werden durch die auswertende Agentur vorgenommen. Hier wäre eventuell zu überlegen, ob die Probanden für eine erneute Kontaktaufnahme zur Verfügung stehen sollten, sodass sie beispielsweise nach dem Test angerufen werden könnten, um solche Unklarheiten auszuräumen.

2. Im asynchronen Videotest werden genauso viele schwere Usability-Probleme aufgedeckt wie im Labortest.

Die Summe an aufgedeckten schwerwiegenden Usability-Problemen (gekennzeichnet mit der Priorität A) war bei beiden Verfahren identisch, auch wenn nicht alle Probleme dieselben waren. Hierbei muss jedoch auch wieder die höhere Stichprobengröße im Videotest berücksichtigt werden. Bei gleicher Anzahl an Probanden dürfte der Videotest in Abhängigkeit von der Qualität der Probanden etwas schlechter abschneiden. Dort ist die Abhängigkeit von den Probanden noch stärker, da diese im Labor noch durch provokante Fragen o. ä. zu mehr lautem Denken und somit zu mehr potenziell wichtigen Erkenntnissen animiert werden können. Es war zwar nicht Bestandteil der Hypothese, ist aber dennoch ein wertvoller, wenngleich scheinbar trivialer Hinweis: Ausschließlich im Labortest kann eine vollumfängliche Diskussion über die schwerwiegenden Usability-Probleme erfolgen (s. o.). Die Nutzer beschreiben das Problem. Daraufhin können sie gefragt werden, wie man das Problem beseitigen könnte. Durch mehrere Rückfrageschleifen können so schon elaborierte Verbesserungsvorschläge mit allen Testpersonen, bei denen das Problem auftrat, erarbeitet werden. Im asynchronen Verfahren des Videotests muss darauf gehofft werden, dass die Probanden von sich aus gute und verständliche Verbesserungsvorschläge machen. Tun sie dies nicht oder erklären sie diese nicht deutlich genug, kann aufgrund der nicht vorhandenen Möglichkeit zur Nachfrage ein herber Informationsverlust entstehen. Liegt jedoch der Fokus auf der Identifikation der schwerwiegenden Usability-Probleme, kann dieses Ziel definitiv auch mit dem Videotest zufriedenstellend erreicht werden. Als Folgeaktion kann dann etwa ein

Workshop stattfinden, in dem die verschiedenen mit der Web-Anwendung betrauten Abteilungen zusammenfinden. Sie können dann anhand der Ergebnisse die Problemstellen diskutieren und selbsttätig Handlungsempfehlungen herleiten und diskutieren. So kann ggf. eine etwa gleichwertige Sammlung an Handlungsempfehlungen entstehen, wie dies der Labortest bereits durch bloße Durchführung des Tests liefern kann.

3. Im Labortest und im asynchronen Videotest werden in der gleichen Erhebungszeit gleich viele Usability-Probleme aufgedeckt.

Diese Hypothese kann nicht abschließend bewertet werden, da hierfür die Testverfahren zu unterschiedlich sind, denn der Labortest war von vornherein auf eine Erhebungsdauer von 30 Minuten und der asynchrone Videotest auf 20 Minuten angelegt. Hier im Labortest z. B. nach 20 Minuten einen Schnitt zu machen und die Problemhäufigkeiten zu vergleichen, ergibt wenig Sinn, da die Tests dann noch an komplett unterschiedlichen Stellen sind und im Videotest bestimmte Aufgaben noch gar nicht begonnen werden konnten. Hier muss dann auch gefragt werden, ob es nur um das Aufdecken der Probleme in der Erhebungszeit geht oder auch um die Menge an zu einem Problem erfassten Daten. Potenziell wäre aber wahrscheinlich die gleiche Problemanzahl pro Zeiteinheit aufzudecken, wenngleich dies aufgrund der Eigenheiten der Methoden nicht unbedingt zielführend ist.

Der Videotest bietet trotz der Begrenzung auf eine Erhebungszeit von ca. 20 Minuten eine relativ hohe Dichte an zu stellenden Aufgaben. So wurden im Studienkonzept für den Labortest und den Videotest die gleichen Aufgaben gestellt. Der Zeitgewinn des Videotests lässt sich durch den Wegfall des retrospektiven Interviews erklären, das dort nicht erfolgt. In kürzerer Zeit können daher etwa gleich viele Aufgaben abgeprüft werden, zu denen potenziell gleich viele Probleme wie im Labortest aufgedeckt werden können. Insbesondere bei Aufgaben, die im Labor schrittweise durchgeführt werden, kann es im Remote-Verfahren zu einem Informationsverlust führen, wenn die Probanden Aufgabenteile auslassen. Hier ist demnach die klare und strukturierte Formulierung der Aufgabenstellungen noch wichtiger als im Labor, wo der Testleiter jederzeit eingreifen kann und eine Aufgabe nochmals anders formulieren kann, sofern diese nicht verstanden wurde.

4. In den testleiterunabhängigen Verfahren ist ein natürlicheres Verhalten der Teilnehmer festzustellen als im Labortest, wo Nutzer etwa Kommentare beeinflusst durch soziale Erwünschtheit abgeben oder unnatürlich später aufgeben, um unbedingt zur Lösung zu gelangen, wenn sie zeitgleich beobachtet werden.

Diese Hypothese hängt eng zusammen mit dem sog. Involvement, also der Art und

Weise, wie stark die Probanden sich während des Tests in das Testobjekt hineinversetzen und wie stark sie durch die Testbedingungen ein vom normalen Alltagsverhalten abweichendes Verhalten an den Tag legen. Dies ist stets eine schwer zu beurteilende Situation, da die Testleitung i. d. R. nur das während des Tests gezeigte Verhalten der Probanden kennt und dieses dann nach eigenem Ermessen in Bezug auf seine Authentizität einschätzen muss. Im Labortest ist die Konzentration der Probanden auf das Testobjekt hoch, da es keine Störeinflüsse gibt und sie sich durch den Testleiter angeregt fühlen, sich mit der Site intensiv auseinanderzusetzen. Hier kann dann angezweifelt werden, dass es eine vergleichbare Situation auch zu Hause geben könne, da es dort stets ablenkende Faktoren gibt oder die Kaufabbrüche bei Unzufriedenheit einen größeren Anteil ausmachen dürften. In den Remote-Verfahren konnte hingegen festgestellt werden, dass viele Probanden Aufgabenteile nicht ausgeführt haben, da sie auch nicht so stark auf die Aufgabenstellungen konzentriert waren. Dies wäre ein im Alltag wahrscheinlich nicht auftretendes Verhalten, da die Nutzer zumindest eine sehr deutliche Vorstellung vom eigenen Kaufbedürfnis haben, bei dem sie wissen, welche Schritte hierfür durchlaufen werden müssen. Allenfalls ein nicht so stark auf die Site konzentriertes Surfverhalten, bei dem dann evtl. Schaltflächen übersehen werden, würde dann das normale Surfverhalten abbilden. Da aber die geringe Beachtung der Aufgabenstellung in den Remote-Verfahren der Aufgabenbearbeitung selbst schon vorangestellt ist, sind die dann in der Erledigung auftretenden Verhaltensweisen schon zu stark durch die vorherige Nichtbeachtung geprägt, sodass gerade wegen dieses geringen Involvements keine Rückschlüsse durch dieses Verhalten möglich werden. Es bleibt demnach unklar, welches Verfahren am ehesten das natürliche Nutzerverhalten abbildet. Bei allen drei Verfahren gibt es Faktoren, die in Frage stellen, dass dies das natürliche Nutzerverhalten ist. Dabei bleibt jeder Usability-Test auch immer ein Test, d. h. eine konstruierte Situation, die Nutzer nicht bei ihrem natürlichen Nutzungsverhalten beobachtet, sondern der Nutzer sich durch das Szenario meist in eine vorgefertigte Situation begibt und dort nicht natürlich handeln wird.

In diesem Zusammenhang muss auch die gewünschte Zielsetzung eines Usability-Tests beachtet werden. Welches Ziel verfolgt die Auftraggeber? Wenn das Ziel ist, das natürliche Nutzerverhalten zu dokumentieren, dürfte ein Verfahren der unbeteiligten Beobachtung wie die Auswertung der Analytics-Daten seines Webangebots als lebensecht anzusehen sein. Dort kann anhand großer Häufigkeiten festgestellt werden, wie sich echte Nutzer der Site verhalten haben. Auch beim Usability-Test wird davon gesprochen, dass man dem echten Nutzer über die Schulter schauen könne. Der eigentliche Fokus bleibt jedoch die Identifikation von Usability-Problemen. Dies können zwar auch die Analytics-Daten zeigen. Durch qualitative Äußerungen der Probanden bzw. über die Analytics-Daten hinausgehende Daten im asynchronen Remote-Test können die Probleme dagegen deutlicher eingegrenzt und damit direkt identifiziert werden. Die Zielsetzung eines jeden Usability-Tests sollte daher eher

sein, dass man weitere Daten sammelt, die über die üblicherweise erfassten Daten der tatsächlichen Site-Besucher hinausgehen. Dadurch erhält man wichtige Aussagen dazu und ein vertieftes Verständnis der Probleme der Anwendung. Ob die Nutzer sich beim Auffinden eher natürlich oder eher unnatürlich verhalten, wird daher für die primäre Zielsetzung des Usability-Tests, nämlich das Identifizieren von Usability-Problemen, eher unerheblich sein. Eine klare Aussage darüber, ob das Verhalten nun natürlich sei oder nicht, kann ohnehin nicht abschließend beantwortet werden. Das tatsächliche Verhalten der Probanden in einem Test muss dementsprechend als deren übliches Verhalten hingenommen werden. Der Fokus soll daher auf dem Identifizieren der Usability-Probleme liegen, die mit verschiedenen Verhaltensweisen der Nutzer aufgedeckt werden können. Je näher diese an deren natürlichen Nutzerverhalten liegen, desto mehr Probleme werden aufgedeckt, die auch im normalen Ablauf der Sitenutzung gefunden werden dürften. Darüber hinaus kann aber auch das Auffinden vermeintlich kleinerer Probleme, die nicht zum üblichen Surfpfad eines Nutzers gehören, zur Verbesserung einer Site beitragen, da auch kleinere Probleme, auf die nicht jeder Nutzer stößt, zu Frustration führen können (vgl. EBERHARD-YOM 2010, S. 131). Die Nähe zum natürlichen Nutzerverhalten bleibt also ein Aspekt, der nicht abschließend mit Vor- und Nachteilen belegt werden kann. Der Fokus sollte daher auf der Identifikation von Usability-Problemen liegen und nicht im Abbild des tatsächlichen Nutzerverhaltens, das durch das Szenario ohnehin so stark gelenkt wird, dass die Nähe angezweifelt werden kann.

5. Der Videotest und der asynchrone Remote-Usability-Test können bei jeder Testperson zu individuelleren Abläufen und Ergebnissen führen, da die Nutzer sich nur an den Aufgabenstellungen orientieren und mit diesen nicht interagiert wird. Im Labortest kann der Nutzer hingegen während des gesamten Testverlaufs gezielt gelenkt werden und z. B. gewollt auf einzelne Aspekte gestoßen werden.

Hier hat sich im Test das Problem ergeben, dass die Probanden in den Remote-Verfahren, insbesondere im asynchronen Remote-Usability-Test die Anweisungen teilweise nicht korrekt umgesetzt haben. So sind dort bei einigen Probanden gar keine richtigen Abläufe entstanden, da die mehrteiligen Aufgaben nicht am Stück abgearbeitet wurden. Je nach Erfahrung der Testpersonen ist eine sehr individuelle Auslegung der Anweisungen ohnehin unwahrscheinlich, da die Nutzer aus vorherigen Tests i. d. R. schon wissen, dass die aneinandergereihte Erledigung von Aufgaben gefragt ist und sie dann den Test auch schneller abschließen können. Im Labor können die Testpersonen von sich aus eine individuelle Nutzungsphase einläuten, die dann im Zweifel durch die Testleitung abgebrochen werden kann, wenn hierfür die Zeit nicht ausreicht oder kein Erkenntnisinteresse für diesen Bereich besteht. Die Wahrscheinlichkeit, beim Remote-Test einen individuelleren Ablauf zu erhalten,

ist also demnach sogar geringer als im Labor. Lediglich in der freien Nutzungsphase kann es, sofern es keine zeitliche Vorgabe hierfür gibt, abhängig von der Aktivität des Probanden, einen ähnlich hohen Informationsgewinn wie im Labor geben. Daher kann die synchrone Betreuung im Labor eher noch als Vorteil angesehen werden, da zwar grundsätzlich auch eine weitestgehend ungelenkte Abarbeitung für die Probanden ermöglicht werden kann, aber jederzeit ein Eingriff möglich ist, falls die Probanden einen Zwischenschritt nicht selbsttätig abschließen können. Innerhalb der Aufgaben wurde aber eher ein negativ individueller Ablauf festgestellt, der sich u. a. im selbsttätigen Abschluss von Kaufvorgängen niederschlug, obwohl es dazu noch keine Anweisung gab. So legte eine Reihe von Probanden ein Produkt bereits in den Warenkorb, obwohl zunächst die Artikeldetailseite zu betrachten war, ohne eine Aktion vorzunehmen. Damit ist insbesondere bei geübten Testern zu rechnen, die mit ihrem gelernten Verhalten in die Tests gehen und daher viele Aktionen nach eigenem Ermessen durchführen, da sie davon ausgehen, dass dies Bestandteil des Tests ist.

Das Potenzial für einen ergebnisoffenen Test ist demnach im asynchronen Verfahren nicht nachweisbar größer, da die Probanden sich ausschließlich an der Aufgabenstellung orientieren können. Wenn sie dann hierbei einzelne Aspekte nicht berücksichtigen, ist der Ablauf zwar individuell, aber auch nicht ergiebiger, da Teilschritte nicht korrekt durchgeführt werden und so die Ergebnisse nicht oder zum falschen Zeitpunkt vorliegen.

6. Der Videotest und der asynchrone Remote-Usability-Test können vielfältige Probleme aufdecken, da jeder Testrechner eine individuelle Konfiguration hat (z. B. unterschiedliche Bildschirmgrößen). In der Aufzeichnung des Videotests können diese Auswirkungen sogar direkt gesehen werden und so bei der Auswertung besser berücksichtigt werden.

Die erhobenen Daten, die das Tool für den asynchronen Remote-Usability-Test Loop11 erhob, lassen keinen hilfreichen Rückschluss auf die individuelle Konfiguration der Computer der Probanden zu. Zwar werden das Betriebssystem und der verwendete Browser abgespeichert, eine Angabe der Bildschirmauflösung oder eine echte Darstellung (Screenshot) eines aufgerufenen Bildschirms mit der vom Probanden verwendeten Auflösung gibt es indes nicht. Im Videotest kann hingegen der Screen der Testperson über die gesamte Testdauer beobachtet werden. So wird u. a. deutlich, was im sofort sichtbaren Bereich bei der verwendeten Bildschirmauflösung und der Browserkonfiguration (z. B. installierte Toolbars, die Höhe einnehmen) angezeigt wird. Je mehr Testpersonen am Test teilnehmen, desto mehr potenziell verschiedene Bildschirmdarstellungen können dann analysiert werden. Auch können ggf. technische Probleme mit den Seiten aufgrund der Rechnerkonfiguration festgestellt werden. Im Labor ist am einmal konfigurierten Rechner nicht mit Fehlern zu rechnen, während am privaten Rechner eines Probanden etwa ein fehlendes Flash-Plugin

bereits für Probleme sorgen kann, sodass bestimmte Seitenbereiche evtl. nicht angezeigt werden. Sollte eine Zielsetzung daher die Erfassung des Nutzererlebnisses bei unterschiedlichen Bildschirmgrößen sein, kann dies am ehesten mit dem Videotest festgestellt werden. Da die Bildschirmauflösung aber nicht als Rekrutierungsvariable verwendet werden kann, ist die Menge an unterschiedlichen Auflösungen zufällig, sodass eine systematische Analyse nicht möglich ist. Als ergänzende Daten können unterschiedliche Bildschirmauflösungen dagegen dennoch hilfreich sein.

7. Der Videotest liefert in kurzer Zeit mit relativ geringem Aufwand für den Auftragnehmer direkt verwertbare Ergebnisse, aus denen dieser auch selbsttätig relativ gut die wichtigsten Erkenntnisse ableiten kann.

Dies kann bestätigt werden, denn es liegen vom Videotest online aufrufbare Aufzeichnungen vor, aus denen das Verständnis gewöhnlicher Nutzer hervorgeht. Dies kann sich ein Auftragnehmer selbst anschauen und dadurch bereits verwerten. Hat er z. B. einen Labortest in Auftrag gegeben, hat er diese Gelegenheit zunächst nur, wenn er die Tests auch mit beobachtet. Ist dies nicht der Fall, muss erst die Auswertung durch das ausführende Unternehmen abgewartet werden bzw. es muss die Auswertungssoftware vorliegen. Da die Labortest-Aufzeichnungen i. d. R. aber länger sind als die der Videotests, bedarf es auch mehr Zeit zur Auswertung. Zusätzlich war für den Labortest auch noch ein Personalbedarf in Form von professionellen Usability-Consultants vonnöten. Zwar sollte auch die Fragekonzeption für einen Videotest relativ professionell erfolgen. Jedoch ist die Hemmschwelle deutlich geringer, auch mit weniger elaborierten Aufgabenstellungen zu testen, da u. a. die Kosten verhältnismäßig gering sind, wenn man es mit dem klassischen Labortest vergleicht.

Die Ableitung der Handlungsempfehlungen aus den Aufzeichnungen wird einer Person mit Expertise im Usability-Bereich leichter fallen als jemandem, der damit keine Erfahrung hat. Dennoch wird es auch einem Laien ermöglicht, das Material zunächst zu sichten und ein Verständnis für das Verhalten des Nutzers zu entwickeln. Die Wahrscheinlichkeit, dass ein unerfahrener Nutzer aus dem Datenbestand des Videotests zu Aussagen über die Usability seines Webangebots kommt, ist hier deutlich höher als bei der Sichtung von Daten aus dem asynchronen Remote-Test, da die Videoaufzeichnungen lebhafter sind und damit direkt nachvollzogen werden können.

8. Die Ergebnisse sind von der verwendeten Testmethode weitgehend unabhängig, da diese eher von anderen Rahmenbedingungen wie dem Studienkonzept und der Interpretation der Ergebnisse im Nachgang abhängig sind.

Für diese Hypothese kann keine abschließende Aussage getroffen werden. Zwar spielt die Konzeption der Studie eine entscheidende Rolle für den Ablauf und die Ergiebigkeit eines Tests. Teilweise erfolgt die Ausrichtung der Studie bereits so, dass die Nutzer auf Probleme gestoßen werden. Es wird so darauf abgezielt, Probleme, die etwa auf Seiten der Usability-Agentur hinlänglich bekannt sind, durch das Ausrichten eines Tests auf das explizite Aufdecken dieses Problems im Test auftreten zu lassen, um ein Problem zu verdeutlichen bzw. erstmalig zu präsentieren, wenn gezeigt wird, dass dies ein Problem ist, das tatsächlich auf Nutzerseite vorhanden ist und nicht nur durch eine expertenbasierte Vorbetrachtung zum Vorschein gekommen ist.

Weiterhin bringt jede am Test-Setting und den Auswertungen beteiligte Person eine subjektive Sichtweise in die Angelegenheit hinein. Ist etwa ein Problem zwar vorhanden, wird aber nicht ausreichend oft durch Probanden als solches beschrieben, kann im Labortest durch nochmaliges Hinsteuern oder explizites Ansprechen des Problems der Nutzer stark beeinflusst werden, sodass er dann doch Äußerungen tätigt, die ein Problem beschreiben, das er eigentlich in seinem persönlichen Nutzererlebnis nicht hatte.

Aufgrund der konstruktiven Unterschiede der Testverfahren darf dennoch nicht angenommen werden, dass deren Ergebnisse vollständig unabhängig von der Methode sind, da die Abläufe methodenspezifisch voneinander abweichen. Wird das Studienkonzept möglichst gut adaptierbar für alle verwendeten Methoden eingesetzt, kann in Ansätzen sichergestellt werden, dass die verwendeten Testmethoden nach ähnlichen Abläufen erfolgen. In der Praxis hat dies i. d. R. keine Relevanz, da die Erhebung in aller Regel nur mit einer Methode durchgeführt wird. Die Frage nach dem Einfluss der Methode kann daher nicht abschließend beantwortet werden bzw. es sollte im Vorfeld eine problematisierende Auseinandersetzung mit den Vor- und Nachteilen der Methode erfolgen. Steht nur das Auffinden einer möglichst großen Anzahl an Usability-Problemen im Vordergrund, muss ohnehin von den Ergebnissen der angewandten Methode ausgegangen werden, da gar keine anderen Erkenntnisse vorliegen. Nur in einem tatsächlich durchgeführten Methodenvergleich gibt es überhaupt derartige Vergleichswerte.

Wird aber der Versuch einer möglichst ergebnisorientierten Studienkonzeption unternommen und auch bei der Auswertung darauf geachtet, dass möglichst alle Erkenntnisse berücksichtigt werden (z. B. durch den Einsatz mehrerer Personen bei der Auswertung), sollte das Potenzial der Methode zufriedenstellend ausgenutzt sein und nur noch in begrenzter Abhängigkeit von der Methode stehen. Schließlich kann nie abschließend festgestellt werden, welche Einflüsse sich auf eine Erhebung ausgewirkt haben. Dies können auch triviale Faktoren sein wie die aktuelle Stimmung einer Testperson, deren Abweichung vom eigentlichen Nutzerverhalten praktisch nicht festgestellt werden kann. Auch technische

Probleme können eine Untersuchung verzögern und dadurch zu verfälschten Ergebnissen führen. Aufgrund der Ansiedlung der Usability-Forschung in der empirischen Sozialforschung lassen sich Einflüsse nie vermeiden und sind als inhärenter Bestandteil von solchen Erhebungen zu sehen. Ein weitgehender Ausschluss derartiger Abweichungen würde daher eher gegen das Interesse solcher Studien stehen, da das Ziel ja das Abbild des natürlichen Nutzerverhaltens sein soll, das ständigen Schwankungen ausgesetzt ist. So haben alle drei Methoden mit diesem Umstand zu tun, werden aber darüber hinaus noch durch ihre spezifischen Eigenschaften als weiterer Einflussfaktor nicht auszuschließen sein.

9 Fazit

Es wurden drei Usability-Testverfahren methodisch miteinander verglichen. Der Vergleich erfolgte an einem Fallbeispiel. Es wurde dabei dennoch versucht, den Vergleich möglichst allgemeingültig zu halten. Daher wurde auch im Text der Arbeit der Fokus eindeutig auf die Potenziale der untersuchten Methoden gelegt und nicht auf die Probleme selbst oder deren Zusammenhang mit dem untersuchten Fallbeispiel. Dieser Teil wird dem Auftraggeber überlassen, der damit das Ziel der Optimierung seines Angebots verfolgt. Für die Erkenntnisse des Methodenvergleichs sind die Anzahl und die Priorisierung der Probleme die zentralen Elemente. Zwar sind diese eindeutig vom Untersuchungsobjekt abhängig. Jedoch wird die Validität dahingehend unterstellt, dass alle drei Verfahren auf dasselbe Untersuchungsobjekt angewandt wurden. Dieses wurde über den gesamten Testzeitraum konstant gehalten, sodass allen Probanden der identische Shop präsentiert wurde und so bei sehr ähnlichen Aufgabenstellungen potenziell die gleichen Ergebnisse für jedes Verfahren denkbar waren. Auf Seiten der Auswertung sind eventuelle Einflüsse aufgrund der Reihenfolge der Verfahren nicht auszuschließen, wenngleich diese auch kaum Relevanz auf die gefundenen Probleme bzw. deren Identifikation und Dokumentation haben. Allenfalls kann ein Versäumnis vermutet werden, dass bereits mehrfach in den anderen Methoden aufgefundene Probleme nicht mehr notiert werden, falls man sich unsicher ist ob diese ggf. bereits dokumentiert sind. Auf der anderen Seite wäre der Einsatz unterschiedlicher Personen für die Dokumentation ebenso wenig zielführend gewesen, da dann nicht klar ist, ob alle die identischen Relevanzkriterien und Schwellen für ein Problem ansetzen und so ggf. unterschiedlich intensiv protokollieren. Durch die personelle Kontinuität im Projekt und die zeitlich nah beieinanderliegenden Erhebungen wurde versucht, mögliche Abweichungen bei den Durchführungen zu vermeiden, sondern möglichst dafür zu sorgen, dass alle Verfahren unter identischen Gesichtspunkten angewandt werden. Daher können die Ergebnisse relativ sicher betrachtet werden und dürfen als Orientierung für die Wahl der adäquaten Methode bei allen ähnlichen Usability-Tests dienen. Wer eine Untersuchung plant, kann sich daher an diesen Empfehlungen orientieren und weiß, mit welchen Vor- und Nachteilen im Hinblick auf die Ergiebigkeit und im Hinblick auf die geplanten Aufwände und Kosten bei den drei untersuchten Verfahren in etwa zu rechnen ist.

Mit einem asynchronen Remote-Usability-Test können einige schwerwiegende Probleme einer Anwendung identifiziert werden. Das Aufsetzen eines solchen Tests ist relativ kurzfristig möglich und die Auswertung nimmt nicht sehr viel Zeit in Anspruch. Der Wahl des Tools zur Erhebung mit dieser Methode sollte besondere Aufmerksamkeit gewidmet werden, damit die Erwartungen an den zu erhaltenden Datenbestand auch erfüllt werden und es hier nicht zu Frustration auf Seiten des erhebenden Unternehmens kommt. Hier sei eine ausgiebige Sichtung verschiedener Tools empfohlen, bevor ein Tool voreilig ausgewählt wird, das nicht über den gewünschten Funktionsumfang verfügt. Das für die Erhebung der Masterarbeit verwendete Tool Loop11 wurde vom Anbieter kostenfrei zur

Verfügung gestellt und daher als Vergleichsinstrument verwendet. Hier sei darauf hingewiesen, dass der asynchrone Remote-Usability-Test der Vollständigkeit halber mit erhoben wurde, aber eher eine untergeordnete Rolle spielte, da der Fokus auf dem Videotest lag. Daher sollte auch bei den Ergebnissen und der Bewertung des aRUT beachtet werden, dass dieser anhand der Anwendung im Rahmen dieser Studie bewertet wurde. Ein anderes Tool hätte zu deutlich anderen Ergebnissen geführt. So sind andere Tools für das asynchrone Remote-Testing am Markt, die über einen weitaus höheren Funktionsumfang verfügen und dann zu besseren, da umfangreicheren Ergebnissen führen können. Nicht ausschließen lässt sich allerdings auch dort, dass die Probanden den Test lieber schnell abschließen wollen und sich daher nicht so viel Mühe geben wie in einem Labortest. Dann kann auch mit einem umfangreichen Tool keine Abhilfe geschaffen werden. Da die Kontrolle über die Nutzer im aRUT grundsätzlich gering ist, darf nicht unterschätzt werden, dass bei vielen Tests erst einige Datensätze aussortiert werden müssen, die nicht den Qualitätskriterien entsprechen. So ist ggf. eine Nachrekrutierung notwendig, für die weitere Kosten entstehen. Wenn der aRUT von einer Website aufrufbar ist und die Teilnehmer z. B. über eine Verlosung entschädigt werden, ist zwar mit höheren Abbruchquoten und insgesamt geringen Teilnehmerzahlen zu rechnen. Dafür kann ein solcher Test über einen sehr langen Zeitraum online bleiben, bis ausreichend gute Datensätze vorliegen. Die anderen Datensätze müssen dann aussortiert werden. Bei einem ausreichend mächtigen Tool und Probanden, die sich beim Test Mühe geben, können daher auch mit dem aRUT die zentralen Erkenntnisse generiert werden. Dass er dennoch deutlich hinter den beiden anderen Verfahren zurückbleibt, ist mit der Einbindung der Probanden in den Test zu erklären, da sie nur in vereinzelt angezeigten Textfeldern die Möglichkeit haben, sich frei zu äußern. Durch die Abstufungen der Skalen und die generelle Auswertung reiner Klickdaten bleibt der Bezug der Probanden zum Testobjekt insgesamt gering.

Der asynchrone Videotest, sofern er mit dieser Benennung treffend beschrieben werden kann, hat sich als respektables Verfahren herauskristallisiert. Bislang gab es keine öffentlich gemachte Einstufung dieses Verfahrens, die methodisch offengelegt war. Vermutlich wird es in Zukunft weitere Studien dieser Art geben, die den Videotest unter bestimmten Gesichtspunkten einstufen werden, etwa unter einem anderen Fokus oder unter Betrachtung bestimmter Untersuchungsobjekte. Im Rahmen dieser Untersuchung wurde der Videotest als relativ starkes Verfahren ermittelt. Die Nutzer sind in einem angemessenen Maß aktiv und geben hilfreiche Kommentare ab. Der Großteil der Usability-Probleme wurde von diesem Testverfahren identifiziert. Insbesondere bei kosmetischen Problemen konnte der Videotest herausstechen, was nur teilweise mit der erhöhten Probandenanzahl begründet werden kann. Der Erkenntnisgewinn wäre auch mit weniger Testpersonen gut gewesen. Weiterhin sticht der Videotest durch seine relativ unkomplizierte Durchführbarkeit heraus, was sich dadurch zeigt, dass bei den meisten Anbietern ausschließlich die Aufgabenstellungen, das ungefähre Abstecken der Zielgruppe und eine intakte Website notwendig sind, bevor der Test dann innerhalb kurzer Zeit durchgeführt werden kann. So kann er theoretisch auch nur von einer Person durchgeführt werden,

während beim Labortest i. d. R. mindestens zwei Personen auf Seiten des erhebenden Unternehmens zwingend notwendig sind. Die Kosten für das Verfahren liegen unter denen eines Labortests. Für die Auswertung muss relativ viel Zeit investiert werden, während dafür die Erhebung ohne Aufwand für den Auftraggeber erledigt wird. Die Stärke des Verfahrens liegt daher in der kurzfristigen Erhebung und dem geringen Personaleinsatz. Durch die aktive Einbindung der Probanden in den Test, versetzen sich diese stärker in die Anwendung als es beim aRUT der Fall ist. Einige kommentieren sehr viel, was zu einer hohen Problemanzahl führt, von denen sich einige dann eher im kosmetischen Segment bewegen. Dies sind aber dennoch wertvolle Erkenntnisse, die im Labortest aufgrund der starken Orientierung am Studienkonzept teilweise unterdrückt werden. Der Videotest gestattet so sehr offene Abläufe und die Gefahr eines Auslassens von Aufgabenteilen durch die Probanden ist nicht auszuschließen. Das Risiko ist aber geringer als im aRUT. Außerdem kann man das Handeln des Probanden durch die dauerhafte Kommentierung jederzeit nachvollziehen und so ggf. verstehen, weshalb ein bestimmter Arbeitsschritt ausgelassen wurde. Im aRUT kann man an den Stellen nicht sicher sein, ob die Teilaufgabe absichtlich ausgelassen oder versäumt wurde. Auch mit einer geringeren Testpersonenanzahl kann der Videotest zu den zentralen Erkenntnissen führen. So kann bereits das Beauftragen eines Tests mit beispielsweise nur drei Nutzern zu drei unterschiedlichen qualitativen Aussagen führen, die kurzfristig vorliegen. Bei den Aufwänden für die Durchführung eines Labortests fallen mehr Testpersonen kaum ins Gewicht, d. h. ein Labortest mit drei Probanden würde sich kaum lohnen, da die Vorbereitung zu aufwändig wäre. Beim Videotest kann problemlos mit wenigen Personen getestet werden. Insbesondere in dieser Konstellation wird deutlich, dass ein so gearteter Videotest auch als sinnvolle Ergänzung für eine andere Methode eingesetzt werden kann. Usability-Tests im Labor stehen häufig allein, aber ein Videotest kann bei geringer Stichprobe durchaus eine sinnvolle Ergänzung zu einem Expert Review (vgl. EBERHARD-YOM 2010, S. 127 f.) oder einer ausgiebigen Auswertung von Analytics-Daten sein (vgl. a. a. O., S. 140 f.) Das Potenzial des Videotests liegt also nicht nur in seiner alleinstehenden Durchführung, sondern offenbart aufgrund größtenteils an den Dienstleister ausgelagerter Aufwände seine Möglichkeiten zur Integration in ein bestehendes Projekt, ohne dass es dafür einer Einbindung einer Agentur bedarf. So ist der Videotest etwa auch für eine Webanwendung denkbar, die ihre Analytics-Daten (vgl. ebd.) ohnehin dauerhaft überwacht und auswertet. Sind dort dann etwa nach einer größeren Änderung der Site unerwartete Kaufabbrüche häufiger festzustellen, kann kurzfristig ein solcher Videotest in Auftrag gegeben werden, der im Gegensatz zu einem Usability-Test im Labor mit sehr viel weniger Vorlauf und insgesamt weniger Aufwand durchführbar ist (vgl. Kapitel 7.6). Wird dann noch für einen solchen kurzfristig anberaumten Zwischentest mit wenigen Probanden getestet, wird die mögliche Sonderrolle dieses Verfahrens deutlich, die sie neben ihrer Rolle als Alternative zum klassischen Usability-Test im Labor außerdem als Eilmethode spielen kann. Die Empfehlung von KRUG (2006, S. 135), im Laufe eines Entwicklungsprozesses oder des Gesamtbetriebs eher häufiger mit kleinen Stichproben zu testen, erhält durch diese Methode tatsächlich die praktikable und mit wenigen Umständen verbundene

Möglichkeit, die bei einem Labortest nicht in der Form gegeben ist, sondern in jedem Fall mit höheren Kosten und v. a. einer längeren Vorlaufzeit verbunden ist. Ist hier nicht ein Dienstleister kurzfristig verfügbar, wirkt sich dies auf die Projektdauer aus. Beim Videotest versprechen die Auftraggeber die Ergebnisse teilweise innerhalb von 24 Stunden (vgl. RAPIDUSERTESTS 2012) vorlegen zu können, sodass bei idealen Bedingungen ein iteratives Testen sowohl bei Live-Anwendungen mit kleinen Veränderungen als auch bei der Entwicklung möglich ist, sobald die Inhalte online zugänglich sind.

Zu vermuten ist, dass der Videotest in Zukunft an Bedeutung gewinnen wird. Aufgrund seiner geringen Kosten und kurzen Vorlaufzeiten ist er relativ problemlos in eine Produktentwicklung oder den Dauerbetrieb einzubinden. Insbesondere für das Abprüfen kleiner Veränderungen kann er kurzfristig anberaumt werden. Die Daten erscheinen deutlich lebhafter als solche aus den bisherigen asynchronen Verfahren, die eher quantitativen Daten entsprechen und damit als Rückschluss auf das tatsächliche Nutzererlebnis zu abstrakt erscheinen. Lediglich die un gelenkte Ausrichtung dieses Testverfahrens darf als definitiver Nachteil betrachtet werden, denn bei einer größeren Stichprobe steht der Erkenntnisgewinn nicht mehr im Verhältnis zu dem Auswertungsaufwand, der nur über die Echtzeitsichtung zu realisieren ist. Daher wird sich dieses Verfahren vermutlich eher als Verfahren für den spontanen Einsatz eignen, worauf auch die von den Betreibern vorgegebene maximale Testdauer von etwa 20 Minuten hindeutet. Ein vollwertiger Ersatz für den Labortest kann der Videotest daher nicht sein, denn er füllt eine Lücke auf, für die es bisher kein vergleichbares Produkt gab, das diese lebensechten Rückmeldungen zu einer Webanwendung zuließ.

Dennoch kann es sein, dass das Verfahren Einfluss auf den Marktanteil des klassischen Usability-Tests nehmen wird, wenn Firmen z. B. statt einem großen Usability-Test im Labor die Aufwände eher auf mehrere Videotests aufteilen oder trotz der vorhandenen Nachteile auf die Durchführung eines Tests im Labor verzichten.

Der Labortest wird nichtsdestotrotz seinen Stellenwert behalten, da er als Komplettpaket zunächst v. a. für Kunden ohne ausreichende Expertise im Usability-Bereich gut geeignet ist. Weiterhin ist er geeignet, wenn in einem Unternehmen keine Ressourcen für das eigenständige Durchführen von Tests vorgesehen sind. Dann kann durch das Beauftragen einer Agentur und einige Koordinationsgespräche der gesamte Aufwand an die Agentur abgegeben werden, sodass hierfür keine eigenen Mitarbeiter zur Verfügung stehen müssen, die dauerhaft angestellt werden müssen. Ein solches Komplettpaket analog zur Agenturarbeit mit Labortests bieten zwar auch einige Videotestanbieter. Dies wirkt sich jedoch wieder deutlich auf die Kosten aus und kann zur Folge haben, dass die Auftraggeber sich das Material gar nicht selbst anschauen, was sie im Labortest durch Teilnahme als Beobachter am Test mit wertvollen Eindrücken versorgt.

Auch aufgrund der möglichen Volumina der Videotests erfahren diese ihre Grenzen in den Einsatzmöglichkeiten. Ein umfangreicher Test im Usability-Labor kann mit realistischen

Aufmerksamkeitsspannen von bis zu 75 Minuten durchgeführt werden (vgl. PUSCHER 2011). Daher sind dort auch ganzheitliche Tests einer Webanwendung möglich, während in den asynchronen Verfahren der Fokus eher auf dem Abprüfen der Funktionstüchtigkeit von Teilaufgaben liegt. Für einen umfangreicheren Test müsste dann eher mit mehreren Testschleifen gearbeitet werden, was sich wiederum negativ auf die Projektdauer auswirkt. Daher kann bei einer umfangreicheren Prüfung der Labortest deutlich praktikabler sein: Er verlangt zwar einen hohen Aufwand, verhältnismäßig hohe Kosten und eine relativ lange Projektdauer. Dafür kann er als ganzheitliches Analyseverfahren zum Abprüfen großer Zusammenhänge eingesetzt werden und so z. B. vor dem Start eines neuen Shops der zielführendere Weg zur Behebung eines Großteils der Usability-Probleme dienen, bevor der Versuch unternommen wird, dies durch ein viel kleinschrittigeres Verfahren zu realisieren, was sich dann noch negativer auf die zeitlichen Vorgaben auswirkt und so den Start eines Shops unnötig heraus zögern könnte. Dann sollte die Wahl eher auf den Labortest fallen, der eine komplette Analyse ermöglicht und eine umfangreiche Problemliste zutage fördert, die bereits über umfangreiche Verbesserungsvorschläge verfügt und so praktisch direkt verwertet werden kann. Mit einer derartigen Liste kann dann sehr viel zielführender an der Site optimiert werden, anstatt dies auf mehrere Kleinprojekte aufzuteilen.

Auch erlaubt der Labortest den Test von noch nicht vollständig ausgestalteten Prototypen, die ggf. einer Erklärung während des Tests bedürfen. So unterstützt der Labortest auch die iterative Vorgehensweise bei der Entwicklung eines neuen Angebots. Der Testleiter kann hier nämlich Hilfestellung geben, sodass bereits einfache Wireframes mit dem Nutzer ausprobiert werden können, bevor das eigentliche Design oder alle Seitenelemente implementiert wurden. Ein Beheben von möglichen Problemen an dieser Stelle erspart daher auch einen großen Aufwand im Nachhinein.

Die spezifische Eignung der Verfahren darf also nicht verneint werden, wenngleich sie durch Inkaufnahme von Nachteilen bzw. den Eingang von Kompromissen alle zu verwertbaren Ergebnissen gelangen und bestimmte Vorgaben wie die zeitliche Komponente und das zur Verfügung stehende Budget schlichtweg als limitierende Faktoren nicht außen vor gelassen werden können. Daher wurden diese Verfahren auch unter Berücksichtigung möglichst viele Faktoren verglichen, um nicht nur deren jeweilige Ergiebigkeit, sondern auch deren Rahmenbedingungen, die vorher betrachtet werden müssen, zueinander ins Verhältnis zu setzen.

Die folgende Tabelle 6 stellt die wichtigsten Eigenschaften der untersuchten Methoden als abschließende Informationen gegenüber. Diese Auflistung ist nicht als erschöpfend zu betrachten, sondern will noch einmal auf sie zentralen Eigenschaften der Methoden eingehen, wie sie auf Seiten der Studienleitung wahrgenommen und als wichtig in Erinnerung behalten wurden.

Tabelle 6: Abschließende Gegenüberstellung von Vor- und Nachteilen der drei Methoden.

<i>Verfahren/Vorteile u. Nachteile</i>	Usability-Test im Labor	Asynchroner Videotest	Asynchroner Remote-Usability-Test
Vorteile	Ganzheitliche Analyse einer Website	Qualitative Äußerungen echter Nutzer (lautes Denken)	Bei großen Stichproben ist die Methode quantitativ und kann die gesamte Nutzerschaft abbilden.
	Testen von erklärungsbedürftigen Prototypen möglich	Relativ ergebnisoffene Gestaltung, da die Probanden sich selbst durch den Test führen	Ergebnisoffene Gestaltung, die Probanden kennen nur die Aufgabenstellung
	Auffinden einer relativ hohen Problemanzahl	In sehr kurzer Zeit durchführbar	Auch mit Site-Nutzern durchführbar
	Der Auftraggeber kann beobachtend am Test teilnehmen und so viele echte Eindrücke mit auf den Weg nehmen.	Es werden bei etwas größerer Stichprobe vergleichbar viele Usability-Probleme erkannt wie im Labortest.	Die Auswertung kann selektiv erfolgen und so problemlos nur für einen Nutzer oder nur für eine Teilaufgabe erfolgen.
	Eye-Tracking-Erhebung möglich	Geringere Kosten als im Labortest	Standardisierte Fragen (Skalen, Checkboxes)
Nachteile	Hoher Aufwand und hohe Kosten	Ergebnissichtung in Echtzeit erforderlich	Je nach Tool-Funktionalität werden nur wenige Probleme gefunden.
	Hoher Zeitaufwand	Keine Nachfrage bei den Nutzern möglich	Geringes Involvement der Nutzer
	Lohnt sich eher für umfangreichere Tests	Begrenzte Erhebungsdauer	Erst durch hohe Stichprobe werden Erkenntnisse generiert
Zusammenfassung	Der Labortest bietet eine umfassende Methode, die auch noch während eines Projekts umgestaltet werden kann, benötigt	Der Videotest ist relativ kurzfristig durchführbar und liefert gute Ergebnisse. Er eignet sich aufgrund der Testdauer am	Das Tool ermöglicht eine einfache Auswertung. Der Erfolg eines Tests ist entscheidend vom Funktionsumfang des

	jedoch auch deutlich mehr Aufwand als die verglichenen Methoden	ehesten für das Abprüfen von einzelnen Schritten.	Tools abhängig, das Involvement der Probanden ist gering.
--	---	---	---

Mit dieser Übersicht wurden so die wichtigsten Punkte noch einmal aufgegriffen. Hier zeigt sich außerdem, dass auch dem asynchronen Remote-Usability-Test nicht gänzlich der Stellenwert abgesprochen werden darf, da er zwar in der Studie kaum Probleme zutage förderte, dies aber weitestgehend auf das verwendete Tool zurückzuführen ist. Dementsprechend muss hier auch auf andere Untersuchungen verwiesen werden, in denen ein anderes Erhebungstool zur Anwendung kam, das mit einem erweiterten Funktionsumfang das Potenzial dieser Methode noch deutlich erhöht.

Der Videotest ist auch gemäß den Informationen in der Tabelle nach wie vor als starkes Verfahren zu sehen, da er kaum Schwächen aufweist bzw. diese durch den geringeren Aufwand ausgleicht.

Als Königsweg zeigt sich hier nochmals der Labortest, dessen Kosten und Aufwände an dieser Stelle nochmals deutlich herausgestellt worden sind. Ein Aspekt, der bisher keine Beachtung fand, in der Tabelle nun erstmals erwähnt wurde, ist die Möglichkeit, den Labortest mit der Erhebung von Eye-Tracking-Daten (vgl. EBERHARD-YOM 2010, S. 139 f.) zu kombinieren. Eine solche Erhebung war nicht Bestandteil dieser Studie, da diese Daten dann für sich alleine gestanden hätten und kein Vergleich zu den beiden anderen Methoden möglich gewesen wäre. Dennoch soll hier noch kurz auf das Potenzial hingewiesen werden: Ausschließlich mit der entsprechenden Technik, und zwar einem sog. „Eyetracker für die Aufzeichnung der Blickbewegungen“ (SARODNICK/BRAU 2011, S. 168) können die Daten der Nutzer zu ihrem Blickverhalten erhoben werden. Dies erfolgt in einem gesonderten Teil der Erhebung und wird anschließend über alle Nutzer ausgewertet, d. h. die Daten können aggregiert werden, um zu erfahren, wie alle Nutzer die Site wahrgenommen haben. Die so erhobenen Daten sind insofern wertvoll, als einfache Aussagen der Probanden über ihr Blickverhalten meist nicht korrekt getätigt werden, sondern diese ihr Blickverhalten nur annähernd korrekt beschreiben. Eine gesonderte Erhebung des Blickverhaltens ist daher nach aktuellem Stand der Technik noch Alleinstellungsmerkmal des Labortests, dessen Ergebnisse so nachprüfbar zeigen, ob bestimmte wichtige Seitenelemente nicht aufmerksamkeitsstark genug platziert sind und daher umgestaltet oder anders positioniert werden können. Führt man einen Test ohne Eye-Tracking durch, ist man auf die Aussagen der Testpersonen angewiesen, wenn diese etwa beschreiben, dass sie eine Schaltfläche nicht finden können oder wenn man diese dabei beobachtet, dass sie längere Zeit nach einem scheinbar gut platzierten Button suchen und diesen einfach nicht klicken. Die Eye-Tracking-Daten bieten dahingehend eine gute Ergänzung und müssen, sofern sie nicht erhoben werden (können), auf Umwegen durch Interpretation ersetzt werden, was nicht immer gelingt. Daneben stellen

sie für viele Kunden eine interessante, da direkt verständliche Visualisierung dar, wenn sie im Anschluss eine sog. Heatmap von einer ausgewählten Seite gezeigt bekommen.

Als finaler Abschluss der Arbeit wird der Versuch eines Ausblicks im kommenden Kapitel 10 unternommen, der weitere Forschungsansätze aufzeigt und eine Prognose über die Entwicklung des Usability-Testings abgibt.

10 Ausblick

Die weitere Entwicklung von Usability-Testmethoden muss in jedem Fall überwacht werden, da hier mit interessanten Entwicklungen zu rechnen ist. Denn der Stellenwert des Nutzertests wird in absehbarer Zeit nicht an Bedeutung verlieren, da er stets die wichtigste Komponente ist, um umfangreiche Aussagen der Nutzer zu einem Produkt zu erhalten.

Der Markt für Tools zur Durchführung von asynchronen Remote-Usability-Tests sollte in jedem Fall beobachtet werden. So kann hier ein neues Produkt mit einem erheblich erweiterten Funktionsumfang für sehr ergiebige Testergebnisse sorgen, die die Methode dann ggf. doch konkurrenzfähig machen können. Hier kommen auch die Firmen ins Spiel, die Webanalyse-Tools anbieten, die teilweise auch schon über erweiterte Funktionen verfügen und so u. a. auch die Mausbewegungen von echten Nutzern nachvollziehbar machen können (vgl. u. a. ETRACKER 2012). Daher könnte der asynchrone Remote-Usability-Test ggf. an Bedeutung verlieren oder muss durch neue Funktionalitäten aufgewertet werden. Für Online-Anbieter mit einem geringen Budget kann er auch weiterhin sehr sinnvoll sein, da diese so ergänzende Erkenntnisse generieren können, die sonst nicht aus den anderweitig erhobenen Daten hervorgehen.

Der Usability-Test im Labor wird vermutlich noch stärker auf seine Kernkompetenzen hin ausgerichtet werden, d. h. von vielen Anbietern deutlich als Premium-Produkt mit reichhaltiger und tiefgreifender Analyse eines Online-Angebots positioniert werden. Dadurch kann seine Stärke deutlich gemacht werden und es entsteht kein so starker Konkurrenzdruck zum Videotest, als wenn der Labortest ebenso als preisgünstige Methode dargestellt würde. Damit diese Strategie erfolgreich ist und bleibt, muss in jedem Fall das Verhalten der Kunden beobachtet werden. Ist beispielsweise ein unerwartet hoher Zuwachs für den Videotest mit negativen Konsequenzen für den Labortest zu verzeichnen, muss diese Strategie überdacht werden oder das Produkt des Labortests in verschiedenen Variationen angeboten werden. So kann beispielsweise nach wie vor das vollumfängliche Premiumpaket beibehalten werden, aber es sollte daneben auch eine weniger umfangreiche Konfiguration des Labortests positioniert werden, die mit deutlich geringerem Budget und Zeitaufwand eine kleine Anzahl an kurzen Tests beinhaltet. Eine solche Variante des Labortests könnte dann, kombiniert mit einem kurzen Ergebnisworkshop, ein direkt verwertbares Ergebnis für den Auftraggeber sein und diesem daher den Aufwand abnehmen, die der Videotest für ihn durch die notwendige Auswertung mitbringt.

Die Entwicklung des Videotests wird vermutlich auch voranschreiten. So werden die Anbieter mit zunehmender Marktdurchdringung ihre Produkte weiterentwickeln können. Auch andere Faktoren wie die Breitbandversorgung der Teilnehmer und die technische Ausstattung der Panelteilnehmer können dafür sorgen, dass ggf. in der Zukunft eine Aufnahme der Gestik und Mimik der Probanden

analog zum Labortest möglich gemacht wird. Dies ist insbesondere für die subjektiven Eindrücke der Beobachter eine sinnvolle Ergänzung.

Je länger die Anbieter der Videotests am Markt sind, desto stärker wächst deren Bekanntheit. So können sie durch eine deutlich steigende Anzahl an durchgeführten Tests mehr Expertise entwickeln und durch Rückmeldungen der Kunden Optimierungen am Produkt vornehmen. Dies kann Maßnahmen in der Bedienung des Tools für die Probanden betreffen, aber auch auf Seiten desjenigen, der das Online-Tool zur Auswertung bedient. Durch eine steigende Bekanntheit wächst auch das Panel mit den Testpersonen an, was sich positiv auf deren Qualität auswirken kann. So können schlecht bewertete Testpersonen aussortiert werden, ohne dass dadurch eine Lücke im Panel entsteht. Je größer das Panel ist, desto wahrscheinlicher ist es, eine spezifische Zielgruppe exakt abbilden zu können. Daneben unterstützt ein größeres Panel auch die kurzfristige Rekrutierung passender Teilnehmer, was ebenfalls die Erhebungsdauer deutlich reduzieren kann. Die Anbieter der Videotests werden daher schon aufgrund der naturgemäß zu erwartenden Entwicklung im Rahmen ihres stärkeren Bekanntwerdens ihre Qualität deutlich steigern können, da sie zusätzliche Erfahrung sammeln und ihre Panels vergrößern können. Im selben Zeitraum kann eine Usability-Agentur, die Labortests anbietet, allein aufgrund des notwendigen Personaleinsatzes keine vergleichbare Anzahl an Projekten betreuen und daher nicht so viele verschiedene Erfahrungen sammeln können.

Daher sind diese Agenturen entsprechend eher in der Positionierung und Kommunikation ihrer Angebote gefragt (s. o.), um die Alleinstellungsmerkmale und das Potenzial ihrer Methoden weiterhin deutlich zu machen und damit Aufträge generieren zu können.

Unabhängig von der Methode, die ein Unternehmen anbietet, wird mit zunehmendem Einsatz des Verfahrens die Erfahrung steigen, wodurch nicht nur die technische Weiterentwicklung vorangetrieben wird, sondern ebenso die methodische Weiterentwicklung. So wurde in den asynchronen Verfahren festgestellt, dass die Probanden bei der Visualisierung der Aufgaben mit einer Art Fortschrittsanzeige (z. B. Aufgabe 2 von 6) dazu neigen, die Aufgaben rasch abschließen zu wollen. Im Labortest wird hingegen mit einer ungefähren zeitlichen Vorgabe gearbeitet, die die Probanden erfahrungsgemäß nicht zu einer schnelleren Aufgabenerledigung animiert. Dadurch sind diese deutlich konzentrierter und erledigen die Aufgaben einzeln nacheinander, ohne bereits in Gedanken bei der nächsten Aufgabe zu sein. Eventuell werden hier die Betreiber von asynchronen Verfahren in Zukunft eher dazu tendieren, den Probanden als Einstieg ebenfalls eine solche ungefähre Zeitvorgabe mitzuteilen und dafür auf die Fortschrittsanzeige verzichten. Dies wäre eine relativ unkomplizierte Maßnahme, um die Konzentration der Probanden zu steigern und auf die einzelnen Teilaufgaben zu lenken und nicht so sehr auf die Abwicklung des gesamten Tests. Es bleibt abzuwarten, ob mit dieser oder ähnlichen Maßnahmen Einfluss auf das Involvement der Probanden genommen werden kann. Denn der Umstand, dass die Probanden im asynchronen Test eher schneller fertig werden wollen als im Labortest wurde im Verlauf der Studie offenbar. Dadurch, dass kein Eingriff während des Tests

möglich ist, sind hier entweder die Auftraggeber solcher Studien gefragt, ihre Aufgabenstellungen so zu formulieren, dass sie nicht nur zu einer schnellen Erledigung verleiten. Oder es kann sinnvoll sein, wenn die technischen Maßgaben dahingehend abgewandelt werden, dass die Probanden ihren aktuellen Fortschrittsstand gar nicht sehen können und so unvoreingenommen die Aufgaben nacheinander ausführen.

Für weitere Vergleichsstudien sollten mehrere konkrete Anbieter von Videotests im Hinblick auf Unterschiede in der Technik und der Probandenauswahl verglichen werden, da hier nur die beispielhafte Betrachtung eines Anbieters von Videotests erfolgen konnte.

Ebenso interessant wäre eine methodisch durchgeführte Analyse weiterer Serviceleistungen solcher Anbieter. So kann ein Kunde auch die Auswertung seiner Nutzervideos in Auftrag geben. Interessant wäre hier zu wissen, wie groß der Informationsgewinn durch eine expertenbasierte Videoauswertung im Vergleich zur Auswertung durch einen Kunden ist, der wahlweise unerfahren oder auch etwas erfahrener im Usability-Kontext sein kann. Vergleichend dazu könnte dies auch für einen Labortest erhoben werden, und zwar durch das Anfertigen eigener Notizen unabhängig von den das Projekt leitenden Usability-Consultants und anschließende Auswertung und Vergleich mit den dokumentierten Ergebnissen der Agentur. Dort käme dann aber wieder der Testleiter als weiterer Faktor mit ins Spiel, der einen deutlichen Einfluss auf den Testverlauf hat, während es im asynchronen Videotest nur die Testperson selbst gibt, die den Verlauf individuell macht.

Für die klassischen Usability-Agenturen kann der Videotest aber nicht nur als Mitbewerber betrachtet werden, sondern er darf auch in Zukunft als sinnvolle Teilmethode betrachtet werden. So kann etwa durch eine Kooperation zwischen einer Usability-Agentur und einem Anbieter von Videotests die Erschließung von Synergien erfolgen. Kunden können einerseits einander vermittelt werden. So kann ein Kunde, der nicht alle Fragestellungen im Videotest beantworten konnte, an eine Usability-Agentur für einen anschließenden Labortest mit verringertem Umfang vermittelt werden, um dort die verbliebenen Fragestellungen endgültig beantwortet zu bekommen. Für die klassische Usability-Agentur kann der Videotest v. a. als ergänzende Teilmethode verwendet werden, die die Erkenntnisse aus einer preisgünstigeren Methode ergänzen kann. Denkbar ist etwa, an eine Experten-Evaluation einer Webanwendung (vgl. SARODNICK/BRAU 2011, S. 144 ff.) einen Videotest mit einer sehr geringen Stichprobe anzuschließen. Dies können auch nur zwei Probanden sein. So wird die Ergiebigkeit eines solchen Kleinstprojekts deutlich gesteigert und die Erkenntnisse der Experten-Evaluation sind anhand der Erkenntnisse des Videotest bereits in der Praxis durch echte Nutzer nachgewiesen. Je nachdem wie die Kooperation zwischen den beiden Agenturen ausgestaltet ist, kann hier der Kunden weitervermittelt werden oder die Usability-Agentur bietet alle Lösungen aus einer Hand und beauftragt nur den Videotest selbst extern. Dann könnte die Agentur die Auswertung auch direkt mit übernehmen. Ein solches Modell wäre für den Kunden am praktikabelsten und würde für diesen auch ein abgeschlossenes Projekt darstellen. Dem Kunden nach einem Expert Review den

Videotest im geringen Umfang zu empfehlen, würde für diesen zusätzliche Umstände bedeuten. Daher sollte eher das direkte Anbieten dieses Verfahrens als zusätzliche Dienstleistung im Projekt dienen und keinen neuen Auftragnehmer einbinden. Ob dies dann im Zuge einer solchen Kooperation zwischen Usability-Agentur und Videotest-Anbieter erfolgt oder die Usability-Agentur einen eigenen Auftrag an den Videotest-Anbieter gibt und die Videos dann selbst auswertet, hängt von der jeweiligen Ausgestaltung ab. Je weniger Aufwand der Kunde mit der Methodenkombination hat, desto eher wird ein solches neues Angebot vermutlich akzeptiert. In jedem Fall ist es als kombinierte Methode empfehlenswert.

Es wurden mehrere Aspekte aufgezeigt, die die mögliche Entwicklung der betrachteten Verfahren skizzieren. Diese Entwicklung gilt es in naher Zukunft zu beobachten, um festzustellen, ob das relativ neue Verfahren des Videotests einen nennenswerten Einfluss auf die Arbeit von Usability-Dienstleistern auf dem deutschen Markt haben wird.

Mit dieser Arbeit wurden das Potenzial des Verfahrens und sein Stellenwert im Rahmen der etablierten Verfahren deutlich, sodass dessen Wert festgestellt werden konnte.

Trotz der allgemeingültigen Ausrichtung bleibt es bei dieser Studie bei einem Fallbeispiel, das einer Reihe an Faktoren unterlag, die die Ergebnisse beeinflusst haben. Es ist daher zu hoffen, dass in naher Zukunft weitere solcher Vergleiche durchgeführt werden, die sich z. B. bestimmten Fallbeispielen widmen. Dies können z. B. Zielgruppen aus einkommenshöheren Schichten sein. Sind diese genauso gut geeignete Teilnehmer für asynchrone Verfahren? Wie gut sind die einzelnen Verfahren in Bezug auf sehr spezielle Untersuchungsobjekte? Der Bezug dieser Arbeit lag definitiv auf dem Einsatzzweck der Evaluation von Onlineshops, da diese eine zentrale Rolle im Internet spielen. Wie steht es aber z. B. um die Bedienbarkeit von Corporate Website von Unternehmen? Wie gut kann etwa auch die Bedienung von Social-Media-Software o. ä. mit solchen Tests bewertet werden? Diverse unterschiedlich geartete Fallbeispiele dürften etwas andere Ergebnisse zutage fördern oder auch noch bestimmte Anwendungsfälle identifizieren, in den der Videotest sich etwa als vollständig ungeeignet zeigt oder in denen der Labortest sogar deutlich weniger Usability-Probleme zutage fördern kann.

Dies sind ausschließlich Vermutungen, weshalb dringend empfohlen wird, die Testverfahren weiteren Vergleichen zu unterziehen, die weitere Aspekte herausarbeiten können oder einen anderen Fokus setzen. Vor allem ist es hierbei wichtig, dass die zu erwartende Forschung ebenso öffentlich gemacht wird wie diese Studie. Häufig ist es der Fall, dass Forschungen, die in der Praxis durchgeführt werden, eher unter Verschluss gehalten werden und nur auszugsweise veröffentlicht werden, um die Ergebnisse mit einer bestimmten Auslegung darzustellen. Auch wenn diese plausibel dargelegt werden, ist doch nicht immer vollständig klar, ob hier nur die positiven Auszüge dokumentiert werden. Daher ist der öffentlichen, methodischen Forschung bei Methodenvergleichen

der Vorzug zu geben. Hier wird also nur der Auftakt gegeben für weitere, ergänzende Forschung, die sich weiteren interessanten Aspekten widmet und andere Schwerpunkte setzen wird.

11 Literatur

ANDREASEN ET AL. 2007

Andreasen, Morten Sieker ; Nielsen, Henrik Villemann ; Schröder, Simon Ormholt ; Stage, Jan: *What Happened to Remote Usability Testing? An Empirical Study of Three Methods*. San Jose (USA) : CHI Proceedings 2007, 28.04.-03.05.2007.

BOSENICK 2009

Bosenick, Tim: Remote User Experience Tests : Messen, befragen und beobachten in natürlicher Testumgebung. In: Reese, Frank (Hrsg.): *Website-Testing : Conversion Optimierung für Landing Pages und Online-Angebote*. Göttingen, BusinessVillage, 2009, S. 279-284.

BRUSH/AMES/DAVIS 2004

Brush, Alice Jane Bernheim ; Ames, Morgan ; Davis, Janet: *A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface*. Wien : CHI Conference on Human Factors in Computing Systems 2004, 24.-29.04.2004.

EBERHARD-YOM 2010

Eberhard-Yom, Miriam: *Usability als Erfolgsfaktor : Grundregeln, User Centered Design, Umsetzung*. Berlin : Cornelsen, 2010.

ECKERT 2008

Eckert, Ann: *Social Software und Usability: Ein empirischer Vergleich der Usability-Testmethoden Labortest, synchroner Remote Test und asynchroner Remote Test im Anwendungsbereich des Web 2.0*. Hildesheim, Universität, Fachbereich Informations- und Kommunikationswissenschaften, Masterarbeit, 2008. Abgerufen von: <http://www.bookrix.de/_PDF-dalarna_1250688191.4449980259.pdf>.

ETRACKER 2012

Etracker GmbH (Hrsg.): etracker Visitor Motion – Mouse Tracking zur Website-Optimierung und Steigerung der Konversionsraten [online]. In: *etracker.com* [Stand: 2012] [Abruf: 2012-08-22]. – URL: <<http://www.etracker.com/de/etracker-produkte-echtzeit-web-controlling-visitor-voice/etracker-visitor-motion.html>>.

GERTH/KRACH/PILOTTEK 2008

Gerth, Norbert ; Krach, Thomas ; Pilottek, Florian: Swordfish : Eine Software zur asynchronen Remote-Evaluation der Usability von Webseiten. In: Brau, Henning (Hrsg.) ; Diefenbach, Sarah (Hrsg.) ; Hassenzahl, Marc (Hrsg.) ; Koller, Franz (Hrsg.) ; Peissner, Matthias (Hrsg.) ; Röse, Kerstin (Hrsg.): *Usability Professionals 2008 : Berichtband des sechsten Workshops des German Chapter der Usability Professionals Association e. V.* Stuttgart : German UPA, 2008. S. 160-163.

GROSS 2010

Groß, Olaf: e-Commerce in Deutschland: Rund 400.000 Unternehmen sind aktiv [online]. In: *shopbetreiber-blog.de* [Stand: 2010-03-10] [Abruf: 2012-08-22]. – URL: <<http://www.shopbetreiber-blog.de/2010/03/10/e-commerce-in-deutschland-rund-400000-unternehmen-sind-aktiv/>>.

KODIYALAM 2003

Kodiyalam, Narayanan: *Remote Usability Evaluation Tool*. Blacksburg (USA), Virginia Polytechnic Institute and State University, Master Thesis, 2003. Abgerufen von <http://scholar.lib.vt.edu/theses/available/etd-05292003-180633/unrestricted/NKodiyalam_Thesis.pdf>.

KRUG 2006

Krug, Steve: *Don't Make Me Think! : Web Usability – Das intuitive Web*. 2. Aufl. Heidelberg : mitp.

LOOP11 2012

Loop 11 (Hrsg.): Clickstream analysis has arrived! [online]. In: *Loop 11 Blog* [Stand: 2012-01-31] [Abruf: 2012-03-14]. – URL: <<http://www.loop11.com/clickstream/>>.

MOHS/NAUMANN/KINDSMÜLLER 2007

Mohs, Carsten ; Naumann, Anja ; Kindsmüller, Martin Christof: Mensch-Technik-Interaktion: intuitiv, erwartungskonform oder vertraut? In: *MMI-Interaktiv* (2007), Nr. 13, S. 25-35.

MOLICH ET AL. 2004

Molich, Rolf ; Ede, Meghan R. ; Kaasgaard, Klaus ; Karyukin, Barbara: Comparative usability evaluation. In: *Behaviour & Information Technology* 23 (2004), Nr. 1, S. 65-74.

NIELSEN 2000

Nielsen, Jakob: Why You Only Need to Test with 5 Users [online]. In: *Jakob Nielsen's Alertbox* [Stand: 2000-03-19] [Abruf: 2012-03-08]. – URL: <<http://www.useit.com/alertbox/20000319.html>>.

PUSCHER 2009

Puscher, Frank: *Leitfaden Web-Usability : Strategien, Werkzeuge und Tipps für mehr Benutzerfreundlichkeit*. Heidelberg : dpunkt.verlag, 2009.

PUSCHER 2011

Puscher, Frank: Usability Remote Tests : Usability im Supermarkt. In: *Internet World BUSINESS* 14 (2011), S. 36.

RAPIDUSERTESTS 2012

Rapidusertests von Userlutions (Hrsg.): Warum verlassen Nutzer Ihre Website? [online]. In: *Web-Usability-Tests – Einfache Conversion-Optimierung* [Stand: 2012] [Abruf: 2012-03-13]. – URL: <<http://rapidusertests.com/static/usability-test-tour/was-sind-rapidusertests.html>>.

RICHTER/FLÜCKIGER 2010

Richter, Michael ; Flückiger, Markus: *Usability Engineering kompakt : Benutzbare Software gezielt entwickeln*. 2. Aufl. Heidelberg : Spektrum Akademischer Verlag, 2010.

SARODNICK/BRAU 2011

Sarodnick, Florian ; Brau, Henning: *Methoden der Usability Evaluation : Wissenschaftliche Grundlagen und praktische Anwendung*. 2. Aufl. Bern : Huber 2011 (Wirtschaftspsychologie in Anwendung).

SELVARAJ 2004

Selvaraj, Prakaash: *Comparative Study of Synchronous Remote and Traditional In-Lab Usability Evaluation Methods*. Blacksburg (USA), Virginia Polytechnic Institute and State University, Master Thesis, 2004. Abgerufen von <http://vtechworks.lib.vt.edu/bitstream/handle/10919/9939/Thesis_Prakaash_Selvaraj.pdf?sequence=2>.

SCHULZ 2008

Schulz, Sebastian: Die Usability von Webshops – Stand und Entwicklung der Messmethoden. In: *transfer Werbeforschung & Praxis* 4 (2008), S. 6-22.

STENFELDT 2012

Stenfeldt, Antje: RapidUserTest vs. Klassischer Usability-Test = Gleiche Ergebnisse [online]. In: *RapidUserTests Blog* [Stand: 2012-08-02] [Abruf: 2012-08-21]. – URL: <<http://rapidusertests.com/blog/2012/08/rapidusertest-vs-klassischer-usability-test-gleiche-ergebnisse/>>.

TULLIS ET AL. 2002

Tullis, Tom ; Fleischman, Stan ; McNulty, Michelle ; Cianchette, Carrie ; Bergel, Marguerite: *An Empirical Comparison of Lab and Remote Usability Testing of Web Sites*. Orlando : Usability Professionals Association Conference, Juli 2002.

UI-CHECK 2012

UI-Check UG (Hrsg.): Remote Usability Tests [online]. In: *Usability Tests in 24 Std. 39 € pro Test. Mit Video und Ton* [Stand: 2012] [Abruf: 2012-03-13]. – URL: <www.ui-check.com>.

WATERSON ET AL. 2002

Waterson, Sarah J. ; Hong, Jason I. ; Sohn, Tim ; Landay, James A. ; Heer, Jeffrey ; Matthews, Tara: *What Did They Do? Understanding Clickstreams with the WebQuilt Visualization System*. New York : AVI '02 Proceedings of the Working Conference on Advanced Visual Interfaces, 2002; S. 94-102.

Ich versichere, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen oder dem Sinn nach entlehnten Passagen sind durch Quellenangabe kenntlich gemacht.

Hamburg, den 24.08.2012

12 Anhang

Digitale Anhänge:

- Problemtabelle der Usabilitytests

Gedruckter Anhang:

- Studienkonzept Labortest (A-1)
- Studienkonzept des asynchronen Videotests (A-2)
- Studienkonzept des asynchronen Remote-Usability-Tests (A-3)

Studienkonzept Szenariobasierter Usability-Test

www.mein-schnulli.de (brand.ding)

*Projekt: „Usability-Test des Shops für einen
Methodenvergleich“*

1. Zeitplanung

- ✓ **KickOff:** 11.06.2012
- ✓ **Studienkonzept & Pretest:** 11.-12.06.2012, 20.06.2012 (Pretest)
- **Datenerhebung:** 20.-21.06.2012 (n=12, Testdauer 30 Min.)
- **Datenauswertung und Ergebnisaufbereitung:** 22.-29.06.2012
- **Problemworkshop:** 22.06.2012 (intern)
- **Präsentation/Ergebnisbesprechung:** 27.07.2012

2. Hinweise und Erläuterungen zum Studienkonzept

- Text in (*kursiv*): Dies sind die Aufgabenstellungen, Testcases bzw. Anweisungen an die Testperson, die aber keine Fragen darstellen.
- Text (normal): Dies sind Hinweise für den Interviewer.
- Text in (**fett**): Dies sind wichtige Themen/Fragestellungen, die im Rahmen der Interviews vom Interviewer berücksichtigt werden.
- TP = Testperson

3. Ablauf des Usability-Tests

- I. Organisatorisches
- II. WarmUp und Erfassung von Anforderungen & Erwartungen
- III. Freie Nutzung
- IV. Szenariobasierte Nutzung: Kauf von personalisierten Produkten
 - i. Test-Case 1: Einstieg Startseite
 - ii. Test-Case 2: Kaufempfehlungen
 - iii. Test-Case 3: Namensänderung
- V. Szenariobasierte Nutzung: Zweites Schnullerpaar
 - i. Test-Case 4: Kauf des gleichen Schnullers
 - ii. Test-Case 5: Passende Schnullerkette
- VI. Szenariobasierte Nutzung: Checkout
 - i. Test-Case 6: Checkout
- VII. Abschließende Bewertung

I. Organisatorisches

- Begrüßung/Angebot eines Getränks
- Bei jeder Testperson (kurz: TP) wird eine Einwilligung zur Aufzeichnung der Aussagen im Rahmen des Interviews per Videokamera und während der Denkprotokolle eingeholt.
- Folgende Hinweise werden vom Versuchsleiter (kurz: VL) gegeben:
 1. Den Testpersonen wird mitgeteilt, dass der Onlineshop und nicht sie als Person getestet werden sollen.
 2. Es wird außerdem darauf hingewiesen, dass die eResult GmbH (Göttingen) nicht an der Gestaltung und Entwicklung der Website beteiligt gewesen ist
(„Wir - die eResult GmbH - haben nichts mit der Entwicklung, Gestaltung und Programmierung der Anwendung zu tun. Sie können frei und unbefangen Ihre Meinung äußern.“).
Abschließend: Kurze Beschreibung des Testablaufs.

Aufzeichnung Morae starten.

II. WarmUp und Erfassung von Anforderungen & Erwartungen

Erkenntnisziele, Fragestellungen und Methoden

Erkenntnisziele:

- Warmup und Heranführung zum Thema, Sensibilisieren
- Was für eine Site wird hinter der URL www.mein-schnulli.de erwartet?
- Background/Erfahrungshorizont der TPen zu Babyartikeln, insbes. Schnullern, selbst darlegen lassen

Methoden:

- Halbstrukturiertes Interview

Aufruf des inhaltsleeren Desktops

„Es geht ja heute um Babyartikel. Wir wollen uns mit einer Website zu diesem Thema auseinandersetzen.“

1. **Kennen Sie den Onlineshop www.mein-schnulli.de? Was erwarten Sie dahinter?**
2. **Wir werden uns heute mit diesem Shop beschäftigen, in dem sie ausgewählte Babyartikel erwerben können.**
3. **Kennen Sie den Onlineshop www.schnuller.de? Was erwarten Sie dahinter? Welche Unterschiede erwarten Sie zu www.mein-schnulli.de?**
4. **Wie sind Ihre persönlichen Erfahrungen mit Babyartikeln?**

Ggf. nachfragen:

- a. **Kaufen Sie solche Artikel für Ihre eigenen Kinder oder für den Freundes- od. Verwandtenkreis?**
- b. **Wie sind Ihre bisherigen Erfahrungen beim Kauf von Babyartikeln im Internet?**

III. Freie Nutzung

Erkenntnisziele, Fragestellungen und Methoden

Erkenntnisziele:

- Wie gehen die Nutzer in der freien Nutzung mit dem Shop mein-schnulli.de um?
- Wie kommen die Nutzer mit den Inhalten und Funktionen zurecht?
- Sind die Inhalte und Funktionen intuitiv verständlich?
- Wie ist der erste Eindruck?

Methoden:

- Freie Nutzung des Shops
- Protokolle lauten Denkens (Thinking-Aloud-Methode)
- Beobachtung & Aufzeichnung des Klickverhaltens
- Aufzeichnung der Gestik & Mimik
- Retrospektives Interview

Anweisung zum „Lauten Denken“ nach/während jeder Aufgabenstellung bzw. Nutzungsphase:

„Während Sie den Shop nutzen, möchte ich Sie bitten, Ihre Gedanken und Empfindungen laut auszusprechen. Sprechen Sie einfach aus, was Ihnen so durch den Kopf geht, was Sie stört, welche Dinge Ihnen gefallen, welche Erwartungen Sie haben, ob diese erfüllt werden und wie Sie die Inhalte und Funktionen verstehen usw.“

Aufruf der Website www.mein-schnulli.de

„Sie haben jetzt die Möglichkeit, den aktuellen Shop von mein-schnulli.de zu sehen.

Bitte nutzen Sie zunächst einige Minuten die Website. Schauen Sie sich dabei alles an, was Sie interessiert.“

Aufruf des inhaltsleeren Desktops

Retrospektives Interview:

5. Was sind Ihre spontanen Eindrücke nach der Nutzung des Shops (Ersteindruck)?

Ggf. nachfragen:

- a. Was gefällt Ihnen gut?
- b. Was gefällt Ihnen weniger gut?
- c. Gibt es etwas, das Ihnen fehlt/was Sie vermissen?

IV. Szenariobasierte Nutzung: Kauf von personalisierten Produkten

Erkenntnisziele, Fragestellungen und Methoden

Erkenntnisziele:

- Wie kommen die Nutzer mit dem Shop zurecht?
- Wie komfortabel ist der Erwerb mehrerer Produkte umgesetzt?
- Wie zufriedenstellend sind die Individualisierungsmöglichkeiten umgesetzt?
- Wie kommen die Nutzer mit Änderungen zurecht?

Methoden:

- Szenariobasierte Nutzung der Klickpfade
- Protokolle lauten Denken (Thinking-Aloud-Methode)
- Beobachtung & Aufzeichnung des Klickverhaltens
- Aufzeichnung der Gestik & Mimik
- Retrospektives Interview

Wiederholte Anweisung zum Lauten Denken.

Anweisung, vor jedem Klick zu beschreiben,

- 1.) Warum ein Link gewählt wird und
- 2.) Welche Inhalte / Seiten / Funktionen erwartet werden (bei Klick)

Test-Case 1: Einstieg Startseite

„Eine gute Freundin von Ihnen hat gerade eineiige Zwillinge bekommen. Sie möchten ihr von NUK Schnuller mit den Namen der Kinder schenken. Da es eineiige Zwillinge sind, sollen die Jungs auch die gleichen Schnuller erhalten.

Bitte schauen Sie zunächst nach dem Schnuller für ein Kind. Wie gehen Sie vor?“

TP sucht sich einen Schnuller aus.

„Sie möchten die Schnuller ja gerne mit dem Namen des Kindes bedrucken lassen. Der Name des ersten Kindes lautet Philipp.“

TP legt Schnuller in den Warenkorb.

TP auf Warenkorb-Zwischenseite lassen

Retrospektives Interview:

1. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

- a. **Sind die Produktinformationen für Sie ausreichend, um eine Kaufentscheidung für einen bestimmten Schnuller zu treffen?**

- b. Was hat Ihnen besonders gut gefallen?
 - c. Was hat Ihnen nicht gefallen?
2. Wie finden Sie die Umsetzung der Individualisierungsmöglichkeit?

Test-Case 2: Kaufempfehlungen

auf Warenkorb-Zwischenseite Test fortsetzen

„Bitte schauen Sie sich die Warenkorb-Zwischenseite an. Gibt es hier eventuell ein Produkt, das Sie interessieren könnte?“

(TP ein weiteres Produkt, idealerweise auch mit Individualisierungsmöglichkeit, kaufen lassen)

3. Wie hat Ihnen die Zwischenseite (Warenkorb-Layer) gefallen?
4. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

- d. Was hat Ihnen besonders gut gefallen?
- e. Was hat Ihnen nicht gefallen?

Test-Case 3: Namensänderung

TP vom Warenkorb auf Artikel zugreifen lassen

„Ihnen fällt auf, dass Sie sich beim Namen für den Schnuller vertippt haben. Der Name des Jungen wird nur mit einem „p“ am Ende, also „Philip“, geschrieben. Bitte korrigieren Sie Ihre Eingabe.“

5. Wie hat Ihnen die Änderung dieser Angabe gefallen?
6. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

- f. War Ihnen klar, dass Sie in jedem Fall auf „Aktualisieren“ klicken müssen, um Ihre Eingabe zu speichern?
- g. Was hat Ihnen besonders gut gefallen?
- h. Was hat Ihnen nicht gefallen?

V. Szenariobasierte Nutzung: Zweites Schnullerpaar

Erkenntnisziele, Fragestellungen und Methoden
Erkenntnisziele: • Wie kommen die Nutzer mit dem Kauf des gleichen Produkts, aber mit anderem Namen zurecht? • Wie gefällt der Kauf von individualisierten Zusatzprodukten? Methoden: • Szenariobasierte Nutzung der Klickpfade • Protokolle lauten Denken (Thinking-Aloud-Methode) • Beobachtung & Aufzeichnung des Klickverhaltens • Aufzeichnung der Gestik & Mimik • Retrospektives Interview

Wiederholte Anweisung zum Lauten Denken.

Anweisung, vor jedem Klick zu beschreiben,

1.) Warum ein Link gewählt wird und

2.) Welche Inhalte / Seiten / Funktionen erwartet werden (bei Klick)

Test-Case 4: Kauf des gleichen Schnullers

„Sie möchten für den Zwillingbruder das gleiche Schnullerpaar verschenken. Sein Name lautet Tom.“

TP von Startseite zum Artikel gehen lassen
--

Retrospektives Interview

7. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

a. Was hat Ihnen besonders gut gefallen?

b. Was hat Ihnen nicht gefallen?

8. War Ihnen zu jeder Zeit klar, dass Sie an einem anderen Artikel (da mit neuem Namen) arbeiten (und dass zwei Paar Schnuller erworben werden müssen, um Schnuller für beide Kinder mit dem jeweiligen Namen zu erhalten)?

Test-Case 5: Passende Schnullerkette

„Tom soll außerdem noch eine Schnullerkette mit seinem Namen dazu bekommen.“

Fortsetzen aus dem Warenkorb heraus

Retrospektives Interview

1. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

- a. Was hat Ihnen besonders gut gefallen?
- b. Was hat Ihnen nicht gefallen?

VI. Szenariobasierte Nutzung: Checkout

Erkenntnisziele, Fragestellungen und Methoden
Erkenntnisziele: • Wie kommen die Nutzer mit dem Checkout-Prozess zurecht? • Ist die Änderung der Versandadresse für die Nutzer klar? • Wie finden die Nutzer die Auswahl der Zahlungsarten? Methoden: • Szenariobasierte Nutzung der Klickpfade • Protokolle lauten Denken (Thinking-Aloud-Methode) • Beobachtung & Aufzeichnung des Klickverhaltens • Aufzeichnung der Gestik & Mimik • Retrospektives Interview

Wiederholte Anweisung zum Lauten Denken.

Anweisung, vor jedem Klick zu beschreiben,

- 1.) Warum ein Link gewählt wird und
- 2.) Welche Inhalte/Seiten/Funktionen erwartet werden (bei Klick)

Test-Case 6: Checkout

„Sie möchten Ihre Bestellung nun abschicken. Wie gehen Sie vor?“

Fortsetzen aus dem Warenkorb heraus

Ohne Registrierung bestellen lassen; Musterdaten: Herr Michael Mustermann, geb. am 01.01.1970, E-Mail: mustermann@test.de, Musterstraße 1, 12345 Musterstadt, Tel.: 0123456; Zahlungsart Vorkasse (*vorher andere Zahlungsarten betrachten lassen*)

„Sie möchten die Bestellung an einen anderen Empfänger senden lassen, die Rechnung möchten aber nach wie vor Sie erhalten. Wie gehen Sie vor?“

Empfänger der Sendung: Frau Michaela Musterfrau, Musterallee 12, 12345 Musterstadt.

Retrospektives Interview

9. Wie hat Ihnen dieser Schritt gefallen?

Ggf. nachfragen:

- a. Was hat Ihnen besonders gut gefallen?
- b. Was hat Ihnen nicht gefallen?

10. Wie empfinden Sie die Zusatzgebühren der jeweiligen Zahlungsmethoden?

11. War Ihnen auf Anhieb klar, wie Sie nur die Empfängeradresse und nicht die Rechnungsadresse ändern?

„Sie möchten Ihre Bestellung jetzt noch abschließen. Bitte fahren Sie fort.“

Zahlungsart Vorkasse wählen lassen, Bestellung im letzten Schritt nicht absenden lassen, sondern abbrechen.

Retrospektives Interview

1. Wie hat Ihnen der Bezahlprozess (Checkout) insgesamt gefallen?

Ggf. nachfragen:

- a. Was hat Ihnen besonders gut gefallen?
 - b. Was hat Ihnen nicht gefallen?
- ### 2. Was sollte man hier Ihrer Meinung nach verbessern?

Aufruf der Startseite mein-schnulli.de
--

VII. Abschließende Bewertung

Erkenntnisziele, Fragestellungen und Methoden
Erkenntnisziele: • Wie bewertet der Nutzer den Shop insgesamt? • Was sollte in jedem Fall verbessert werden? • Wo haben die Nutzer Probleme? Methoden: • Halbstrukturiertes Interview

„Wir sind nun am Ende des Tests angekommen. Sie haben den Shop mein-schnulli.de nun ausgiebig kennen gelernt.“

2. **Welchen Gesamteindruck hat der Shop bei Ihnen hinterlassen?**
3. **Wurden Ihre Erwartungen an einen Shop für Babyartikel erfüllt?**
4. **Was hat Ihnen besonders gut gefallen?**
5. *Wie hat Ihnen der Kauf gleicher Artikel mit unterschiedlichen Namen gefallen?*
6. **Worin sehen Sie noch Probleme? Was sollte unbedingt noch verbessert werden?**

Ggf. nachfragen:

- a. **Haben Sie irgendwelche Services oder Funktionen vermisst?**

Danksagung, Feedback zum Test, Übergabe der Incentivierung

- Browsercache leeren

A-2 Studienkonzept asynchroner Videotest

Studienkonzept Videotest rapidusertests.com

Eckdaten:

- Teilnehmerzahl: n=20
- Erhebung: KW 26
- Testdauer je Test: 20 Minuten

Ablauf des Tests

1. Auftakt: Bevor Seite aufgerufen wird (*Start auf weißer Seite*):

Anweisung: „Heute geht es um Shops für Babyartikel: Was erwarten Sie hinter der Internet-Adresse www.mein-schnulli.de für einen Shop? Was erwarten Sie im Gegenzug hinter der Internet-Adresse www.schnuller.de für einen Shop?“

Szenario: „Stellen Sie sich vor, eine gute Freundin von Ihnen hat gerade eineiige Zwillinge bekommen. Sie möchten ihr Schnuller mit den Namen der Kinder schenken.“

2. Seite aufrufen

„Bitte rufen Sie nun die Site www.mein-schnulli.de auf. Schauen Sie sich die Seite kurz an und schildern Sie Ihren ersten Eindruck.

Sie möchten nun für die Zwillinge Ihrer Freundin Schnuller der Marke NUK mit den Namen der Kinder schenken. Da es eineiige Zwillinge sind, sollen die Jungs auch die gleichen Schnuller erhalten. Bitte schauen Sie zunächst nach dem Schnuller für das erste Kind. Der Schnuller soll mit dem Namen „Philipp“ bedruckt werden.“

3. „Bitte fügen Sie das gewählte Produkt Ihrem Warenkorb hinzu. Schauen Sie sich hierbei auch die Bestätigungsseite an, die angezeigt wird, wenn Sie das Produkt in den Warenkorb gelegt haben. Legen Sie von dort aus gerne eines der Ihnen auf dieser Seite empfohlenen Produkte zusätzlich in den Warenkorb.“

4. Namensänderung

„Ihnen fällt auf, dass Sie sich beim Namen des Kindes vertippt haben, dieses schreibt sich am Ende mit nur einem *p*. Bitte ändern Sie den Schnulleraufdruck in *Philip*.“

5. Zweites Schnullerpaar

„Sie möchten nun für den Zwillingenbruder genau das gleiche Schnullerpaar verschenken. Sein Name lautet *Tom*. Außerdem soll Tom noch eine Schnullerkette mit seinem Namen dazu bekommen. Bitte wählen Sie diese Artikel aus und legen Sie sie in den Warenkorb.“

6. Zur Kasse

„Im Anschluss können Sie zur Kasse gehen und den Kaufprozess ohne Registrierung durchgehen. Die Bestellung läuft auf den Namen Michael Mustermann (geboren am

01.01.1970, E-Mail: mustermann@mein-schnulli.de, wohnhaft Musterstraße 1, 12345 Musterstadt, Tel.: 0123456). Die Lieferung soll allerdings bitte an Frau Michaela Musterfrau (wohnhaft Musterallee 12, 12345 Musterstadt) geliefert werden.

Schauen Sie sich dann bitte die Zahlungsarten an. Welche würden Sie wählen? Wählen Sie schließlich Vorkasse aus und gehen Sie zum nächsten Schritt. Schicken Sie die Bestellung dann bitte nicht ab. Sie haben das Ende des Tests erreicht.“

Nachbefragung:

- Welchen Gesamteindruck hat der Shop bei Ihnen hinterlassen?
- Wurden Ihre Erwartungen an einen Shop für Babyartikel erfüllt?
- Was hat Ihnen besonders gut/schlecht gefallen?
- Wie fanden Sie den Kauf des gleichen Produkts, aber mit unterschiedlichen Namen?
- Was sollte unbedingt noch verbessert werden? Haben Sie irgendwelche Services und Funktionen vermisst?

A-3 Studienkonzept asynchroner Remote-Usability-Test

Studienkonzept asynchroner Remote-Usability-Test Loop11

Eckdaten:

- Teilnehmerzahl: n=40
- Erhebung: KW 26
- Testdauer je Test: 20 Minuten

Ablauf des Tests

1. „Vielen Dank, dass Sie sich die Zeit nehmen, uns bei unserer Studie zu unterstützen. Es wird heute um Babyartikel gehen. Sie werden den Kauf von Artikeln in einem Shop ausprobieren. Im Anschluss an den Test werden Ihnen abschließende Fragen gestellt. Auf der folgenden Seite erhalten Sie Informationen zum Ablauf des Tests, der in einzelne Teilaufgaben unterteilt ist.“
2. Erklärung der Bedienung des Testtools
3. Bevor Seite aufgerufen wird: Fragefeld „*Open ended (Comments box)*“
 - „Was erwarten Sie hinter der Adresse www.mein-schnulli.de für einen Shop? Was erwarten Sie im Gegenzug hinter der Adresse www.schnuller.de für einen Shop? Bitte antworten Sie einfach spontan, ohne sich die Seiten vorher anzuschauen. Falls Sie eine oder beide Seiten bereits kennen, können Sie das hier auch gerne erwähnen.“
 - „Bitte erläutern Sie kurz Ihre persönlichen Erfahrungen mit Babyartikeln. Kaufen Sie solche Artikel für Ihre eigenen Kinder oder als Geschenke für den Freundes- und Verwandtenkreis? Wie sind Ihre Erfahrungen beim Kauf von Babyartikeln im Internet?“
4. Seite aufrufen (Startseite www.mein-schnulli.de)

Bitte rufen Sie nun die Site www.mein-schnulli.de auf. Schauen Sie sich die Seite kurz an. Dann gehen Sie bitte nach folgender Aufgabenstellung vor:

Aufgabenstellung: „Eine gute Freundin von Ihnen hat gerade eineiige Zwillinge bekommen. Sie möchten ihr von NUK Schnuller mit den Namen der Kinder schenken. Da es eineiige Zwillinge sind, sollen die Jungs auch die gleichen Schnuller erhalten. Bitte schauen Sie zunächst nach dem Schnuller für das erste Kind. Der Schnuller soll mit dem Namen „Philipp“ bedruckt werden. Bitte fügen Sie das gewünschte Produkt Ihrem Warenkorb hinzu. Schauen Sie sich hierbei auch die Bestätigungsseite an, die angezeigt wird, wenn Sie das Produkt in den Warenkorb gelegt haben. Von dort können Sie auch gerne ein weiteres Produkt zu Ihrem Einkauf hinzu wählen.“
5. Wie hat Ihnen dieser Schritt gefallen?
6. Gibt es noch etwas, das Sie zu diesem Schritt loswerden möchten (optional)?
7. Namensänderung

Ihnen fällt auf, dass Sie sich beim Namen des Kindes vertippt haben, dieses schreibt sich am Ende mit nur einem „p“. Bitte ändern Sie den Schnulleraufdruck in „Philip“.

8. Wie hat Ihnen dieser Schritt gefallen?

9. Gibt es noch etwas, das Sie zu diesem Schritt loswerden möchten (optional)?

10. Zweites Schnullerpaar (ausgehend von der Startseite)

„Sie möchten nun für den Zwilling Bruder genau das gleiche Schnullerpaar verschenken. Sein Name lautet Tom. Außerdem soll Tom noch eine Schnullerkette mit seinem Namen dazu bekommen.“

11. Wie hat Ihnen dieser Schritt gefallen?

12. War Ihnen zu jeder Zeit klar, dass Sie an einem anderen Artikel arbeiten als mit dem, der bereits im Warenkorb liegt?

13. Zur Kasse

„Im Anschluss können Sie zur Kasse gehen und ohne Registrierung den Kaufprozess durchgehen. Die Bestellung läuft auf den Namen Michael Mustermann (geboren am 01.01.1970, E-Mail: mustermann@test.de, wohnhaft Musterstraße 1, 12345 Musterstadt, Tel.: 0123456). Die Lieferung soll allerdings an Frau Michaela Musterfrau (wohnhaft Musterallee 12, 12345 Musterstadt) geliefert werden. Schauen Sie sich die Zahlungsarten an. Welche würden Sie wählen? Wählen Sie dann Vorkasse aus und gehen Sie zum nächsten Schritt. Schicken Sie die Bestellung dann bitte nicht ab. Sie haben das Ende des Tests erreicht.“

14. Wie finden Sie die Preisaufschläge bei einigen der angebotenen Zahlungsarten?

15. Welche Zahlungsart(en) bevorzugen Sie normalerweise (mehrere Optionen wählbar)?

Nachbefragung:

- Welchen Gesamteindruck hat der Shop bei Ihnen hinterlassen?
- Wurden Ihre Erwartungen an einen Shop für Babyartikel erfüllt?
- Was hat Ihnen besonders gut/schlecht gefallen?
- Wie fanden Sie den Kauf des gleichen Produkts, aber mit unterschiedlichen Namen?
- Was sollte unbedingt noch verbessert werden? Haben Sie irgendwelche Services und Funktionen vermisst?