

Literaturdokumentation

„Lautes Denken“

Felix Schmidt, Meike Schröder, Daniel Postert, Theresa Jarchow

Autor: Norgaard, M., Hornbaek, K.

Titel: What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing

Jahr: 2006

Laufende Nr. im Literaturverzeichnis: 19

Schlagwörter: Usability evaluation, think aloud testing, industrial software development, user-centered design

Kurzbeschreibung:

Das explorative Paper untersucht inwieweit Thinking Aloud in der Praxis eingesetzt wird. Dazu wurden 14 Usability Sessions aus verschiedenen Firmen aufgezeichnet und analysiert. Die Studie zeigt, dass strukturierte Analysen direkt im Anschluss an die Erhebung eher selten sind. Die Autoren kommen zu dem Schluss, dass die meisten Versuchsleiter versuchen ihre Meinung zu bestätigen und empirische Beweise für bereits bekannte Probleme zu generieren und dadurch häufig unbekannte Probleme übersehen oder nicht genug beachten. Desweiteren zeigt die Studie auf, dass einige Faktoren, welche in der Forschung selten diskutiert werden, die Umsetzung von usability Evaluationen und Tests beeinflussen (z.B. unfertige Prototypen oder fehlende Expertise mit dem System seitens des Versuchsleiters)

Es ist schade, dass die Autoren nicht den gesamten Prozess von Anfang bis Ende untersucht haben, da es interessant gewesen wäre herauszufinden ob und wann eine strukturierte Analyse der Usability Tests in der Praxis stattfindet.

Erstellt von: Felix Schmidt (346741)

Autor(en): Lynne Cooke

Titel: Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach

Jahr: 2010

Laufende Nr. im Literaturverzeichnis: 5

Schlagwörter: Cognitive psychology, communication, technical communication

Kurzbeschreibung:

In diesem Paper zeigt die Autorin durch eine vergleichende Analyse mit Eye-tracking das Thinking Aloud in 80% der Fälle akkurat wieder spiegelt was Benutzer gerade Denken. Sie argumentiert, dass die restlichen 20% der Fälle in denen das Eye-tracking nicht mit den Äußerungen übereinstimmt, dadurch zu erklären sei, dass visuelle Informationen schneller verarbeitet werden können als sie verbalisiert werden können. Desweiteren wurde untersucht welche „Content“ Kategorien in den Verbalisierungen enthalten waren und was Augenbewegungen während den Phasen des Schweigens über das Verhalten der Benutzer aussagen können. Dazu wurden die Augenbewegungen und Verbalisierungen von 10 Benutzern aufgezeichnet und analysiert, während diese verschiedene Aufgaben auf einer Webseite bearbeiteten. Der Großteil der Verbalisierungen enthielten Wörter, Phrasen oder Sätze, welche die Benutzer auf dem Bildschirm lesen konnten oder Verbalisierungen über das momentane Verhalten bzw. Ziele. Es wurde gezeigt, dass Lückenfüller und Phasen des Schweigens dem Benutzer ermöglichen die dargestellten Informationen zu analysieren und verarbeiten. Das Paper zeigt also, dass Thinking Aloud eine wertvolle Methode ist und wesentlich zum Erkenntnisgewinn von Usability Tests beiträgt. Es wird darauf hingewiesen, dass Phasen des Schweigens mit erhöhter kognitiver Last einhergehen könnten und Usability Experten deshalb ein besonderes Augenmerk auf solche Situationen richten sollten.

Eine sehr interessante Studie, welche die Validität und Reliabilität der Thinking Aloud Methode untersucht und einen positiven Ausblick gibt.

Erstellt von: Felix Schmidt (346741)

Autoren: Olmsted-Hawala, E.L., Murphy, E.D., Hawala, S., Ashenfelter K.T.

Titel: Think-Aloud Protocols: A Comparison of Three Think-Aloud Protocols for use in Testing Data-Dissemination Web Sites for Usability

Jahr: 2010

Laufende Nr. im Literaturverzeichnis: 21

Schlagwörter: Think aloud, user testing, verbalization, user-centered design, usability testing, experimental design

Kurzbeschreibung:

In dieser Studie untersuchen die Autoren welche Effekte die Anwendung verschiedener Thinking Aloud Methoden auf die Effektivität und Zufriedenheit von Testpersonen hat.

Dazu wurden 80 Personen in einer double blind, between subject design Studie getestet. Drei verschiedene Thinking Aloud Protokolle wurden getestet: ein traditionelles, ein "speech-communication" Protokoll, sowie ein Protokoll das „Coaching“ seitens des Versuchsleiters enthielt. Desweiteren gab es noch eine Kontrollgruppe, welche nicht laut denken sollte. Die Ergebnisse zeigen, dass sowohl die Rate der Task-completion als auch die subjektive Zufriedenheit der Testpersonen in der „Coaching“ Bedingung am höchsten war. Dies ist problematisch, da es verfälschte Ergebnisse liefert, welche ein Nutzer ohne Unterstützung nicht erreichen würde. Denn in den Bedingungen ohne Unterstützung, welche gleichzeitig näher an der Realität der Benutzung sind, waren die Testpersonen weniger zufrieden. In Bezug auf die Effektivität konnte kein signifikanter Unterschied zwischen den verschiedenen Bedingungen festgestellt werden. Das bedeutet das, dass „laute Denken“ keinen Einfluss auf die Dauer hatte, die nötig war um eine Aufgabe zu bewältigen. Dieses Ergebnis wird allerdings dadurch geschwächt, dass nur acht **einfache** Aufgaben bearbeitet wurden. Der Effekt von Thinking Aloud auf die Effektivität der Bearbeitung von schwierigen oder komplexen Aufgaben bleibt also weiterhin unbeantwortet. Die Autoren schließen mit dem Ratschlag, entweder die traditionelle oder die „speech-communication“ Methode zu verwenden, da beide realistischere Ergebnisse liefern als die „Coaching“ Methode.

Erstellt von: Felix Schmidt (346741)

Autoren: Yen,P.Y., Bakken,S.

Titel: A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol - An Example from a Web-based Communication Tool for Nurse Scheduling

Jahr: 2009

Laufende Nr. im Literaturverzeichnis: 25

Schlagwörter: Usability evaluation, heuristic evaluation, think-aloud protocol, staffing and scheduling, nursing

Kurzbeschreibung:

Diese Studie zeigt, dass Heuristic Evaluation und Thinking Aloud kombiniert werden sollten. Dazu wurden sechs Usability Experten und einige (keine genaue Zahl genannt) Endnutzer beobachtet während Sie mit dem Programm interagierten. Die Studie zeigt, dass die Heuristic Evaluation eher generelle Probleme aufdeckt, wohingegen die Thinking Aloud Methode eher Probleme bei der Aufgabenbearbeitung enthüllt. In anderen Worten:

Usability Experten finden mehr Probleme zum Thema „ease of use“ wohingegen Endnutzer (während Sie laut Denken) eher oder Probleme finden die mit der „usefulness“ zu tun haben.

Daher wird eine Kombination von beiden Methoden als „Best practice“ vorgeschlagen. Als Schwäche dieser Studie kann die geringe Anzahl von Versuchspersonen angeführt werden, auch finde ich Problematisch dass nirgendwo die Anzahl der Versuchspersonen für den Thinking Aloud Teil der Studie genannt werden.

Erstellt von: Felix Schmidt (346741)

Autor(en): Guan, Z., Lee, S., Cuddihy, E., Ramey, J.

Titel: The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking

Jahr: 2006

Laufende Nr. im Literaturverzeichnis: 10

Schlagwörter: Retrospective think aloud, validity, reliability, verbalization, eye tracking, usability research.

Kurzbeschreibung:

Gemäß den Autoren ist “Retrospective Think Aloud” (RTA) eine usability Methode welche die Äußerungen eines Benutzer im Bezug auf vorher bearbeitete Aufgaben sammelt. Dies ist eines der weniger Paper, welche die Validität und Reliabilität dieser Methode (experimentell) untersucht. Dazu wurden die Augenbewegungen der Probanden und die später gemachten Äußerungen analysiert. Die Autoren kommen zu dem Schluss, dass RTA eine valide und nützliche Methode ist, deren Validität von der Aufgabenkomplexität unberührt bleibt. Desweiteren scheint RTA zusätzliche Informationen bezüglich der Lösungsstrategien und Annahmen des Nutzers zu liefern.

Die Autoren analysierten auch Phasen des Schweigens und fanden heraus, dass diese häufig mit Schwierigkeiten bei der Bearbeitung von Aufgaben einhergingen. Dies bestätigt andere Arbeiten in denen ebenfalls gefunden wurde, dass Phasen des Schweigens Indikatoren für Usability Probleme sein können.

Die Studie unterstützt und ermutigt den Einsatz der RTA Methode.

Erstellt von: Felix Schmidt (346741)

Autoren: Rogers, Y., Sharp, H., Preece, J

Titel: Interaction Design (3rd edition), John Wiley & Sons, 2011, S. 256-258 und S. 286-290

Jahr: 2011

Laufende Nr. im Literaturverzeichnis: 22

Schlagwörter: Thinking Aloud, Analysis, Data Gathering

Kurzbeschreibung:

In Ihrem Buch: Interaction Design-Beyond Human Computer Interaction geben Preece et al. Eine Einführung in die Methode der Think Aloud Technique. Dabei werden sowohl die Phase der Datenerhebung als auch die darauffolgende Analyse mit Hilfe von Beispielen dargestellt und erklärt.

Erstellt von: Felix Schmidt (346741)

Autor: Freeman, B.

Titel: "Triggered Think - Aloud protocol: using eye tracking to improve usability test moderation"

Jahr: 2011

lfd. Nummer:1

Schlagwörter: Triggered Think Aloud; Eyetracking, Usability methods, ConcurrentThink Aloud

Kurzbeschreibung:

Die Autoren stellen "Triggered Think-Aloud" als eine neue Methode vor, welche Think Aloud effektiver und effizienter machen soll.

Im Gegensatz zum Concurrent Think-Aloud wird die Versuchsperson beim Triggered Think Aloud nicht aufgefordert, ununterbrochen zu verbalisieren, sondern nur dann zu berichten, wenn ihnen etwas auffällt, das sie gerne mitteilen möchten.

Mittels Eyetracker beobachtet der Tester dann während des Tasks in Echtzeit, mit welchen Elementen des Testobjekts sich die Versuchsperson zur Zeit befasst. Der Tester kann bei interessierenden Themen gezielt nachfragen, während uninteressante Themen nicht überflüssigerweise verbalisiert werden. Das Think Aloud wird somit effizienter. Die kognitive Belastung für die User wird gering gehalten, Fragen wie "Worauf achten sie jetzt gerade" oder "Haben sie diesen Button bemerkt" werden redundant.

Der (nicht erwünschte) Einfluss des Think Aloud auf das Verhalten der User soll so

minimiert werden.

Im Gegensatz zum Retrospective Think Aloud, bei dem der Einfluss ebenfalls minimal ist (da komplett auf Verbalisierung während der Aufgabe verzichtet wird), kommt es bei TTA auch nicht zu unerwünschten Verzerrungen durch Memory-Effekte oder zu Verfälschung (da Probanden, nicht wie bei Retrospective Think Aloud, wenig Zeit haben, über Gründe für ihr Verhalten nachzudenken).

Triggered Think Aloud verbindet also die Vorteil von Concurrent Think Aloud und Retrospective Think Aloud.

Leider gibt es zu diesem spannenden Ansatz bisher nur eine kleine Vorstudie und keine empirischen Ergebnisse.

Auch setzt sich der Autor wenig mit der Mehrbelastung des Versuchsleiters auseinander, der bei dieser Methode ununterbrochen darauf achten muss, wo der Blick der User gerade liegt. Verpasst er den richtigen Augenblick um gezielt nachzufragen, bringt auch eine Videoaufnahme nichts mehr. Zudem stellt sich die Frage, wie gravierend eine Unerwartete Unterbrechung des Denkprozesses der User ist, im Vergleich zur ständigen Verbalisierung.

Erstellt von: Meike Schröder

Autor: Elling S., Lentz L. & de Jong M.

Titel: Retrospective think-aloud method: using eye movements as an extra cue for participants' verbalizations

Jahr: 2011

Lfd.Nr.: 2

Schlagwörter: Retrospective Think Aloud, Eyetracking, Gaze Replay

Zusammenfassung:

In der vorliegenden Studie beschäftigen sich Elling et.al mit der Frage, ob der Einsatz von Eytracking beim Retrospective Think Aloud sinnvoll ist, um den User beim nachträglichen Verbalisieren zu unterstützen, d.h. um die Erinnerung daran zu erleichtern, wo im jeweiligen Moment der Fokus lag.

Untersucht wird, ob Retrospective Think Aloud *mit* Eyetracking zu mehr oder anderer Verbalisierung führt als Retrospective Think Aloud *ohne* Eyetracking. Bereits existierenden Studien zu der selben Fragestellung mangelt es laut Elling et.al an Gütekriterien, vorrangig auf Grund zu kleiner Stichproben und zu einfachen Tasks. Aufgabe der insgesamt 40 User war es bei ihrer Studie, auf der Homepage einer Stadt

herausfinden, wie staatliche Zuschüsse beantragt werden können. Es wurden dabei zwei Gruppen miteinander verglichen: in der ersten Gruppe wurde Retrospective Think Aloud mit dynamicscreenreplay angewandt (die User sahen im Nachhinein noch einmal ihre Aktionen auf der Webseite), in der zweiten Retrospective Think Aloud mit dynamicgazereplay (zusätzlich wurden hier die Blickbewegungen der User eingeblendet). Erhoben wurde die Gesamtzahl entdeckter Usabilityprobleme, die Art der Usability-Probleme (Navigation, Inhalt, Design), sowie die Meinung der User zu dem Versuch

Es wurden keine signifikanten Unterschiede hinsichtlich der Anzahl und Art der entdeckten Probleme gefunden, es scheint also für die Entdeckung von Usabilityproblemen nicht notwendig zu sein, Retrospective Think Aloud mit Eyetrackingdaten zu unterstützen. Mögliche Gründe dafür könnten sein, dass der erwartete Vorteil durch die Faszination von Laien gegenüber Eyetracking relativiert wird. Methodisch unsauber an der Studie ist die Tatsache, dass User, für die aufgrund zu schneller Augenbewegungen nur unvollständige Eyetracking Daten vorlagen, daraufhin einfach der anderen Bedingung (also Screenreplay) zugeordnet wurden, wodurch es sich nicht mehr um randomisierte Stichproben handelt.

Die Einführung der User in Think Aloud und Beispiele fanden erst *nach* Bearbeitung der Aufgabe und *direkt vor* dem Retrospective Think Aloud statt, was das Gedächtnis der User zusätzlich belastet haben könnte.

Für eine Studie zur Untersuchung einer Methode würde es sich evtl. auch besser eignen, eine Seite zu testen, bei der bekannt ist, wieviele Usabilityprobleme sie hat, um z.B. sicher gehen zu können, dass keine Deckeneffekte vorliegen.

Positiv hervorzuheben ist die große und repräsentative Stichprobe.

Erstellt von: Meike Schröder

Autor: McDonald, S., Edwards, H. & Zhao, T.

Titel: Exploring think-alouds in usability testing: an international survey.

Jahr: 2012

lfd. Nr.:3

Schlagwörter: Think aloud, Survey, Usability-Testing

Zusammenfassung:

Das Paper von McDonald gibt einen Überblick darüber, wo, wie und warum die Thinkaloud-Technik bei Usabilitytests angewandt wird.

Es scheint in den letzten Jahren ein Wandel bei dem Einsatz der Methode des TA zu geben, weg vom klassischen hin zu einem "relaxteren" TA. Dies könnte negative Konsequenzen für die Reliabilität und Validität haben. Um festzustellen ob diese Entwicklung tatsächlich stattfindet, und wenn ja, was mögliche Ursachen sein könnten, untersuchen die Autoren systematisch den Einsatz der Methode, um ein reliables Bild der aktuellen Situation zu erlangen.

Befragt wurden Usabilityfirmen, einzelne Personen, Teilnehmer bei Usabilitykonferenzen u.Ä.. An der anonymen Onlineumfrage, welche zwischen November und Dezember 2010 stattfand, beteiligten sich 207 Usabilitytester.

Die Fragen-gebiete lassen sich unterteilen in: 1. Welche Methode wird benutzt, 2. wie ist der Ablauf, 3. wie ist die Interaktion zwischen User und Tester und 4. wie wird ausgewertet.

Es handelte sich um 22 Fragen, zusammengesetzt aus offenen, geschlossene, und Likert-Skalen. Die Dauer der Umfrage lag bei 10 Minuten.

Eine erste Pilotstudie wurde zum "offputting" genutzt, also um die Anzahl offener Fragen, zu reduzieren.

Es folgt ein Auszug der Ergebnisse:

98/ der befragten benutzen Think Aloud mindestens *manchmal*, 71% mindestens *regelmäßig*.

Alle Teilnehmer halten es für die beste Methode, Probleme und Ursachen der Probleme zu finden (im Vgl. zu Interview, Verhaltensbeobachtung, Perfromancedata).

Die am häufigsten verwendete Art des TA ist Concurrent Think Aloud, 98% haben diese Methode schon mind. Einmal verwendet, bei 89% ist Concurrent Think Aloud die am häufigsten eingesetzte Art des Think Aloud. Ein häufig genannter Grund gegen Retrospective Think Aloud ist die längere Dauer

Instruktionen:: 68% der Befragten gaben an, ihren User eine Instruktion gemäß Ericsson und Simon zu geben.

Einüben: 52% trainieren Think Aloud überhaupt nicht, 15% nur sehr selten mit ihren User, bevor der eigentliche Task losgeht. Dies begründendie Befragten damit, dass die User mit Think vertraut sei und ein Training nicht notwendig. Die Autoren geben zu bedenken, dass auch der hohe Zeitdruck ein Grund für das Weglassen des Trainings sein könnte.

Während Ericsson und Simon raten, Interaktionen wenn möglich zu vermeiden, befolgen dies nur etwas ein Drittel der Befragten. Rund die Hälfte verfolgt einen "relaxteren" Ansatz. Interessant ist hierbei eine Interaktion zw. Experience und Unterbrechungen, Experten unterbrechen deutlich weniger als weniger Erfahrene

Nur 6 % der Befragten richten sich komplett nach den Empfehlungen von Simon und Ericsson, was die Vermutung eines starken Unterschiedes zwischen Think Aloud-Theorie und Praxis bestätigt.

Web basierte Unfragen haben natürlich immer den Nachteil, das sie freiwillig sind, es antworten generell eher Leute, die sich für die Inhalte der Studie interessieren, d.h. In der Realität könnte es um einiges schlechter bezüglich der Güte der eingestetzten Think-Aloud-Methoden stehen. Zudem ist nichts über den Wahrheitsgehalt der gemachten Aussagen bekannt.

Über mögliche kulturelle Unterschiede wurde leider keine Aussage gemacht.

Auch unklar ist, wieviele Anfragen rausgeschickt wurden, das heißt die Rücklaufquote, die doch recht interessant ist, bleibt unbekannt.

Erstellt von: Meike Schröder

Autor: Strain P., Dawn Shaikh A., Boardman R.

Titel: Thinking but not seeing: think-aloud for non-sighted users

Jahr: 2007

Lfd. Nr: 4

Schlagwörter: Think aloud, User testing, blind user, guideline

Kurzbeschreibung: Technik soll für alle Menschen zugänglich sein, und dies schließt natürlich auch blinde Menschen/ visuell beeinträchtigte Menschen ein. Worauf bei Usability-Tests, insbesondere bei der Think Aloud-Methode, mit dieser Nutzergruppe geachtet werden sollte, erläutert dieser Artikel.

Blinde Menschen navigieren mittels Screenreader, welche den Inhalt einer Seite vorlesen. Dies macht es schwer, Concurrent Think Aloud anzuwenden.

Der Autor stellt daher drei Abwandlungen des Think Aloud zum Einsatz mit blinden Menschen vor:

Synchronized Concurrent Think Aloud: Der User hat die Möglichkeit, den Screenreader auf Pause zu stellen, wann immer er etwas mitteilen möchte. So kommt es nicht zu Konflikten zwischen Reden und Zuhören, allerdings wird die Aufgabe dabei unterbrochen. Diese Abwandlung funktionierte in einer Vorstudie gut bei nicht zu komplexen Webseiten. Bei kurzen Aufgaben oder kognitiv anspruchsvollen bewährte sich Retrospective Think Aloud.

Eine weitere Modifikation ist das „Modified Stimulated Retrospective Think Aloud“: Nach Beenden der Aufgabe erfolgt ein Walkthrough, d.h. die Person geht das Interface noch

einmal durch, und erklärt dabei, was passiert bzw. in ihr vorgeht. Der Versuchsleiter kann dabei den Screenreader jederzeit stoppen und genauer nachfragen, wenn weitere Infos von Interesse sein könnten.

Weitere Probleme bei Think Aloud mit blinden Menschen sind die Geschwindigkeit des Screenreaders, den blinde User sehr schnell stellen, was zur Folge hat, dass der Tester starke Verständnisschwierigkeiten hat. Dazu kommen die komplexen Keystrokes, mit denen Blinde navigieren (z.B. "H" for Header), denen der Tester nur schwer folgen kann. Dem kann man z.B. entgegenwirken, indem der aktive Bereiche des Bildschirms visuell hervorgehoben werden, so dass die Tester nachvollziehen kann, wo der User sich gerade befindet

Leider handelt es sich hierbei nur um eine Vorstudie mit eher anekdotischer Evidenz. Dennoch behandelt das Paper ein wichtiges und noch oft zu kurz gekommenes Thema, ist sehr Praxisnah und liefert einfache Handlungsvorschläge, die ohne zu großen technischen Aufwand realisierbar sind.

Erstellt von: Meike Schröder

Autor: Als B.S., Jensen J.J., Skov M.B.

Titel: Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children

Jahr: 2005

Lfd. Nr.: 5

Schlagwörter: Constructive Interaction, Think Aloud, Usability with Children, Comparison

Kurzbeschreibung: Bei der Methode der Constructive Interaction arbeiten zwei User zusammen an einer Aufgabe, was im optimalen Fall zu "Natürlichem Think Aloud" führt. Dies scheint gerade für Kinder, die den Instruktionen für das klassische Laut-Denken evtl. noch nicht folgen können und für die traditionelles Think Aloud eine sehr ungewohnte Situation ist, eine gute Alternative zu sein.

Ob sich die Ergebnisse eines Usabilitytest mit Kindern hinsichtlich der zwei Methoden unterscheidet, wird in dieser Studie untersucht. Zusätzlich wird untersucht, welchen Einfluss die Zusammenstellung der Interaktionspartner hat.

60 Kinder wurden dafür in drei Versuchsgruppen aufgeteilt: Think Aloud, Constructive Interaction mit Kindern die sich schon kennen und Constructive Interaction mit Kindern die sich noch nicht kennen.

Die User bzw. die Dyade musste verschiedene Aufgaben an einem ihm unbekanntem Mobiltelefon bearbeiten und dabei laut denken bzw. "konstruktiv interagieren". Im

Anschluss wurde der NASA-TLX zu Erfassung der kognitiven Workloads ausgefüllt. Die gefundenen Probleme wurden in „kritisch“, „ernst“ und „kosmetisch“ kategorisiert. Die Ergebnisse zeigen, dass bei Constructive Interaction (bekannte und unbekannte Interaktionspartner) mehr Probleme gefunden werden als bei Think Aloud. Bei Partnern, die sich schon vorher kannten wurden zudem mehr Fehler gefunden als bei unbekanntem, insbesondere in der Kategorie kritische Fehler

Zu kritisieren ist das Alter der Versuchsperson (13-14 Jahre). Es handelt sich dabei eher um Teenager als um Kinder, und in diesem Alter ist Think-Aloud sicherlich schon verständlich. Eine Studie mit jüngeren Versuchspersonen wären sicher interessanter und aufschlussreicher gewesen, Zudem war die Anzahl der VP je Bedingung unterschiedlich, 12 Personen in der Think-Aloud-Bedingung im Gegensatz zu 12 Dyaden, also 24 Personen in der Constructive-Interaction-Bedingung. Das in der zweiten Bedingung mehr Fehler gefunden wurden, kann also auch auf die größere Gruppe zurückgeführt werden. Auch ist es schade, dass nur gleichgeschlechtliche Paare untersucht wurden.

Erstellt von: Meike Schröder

Autor: Hertzum M., Hansen K.D & Andersen H.H.K.

Titel: Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload?

Jahr: 2009

lfd. Nr.: 6

Schlagwörter: Think aloud, relaxed Think Aloud, comparison, mental workload, behaviour,

Kurzbeschreibung: Immer häufiger wird Think Aloud in einer "relaxten" Form statt wie im Original durchgeführt. In der Praxis heißt das, der User wird während der Aufgabenbearbeitung öfter unterbrochen, und der Tester fragt nach Kommentaren und oder Erklärungen für das aktuelle Handeln. Ericsson und Simon raten in ihrer „Anleitung“ zum Think Aloud hingegen, nur bei längerem Schweigen den User daran zu erinnern, weiterhin laut zu denken.

Welchen Einfluss dies auf die Validität der Methode hat, und wie sich diese Art des Think-Aloud auf den mentalen Workload auswirkt, untersuchen Hertzum, Hansen und Andersen.

Es werden die Einflüsse von Think Aloud, „relaxtem“ Think Aloud und keinem Think Aloud auf des Verhalten der Users untersucht, im speziellen auf die Bearbeitungsdauer, Augenbewegungen, mentalen Workload (mittels NASA-TLX) und darauf, ob die gegebenen Aufgaben korrekt bearbeitet wurden.

Acht Versuchspersonen wurden gebeten, Aufgabe auf einer Homepage eines TV-Senders und in einem Webshop zu bearbeiten.

Die Daten ergaben, dass „klassisches“ lautes Denken kaum oder nur sehr geringe Verhaltensunterschiede zum Nicht-laut-denken aufweist, während „relaxte Laut-Denker“ im Allgemeinen länger für die Bearbeitung der Aufgaben brauchten, die Aufgabenstellung häufiger missachteten, und einen subjektiv höheren mentalen Workload hatten.

Dass sich ein wissenschaftlicher Artikel damit auseinander setzt, wie die Praxis von der Theorie abweicht, und welche Folgen das hat, passiert leider nur selten, daher ist dieser Artikel sicher eine Bereicherung.

Methodisch zu bemängeln ist allerdings, das alle User zu Beginn ein Training bekommen, in dem sie lernten, wie man laut denkt. Dies ist in der Praxis aber leider nicht immer der Fall, so dass die externe Validität der Ergebnisse dieser Studie u.U. nicht gegeben ist.

Zudem ist die Stichprobe mit nur 8 Versuchspersonen doch recht klein und die Reihenfolge einzelner Versuchsbedingungen wurde nicht ausbalanciert, was ebenfalls einen Einfluss auf die Ergebnisse haben könnte.

Erstellt von: Meike Schröder

Autor(en): Ericsson K. A., Simon H.A.

Titel: *Protocol analysis: Verbal report as data,*

Jahr: 1984

Laufende Nummer im Literaturverzeichnis:

Schlagwörter:

Kurzbeschreibung: In ihrem Paper (aus dem später ein ganzes Buch wurde) wollen Simon und Ericsson zeigen, das verbaler Report eine gute Quelle für wertvolle und Valide Informationen über kognitive Vorgänge sein kann. Auch wenn die Verbalisierungen nicht immer vollständig sind, geben sie dennoch einen guten Einblick.

Die Autoren kategorisieren Verbalisationen in drei Kategorien: Level 1 (direkt aus dem Kurzzeitgedächtnis abgerufen), Level 2 (direkte Verbalisierung von Informationen, die in nicht-sprachlicher Form vorliegen, wie z.B. Bilder), und Level 3 (gefilterte/ inferierte Informationen). Level 1 und Level 2 gelten als Valide, da sie direkt aus dem Kurzzeitgedächtnis abgerufen werden, während Level 3 Verzerrungen unterliegt)das aud

dem Langzeitgedächtnis abgerufen). Level 1 und Level 2-Verbalisierungen haben laut den Autoren keinen Einfluss auf die Performance, im Gegensatz zu Level 3-Verbalisierungen, da diese die cognitive Load erhöhen.

Weiterer Inhalt des Artikels ist die Vorstellung eines Vorgehensmodells für Think-Aloud, wobei der wichtigste Punkt sicherlich ist, dass Probanden nie nach gezielten Informationen gefragt werden sollten, da es sich hierbei um nicht-valide Level-3-Verbalisierungen handelt.

Bei dieser Studie handelt es sich selbstverständlich um einen Klassiker, dennoch ist es keine „Studie“ im eigentlichen Sinne, es stellt lediglich die Entwicklung eines theoretischen Modells zum Think-Aloud anhand vorheriger Studien dar. Empirisch überprüft wird das Modell hingegen leider nicht.

Erstellt von: Meike Schröder

Titel: Cultural Cognition in Usability Evaluation

Autor(en): Clemmensen, T., Hertzum, M., Hornbaek, K., Shi, Q., Yammiyavar, P.

Jahr: 2009

Laufende Nummer im Literaturverzeichnis: 4

Schlagwörter: cultural differences, thinking aloud, usability, international systems development

Kurzbeschreibung:

In diesem Paper wurde analysiert wie der kulturelle Hintergrund einen Effekt auf Usability-Evaluationen hat, basierend auf der thinkaloud - Methode. Es handelt sich hierbei um eine Literaturstudie von 2009, die die Notwendigkeit der Berücksichtigung des kulturellen Einflusses darstellt.

Grundlegend wird angenommen, dass sich Personen mit verschiedenen kulturellen Hintergründen in ihrer kulturellen Kognition unterscheiden. Hierbei werden die westliche Kultur (Europäer, Amerikaner und Bürger des Commonwealth) und die östliche Kultur (prinzipiell Chinesen, Koreaner und Japaner) gegenübergestellt. Als Beispiele für in Studien gefundene Unterschiede nennen die Autoren unter anderem worauf sich Personen beziehen mit der Umwelt als östlichem Beispiel und Objekten in der westlichen Kultur. Zudem charakterisieren sie westliche Personen als analytische Denker -

"geradlinig denkend" - wohingegen sie östliche Personen als ganzheitlich denkend beschreiben - "in einem Kreis denkend".

Es werden 4 Elemente genannt, die bei der Beachtung von kulturellen Effekten eine Rolle spielen. Diese sind: Instruktionen & Aufgaben, Verbalisation, "Lesen" des Nutzers und allgemeine Beziehung zwischen dem Nutzer und dem Evaluator. Für diese 4 Aspekte nennen die Autoren folgende, zu beachtende kulturelle Einflüsse. Für West-Kulturelle findet sich kein Unterschied ob Aufgaben als reine Instruktionen oder eingebettet in Szenarios präsentiert werden. Ost-Kulturelle hingegen sprechen eher auf die kontextuellen Informationen an. Während Ost-Kulturelle durch thinkingaloud in ihrer Leistung eher eingeschränkt werden, scheint dies kein Effekt auf West-Kulturelle zu haben bzw. die Leistung eher zu verbessern. Westliche Nutzer sind für den Evaluator leichter zu "lesen" da sie eher dazu tendieren, mehr Überraschung zu zeigen und somit mehr Informationen zu ihrem erlebten Problem preisgeben. Nutzer aus dem Osten hingegen tendieren eher dazu Aspekte mit der testenden Person indirekt und mit Finesse zu klären um die soziale Harmonie während der Evaluation zu wahren. Aufgrund dessen wird die Evaluation substantiell dadurch beeinflusst, ob die beteiligten Personen den gleichen kulturellen Hintergrund haben oder nicht.

Die Studie zeigt einen wichtigen Aspekt in der Usability-Evaluation auf, dem es sich in weiterer Forschung eingehender zu widmen gilt.

Erstellt von: Daniel Postert (346628)

Titel: Trust in Wikipedia: How Users Trust Information from an unknown Source

Autor(en): Lucassen, T., Schraagen, J.M.

Jahr: 2010

Laufende Nummer im Literaturverzeichnis: 15

Schlagwörter: experimentation, human factors, reliability, verification

Kurzbeschreibung:

Die Autoren der Studie beschäftigen sich mit Wikipedia als Informationsquelle mit stark steigender Bedeutung. Da Wikipedia eine freie Enzyklopädie ist, in der jeder Einträge verfassen kann, sollte die Glaubwürdigkeit von Artikeln verschiedener Qualität von den Teilnehmern (Laien) eingeschätzt werden.

Es wurde ein 2 (Vertrautheit mit Thema) x 2 (Qualität des Artikels) within-Design gewählt,

mit Vertrautheit und Qualität als unabhängige Variablen. Als abhängige Variablen wurden Comentare per thinkinaloud - Methode, ein Vertrautheits-Rating sowie ein Glaubwürdigkeits-Rating (beide per 7-Punkt-Likert-Skala) erfasst. Die Teilnehmer waren n=15 Studenten deutscher und holländischer Herkunft mit im Schnitt 5 Jahren Erfahrung im Gebrauch von Wikipedia. Die Seiten wurden hinsichtlich ihrer Qualität nach dem Wikipedia Editorial Team Assessment eingestuft. Die Teilnehmer wurden vor der Testung hinsichtlich ihrer Vertrautheit zu den verschiedenen Themen befragt und in der Anwendung des Wikipedia Screening Task und der thinkinaloud - Methode unterwiesen.

Es wurden keine statistischen Unterschiede zwischen guten und schlechten Artikeln sowie vertrauten und nicht vertrauten Themen bezüglich der Glaubwürdigkeit gefunden. Hinsichtlich der Methode wird jedoch berichtet, dass es einen Unterschied hinsichtlich der Angabe von Kommentaren zu den Artikeln zwischen dem thinkinaloud - Protokoll (1147 Kommentare) und den Fragebögen (456 Kommentare) gegeben hat. Die Studie basierte im Folgenden auf den thinkinaloud - Aussagen, da ein wichtiger Vorteil darin besteht, dass mehr Informationen hinsichtlich der kognitiven Prozesse erhalten werden im Vergleich zu den Fragebögen. Die Autoren geben keine Information hinsichtlich der Qualität der Kommentare, die durch die thinkinaloud - Methode erfasst wurden.

Erstellt von: Daniel Postert (346628)

Titel: Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning

Autor(en): Kim, B., Park, H., Baek, Y.

Jahr: 2009

Laufende Nummer im Literaturverzeichnis: 13

Schlagwörter: media in education, interactive learning environments, teaching/learning strategies, virtual reality

Kurzbeschreibung:

Ziel der Studie war es, Effekte von meta-kognitiven Strategien auf die Leistung bezüglich des Lernens und des Spielens über soziale Problemlösefähigkeiten zu finden. Dabei wird angenommen, dass meta-kognitive Strategien eine wichtige Rolle bezüglich Problemlösesituationen spielen.

Als spielebasierte Lernumgebung wurde das kommerzielle Online-Rollenspiel "Gersang" genutzt, welches mit seinen wirtschaftlichen Aspekten die Spieler dazu anhält zu

kollaborieren, Strategien zu entwickeln, planen und mit Objekten, Ressourcen und andere Spielern zu interagieren. Akademische Leistung sowie der Spielpunktstand wurden als abhängige Variablen festgelegt. Als meta-kognitive Strategien (und unabhängig Variablen) wurden self-recording (schriftlich), modeling und thinkingaloud in verschiedenen Gruppen realisiert. Zudem wurde die soziale Problemlösefähigkeit als mediierende Variable erfasst.

Die Stichprobe bestand aus 132 Schülern der neunten Klasse in Südkorea. Spielbezogen wurde für ein einheitliches Ausgangsniveau gesorgt. Spielpunktstand sowie soziale Problemlösefähigkeit (gemessen über das socialproblemsolvinginventory-revised) wurden als pre- und post-Test erhoben. Dazwischen erfolgte eine 10-wöchige Testphase in der die Teilnehmer 2 mal wöchentlich 45 Minuten spielten unter Anwendung der jeweiligen meta-kognitiven Strategie, in die sie vor Beginn instruiert wurden. Die Ergebnisse zeigen, dass die thinkingaloud - Methode den größten gefundenen Einfluss auf die soziale Problemlösefähigkeit aufweist. Diese weist einen großen Zusammenhang mit sowohl der akademischen Leistung als auch mit dem Spielpunktstand auf. Ähnliche Ergebnisse wurden bezüglich des modelings gefunden, lediglich self-recording weist einen wesentlich geringeren Zusammenhang mit sozialen Problemlösefähigkeiten auf. Dies wird dahingehend interpretiert, dass die self-recording-Methode an sich im Vergleich zu beiden anderen keine soziale Interaktion beinhaltet.

Verschiedene Aspekte der Studie sind kritisch zu betrachten. Die Studie basiert auf kommerziellen Spielen und nicht auf Lernspielen. Bei kommerziellen Spielen kann eine grundlegend andere intrinsische Motivation des Spielers nicht ausgeschlossen werden. Die gefundenen Effekte des hohen Effektes von sozialen meta-kognitiven Strategien auf die Leistung der Teilnehmer könnte auch auf eine Peer-Interaktion und somit Mediation zurückzuführen sein. Außerdem sind große kulturelle Unterschiede nicht auszuschließen, da Videospiele in der sudkoreanischen Gesellschaft eine bedeutend andere Rolle spielen als in der europäischen Kultur. Somit ist der Generalisierungsgrad der Ergebnisse eingeschränkt zu betrachten.

Erstellt von: Daniel Postert (346628)

Titel: Metaphors of Human Thinking for Usability Inspection and Design

Autor(en): Frokjaer, E., Hornbaek, K.

Jahr: 2008

Laufende Nummer im Literaturverzeichnis: 9

Schlagwörter: usability evaluation, metaphors of human thinking, thinking aloud testing, usability inspection

Kurzbeschreibung:

In der Studie wurde eine neue Usability Technik "metaphors of human thinking" (MOT) getestet. Die Metaphern beinhalten Gewohnheiten, Gedankenfluss, Aufmerksamkeit, Assoziationen und Wissen.

Die Testung fand in 3 Experimenten statt. Die erste Testung zeigte, dass Usability Probleme gefunden durch MOT tiefergreifender und schwieriger zu reparieren sind, als solche die durch heuristische Evaluation gefunden wurden. Die Ergebnisse des 2. Experiments zeigen, dass mit MOT mehr Probleme gefunden werden als mit cognitivewalkthrough. Das dritte Experiment vergleicht die Methoden MOT, cognitivewalkthrough (CW) und thinkingaloud (als Gold-Standard des Usabilitytestings).

An diesem Experiment nahmen n=55computerscienceStudenten teil, welche 2 verschiedene nicht-traditionelle User Interfaces (naturallanguageinterface&mobilephoneinterface) evaluieren sollten. Jeder Teilnehmer wandte jeweils 2 der 3 Methoden an und evaluierte beide Interfaces. Nachdem sie eine individuelle Liste der Usability-Problemen erstellt hatten, wurden diese in Gruppen von 2-3 Personen zu einer Gruppenliste zusammengetragen. Letztendlich wurden 2 Ziellisten kreiert, ein für jedes Interface und diese wurden von Clients (Mitarbeiter an dem jeweiligen Interface) nach Schwere und Komplexität bewertet. Als unabhängige Variablen galten die 3 Methoden MOT, CW und thinkingaloud. Die abhängigen Variablen waren Anzahl der gefundenen Probleme, Ernsthaftigkeit der Probleme eingeschätzt von den Probanden und Rating der Clients hinsichtlich Schwere und Komplexität sowie Bevorzugung durch Teilnehmer und Evaluatoren.

Die Ergebnisse zeigen, quantitativ mehr Probleme durch MOT gefunden werden, allerdings mit nur kleinen Effekten. Mit der thinkingaloud Methode werden am meisten Probleme gefunden, die von den Teilnehmern als schwerwiegend kategorisiert werden (22%) im Vergleich zu MOT (13%). Hinsichtlich der Einschätzung der Clients konnte kein signifikanter Unterschied gefunden werden. N=18 Teilnehmer bevorzugten thinkingaloud über MOT (n=3 MOT über TA). Gründe hierfür sind zum einen, dass Ergebnisse der Methode leichter zu interpretieren sind. Außerdem geben die Teilnehmer an, mit der

thinkingaloud Methode in der Lage waren solche Probleme zu identifizieren, wozu sie ansonsten nicht in der Lage gewesen wären.

Die Studie zeigt eine interessante Methode auf, die besonders im Vergleich zu cognitivewalkthrough und der heristischen Evaluation eine neue Denkweise mit guten Ergebnissen bietet. Im Vergleich mit der thinkingaloud Methode jedoch, treten diverse Schwachpunkte in den Fokus.

Erstellt von: Daniel Postert (346628)

Autor(en): Birns, J. H., Joffre, K. A., Leclerc, J. F., Paulsen, C. A.

Titel: Getting the whole Picture: Collecting Usability Data using two Methods – Concurrent Think Aloud and Retrospective Probing

Jahr: 2002

Laufende Nr. im Literaturverzeichnis: 2

Schlagwörter: concurrent think aloud, retrospective probing, usability, methods

Kurzbeschreibung: Diese Arbeit fasst Stärken und Schwächen der Methoden des Concurrent Think Aloud (CTA) und des Retrospective Probings basierend auf Forschungsergebnissen und praktischer Erfahrung zusammen.

CTA stellt hierbei die beliebtere und häufiger durchgeführte Methode dar. Die Stärken dieser Methode liegen darin, dass Versuchsleiter unmittelbar den Denkprozessen der Testpersonen folgen können und durch dieses Feedback einen guten Einblick in die Nutzererfahrung erhalten. Die Verbalisierung, welche simultan stattfindet stellt jedoch gleichzeitig auch die größte Schwäche dar, da dadurch eine zusätzliche Anforderung an das Arbeitsgedächtnis und somit den Proband gestellt wird, welche einen Einfluss auf die Aufgabenbearbeitung haben kann.

Im Gegensatz dazu stellen die Autoren das Retrospective Probing vor. Dabei werden der Testperson direkt nach Aufgabenbearbeitung Fragen zur Interaktion mit dem entsprechenden System gestellt. Da keine Verbalisierung während der Bearbeitung stattfindet fällt die Schwäche der CTA weg. Im Nachhinein gegebenes Feedback fokussiert sich dabei meistens auf größere Systemschwächen bezüglich der Usability. Der größte Vorteil hierbei ist, dass eine ganzheitliche, reflektierte Aussage zu der Anwendung getroffen werden kann und nicht immer nur Teilaspekte bewertet werden. Nachteilig wird erwähnt, dass Personen dazu tendieren, ihr Verhalten durch eigene (Mis-)Konzepte und Theorien zu erklären, wodurch die Aussagengüte verfälscht werden kann.

Die Autoren geben einen gut Überblick über beide Verfahren bezüglich ihrer Stärken,

Schwächen und warum es sinnvoll sein kann beide zu kombinieren. Leider werden die Verfahren nicht detailliert beschrieben und keine konkreten Forschungsbeispiele bezüglich der Annahmen gegeben. Somit dient die Ausarbeitung als grundlegende theoretische Ansicht der beiden vorgestellten Verfahren.

Erstellt von: Daniel Postert (346628)

Autor(en): van den Haak, M. J., de Jong, M. D. T., Schellens, P. J.

Titel: Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue

Jahr: 2003

Laufende Nr. im Literaturverzeichnis: 24

Schlagwörter: retrospective think-aloud, concurrent think-aloud, usability

Kurzbeschreibung: In dieser Studie werden retrospective (RTA) und concurrent think-aloud (CTA) Protokolle anhand verschiedener Usability-Aspekte verglichen. Das Testobjekt ist hierbei der Online Katalog der Vrije Univeriteit, Amsterdam. Der Online-Katalog wurde gewählt, da er Charakteristika einer Suchmaschine mit den Online-Eigenschaften einer Website kombiniert.

Die Stichprobe bestand aus n=40 Teilnehmern, die alle Studenten der Kommunikationswissenschaften der Universität Twente waren. Sie wurden in gleicher Gruppengröße den zwei think-aloud Bedingungen zugewiesen. In jeder Bedingung wurde eine Aufgabenbatterie mit sieben Aufgaben durchgeführt. In der CTA-Bedingung wurden die Teilnehmer dazu aufgefordert die Aufgaben zu bearbeiten und dabei laut zu denken. Der Versuchsleiter war anwesend stand jedoch für Rückfragen nicht zur Verfügung und griff nur in den Versuchsablauf ein, wenn die Person seit einiger Zeit nichts mehr gesagt hatte. In der RTA-Bedingung wurden die Teilnehmer wurden die Probanden ebenfalls gebeten die Aufgaben zu bearbeiten, hier jedoch ohne jegliche Verbalisierung. Sie wurden dabei gefilmt und kommentierten ihr Verhalten im Nachhinein, als sie die Aufzeichnungen betrachteten. Drei Fragestellungen wurden verfolgt: (1) Unterschiede bezüglich der Häufigkeit und Art der gefundenen Usability-Probleme, (2) Unterschiede in der Aufgaben-Performanz und (3) Unterschied in dem Erleben der Teilnehmer. Hinsichtlich der Menge der gefundenen Usability-Probleme waren beide Arten miteinander vergleichbar. Jedoch wurden Unterschiede in der Art der auftretenden Probleme gefunden. So wurden mit der CTA-Methode mehr Probleme identifiziert, welche sich auf die Beobachtung bezogen, mit der RTA-Methode wurden eher Probleme

gefunden, die nur durch die anschließende Verbalisierung zu erkennen waren nicht jedoch durch alleinige Beobachtung. Bezüglich der Aufgabenbewältigung zeigte sich, dass Teilnehmer der CTA-Methode mehr Fehler machten und weniger erfolgreich waren, die sieben Aufgaben erfolgreich abzuschließen. Dies wird der Anforderung des zusätzlichen Verbalisierens während der Aufgabenbearbeitung zugeschrieben. In beiden Fällen bewerteten die Teilnehmer ihr Erleben bezüglich des Lauten Denkens eher neutral, die größere Beanspruchung bei der CTA-Methode wurde also subjektiv durch die Teilnehmer nicht bestätigt.

Die Studie wurde hinsichtlich der Methodik sehr sauber durchgeführt, die thinking-aloud Durchführung war möglichst standardisiert, die Gruppen wurden randomisiert zugewiesen und unterschieden sich nicht in deskriptiven Aspekten. Der Vergleich der beiden Methoden stellt eine wichtige Forschungs- und Anwendungsfrage dar und die Ergebnisse zeigen dass es sich um äquivalente, jedoch hinsichtlich der gefundenen Artefakte unterschiedliche Evaluationsmethoden handelt. Aufgrund der Einschränkung des Vergleichs auf ein Objekt, kann die Generalisierung der Ergebnisse jedoch hinterfragt werden und verlangt nach weiterer Gegenüberstellung.

Erstellt von: Daniel Postert (346628)

Autor(en): Hoppmann, T. K.

Titel: Examining the „point of frustration“. The think- aloud method applied to online search task

Jahr: 2009

Laufende Nr. im Literaturverzeichnis: 12

Schlagwörter: ThinkingAloud, Methodenkombination, Rechercheprozess, Stimmungen, Frustration

Kurzbeschreibung: In diesem Paper wird ein Experiment vorgestellt bei dem der ThinkingAloud Test, kombiniert mit einem standardisierten Fragebogen (Ermittlung von Computerkenntnissen, Informationsbedarf und Erwartungen an die Website) und einem nachträglichen Interview, dazu verwendet wird um den Prozess der Informationssuche von Experten auf einer European Union Website zu untersuchen. Zudem sollen Stimmungsverläufe, wie die Zufriedenheit oder Unzufriedenheit der Testpersonen während der Bearbeitung von Rechercheaufgaben, analysiert werden. Dazu werden zehn

Experten darum gebeten auf der Website arbeitsrelevante Informationen zu recherchieren. Dabei sollen sie alles verbalisieren, was ihnen durch den Kopf geht. Ihnen ist es selbst überlassen, an welcher Stelle sie mit den Ergebnissen zufrieden sind und die Aufgabe als beendet ansehen. Während der Aufgabenbearbeitung ist ein Assistent der Versuchsleitung anwesend, beobachtet das Vorgehen und macht sich Notizen über die Stimmung der Testpersonen. Falls notwendig regt er die Testpersonen zum Weiterreden an. Die Verbalisierungen werden protokolliert und auf Video aufgezeichnet. Jede Session dauert bis die Testpersonen zufriedenstellende Ergebnisse gefunden haben oder einen Frustrationspunkt erreicht haben, an dem sie die Aufgabe abbrechen wollen. Um die getätigten Äußerungen besser nachzuvollziehen, werden sie anschließend darum gebeten ihre Strategie bei der Aufgabenbearbeitung zu erläutern. Zur Auswertung wurden sowohl die Verbalisierungen als auch die nachträglichen Erklärungen transkribiert. Zur Interpretation der Äußerungen wurden zusätzlich die Notizen der Assistenten und die Videoaufzeichnungen hinzugezogen. Die jeweiligen Vorgehensweisen wurden schließlich hinsichtlich unterschiedlicher Kategorien (Rechercheprozess, positive/negative Bemerkungen, Stimmung) kodiert und die jeweiligen Strukturen analysiert. Mithilfe dieses Vorgehens konnten die unterschiedlichen Prozesse der Informationssuche rekonstruiert und in einem Modell verallgemeinert werden. Auch wurde ersichtlich an welchen Stellen des Prozesses die Wahrscheinlichkeit für Frustrationen stieg. In diesem Fall war die Frustration am größten, wenn keine zufriedenstellenden Informationen gefunden wurden.

Das Experiment zeigt, dass mit der ThinkingAloud Methode nicht nur der Prozess einer Informationssuche nachgebildet, sondern auch Stimmungsänderungen identifiziert werden können.

Es wird sehr ausführlich beschrieben, wie die erhobenen qualitativen und quantitativen Daten ausgewertet und schließlich zu Ergebnissen integriert werden. Auch wird deutlich, dass sich die zusätzliche Verwendung von nachträglichen Interviews und Videoaufzeichnungen als vorteilhaft zur Vervollständigung der Verbalisierungen des ThinkingAloud Tests erweisen.

Erstellt von: Theresa Jarchow (309960)

Autor(en): Moritz, F., Meinel, C.

Titel: Mobile web usability evaluation- combining the modified think aloud method with the testing of emotional, cognitive and conative aspects of the usage of a web

Jahr: 2010

Laufende Nr. im Literaturverzeichnis: 17

Schlagwörter: Thinking-aloud Methode, Feldstudie, Kombination

Kurzbeschreibung: In diesem Paper wird eine Methode eines Usability Tests vorgeschlagen, bei der der ThinkingAloud Test mit einer Feldstudie kombiniert werden sollte. Das angestrebte Ziel dieser Kombination ist es dabei, die Probleme eines unnatürlichen Laborsettings, in dem ThinkingAloud Tests normalerweise stattfinden, zu überwinden. Die Autoren schlagen vor, dass Experimentalsetting in die natürliche Umgebung der Probanden zu verlegen und diese mit diese dabei mit ihnen bekannten Gerätschaften arbeiten zu lassen. Dabei arbeitet die Testperson beispielsweise an ihren eigenen Laptop eine vorgegebene Aufgabe und spricht dabei laut ihre Gedanken aus, während sie von der integrierten (unsichtbaren) Video- und Soundtechnik aufgezeichnet wird. Die Testperson fühlt sich dabei weniger beobachtet und verhält sich natürlicher. Auch halten die Autoren die Aufzeichnung des Screens, den der Testuser während der Bearbeitung sieht, für sinnvoll. Es werden anschließend Vor- und Nachteile von Laboruntersuchungen und Feldstudien gegenübergestellt. In einem Laborsetting sitzt der Proband allein am System verbalisiert seine Gedanken ohne wirklich mit jemanden zu kommunizieren und wird von sichtbaren Kameras aufgezeichnet. Der neue Ansatz sieht vor, dass der Versuchsleiter neben dem Probanden sitzt und jederzeit Fragen stellen darf. Die Atmosphäre soll so angenehmer und entspannter gestaltet, um das natürliche Verhalten der User im Umgang mit dem zu testenden Produkt aufzuzeichnen. Diese Methodenkombination stellt in diesem Paper jedoch lediglich einen Vorschlag dar und wird nicht experimentell mit anderen Methoden verglichen, weshalb die angestrebten Ziele nicht empirisch belegt werden können. Allerdings handelt es sich um einen guten Ansatz, der in zukünftigen Forschungstätigkeiten implementiert werden kann.

Erstellt von: Theresa Jarchow (309960)

Autor(en):Norgaard, M., Hornbaek, K.

Titel:What do usability evaluators do in practice? An explorative study of think-aloud testing

Jahr:2006

Laufende Nr. im Literaturverzeichnis: 19

Schlagwörter: Usability Evaluation, ThinkingAloud, User-zentriertes Design, praktische Umsetzung

Kurzbeschreibung: In diesem Paper werden die Ergebnisse einer Analyse von

Audioaufnahmen aus 14 ThinkingAloud Sessions vorgestellt, mit denen die tatsächliche Umsetzung des Verfahrens in der Praxis untersucht wurde. Zur Auswertung wurden die Inhalte der Verbalisierungen in sechs Bereiche aufgeteilt: Analyse der Testergebnisse, Fokus auf die Bestätigung bekannter Probleme, Einflussfaktoren während des Tests, gestellte Fragen während des Tests, Einhaltung wissenschaftlicher Standards und Aufdeckung von Usability Problemen vs. Brauchbarkeit.

Dabei zeigte sich, dass eine zeitnahe Auswertung der Beobachtungen (wenn die Erinnerungen noch sehr frisch sind) nur sehr selten, manchmal auch gar nicht, vorgenommen wird. Die meisten Versuche wurden hinsichtlich des Aufgaben-Designs und der Fragestellungen dazu ausgelegt, bereits bekannte Usability Probleme zu bestätigen. Hypothetische und abstrakte Fragen der Versuchsleiter zielten eher auf erwartete Probleme statt auf tatsächlich erfahrene Probleme mit dem System ab. In vielen Untersuchungen wurde der Workflow der Probanden zudem durch technische Probleme oder störende Umgebungen unterbrochen. Oft bereitete auch eine unzureichende Planung der Stichprobengröße Probleme.

Die Studie zeigt, dass die theoretische Vorgehensweise eines ThinkingAloud Tests, wie sie in der Literatur beschrieben wird, relativ selten in der Praxis umgesetzt wird. Der Fokus während des Versuchs lag in den meisten Fällen zu stark auf bereits bekannte Probleme. Zudem wurden die Erfahrungen der Testuser zu wenig beachtet und ausgewertet. Die Testleiter priorisierten dabei eher Usability-Probleme auf Kosten von Utility-Problemen. Mithilfe dieser explorativen Studie können die Autoren schließlich erhebliche Defizite des praktischen Vorgehens von Usability Experten aufdecken. Auch leiten sie daraus Implikationen für zukünftige Arbeiten ab.

Dennoch lassen sich die Ergebnisse dieser Studie nur bedingt verallgemeinern. Für die Untersuchung wurden sieben dänischen Unternehmen herangezogen, welche Usability Evaluationen entweder als Beratung oder für die Integration in die Systementwicklung betrieben. Davon verwendeten drei Unternehmen ihre Forschungsergebnisse für auswärtige Kunden, während der Rest die Tests für interne Auftraggeber durchführte. Dies kann dazu beitragen, dass die Durchführung der Tests durch die Erwartungen und Meinungen der Usability Experten stark beeinflusst wird.

Erstellt von: Theresa Jarchow (309960)

Autor(en): Obrist, M., Bernhaupt, R., Beck, E., Tscheligi, M.

Titel: Focusing on elderly: an iTV usability evaluation study with eye-tracking

Jahr:2007

Laufende Nr. im Literaturverzeichnis: 20

Schlagwörter: interaktives TV, Usability, Eye- Tracking, Thinking Aloud

Kurzbeschreibung: In dieser Studie wird die Benutzerfreundlichkeit eines interaktiven Fernsehsystems untersucht. Ausgangspunkt ist die Beobachtung, dass es älteren Menschen immer schwerer fällt mit der schnellen Entwicklung von technischen Systemen hinterher zu kommen und sie häufig Probleme haben, die innovative Technik zu bedienen. Mithilfe einer Usability Studie sollte das Design und die Navigation eines interaktiven Fernsehsystems (*iTV*) evaluiert werden. Neben der untersuchten Benutzerfreundlichkeit war es ein weiteres Studienziel mögliche Gründe für die Schwierigkeiten älterer User im Umgang mit solchen Systemen zu analysieren. Zum Experiment wurden deshalb zwei unterschiedliche Altersgruppen gebildet, bestehend aus 20-30-jährigen und über 50jährigen Testpersonen (N = 16). Zur Erhebung der Daten wurde neben der ThinkingAloud Methode auch ein Eye- Tracking Verfahren eingesetzt sowie nachträgliche Befragungen durchgeführt.

Während des Experiments nahmen die Testpersonen vor dem TV-Gerät Platz und erhielten insgesamt sechs Aufgaben zur Bearbeitung. Dabei sollten sie alle Gedanken und Eindrücke aussprechen. Nach jeder Aufgabe bewerteten die Testpersonen dann die Schwierigkeit der Bewältigung auf einer Skala von 1 (leicht) bis 5 (sehr schwierig). Des Weiteren wurden sie um ein kurzes Feedback zur Verbesserung der Technik gebeten. Zum Schluss füllten alle Probanden einen Fragebogen aus, mit dem die User-Experience erfasst wurde, und beantworteten allgemeine Fragen über ihre Eindrücke vom System. Zusätzlich wurden die Anzahl der gelösten Aufgaben, klassifiziert nach gelöst/mit Hilfe gelöst/ ungelöst, und die dafür benötigte Zeit erhoben. Der Versuchsleiter durfte helfend eingreifen, wenn die Testpersonen länger als eine Minute keine richtige Lösung fanden und Hinweise geben.

Die älteren Probanden hatten, verglichen mit den jüngeren Probanden, insgesamt mehr Probleme damit die Aufgaben selbstständig und korrekt zu lösen. Die Blickverläufe verrieten, dass sie außerdem Probleme hatten die Bedeutung verschiedener Buttons oder die Informationsseite zu verstehen. Auch gestalteten sich die Schriftgröße und die Farbgebung des Interfaces ungünstig für die älteren User. Die Auswertung der benötigten Zeit indizierte, dass die betagteren Testpersonen, verglichen mit der jüngeren Stichprobe, stets langsamer hinsichtlich ihres Blick-, Such- und Identifikationsverhaltens waren. Sie hatten zudem erhebliche Schwierigkeiten mit der Navigation.

Die Auswertung der erhobenen Daten legt also nahe, dass die Gruppe der älteren

Probanden (> 50 Jahre) mehr Probleme im Umgang mit dem interaktiven Fernsehsystem aufwies, als die jüngere Gruppe. Bei der Auswertung wurde jedoch nicht beachtet, dass diese Gruppe ausschließlich aus Personen bestand, die wenig bis keine PC-Kenntnisse aufwies oder zuvor bereits mit solchen interaktiven Systemen gearbeitet hat. Die jüngeren Probanden hingegen wiesen mehrheitlich diesbezügliche Kenntnisse auf, was ihnen erhebliche Vorteile bei der Aufgabenbearbeitung im Experiment einbrachte. Die gefundenen Effekte können also nur bedingt mit Altersunterschieden begründet werden. Spezifische Usability Probleme (z.B. Schriftgröße, Farbgebung) können so allerdings identifiziert und für eine Software-Entwicklung unter Einbezug älterer Zielgruppen berücksichtigt werden.

Das Eye-Tracking Verfahren stellte sich zudem als hilfreiche Methode heraus, um die verbalen Informationen der Probanden zu vervollständigen und die Interpretation zu erleichtern.

Erstellt von: Theresa Jarchow (309960)

Autor(en): Zhao, T., McDonald, S., Edwards, H. M.

Titel: The impact of two different think-aloud instructions in a usability test: a case of just following orders?

Jahr: 2012

Laufende Nr. im Literaturverzeichnis: 26

Schlagwörter: Usability Test, Thinking Aloud, Instruktion

Kurzbeschreibung: In diesem Paper wurde untersucht, inwiefern explizite Instruktionen zum Verbalisieren von Erfahrungen, Erwartungen und Gefühlen, verglichen mit der allgemeinen Aufforderung zum lauten Denken, einen Effekt auf das Verhalten hat. Dazu wurden die Ergebnisse beider Vorgehensweise miteinander verglichen.

Am Experiment nahmen insgesamt 16 Probanden teil. Mithilfe des ThinkingAloud Verfahrens sollte eine Webseite zur Reisebuchung evaluiert werden. Für die Aufgaben wurden insgesamt fünf Szenarien präsentiert, für die es jeweils eine richtige Lösung gab (z.B. für ein bestimmtes Datum den preiswertesten Flug nach Paris finden). Zur Untersuchung des Effekts der unabhängigen Variable Instruktion wurden zwei Gruppen gebildet. Die Instruktion der ersten Gruppe war an das klassische Vorgehen von Ericsson und Simon angelehnt und erforderte lediglich das Verbalisieren der eigenen Gedanken (*neutrale Bedingung*). Die Probanden aus der zweiten Gruppe wurden gebeten, während der Aufgabenbearbeitung ihre Gedanken auszusprechen und außerdem zu berichten, was sie mit ihren jeweiligen Verhalten erwarteten, ob diese Erwartungen erfüllt wurden und

was sie dabei empfanden (*Instruktions-Bedingung*).

Als abhängige Variablen wurden die Aufgabenbearbeitung (benötigte Zeit, Erfolg, Abbruchrate), Verbalisierungen (Qualität und Quantität der Äußerungen, Anzahl von Hinweisen zur Verbalisierung), Usability Probleme (Schweregrad, Ursache, Art des Problems) ausgewertet. Zusätzlich bearbeiteten die Probanden nach der Aufgabenbearbeitung einen workload- Fragebogen (*NASA`s-TLX*) und einen Likert-Skala-basierten Fragebogen, um den Fokus ihrer verbalisierten Gedanken zu verdeutlichen.

Die Auswertung der Daten ergab zunächst keine Effekte der variierten Instruktion auf die Qualität der Aufgabenbearbeitung. Jedoch wirkte sich die Instruktion negativ auf die mentale Belastung (workload) aus, indem die Probanden aus dieser Gruppe über mehr workload berichteten. Bezüglich der anderen Faktoren erwies sich die Instruktions-Bedingung jedoch als überlegen. Es wurden mehr Äußerungen über die eigenen Erfahrungen und Erwartungen gemacht, was sich wiederum positiv auf die Anzahl der identifizierten Usability Probleme auswirkte.

Die Ergebnisse weisen also darauf hin, dass die explizite Instruktion zur Verbalisierung von Erwartungen, Erfahrungen und Gefühlen, verglichen mit der allgemeinen Instruktion einfach nur laut zu denken, einen Effekt auf die Qualität und die Quantität der erhobenen Daten hat und zu mehr Informationen führt. Auch wenn dies auf Kosten eines erhöhten mental workload der Probanden geschieht.

In dieser Studie werden das Studiendesign sowie der Analyse-Prozess sehr detailliert beschrieben. Auch werden aus den Ergebnissen Implikationen für die zukünftige Forschung abgeleitet.

Eventuell ist die Verwendung von unabhängigen Stichproben kritisch zu sehen.

Insbesondere aufgrund der kleinen Stichprobengröße können die Ergebnisse nur bedingt verallgemeinert werden.

Erstellt von: Theresa Jarchow (309960)

Autor(en): Nielsen, J.

Titel: Estimating the number of subjects needed for a thinking aloud test

Jahr:1994

Laufende Nr. im Literaturverzeichnis: 18

Schlagwörter: ThinkingAloud, Stichprobengröße, Usability Test, Usability Probleme

Kurzbeschreibung: In dieser Studie wird anhand von zwei durchgeführten

Experimenten gezeigt, dass ein ThinkingAloud Test keine große Anzahl von Versuchspersonen benötigt, um viele Usability Probleme zu identifizieren. Zwei Gruppen unterschiedlicher Größe (24 und 30 Versuchspersonen) explorierten jeweils ein System und sollten dabei auftretende Probleme verbalisieren. Ein anwesender Versuchsleiter notierte die genannten Probleme. Die kleinere Gruppe konnte insgesamt neun, die größere Gruppe entdeckte 14 Probleme im Umgang mit dem Interface identifizieren. Anhand der Proportion gefundener Usability Probleme zeigte sich, dass für das erste System durchschnittlich 3.2 Probanden, für das zweite System durchschnittlich 3.7 Probanden nötig waren, um ein Problem zu identifizieren. Die Auswertung der Daten zeigte, dass sich mit dem Anstieg der Stichprobengröße zwar die Anzahl gefundener Probleme vergrößerte, dieser Anstieg jedoch mit zunehmender Probandenzahl abflachte. Während also beispielsweise mit einem zweiten Proband 20% mehr Probleme gefunden wurden, resultierte mit einem sechsten Probanden nur noch 5% zusätzlicher Informationsgewinn. Anhand dieser Proportionen lässt sich die benötigte Stichprobengröße für ein Experiment gut abschätzen. Sollten zum Beispiel die beiden letzten Testpersonen keine neuen Probleme mehr finden, so kann man davon ausgehen, dass bereits alle Probleme identifiziert wurden. Zur Abschätzung der Stichprobengröße kann also entweder die durchschnittliche Fehleranzahl pro Proband (Anzahl aller gefundenen Probleme durch Anzahl der Probanden) berechnet werden oder man teilt die Anzahl neuer Probleme durch die Anzahl aller Probanden. Dabei erweist sich, nach Ermittlung der Standardabweichungen, das letztere Vorgehen als bessere Methode zur Schätzung.

Das durchgeführte Experiment zeigt, dass mit bereits fünf Probanden 77-85% der Usability Probleme identifiziert werden können. Kritisch zu bemängeln ist dabei allerdings, dass die Experimentatoren in dieser Studie alle sehr unerfahren waren und teilweise das erste Mal einen solchen Test durchführten, was unter Umständen die Ergebnisse verfälschen kann.

Abschließend werden weitere Gründe aufgeführt, die die Verwendung kleiner Stichproben für einen ThinkingAloud Test plausibel machen.

Erstellt von: Theresa Jarchow (309960)