

Die Usability von Webshops – Stand und Entwicklung der Messmethoden*



Dr. Sebastian Schulz

Assistent des Vorstandes bei
der Hannoversche
Lebensversicherung AG

✉ sebastian.schulz@hannoversche-leben.de

* Der Autor dankt Herrn Prof. Dr. Günter Silberer für die Betreuung seiner Dissertation, die die Grundlage dieses Artikels darstellt, und für die Unterstützung während der gemeinsamen Zeit am Institut für Marketing und Handel.

Die Usability von Websites stellt unabhängig vom Web 1.0 oder Web 2.0 immer noch einen entscheidenden Faktor des Online-Erfolges von Anbietern dar. Trotz unzähliger Anforderungskataloge und Erfolgsfaktoren zeigen Studien, dass nicht auf eine empirische Usability-Evaluation verzichtet werden kann. Unternehmen stehen dafür mittlerweile zahlreiche Methoden zur Verfügung, deren Leistungsfähigkeit aber auch für Experten nur schwer zu bewerten ist. Dieser Beitrag stellt einen Klassifikationsansatz für Usability-Evaluationsmethoden vor und gibt einen Überblick über die verschiedenen Evaluationsverfahren. Anschließend werden prozessorientierte Verfahren der Usability-Evaluation im Detail vorgestellt sowie ein empirischer Methodenvergleich anhand eines systematisch variierten Webshops durchgeführt. Die Ergebnisse geben Aufschlüsse über die Leistungsfähigkeit der heuristischen Evaluation, der Blickregistrierung und verbaler Protokolle. Der Beitrag schließt mit Implikationen für Forschung und Praxis.

Schlagworte: > Usability > Usertracking > verbale Protokolle > Blickregistrierung > heuristische Evaluation

1. Einleitung

Glaubt man zahlreichen Veröffentlichungen aus Wissenschaft und Praxis hat das Internet den nächsten Evolutions-schritt vollzogen. Speziell unter dem Schlagwort „Web 2.0“ fallen der Fokus der Online-Forschung und das Interesse der Praxis in letzter Zeit zunehmend auf den Content und die Content-Erstellung durch den User (z. B. Killian et al. 2007). Dabei ist zu beobachten, dass grundlegende Anforderungen an eine Website wie die Utility (Nützlichkeit) und Usability (Benutzerfreundlichkeit) in den Hintergrund treten und bei der Konzepterstellung nur eine untergeordnete Rolle spielen. Sowohl bei Web 2.0 Angeboten als auch bei klassischen Websites wird sich oftmals nur an grundlegenden Prinzipien des Website-Designs orientiert, eine empirische Analyse der Usability oder systematische Integration des Usability-Engineering in die Entwicklung findet zu Unrecht nur selten statt (Schulz 2008). Im täglichen Betrieb der Website begnügen sich Unternehmen und Online-Marketing-Manager anschließend mit serverseitigen Log-File-Analysen oder dem Einsatz von Online-Analysetools

wie z. B. Google Analytics. Diese Tools sind jedoch nur bedingt für die Überwachung und Optimierung der Website aus Usability-Sicht geeignet.

Dabei ist erwiesen: Websites die nicht usable sind verlässt der Nutzer und kommt meistens nicht zurück (Nielsen/Loranger 2006). Möchte ein Konsument einen Einkauf im Internet tätigen, wird er mit den Designelementen und der damit verbundenen Usability konfrontiert und muss damit interagieren (Nielsen 2000, S. 10). Ist der Webshop nicht usable, bricht der Konsument den Kaufprozess ab und generiert keinen Umsatz. In einer Zeit in der ca. 60 % der Europäer einen online Zugriff besitzen, mehr als 31 Stunden im Monat im Netz verbringen und dabei im Schnitt 1500 Seitenaufrufe aufweisen, stellt die Usability den Schlüssel zu einem erfolgreichen Online-Geschäft dar (Waiguny et al. 2008, S. 2). Daher sollte auch in der Zeit des Web 2.0 ein effektives und effizientes Usability-Engineering an erster Stelle stehen. Dafür stehen Unternehmen mittlerweile zahlreiche Methoden zur Verfügung, deren Leistungsfähigkeit und Erkenntnisgewinn auch für Experten nur schwer zu

bewerten sind. Dieser Beitrag soll daher einen Überblick über existierende Meßmethoden geben und damit einen gezielten Einsatz ermöglichen. Anschließend werden ausgewählte Tracking-Methoden vorgestellt und Forschungsfragen für einen Methodenvergleich prozessorientierter Verfahren definiert. Diese Forschungsfragen werden anschließend im Rahmen einer empirischen Studie beantwortet. Der Beitrag schließt mit Implikationen für Forschung und Praxis.

2. Stand der Usabilityforschung und Messmethoden

2.1 Grundlagen und Definition von Usability

Usability wird oftmals als Sammelbegriff für die Benutzerfreundlichkeit bzw. Nutzbarkeit einer Website oder als die Gebrauchstauglichkeit von einem Internetauftritt oder Programm definiert. Wie verschiedene Autoren feststellen, sind diese Definitionen hinreichend für ein grundlegendes Verständnis, aber den eigentlichen Kern von Usability treffen sie nicht (Sarodnick/Brau 2006). Usability fokussiert nicht nur auf eine komfortable Nutzung eines Systems (wie „user friendly“), sondern beinhaltet auch die Forderung nach einer geeigneten Unterstützung des Nutzers durch das Interface (Sarodnick/Brau 2006, S. 17; Dzida 1994, S. 386). Unstrittig ist, dass Usability ein Qualitätsmerkmal für Interfaces darstellt. Es ist als Konzept aufzufassen, das selbst noch kein konkretes Evaluationskriterium darstellt, sondern einen Rahmen für die Ableitung von Kriterien bietet (Gediga/Hamborg 2002, S. 3). Dabei sind Definitionen so vielfältig wie Veröffentlichungen zur Usability. Neben den Unterschieden im Detail wird Usability von den meisten Autoren als mehrdimensionales Konstrukt (Effektivität, Effizienz, Zufriedenheit) aufgefasst (z. B. Nielsen 1993; Gray/Salzman 1998). Dabei hat sich in den letzten Jahren sowohl in der wissenschaftlichen Literatur als auch in der Praxis das Verständnis der DIN ISO-NORM 9241 durchgesetzt (Oppermann/Reiterer 1997; Konradt et al. 2003; Sarodnick/Brau 2006; Schulz et al. 2006).

Usability wird demzufolge definiert als Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen (Oppermann/Reiterer 1997; DIN ISO-NORM 9241). Unter Effektivität wird die grundlegende Zielerreichung (Wirksamkeit, Output) verstanden. Im WWW kann sie als der Präzisionsgrad und die Vollständigkeit, mit der der User sein individuelles Ziel in der Interaktion mit der Website erreichen kann, aufgefasst werden. Die Effizienz ist ein

Maß für die Wirtschaftlichkeit (Kosten-Nutzen-Relation). Im WWW stellt sie den notwendigen Aufwand in Relation zur individuellen Zielerreichung dar. Die Zufriedenheit drückt hingegen eine positive Einstellung gegenüber dem Interface aus (Oppermann/Reiterer 1997, S. 233).

Haben sich die Messung und Berücksichtigung der kognitiven Komponente von Zufriedenheit bzw. einer positiven Einstellung bereits etabliert, spielt die emotional-affektive Komponente bisher nur eine untergeordnete Rolle (Yom 2003). Dieses in der Literatur oftmals als Joy of Use thematisierte Konstrukt sollte in der Zukunft auch bei der Usabilitymessung mehr Beachtung finden. Abschließend ist zu beachten, dass Usability immer vor einem Kontext (Nutzer, Situation etc.) sowie einer konkreten Aufgabe und den damit verbundenen Zielen betrachtet und untersucht werden muss (Schweibenz/Thissen 2003). Dies macht eine weitere Konkretisierung der Definition bzw. der Usability in Abhängigkeit der Untersuchungssituation und der Nutzer nötig.

2.2 Messmethoden im Überblick

Der Wissenschaft und Praxis stehen zahlreiche Methoden der Usability-Evaluation zur Verfügung, die in den verschiedenen Phasen eines Usability-Engineering zum Einsatz kommen können. In der wissenschaftlichen Literatur finden sich vor allem drei Ebenen, auf denen Usability-Evaluationsmethoden differenziert bzw. klassifiziert werden können: Dem Evaluationsansatz, Evaluationszeitpunkt und Evaluationstyp (Sweeney et al. 1993, S. 694; Yom 2003, S. 107). Bei der Differenzierung auf der Ebene des Evaluationstyps wird das Kriterium der Evaluationszielsetzung zur Differenzierung von Methoden herangezogen. Danach wird unterschieden in Methoden der summativen Evaluation – Ziel ist es, ein Interface anhand eines abschließenden Wertes (Overall Score) zu bewerten – und der formativen Evaluation – Ziel ist es, konkrete Verbesserungsvorschläge aus dem Nutzungsprozess abzuleiten (Christophersen 2007, S. 29). Der Zeitpunkt des Methodeneinsatzes im Entwicklungsprozess wird bei der Differenzierung auf der Ebene des Evaluationszeitpunkts herangezogen. Eine Evaluation kann während der Konzeptphase beim simulierten Prototypen, dem High Fidelity Prototypen oder bei der Beta-Version vorgenommen werden und damit in vier Phasen eingeteilt werden (Christophersen 2007, S. 29).

In Wissenschaft und Praxis hat sich jedoch vor allem durch den Bezug zur Datenquelle die Klassifikation auf der Ebene des Evaluationsansatzes durchgesetzt. Diese Ebene soll auch hier für die Klassifikation von Methoden herangezo-

Tabelle 1: Klassifikation von Usability-Evaluationsverfahren nach dem Evaluationsansatz

Theoriebasierte Verfahren	Expertenbasierte Verfahren	Benutzerbasierte Verfahren
(formal-analytische)		(empirische)
<ul style="list-style-type: none"> › Kognitive Modelle und Architekturen Speziell: die verschiedenen Arten der GOMS (Goals, Operators, Methods and Selection Rules) 	<ul style="list-style-type: none"> › Heuristische Evaluation › Checklisten, Expertenleitfäden › Cognitive Walkthrough 	<ul style="list-style-type: none"> › Blickregistrierung › Logfile Erfassung › Befragungen › Fokussierte Interviews › Fokusgruppen › Verbale Protokolle <ul style="list-style-type: none"> – Protokolle Lauten-Denkens – Nachträgliche Protokolle – Videokonfrontation

Quelle: Schulz et al. 2006; Sarodnick/Brau 2006; Yom/Wilhelm 2004; Mack/Nielsen 1994

gen werden. Auf der Ebene des Evaluationsansatzes finden sich in der Literatur bereits verschiedene Unterteilungen für Methoden. Weit verbreitet ist die Unterteilung in theoriebasierte, expertenbasierte (analytische) und benutzerbasierte (empirische) Evaluationsmethoden (Schulz et al. 2006, S. 25; Yom/Wilhelm 2004, ▶ Tabelle 1).

Die theoriebasierte Evaluation arbeitet mit kognitiven Modellen. Diese Modelle versuchen die Prozesse, welche bei der Lösung einer interaktiven Dialogaufgabe auftreten, abzubilden und quantitativ zu prognostizieren (Yom 2003). Beispiele dafür sind die verschiedenen GOMS-Methoden. Im Bereich der praktischen Usability-Evaluation von Websites spielen die kognitiven Modelle auf Grund ihrer komplizierten Anwendung und des benötigten Fachwissens nur eine untergeordnete Rolle. Der größte Vorteil expertenbasierter Verfahren ist, dass sie vergleichsweise schnell durchzuführen sind und dass eine kleine Gruppe an Experten bis zu 75 % der Usabilityprobleme eines Interfaces finden können (Nielsen 1992). Der größte Nachteil ist, dass Experten immer vor dem Hintergrund ihrer Fähigkeiten und ihres Wissens evaluieren, „they cannot step back behind what they already know. So they will always be surrogate users (expert evaluators who emulate users)” (Harms/Schweibenz 2000). Expertenorientierte Verfahren können in allen Phasen des Usability-Engineering Prozesses zum Einsatz kommen und verfolgen vorwiegend das Ziel einer formativen Evaluation eines Interfaces (Holzinger 2005, S. 72). Expertenorientierte Verfahren brauchen des Weiteren kaum technische Hilfsmittel, was sie oftmals preiswerter als andere Verfahren macht. Benutzerbasierte Verfahren ermöglichen es, eine große Menge qualitativer und quantitativer Daten auf Basis des tatsächlichen Interaktionsverhaltens des

Nutzers mit dem Interface zu generieren. Ein Nachteil ist es, dass sich die Evaluationen oftmals in einem Labor bzw. einer Laborsituation abspielen und die Methoden einen großen planerischen und technischen Aufwand voraussetzen, was sie überdies teuer und zeitintensiv macht (Harms/Schweibenz 2000).

Bei der Bewertung der verschiedenen Methoden werden oft nur praxisorientierte Kriterien zugrunde gelegt (Schulz 2008, S. 102), Bewertungskriterien mit wissenschaftlicher Relevanz, wie Objektivität, Reliabilität und Validität, werden noch zu selten berücksichtigt. Analysiert man Studien, die wissenschaftliche Kriterien zugrunde legen, sind es vor allem Befragungsinstrumente (z. B. UFOS, SUMI etc.), die Blickregistrierung und nachträgliche verbale Protokolle, die die Anforderungen der Objektivität, Reliabilität und Validität erfüllen (Holling et al. 2006). Überdies existieren Studien zu Expertenleitfäden und zu GOMS-Techniken, die den beiden Methoden Validität bescheinigen. Aufgrund der Einschränkung bei der Objektivität schneiden viele expertenorientierte Verfahren (z. B. heuristische Evaluation, Walkthrough-Methoden) schlecht beim Zugrundelegen von wissenschaftlichen Kriterien ab (Holling et al. 2006). Die einzelnen Methoden der Tabelle 1 sollen, bis auf die ausgewählten Tracking-Methoden im nächsten Abschnitt, hier nicht im Detail erläutert werden (für einen Überblick vgl. Schulz 2008, S. 73 ff.).

2.3 Vorstellung ausgewählter Tracking-Methoden

Sowohl in der Forschung als auch in der Praxis hat sich die Meinung gefestigt, dass der Einsatz von prozessorientierten Tracking-Methoden, wie z. B. verbale Protokolle, und die

damit gewonnen Daten einen deutlichen Erkenntnisgewinn, im Gegensatz zu klassischen Usability-Befragungen, bringen können (Schmeißer/Sauer 2005, S. 75). Im Rahmen dieses Beitrags fiel die Wahl der Methoden auf die heuristische Evaluation, verbale Protokolle sowie die Blickregistrierung. Bei der heuristischen Evaluation handelt es sich dabei um das am meisten genutzte Verfahren der expertenorientierten Evaluierung (Nielsen/Molich 1990, S. 255). Als vergleichbare Methode der benutzerorientierten Verfahren bieten sich die verbalen Protokolle an. An Stelle von Experten interagieren bei verbalen Protokollen typische Nutzer und äußern wie bei der heuristischen Evaluation ihre Gedanken. Um visuelle Zuwendungen, Aufmerksamkeitsverläufe und Informationsverarbeitung im Interaktionsprozess zu messen wird als wichtige Methode der Usability-Forschung zudem immer die Blickregistrierung genannt. Jedoch ist ihr eigentlicher Beitrag im Rahmen einer Evaluation nur in Ansätzen überprüft (Burmester/Weinhold 2007).

Heuristische Evaluation

Kantner und Rosenbaum (1997, S. 154) sagen zur Stärke der heuristischen Evaluation: „Heuristic evaluation is valuable when time and resources are short. Skilled evaluators can produce high quality results...“. Bei der heuristischen Evaluation interagieren Experten, anstelle des Nutzers, mit der Website und evaluieren diese. Die Experten versuchen sich dabei in den typischen Nutzer zu versetzen und unter Rückgriff auf ihr Expertenwissen ein Interface zu bewerten (Sarodnick/Brau 2006, S. 135). Die Grundlage der Evaluation stellen dabei Heuristiken, d. h. anerkannte Prinzipien und Richtlinien für eine benutzerfreundliche Gestaltung dar (Schweibenz/Thissen 2003). Ausgehend von bestehenden Heuristiken ist eine Anpassung und Erweiterung grundlegender Prinzipien auf spezielle Untersuchungsgegenstände (z. B. einen speziellen Webshop) sinnvoll und gewünscht, wenn dies die qualitative und quantitative Evaluation verbessert (Smith/Mosier 1986, S. 8). Heuristiken können sehr komplex sein, wie die 944 Guidelines von Smith und Mosier (1986) oder die 209 Research-Based Web Design & Usability Guidelines des HSS (2006) zeigen. Jedoch ist es auch möglich, dass sie aus verdichteten, oft allgemeiner gehaltenen Prinzipien bestehen, wie die neun Heuristiken von Nielsen und Molich (1990) oder von Sarodnick und Brau (2006).

Ziel der heuristischen Evaluation ist es, mit einer geringen Anzahl von Experten Verstöße gegen die Prinzipien und damit Usabilityprobleme während des Interaktionsprozesses aufzudecken (Nielsen 1993). Als Experten kommen bei der heuristischen Evaluation sowohl Interface-Experten,

Fach-Experten, Usability-Experten als auch Doppel-Experten in Betracht (Schweibenz/Thissen 2003, S. 102). Nielsen und Molich (1992) empfehlen auf Basis von durchgeführten empirischen Studien den Einsatz von fünf Usability-Experten oder drei Doppel-Experten um angestrebte 75% der Usabilityprobleme eines Interfaces aufzudecken. Laut Nielsen (1992) sind Doppel-Experten (Fach- und Usability-Experten) mit einer Aufdeckung von max. 61% normalen Experten (41% Aufdeckung) überlegen. Durch eine Gesamtliste aller gefundenen Probleme der Experten können zudem Rückschlüsse auf die Schwere eines Problems gezogen werden (Yom 2003). Bei Nielsen (1994) findet sich überdies der so genannte Severity-rating Ansatz, der Probleme nach der Häufigkeit des Auftretens, der Auswirkung beim Auftreten und der Persistenz des Problems bewertet und so die Schwere des Fehlers einstuft.

Die Durchführung einer heuristischen Evaluation besteht idealtypisch aus einer Vorbereitungsphase, der Durchführung der Evaluation, einer anschließenden Besprechung und der Bewertung bzw. Einstufung der gefundenen Probleme (Schweibenz/Thissen 2003). Vor allem die Kenntnis und das Verständnis der anzuwendenden Heuristiken sind von großer Bedeutung. Die Evaluation eines Webinterfaces kann direkt am PC durchgeführt werden. Dabei ist es möglich, den Screen-Content computerseitig aufzuzeichnen und gleichzeitig alle Kommentare des Experten aufzunehmen.

Verbale Protokolle

Unter dem Begriff verbale Protokolle werden in der Forschung verschiedene Methoden subsumiert. Unterscheiden lassen sich drei Arten: Protokolle Lauten-Denkens (PLDs; engl. Concurrent thinking aloud), nachträgliche Gedankenprotokolle (engl. Retrospective thinking aloud) und die Videokonfrontation (engl. Video confrontation) (Ericsson/Simon 1993, S. 16; Silberer 2005, S. 264). Im Fokus dieses Beitrags stehen die PLDs, da sie sich in verschiedenen Studien als das effektivere Verfahren herausgestellt haben. So zeigte sich in der Studie von Schulz et al. (2007), dass die Situation bei PLDs als natürlicher von den Probanden empfunden wird, als bei einer videogestützten Gedankenrekonstruktion (VGR). Zudem war der Handlungsbezug der geäußerten Gedanken im Rahmen der PLDs größer und es konnten kein Einfluss der Methode auf das Surfverhalten oder die Aufgabenerfüllung ermittelt werden (auch van Den Haak et al. 2003).

Bei einem Protokoll Lauten-Denkens (PLD) soll die betreffende Person während der Interaktion mit der Website alles, was ihr durch den Kopf geht, sofort bzw. parallel zu ihrem

Denken gut vernehmbar und möglichst verständlich artikulieren (Silberer 2005, S. 265). Dies gilt dementsprechend auch für Gedanken und Probleme zur Benutzerfreundlichkeit. Ob der Testperson konkrete Aufgaben gegeben werden, die sie abarbeiten soll, hängt von der Zielsetzung der Untersuchung ab. Nielsen (1993, S. 195) beschreibt Protokolle Lauten-Denkens wie folgt: „Thinking aloud may be the single most valuable usability engineering method. Basically, a thinking-aloud test involves having a test subject use the system while continuously thinking out loud [...]“. Durch die Verbalisierung der Gedanken bzw. Probleme während des Interaktionsprozesses liegen diese in sequentieller Abfolge vor und lassen sich zusammen mit der Interaktion untersuchen. Dem Testleiter ist es damit möglich, zu verstehen, wie die Website aus Nutzersicht wahrgenommen, verstanden und genutzt wird. Dies macht es wiederum leichter, schwerwiegende Fehler bzw. Missverständnisse zu identifizieren und nachzuvollziehen (Nielsen 1993). Trotzdem weisen PLDs Schwächen auf: So müssen Nutzer, die ihre Gedanken äußern, diese zuerst wahrnehmen, anschließend die wahrgenommenen Gedanken speichern und abschließend auf das Gespeicherte zugreifen und die Gedanken in Worte fassen können (Genest/Turk 1981, S. 243; Silberer 2005). Zudem besteht die Möglichkeit, dass die Versuchsperson nicht bereit ist, ihre Gedanken zu äußern, z. B. weil es sich um Tabuthemen handelt oder es ihr peinlich ist, dass sie eine bestimmte Funktion nicht bedienen kann. Die damit angesprochene Auskunftsfähigkeit und Auskunftsbereitschaft wirken sich dabei negativ auf die Validität der Methode aus (Silberer 2005, S. 265). Weitere Probleme bezüglich der Validität können auftreten, da durch das verbalisieren während eines PLDs eine erhebliche Beeinflussung der Verhaltenssituation einhergeht (van Den Haak et al. 2003, S. 347). Die externe Validität kann überdies von Selektivitäts- und Reaktivitätseffekten beeinflusst werden (Büttner/Silberer 2008, S. 233). Das Problem der Auskunftsbereitschaft und Auskunftsfähigkeit betrifft aber alle Arten der verbalen Protokolle.

Im Gegensatz zu PLDs, wird bei einem nachträglichen Protokoll der Proband gebeten nach der Interaktion mit der Website über seinen Gedankenprozess zu berichten (Silberer 2005). Dies ist eine Möglichkeit, der Reaktivität durch das begleitende Verbalisieren zu begegnen (Silberer/Büttner 2008, S. 1112). Doch ist die Vollständigkeit solcher Protokolle durch die begrenzte Gedächtnisleistung der Versuchspersonen gefährdet und eignet sich nur für die Erfassung zentraler Gedanken (Ericsson/Simon 1993). Eine modifizierte Form der nachträglichen Protokolle stellt die VGR dar, bei der die nachträgliche Verbalisierung durch ein

Video des Interaktionsprozesses unterstützt wird. Die Videokonfrontation (VK) ist eine Erweiterung der nachträglichen Protokolle. Bei der Methode der VK werden nicht nur der Interaktionsprozess mit der Website, sondern auch die Tastatur und das Gesicht der Versuchsperson aufgezeichnet (Hamborg 2002). Anders als beim nachträglichen lauten Denken werden den Probanden bei dieser Methode während der Betrachtung des Videos konkrete Fragen zu ihrem Vorgehen und zu kritischen Situationen gestellt. Die Berücksichtigung und Beachtung von drei Videobildern durch die Versuchsperson bindet bei der Verbalisierung jedoch mehr kognitive Kapazitäten. Eine Studie von Hamborg (2002) zeigt, dass drei Stimuli zu einer Reduzierung der Anzahl der geäußerten Gedanken führen.

Blickregistrierung

Die Methode der Blickregistrierung stellt ein apparatives Verfahren der Blickmessung dar. Durch die Erfassung des Blickfeldes (Videosignal 1) und des Auges bzw. der Pupillenbewegung (Videosignal 2) kann nach einer Kalibrierung, d. h. der Abstimmung der beiden Videosignale, die Blickbewegung im Blickfeld digital auf einen PC abgebildet und aufgezeichnet werden. Nach der Kalibrierung hat man bei der Blickregistrierung die Möglichkeit, den sequentiellen Blickpfad (Sakkaden, Fixationen) über die Website aufzuzeichnen.

An dieser Stelle soll kurz auf den Zusammenhang zwischen Blickregistrierung und Usability eingegangen werden. Dieser wird in Veröffentlichungen oftmals nur unzureichend oder gar nicht herausgearbeitet (Burmester/Weinhold 2007, S. 43). Die Blickregistrierung bietet durch die Messung visueller Zuwendungen folgende Einsatzmöglichkeiten bei der Evaluation einer Website (Heinsen/Vogt 2003, S. 159):

- Darstellung des Blickverlaufes über die Website, z. B. Reihenfolge der betrachteten Elemente, auch Areas of interest (AoI) genannt.
- Ermittlung der Betrachtungszeit von AoI.
- Ermittlung von Übergangswahrscheinlichkeiten zwischen den AoI.
- Identifikation toter Zeiten, in denen nicht geklickt oder gescrollt wird.
- Benötigte Zeit für das Scanning einer Website

Folgende darauf aufbauende Möglichkeiten der Blickregistrierung finden sich in der Literatur:

- Es können Rückschlüsse auf den Aufmerksamkeitsverlauf und die Informationsverarbeitung gezogen werden.

- › Der Verlauf der Aufmerksamkeit kann quantitativ erfasst werden.

Um den Zusammenhang zwischen der Usability und dem Blickverlauf zu konkretisieren, ist an dieser Stelle der Rückgriff auf die Definition der Usability eines Webshop nötig. Ein Webshop ist dann usable, wenn er eine effektive, effiziente und zufriedenstellende Informationssuche, Informationsverarbeitung und Durchführung von Transaktionen ermöglicht. Diesen Anforderungen muss auch das visuelle Design des Webshops genügen. Daher sollte das Design des Webshops über die entsprechenden Design-Maßnahmen so optimiert werden, dass die visuelle Zuwendung (der Blick) des Nutzers in kurzer Zeit auf die relevanten Bereiche bzw. Elemente, wie z. B. Suchfunktion, Navigation, etc. geführt wird (Heinsen/Vogt 2003, S. 155). Die klare Führung des Blickes über die Website mit der damit verbundenen Möglichkeit des schnellen Zugriffs auf wichtige Funktionen scheint dabei eine notwendige, aber nicht hinreichende Bedingung für eine gute Usability zu sein. Erkenntnisse über den Klickprozess oder emotionale Prozesse bzw. weitere Einschätzungen der Website, die bei der Bewertung der Usability eine wichtige Rolle spielen, gewinnt man nicht. Auch der Rückschluss auf kognitive Prozesse der Informationsverarbeitung ist teils subjektiv und nicht objektiv messbar, da z. B. auf Basis einer vorher festgelegten Fixationszeit in Millisekunden (genormte Werte liegen dabei nicht vor) auf die Aufmerksamkeit oder Informationsverarbeitung geschlossen wird. Es scheint daher sinnvoll und angebracht im Rahmen der Blickregistrierung lediglich von visueller Zuwendung zu sprechen, da auf Aufmerksamkeitsverläufe oder gar Informationsverarbeitung nur subjektiv geschlossen werden kann und ein Zusammenhang wissenschaftlich bisher nicht abschließend bestätigt ist. Trotzdem sind viele Experten überzeugt, dass die Blickregistrierung im Bereich der formativen Usability-Evaluation eine wichtige Methode darstellt (Straub 2006).

2.4 Offene Forschungsfragen

Die Vorstellung der Methoden hat gezeigt, dass sowohl die heuristische Evaluation als auch die PLDs in der Literatur als leistungsfähige Methoden der Usability-Evaluation beschrieben werden. Beide sind jedoch von der Ausrichtung und dem angestrebten Erkenntnisgewinn sehr ähnlich angelegt. Dabei fehlt es bisher an fundierten wissenschaftlichen Vergleichen der beiden Methoden. So stellt sich die Frage, ob der Einsatz von Experten oder Nutzern sich in einem Unterschied in der Anzahl und der Qualität der identifizierten Usability-Probleme äußert oder wie viel Teilnehmer in

der entsprechenden Versuchsanordnung benötigt werden um einen vergleichbaren Erkenntnisgewinn zu erzielen. Daher lautet die erste Forschungsfrage wie folgt:

- F1:** Welche prozessorientierte Methode (PLD vs. heuristische Evaluation) erweist sich anhand der Kriterien Anzahl der Aussagen, identifizierte Probleme, identifizierte schwerwiegende Probleme, Aufwand und benötigte Interaktionszeit als die Effektivere?

Zudem konnte bereits bei der Vorstellung der Methoden aufgezeigt werden, dass der tatsächliche Beitrag der Blickregistrierung im Rahmen der Usability-Evaluation noch offen ist. Daher soll im Rahmen der hier durchgeführten Studie der erzielbare Erkenntnisgewinn sowie Nutzen der Methode im Rahmen der Usability-Evaluation näher untersucht und beziffert werden:

- F2:** Welchen Erkenntnisbeitrag liefert die Blickregistrierung im Rahmen der Usability-Evaluation?

3. Empirische Studie zum Vergleich ausgewählter Methoden

3.1 Anlage der Studie

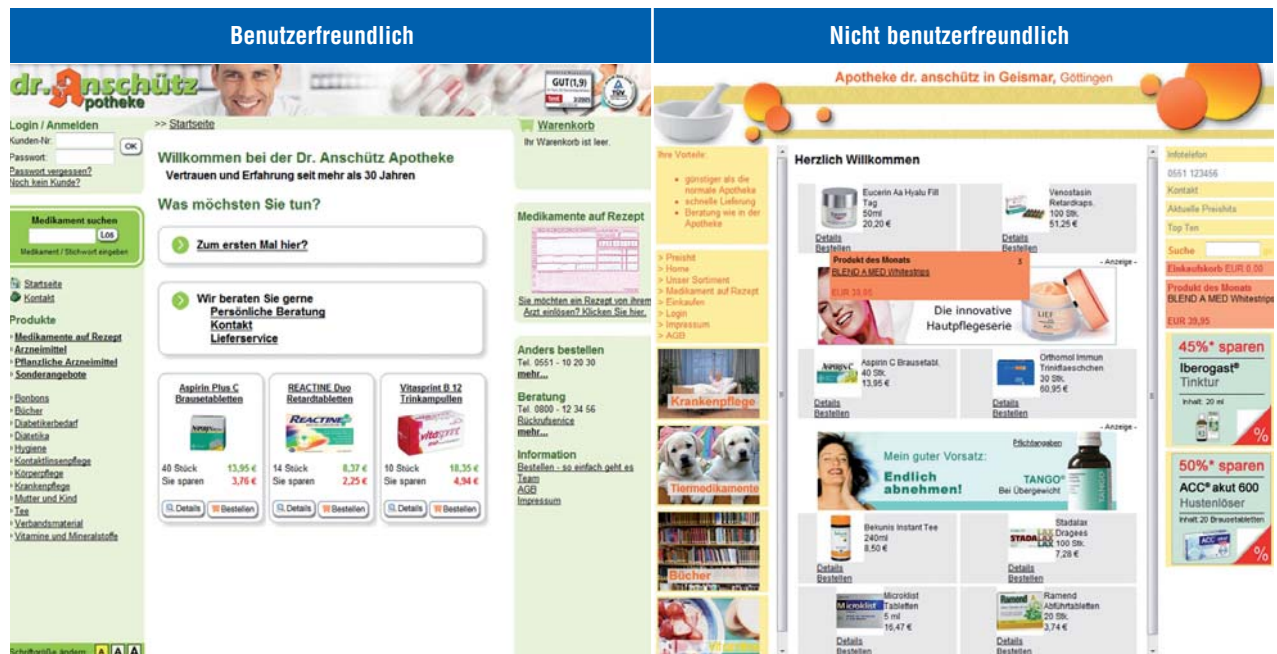
Erhebungssituation

Der empirische Methodentest besteht aus insgesamt vier Teilstudien, von denen eine im Feld und drei im Labor durchgeführt wurden. Bei der Feldstudie handelt es sich um Online-Befragungen zur Usability der Stimuli-Shops, die zum Zwecke eines Pre-Tests durchgeführt wurden. Im Labor wurden die PLDs und die Blickregistrierung mit Nutzern sowie die heuristische Evaluation mit Experten durchgeführt.

Erhebungsobjekt und Mittel der Datenerhebung

Um einen Methodenvergleich sinnvoll durchzuführen, wurden zwei verschiedene Varianten eines Internetwebshops (Internetapotheke) mit unterschiedlich gut ausgeprägter Usability entwickelt und in einem Pre-Test mit N=1002 Nutzer überprüft. Die Designs beruhen auf dem Internetapotheken-Webshop der SavIT GmbH (Kiel). Alle durchgeführten Datenauswertungen und Analysen beruhen auf der Interaktion der Probanden mit diesen Internetshops. Die Datenerhebung erfolgte mittels Online-Fragebögen, die mit dem Globalpark Umfragecenter in der Version 5.1 erstellt wurden und online zur Verfügung standen. Im Rahmen der Benutzertests im Labor kamen PLDs in Kombination mit einem Screen-Recording (Camtasia 3.0

Abbildung 1: Startseite der beiden Stimuli-Webshops



Insgesamt wurden 20 Aspekte (z. B. Kontraste, Textdesign, Scrolling etc.) mit 66 Einzelvariationen in den beiden Webshops umgesetzt. In der benutzerfreundlichen Version wurden die Aspekte möglichst perfekt und in der nicht benutzerfreundlichen Version möglichst schlecht, aber realitätsnah (dies wurde im Fragebogen überprüft) ausgestaltet (vgl. Abbildung 1). Im Rahmen der Online-Befragung (Pre-Test) wurden die Designs unter Verwendung des Usability Questionnaire for Online Shops 2.0 (UFOS; Christophersen 2007) überprüft. Die beiden Designs unterschieden sich signifikant ($F(df=3)=26,24; p=.000$) mit Mittelwerten von Benutzerfreundlich= 5.49 zu Nicht benutzerfreundlich= 3.11 (Skala: 1-7).

der Firma TechSmith) und eine Blickregistrierung zum Einsatz (Kopfgetragenes System der Firma SMI). Der Expertentest wurde in Form einer heuristischen Evaluation mit anschließender Befragung durchgeführt.

Entwicklung der Stimuli-Shops

Im Folgenden soll kurz auf die Entwicklung der Internet-apotheken-Webshops eingegangen werden. Ziel war es die Methoden (Erkenntnisgewinn usw.) anhand eines systematisch variierten Webshops zu prüfen. Dafür wurde eine Internetapotheke entsprechend variiert (benutzerfreundlich vs. nicht benutzerfreundlich). Die Dimensionen der Variation stellten erweiterte Heuristiken der ISO-Norm dar (Erwartungskonformität, Steuerbarkeit, Selbstbeschreibungsfähigkeit, Lernförderlichkeit, Individualisierbarkeit, Aufgabenangemessenheit, Prozessangemessenheit, System und Datensicherheit, Fehlertoleranz, Wahrnehmungssteuerung, Dispositionsspezifischer Aspekt, Joy of Use). Um diese Dimensionen auszugestalten und damit die Variation durchzuführen wurden Quellen für Ansatzpunkte identifiziert und auf Basis einer umfassenden Literaturrecherche

festgelegt (Schulz 2008). Berücksichtigt wurden bei der Ausgestaltung Erkenntnisse der folgenden Bereiche:

- ▶ Allgemeine Definitionen von Usability
- ▶ Empirische Studien zur Wirkung einzelner Website-Komponenten
- ▶ Studien zur Erwartungskonformität einzelner Website-Komponenten
- ▶ Prinzipien, Richtlinien und Kataloge von Empfehlungen
- ▶ Heuristische Regeln
- ▶ Dimensionen standardisierter Usabilityfragebögen

Operationalisierung der Variablen

Die Gedanken zur Usability bzw. die verbalisierten Usability-Probleme wurden bei den Nutzern und Experten prozessbezogen während des Surfens erhoben (per Screen-recording mit Audiokommentaren) und für eine quantitative Analyse entsprechend codiert. Das Kategoriensystem wurde dabei deduktiv ausgehend von der Forschungsfrage erstellt. Die Erstellung wurde zudem von einer Testcodierung unterstützt (Bortz/Döring 2002). So wurde in einem ersten Schritt

jede Aussage entsprechend ihrer Wertung (positiv/negativ) codiert. Auf der nächsten Ebene erfolgt die Codierung nach dem Websiteelement, das von der Testperson angesprochen wurde. Im letzten Schritt erfolgt die Zuordnung zu einem entsprechenden Usabilityprinzip, das im Rahmen des verbalisierten Gedanken angesprochen wurde. Bei der Codierung des Blickverlaufs wurden zwei unabhängige Codierungen durchgeführt. Im Rahmen der ersten Codierung erfolgte die Zuordnung des Blickverlaufs zu den fixierten Elementen auf der Startseite für die ersten 25 Sekunden jeder Session. In der zweiten Codierung wurden die gesamten Blickvideos der Probanden codiert. Als Kontrollvariablen wurden darüber hinaus die Interneterfahrung, die Erfahrung mit Einkäufen im Internet, Erfahrungen mit dem Medikamentenkauf im Internet, Soziodemografika (Alter, Beruf, Familienstand etc.) sowie die Usability mit dem UFOS 2.0 per Fragebogen erhoben.

3.2 Stichproben

Pre-Test

Die Stichprobengröße des Online-Pretest beläuft sich auf $N=1002$ Probanden. Das Alter der Teilnehmer beträgt im Durchschnitt 30,6 Jahre ($SD=11,3$), wobei die jüngste Person 13 Jahre und die älteste Person 77 Jahre alt waren. Die größte Berufsgruppe in der Stichprobe stellen die Studenten mit 41,9 %, gefolgt von den Angestellten mit 35,9 % dar. 47,9 % der Personen in der Stichprobe sind weiblich, 48,6 % männlich und 3,5 % machten keine Angabe. Pro Woche nutzten die Probanden das Internet $M=19,61$ Stunden ($SD=17,0$; Median 15,0) und hatten bis zum Zeitpunkt der Untersuchung im Schnitt 30 Produkte im Internet gekauft.

PLDs, Blickregistrierung und heuristische Evaluation

Die Stichprobengröße der Laboruntersuchung beläuft sich auf $N=48$ Nutzer und $N=12$ Experten. Das Alter der Nutzer beträgt im Durchschnitt 25,88 Jahre ($SD=4,2$) und das Alter der Experten 29,09 Jahre ($SD=4,9$). Bei den Nutzern handelt es sich zu 75 % um Studenten und zu 25 % um Angestellte, bei den Experten dagegen um 91,7 % Angestellte und 8,3 % Selbstständige. In der Stichprobe der Nutzer sind 31,3 % weiblich, bei den Experten 20,8 %. Die Nutzer schätzen ihre Interneterfahrung (Skala: 1 viel bis 6 keine Erfahrung) mit einem Mittelwert von $M=1,89$ ($SD=0,89$) hoch ein. Sie verbringen pro Woche $M=16,43$ Stunden im Internet und hatten im Mittel bereits 66 Produkte über das Internet gekauft. Vier Nutzer hatten bereits Medikamente über das Internet erworben. Die Experten schätzen ihre Interneterfahrung mit einem Wert von $M=1,08$ ($SD=$

$0,31$) sehr hoch ein. Die Experten verfügen über eine durchschnittliche Berufserfahrung von $M=5,6$ Jahren. Sieben der zwölf Experten hatten sich bereits mit dem Thema Internetapotheken auseinandergesetzt. Die Experten verbringen $M=35,3$ ($SD=12,2$) Stunden pro Woche im Internet. Aus dem Bereich universitäre Usability- und Website-Forschung kamen fünf Experten, drei Experten waren in ihren Institutionen für ein Content-Management-System (CMS) zuständig und vier arbeiteten in Firmen, die sich professionell mit Website-Design und Usability-Testing beschäftigen.

3.3 Ergebnisse

Bevor nun im Folgenden auf die Ergebnisse zu den beiden Forschungsfragen eingegangen wird, soll zunächst kurz auf den im Rahmen dieser Studie angewendeten Usability Fragebogen UFOS 2.0 eingegangen werden. Die Methode wurde im Rahmen des Pre-Tests und als Ergänzung nach der Durchführung der prozessorientierten Verfahren eingesetzt. Die Fragebögen ermöglichten auf schnellem und unkompliziertem Wege eine zuverlässige summative Bewertung der Designs (► [Tabelle 2](#)). Jedoch ließ der reine Usability-Score, auch bei einer Analyse der einzelnen Items, keine detaillierten Aussagen über die Gründe zu, die zu der Bewertung führten. Die Integration von offenen Textfeldern in den Usability-Fragebögen erwies sich dabei als eine erste Möglichkeit, weiterführende Informationen zu erhalten. So machten 95 % der Probanden Angaben zur Usability in den offenen Textfeldern, von denen 41 % brauchbar waren. Um weitere Aufschlüsse zu erhalten musste jedoch eine Kategorisierung der Aussagen durchgeführt werden. Dabei lag die Anzahl und Qualität der Hinweise und Problembeschreibungen deutlich unter denen der prozessorientierten Methoden.

Nun sollen im Folgenden die heuristische Evaluation und die PLDs im Mittelpunkt der Betrachtung stehen. In diesem Zusammenhang soll die Effektivität der beiden prozessorientierten Methoden genauer betrachtet und damit die erste Forschungsfrage beantwortet werden:

- F1:** Welche prozessorientierte Methode (PLD vs. heuristische Evaluation) erweist sich anhand der Kriterien Anzahl der Aussagen, identifizierte Probleme, identifizierte schwerwiegende Probleme, Aufwand und benötigte Interaktionszeit als die Effektivere?

Betrachtet man zunächst die Validität der beiden Methoden, so gelingt es beiden Methoden zwischen den beiden

Tabelle 2: Reliabilitäten der UFOS-Skala 2.0

UFOS-Skala	Cronbach's α
Pre-Test	0,94
Expertenbefragung (Labor)	0,97
Nutzerbefragung (Labor)	0,96

Designs der Internetapotheken-Webshops zu differenzieren. Sowohl bei den Experten als auch bei den Probanden unterscheiden sich wie erwartet die Anzahl der Aussagen zur Usability und die Anzahl der identifizierten Probleme signifikant in Abhängigkeit der beiden Designs (► **Tabelle 3**).

Überprüft man darüber hinaus den Zusammenhang zwischen der Anzahl der getätigten Aussagen und dem mit dem UFOS-Fragebogen gemessenen Usability-Gesamtscore zeigt sich, dass sowohl die Anzahl der negativen Aussagen der Probanden im Rahmen der PLDs als auch der Experten im Rahmen der heuristischen Evaluation mit dem Gesamtscore hoch und signifikant korrelieren. Dies bedeutet, dass mehr negative Aussagen sich entsprechend auch in einer negativen Bewertung der Usability in der Befragung niederschlugen (► **Tabelle 4**).

Im nächsten Schritt soll kurz der Aufwand im Rahmen der Vorbereitung verglichen werden, da auch dies Kosten verur-

sacht und im Rahmen einer Entscheidung für die eine oder andere Methode eine Rolle spielen kann. Es mussten im gleichen Umfang Unterlagen (Versuchsablauf, Anschreiben etc.) erstellt werden. Der technische Aufwand zumindest bei einer Kombination mit einer Verhaltensbeobachtung (Screenrecording) liegt im mittleren Bereich verglichen mit einer Befragung oder einer Blickregistrierung. Die Akquise geeigneter Versuchsteilnehmer erwies sich als schwieriger bei der heuristischen Evaluation, da für die Experten keine Aushänge oder ein Panel in Anspruch genommen werden konnten. Die Gesamtdauer der Untersuchung belief sich bei der heuristischen Evaluation im Schnitt auf 1,05 h da sowohl für die Vorbereitungen (Besprechung der zugrunde gelegten Heuristiken; Definition der typischen Nutzer etc.) als auch die Nachbesprechung deutlich mehr Zeit in Anspruch nahmen. Für die PLDs mussten im Schnitt 32 Minuten aufgewendet werden. Da es sich bei der durchgeführten Studie um eine wissenschaftliche Arbeit handelte, können anfallende Kosten hier nicht beziffert werden.

Tabelle 3: T-Test der Anzahl der Aussagen zur Usability in Abhängigkeit der Shop-Designs

	T-Test – Benutzerfreundliches vs. nicht benutzerfreundliches Design				
	M (SD) Gut	M (SD) Schlecht	T	df	p
PLDs (Probanden)	3,20 (2,30)	10,82 (6,45)	-3,55**	20	0,001
Heuristische Evaluation (Experten)	5,25 (2,00)	19,50 (6,36)	-7,40***	22	0,000

Tabelle 4: Korrelation zwischen negativen Aussagen zur Usability und UFOS-Gesamtscore

	Korrelation nach Pearson (r)	
	Anzahl der Aussagen PLD	Anzahl der Aussagen HE
UFOS Usability-Gesamtscore	0,695**	0,721**

** p < 0,01; einseitiger Test

Tabelle 5: Vergleich der Ergebnisse der Usability-Evaluation zws. Experten und Nutzern

	Heuristische Evaluation (Experten)		PLDs (Nutzer)	
	Benutzerfreundlich	Nicht benutzerfreundlich	Benutzerfreundlich	Nicht benutzerfreundlich
Aussagen zur Usability insgesamt	63	234	32	130
Identifizierte Probleme	30	61	14	30
Identifizierte schwerwiegende Probleme ¹	4	23	0 (3) ²	9 (17) ²
Verhältnis Probleme zu gesamten Aussagen	47,6 %	26,1 %	43,8 %	23,1 %
Interaktionszeit	17,5 Minuten		11,7 Minuten	

¹ Schwerwiegend: mindestens 30 % der Experten identifizierten das entsprechende Problem

² Die Nutzer identifizierten 3 von 4 bzw. 17 der 23 schwerwiegenden Probleme nach Aussagen der Experten

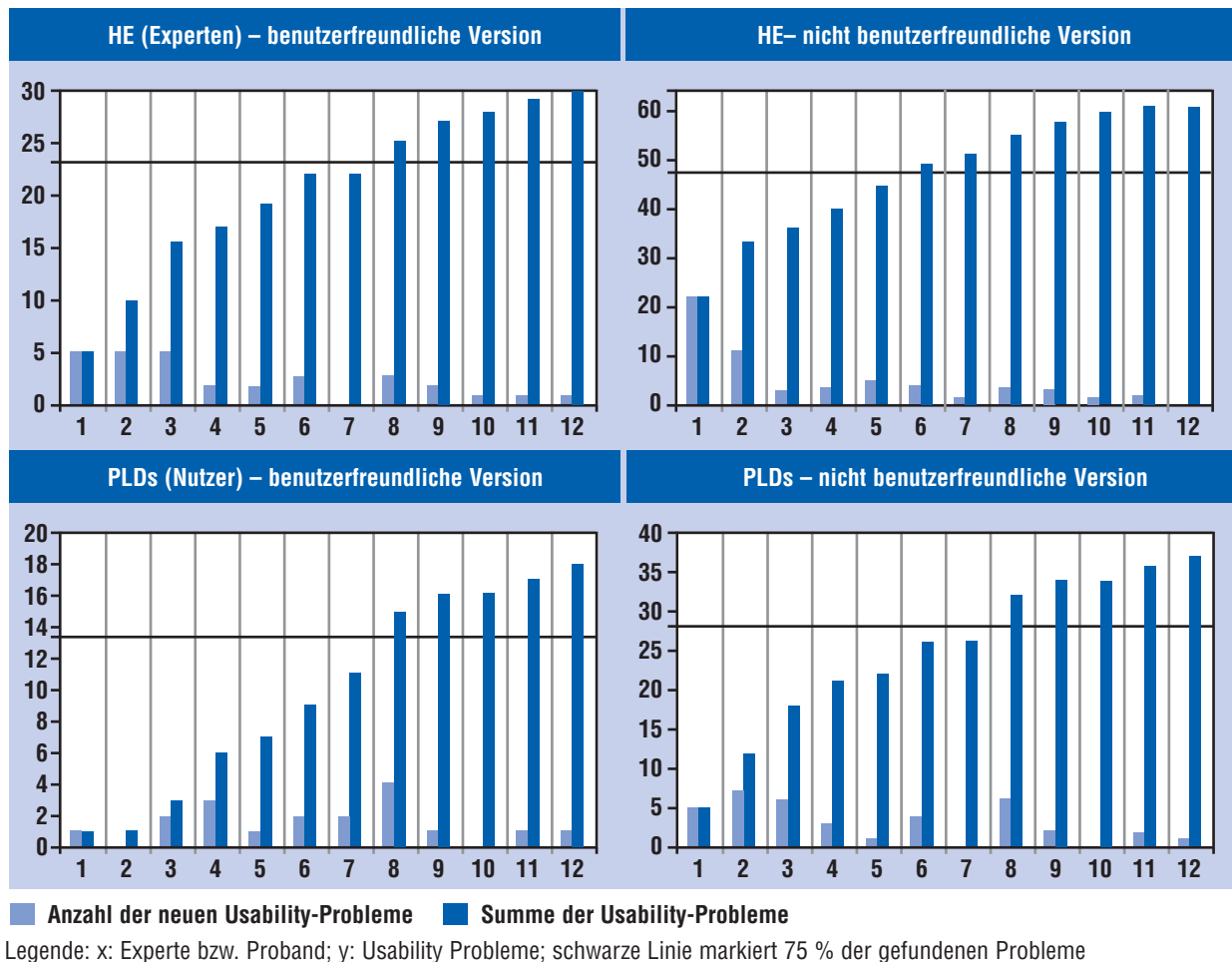
Für die Bewertung und den Vergleich der heuristischen Evaluation und der PLDs wurden im Anschluss an die Kodierung verschiedene Auswertungen durchgeführt. Zunächst folgte eine qualitative Analyse der Aussagen anschließend wurden quantitative Auswertungen der angesprochenen Elemente der Website, den betroffenen Usability-Prinzipien sowie dem Zusammenhang zwischen Elementen und Prinzipien durchgeführt. Beide prozessorientierten Methoden liefern einen erheblichen Beitrag im Rahmen der Usability-Evaluation, wobei sie jedoch überlappende Ergebnisse produzierten. Dabei wurden alle durch die PLDs identifizierten Probleme auch von der heuristischen Evaluation aufgedeckt. An dieser Stelle ist kurz anzumerken, dass trotz der umfassenden Literaturrecherche und einer möglichst perfekten Ausgestaltung noch vier schwerwiegende Probleme in der benutzerfreundlichen Version des Webshops gefunden wurden.

Insgesamt äußerten die Experten im Rahmen der heuristischen Evaluation mehr Gedanken zur Usability, identifizierten mehr Probleme und mehr schwerwiegende Probleme zur Benutzerfreundlichkeit. Dabei gilt ein Problem als schwerwiegend wenn es von mindestens 30 % der Teilnehmer angesprochen wurde. Vergleicht man die durch die Experten identifizierten Probleme mit denen, die durch die Nutzer aufgedeckt wurden, ließen sich insgesamt 17 der 23 schwerwiegenden Probleme auch durch PLDs ermitteln. Betrachtet man die Effizienz der beiden Methoden in dem Sinne, dass man die gesamten kodierten Aussagen ins Verhältnis zu den identifizierten Problemen setzt, liegen die Werte mit 47,6 % zu 43,8 % bei der benutzerfreundlichen Version und 26,1 % zu 23,1 % bei der heuristischen

Evaluation höher (► [Tabelle 5](#)). Als ein Nachteil der heuristischen Evaluation ist anzumerken, dass diese deutlich mehr Zeit in Anspruch nahm. So nahm die Interaktion mit der Website bei der heuristischen Evaluation ca. 50 % mehr Zeit in Anspruch als bei den PLDs (► [Tabelle 5](#)).

Im letzten Schritt soll nun noch die benötigte Anzahl an Versuchsteilnehmern betrachtet werden. Insgesamt evaluierten sowohl zwölf Experten als auch Nutzer die beiden Designs der Internetapotheke. Dabei ist der Beitrag jedes weiteren Experten und Nutzers während der Evaluation interessant, um Empfehlungen zur optimalen Anzahl von benötigten Teilnehmern auszusprechen. [Abbildung 2](#) zeigt die Anzahl der neu identifizierten Usability-Probleme pro zusätzlichen Teilnehmer und die Summe der gefundenen Probleme nach jedem Teilnehmer. Die eingezeichneten Linien markieren in jeder der vier Abbildungen, wann 75 % der Usability-Probleme gefunden wurden. Dem liegt die Prämisse zugrunde, dass nach den 12 Experten bzw. Nutzer (annähernd) 100% der vorhandenen Probleme aufgedeckt werden konnten. Die Abbildungen zeigen, dass dies sowohl bei Experten als auch Nutzern nach sieben bis acht Teilnehmern der Fall ist (einmal bereits nach sechs; ► [Abbildung 2](#)). Überdies ist zu beachten, dass alle schwerwiegenden Usability-Probleme bereits nach dem achten Teilnehmer als solche identifiziert werden konnten bzw. feststanden. Daher kann auf Basis der hier gewonnenen Erkenntnisse die Durchführung von PLDs oder einer heuristischen Evaluation mit acht Teilnehmern empfohlen werden. Dabei ist immer noch zu beachten, dass im Rahmen der heuristischen Evaluation in der Summe deutlich mehr Probleme identifiziert werden konnten als bei den PLDs.

Abbildung 2: Neu aufgedeckte und aufsummierte Usability-Probleme pro Experte bzw. Nutzer



Im Folgenden soll nun die zweite Forschungsfrage beantwortet werden:

F2: Welchen Erkenntnisbeitrag liefert die Blickregistrierung im Rahmen der Usability-Evaluation?

Um die Qualität und den Erkenntnisbeitrag der Methode einschätzen zu können wurden die relative Fixationsdauer und Fixationsanzahl von Elementen auf den beiden Webshops und die grafische Auswertung der visuellen Zuwendung näher betrachtet bzw. ausgewertet (für Beispiele ▶ [Abbildung 3](#)). Zudem wurden noch Übergangswahrscheinlichkeiten zwischen fixierten Flächen berechnet. Wie erwartet unterschieden sich die Ergebnisse zwischen den beiden Designs des Internetapotheken-Webshops deutlich: Der visuelle Zuwendungsverlauf beim benutzerfreundlichen Design war ruhiger, wichtige Bereiche des benutzerfreundlichen Designs konnten schneller visuell erfasst werden, die Ablenkung war geringer und wichtige Inhalte (z. B. der

Content) waren schneller zugänglich. Aus Usability-Sicht konnten zudem unbeachtete Bereiche und damit unbeachtete Informationen erkannt werden. Des Weiteren konnte in der nicht benutzerfreundlichen Version des Webshops die Ablenkungswirkung der Bilder identifiziert werden. Jedoch mussten dazu die erfassten visuellen Zuwendungsverläufe subjektiv interpretiert werden. Vergleicht man die Blickregistrierung mit den beiden eingesetzten prozessorientierten Methoden, lässt sich festhalten, dass durch die Blickregistrierung keine weiteren Usability-Probleme identifiziert werden konnten, die nicht bereits durch die PLDs bzw. die heuristische Evaluation in Kombination mit Verhaltensbeobachtung identifiziert werden konnten. Man erhält lediglich weiterführende Erkenntnisse zum Wahrnehmungsverlauf und spezifischen visuellen Zuwendungen auf den verschiedenen Seiten.

Zusammenfassend lässt sich daher festhalten, dass die Blickregistrierung zur Identifikation von Usability-Problemen nur in geringem Maße geeignet ist. Dies stellen auch

genen Evaluationsmethoden scheint dagegen nicht sinnvoll, da sie deutlich überlappende Ergebnisse produzierten. Im Rahmen dieser Studie konnte zudem beiden Methoden Validität bescheinigt werden.

Es hat sich zudem gezeigt, dass der Beitrag der Blickregistrierung eher im Bereich der weiterführenden Ursachenforschung bereits aufgedeckter Probleme zu sehen ist. Die Blickregistrierung kann z. B. Antworten darauf geben, ob ein vorhandenes Element (z. B. die Schriftgrößenänderung) nicht wahrgenommen wurde oder ob der Sinn von den Nutzern nicht verstanden wurde. Ein Einsatz empfiehlt sich auf Basis der hier gewonnen Erkenntnissen nur, wenn klare wahrnehmungspsychologische Fragestellungen vorhanden sind (auch Burmester/Weinhold 2007, S. 43).

Abschließend ist anzumerken, dass im Bereich der Methoden bzw. der Methodenvergleiche weitere Forschungen notwendig sind: Nur selten finden sich in wissenschaftlichen Veröffentlichungen objektive Kriterien, die es ermöglichen,

Evaluationsmethoden sinnvoll zu vergleichen. Des Weiteren stehen Kriterien mit praktischer Relevanz deutlich im Vordergrund und wissenschaftliche Kriterien, wie Objektivität, Reliabilität oder Validität werden zu Unrecht vernachlässigt. Der hier vorgestellte Beitrag kann nur als ein erster Schritt in diese Richtung angesehen werden. Limitationen der Studie ergeben sich zudem daraus, dass nur ein Typ von Webshop (Internetapotheken) betrachtet wurde und die Stichprobe der „normalen“ Nutzer deutlich studentisch geprägt war.

Literatur

Bortz, J.; Döring, N. (2002): Forschungsmethoden und Evaluation für Sozialwissenschaftler, 3. Auflage, Berlin, Heidelberg, New York.

Büttner, O.; Silberer, G. (2008): Video-cued Thought Protocols: A Method for Tracing Cognitive Processes at the Point of Purchase, in: Brick & Mortar Shopping in the 21st Century, Hrsg.: Lowrey, T. M., Mahwah, S. 221 - 241.

Burmester, M.; Weinhold, T. (2007): Liefert Eye-Tracking einen Mehrwert im Rahmen formativer Usability-Tests?, in: i-com, 6. Jg., H. 2, S. 39 - 43.

Christophersen, T. (2007). Usability im Online-Shopping: Entwicklung eines Fragebogen Instrumentes (ufosV2) unter Berücksichtigung formati-

LITERATUR FÜR MEHR DURCHBLICK

Management Book Service bietet Wirtschaftsprofis und denen, die es werden wollen, die ganze Welt der Wirtschaft und vieles mehr!

UNSERE SERVICELEISTUNGEN – IHR VORTEIL

- Besorgung aller lieferbaren Titel aus dem In- und Ausland
- 16.000 lagernde Fachbücher aus allen Bereichen der Wirtschaft
- Internationale Buchdatenbanken mit mehr als 2,8 Millionen Büchern
- Eine große Auswahl an Belletristik-Titeln, Videos, Magazinen und Hörbüchern
- Regelmäßige Informationen über Novitäten am Fachbuchmarkt
- Prompter Buchversand
- Öffnungszeiten: MO – FR 9.00 bis 18.30, SA 9.00 – 13.00

MANAGEMENT
MBS
BOOK SERVICE



1090 Wien
Augasse 5-7
Tel.: 01/317 89 11
www.mbsbuch.at

- ver und reflektiver Messmodelle, erschienen in der Reihe: Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business, URL: [http://www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/new/fileadmin/publications/pdf/dissertation_christophersen.pdf].
- DIN EN ISO 9241-10 (1996):** Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 10 Grundsätze der Dialoggestaltung. Berlin.
- Ericsson, A. K.; Simon, H. A. (1993):** Protocol Analysis: Verbal Reports as Data, Cambridge, Massachusetts.
- Dzida, W. (1994):** Qualitätssicherung durch software-ergonomische Normen, in: Einführung in die Software – Ergonomie, 2. Auflage, Hrsg.: Eberleh, E.; Oberquelle H.; Oppermann E., Berlin, S. 372 - 406.
- Gediga, G.; Hamborg, K.-C. (2002):** Evaluation in der Software-Ergonomie: Methoden und Modelle im Software-Entwicklungsprozess, in: Zeitschrift für Psychologie, 210. Jg., H. 1, S. 40 - 57.
- Genest, M.; Turk, D.C. (1981):** Think-aloud Approaches to Cognitive Assessment, in: Cognitive Assessment, Hrsg.: Merluzzi, T.V.; Glass, C. R.; Genest, M., New York, S. 233 - 269.
- Gray, W. D.; Salzman, M. C. (1998):** Damaged merchandise? A Review of Experiments that Compare Usability Evaluation Methods, in: Human Computer Interaction, 13. Jg., H. 3, S. 203 - 261.
- Hamborg, K.-C. (2002):** Gestaltungsunterstützende Evaluation von Software: Zur Effektivität und Effizienz des IsoMetricsL Verfahrens, in: Mensch & Computer 2002: Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten, Stuttgart, S. 303 - 312.
- Harms, I.; Schweibenz, W. (2000):** Usability Engineering Methods for the Web. Results from a Usability Study, in: Informationskompetenz – Basiskompetenz in der Informationsgesellschaft. Proceedings des 7. Internationalen Symposiums für Informationswissenschaft, Hrsg.: Knorz, G.; Kuhlen, R., Konstanz, S. 17 - 30.
- Heinsen, S.; Vogt, P. (2003):** Usability praktisch umsetzen. Ein Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte, München.
- Holling, H.; Freund, P. A.; Kuhn, J.-T.; Salascheck, M. (2006):** Benutzbarkeit von Software: Wie usable sind Evaluations-Verfahren?, Arbeitsbericht Nr. 41 des Kompetenzzentrums Internetökonomie und Hybridität, Münster.
- Holzinger, A. (2005):** Usability Engineering Methods for Software Developers, in: Communication of the ACM, 48. Jg., H. 1, S. 71 - 74.
- HSS (2006):** Researched-based Web Design & Usability Guidelines, Health Department of Health and Human service, URL: [<http://www.usability.gov/pdfs/guidelines.html#1>], Washington.
- Kantner, L.; Rosenbaum, S. (1997):** Usability Study of WWW Sites: Heuristic Evaluation vs. Laboratory Testing, in: Proceedings of the 15th SIGDOC 97 conference, NY, S. 153 - 160.
- Kilian, T.; Hass, B.; Walsh, G. (2007):** Web 2.0 – Neue Perspektiven im E-Business, Heidelberg.
- Konradt, U.; Wandke, H.; Balzas, B.; Christophersen, T. (2003):** Usability in Online Shops: Scale Construction, Validation and Influence on the Buyers Intention and Decision, in: Behaviour & Information Technology, 22. Jg., H. 3, S. 61 - 73.
- Mack, R. L.; Nielsen, J. (1994):** Executive Summary, in: Usability Inspection Methods, Hrsg.: Nielsen, J.; Mack, R. L., New York, S. 1 - 23.
- Nielsen, J. (1992):** Finding Usability Problems Through Heuristic Evaluation, in: Proceedings of the CHI 92 Conference, Monterey, CA, S. 373 - 380.
- Nielsen, J. (1993):** Usability Engineering, Boston, San Diego, New York.
- Nielsen, J. (1994). Heuristic Evaluation,** in: Usability Inspection Methods, Hrsg.: Nielsen, J.; Mack, R. L., New York, S. 25 - 62.
- Nielsen, J. (2000):** Designing Web Usability: The practice of Simplicity, USA.
- Nielsen, J.; Loranger, H. (2006):** Web Usability, München, Boston, San Francisco etc.
- Nielsen, J.; Molich, R. (1990):** Heuristic Evaluation of User Interfaces, in: Proceedings of Conference on Human Factors in Computing Systems 1990, URL: [<http://portal.acm.org/citation.cfm?id=97281&dl=ACM&coll=portal>], (Abruf: am 01.08.2008), S. 249 - 256.
- Oppermann, R.; Reiterer, H. (1997):** Software Evaluation using the 9241 Evaluator, in: Behaviour & Information Technology, 16. Jg., H. 4/5, S. 232 - 245.
- Sarodnick, F.; Brau, H. (2006):** Methoden der Usability Evaluation – Wissenschaftliche Grundlagen und praktische Anwendung, Bern.
- Schmeißer, D. R.; Sauer, H. M. (2005):** Usability-Forschung im E-Commerce – Methoden und Ergebnisse am Beispiel des Online-Versandhandels, in: Planung & Analyse, o. Jg., H. 1, S. 75 - 81.
- Schulz, S. (2008):** Surfverhalten und -erleben in Webshops - Der Einfluss der Usability, des Risikos und des Kundenstatus auf den Surfprozess sowie die Auswirkungen auf das Vertrauen am Beispiel von Internetapotheken, Dissertation Institut für Marketing und Handel, Göttingen, (erscheint Anfang 2009).
- Schulz, S.; Beulshausen, L.; Silberer, G. (2007):** Measurement of Webshop Usability: A Question of the Right Method?, Vortrag auf der 9th General Online Research Conference, Leipzig, Deutschland.
- Schulz, S.; Mau, G.; Silberer, G. (2006):** Benutzerfreundlichkeit von Reisekatalogen - Die Anwendung von Usability-Fragebögen auf Printkataloge, in: transfer - Werbeforschung und Praxis, 51. Jg., H. 4., S. 24 - 32.
- Schweibenz, W.; Thissen, F. (2003):** Qualität im Web: Benutzerfreundliche Webseiten durch Usability Evaluation, Berlin, Heidelberg.
- Silberer, G. (2005):** Videogestützte Rekonstruktion kognitiver Prozesse beim Ladenbesuch, in: Marketing ZFP, 27. Jg., H. 4, S. 263 - 280.
- Silberer, G.; Büttner, O. B. (2008):** Marktforschung am Point of Sale, in: Handbuch Marktforschung: Methoden - Anwendungen – Praxisbeispiele, Hrsg.: Herrmann, A.; Homburg, C.; Klarmann, M., 3. Aufl., Wiesbaden, S. 1097 - 1123.
- Straub, K. (2006):** Heatwave! Leveraging Heat Maps (and other Eye Tracking Data) to Refine your Information Architecture, in: UI Design Newsletter, URL: [<http://www.humanfactors.com/downloads/aug06.asp>], (Abruf: am: 03.08.2008).
- Smith, S. L.; Mosier J. N. (1986):** Guidelines for Designing User Interface Software, Technical Report MTR-10090, Bedford, online verfügbar, URL: [<http://www.hcibib.org/sam/>].
- Sweeney, M.; Maguire, M.; Shackel, B. (1993):** Evaluating User-computer Interaction: A Framework, in: International Journal of Market Man-Machine, 38. Jg., H. 4, S. 689 - 711.
- Van Den Haak, M. J.; De Jong, M. D. T.; Schellens, P. J. (2003):** Retrospective vs. Concurrent Think-aloud Protocols: Testing the Usability of an Online Library Catalogue, in: Behaviour & Information Technology, 22. Jg., H. 5, S. 339 - 351.
- Waiguny, M.; Matzler, K.; Toschkov, A.; Hoppe, M. (2008):** Usability-Dimensions and their Emotional Outcome, in: Proceedings of the 37th EMAC, Brighton, CD-Rom.
- Yom, M. (2003):** Web-Usability von Online-Shops, Göttingen.
- Yom, M.; Wilhelm, T. (2004):** Methoden und Erkenntnisse der Web-Usability-Forschung, in: Usability von Internet-Angeboten. Grundlagen und Fallstudien (Stuttgarter Beiträge zur Medienwissenschaft Nr.10), Hrsg.: Zerfuß, A.; Zimmermann, H., Baden-Württemberg.