

**Lehrstuhl für Wirtschaftsinformatik
Insbesondere Systementwicklung**

Prof. Dr. Werner Mellis

**Hauptseminar Wirtschaftsinformatik
im Sommersemester 2006**

Thema Nummer 7

Usability

Vorgelegt von:

Wittek, Thomas
Hildebrandstr. 1
50739 Köln
Tel.: 0221 / 49 25 121
Matrikel-Nr. 3776263

16. Mai 2006

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Abbildungsverzeichnis	III
1 Einleitung	1
2 Definition der Usability	2
2.1 Problematik der Definition der Usability	2
2.2 Definitionen und Modelle	3
2.2.1 Shackel (1991)	3
2.2.2 Nielsen (1993)	3
2.2.3 ISO 9241-11 (1998)	4
2.2.4 ISO 9126-1 (2001)	5
3 Methoden zur Messung der Usability	8
3.1 Klassifizierung der Methoden	8
3.2 Beschreibung ausgewählter Methoden	8
3.2.1 Testing: Thinking Aloud	9
3.2.2 Inspection: Heuristic Evaluation	10
3.2.3 Inquiry: Questionnaires	11
4 Probleme bei der Messung der Usability	12
4.1 Ursachen für die Probleme	12
4.2 Beschreibung ausgewählter Probleme	13
4.2.1 User Effect	13
4.2.2 Evaluator Effect	14
4.2.3 Task-Selection Bias	14
5 Fazit	16
Thesenpapier	18
Literaturverzeichnis	21

Abkürzungsverzeichnis

CW	–	Cognitive Walkthrough
HE	–	Heuristic Evaluation (Heuristische Beurteilung)
ISO	–	International Standards Organization
PUTQ	–	Purdue Usability Testing Questionnaire
QUIS	–	Questionnaire for User Interaction Satisfaction
SUMI	–	Software Usability Measurement Inventory
UI	–	User Interface (Benutzungsoberfläche/-schnittstelle)

Abbildungsverzeichnis

Abb. 2-1: Shackels Definition der Usability	3
Abb. 2-2: Niensens Definition der Usability	4
Abb. 2-3: Usability in ISO 9241-11	5
Abb. 2-4: Definitionsrahmen der Usability in ISO 9241-11	5
Abb. 2-5: Verschiedener Umfang der Softwarequalität	6
Abb. 2-6: Kommentiertes Qualitätsmodell nach ISO 9126-1	7
Abb. 3-7: Einordnung einiger Methoden zur Beurteilung der Usability in ein Klassifikationsschema	9
Abb. 4-8: Binomialmodell der Fehlerfindung für $p = 0.3$, $p = 0.2$ und $p = 0.1$	13

1 Einleitung

Usability hat eine große praktische Bedeutung. Sie gilt als wesentlicher Faktor bei der Beurteilung der Softwarequalität.¹

Insbesondere die schon sehr breite und weiterhin wachsende Nutzung von Computersystemen durch Personen ohne eine technische Ausbildung macht eine einfache Benutzung dieser Systeme zu einem wichtigen Qualitätskriterium². Dementsprechend stellt sie einen wesentlichen Wettbewerbsfaktor dar.

Zudem führt die zunehmende Verbreitung des Internets zu einer stärkeren Dezentralisation der Anwendungen, was einen Zugriff der Entwickler auf die Benutzer wesentlich erschwert oder gar unmöglich macht und so Prüfungen der Usability im Vorfeld zwingend erforderlich werden lässt.³ Um zielgerichtet eine gute Usability der zu entwickelnden Anwendung⁴ zu erreichen, ist es erforderlich, die Usability messen zu können.

Jedoch ist die Durchführung der Usability-Messung in der Praxis nicht unproblematisch.⁵ Deswegen müssen zur geeigneten Gestaltung der Messung der Usability die möglichen Probleme verstanden und bei der Methodenauswahl sowie bei der Durchführung der Messung beachtet werden.

Ziel dieser Arbeit ist es, grundlegende Konzepte der Usability zu skizzieren, die Methoden zur Messung voneinander abzugrenzen sowie bei der Messung auftretende Probleme zu beschreiben.

¹Vgl. Abran, Khelifi, Suryn /Usability Iso/ 325

²Vgl. Lin, Choong, Salvendy /Proposed Index/ 267

³Vgl. Abran, Khelifi, Suryn /Usability Iso/ 325

⁴Unter dem Begriff *Anwendung* kann man klassische Software, aber auch eine Website verstehen

⁵Vgl. Kapitel 4

2 Definition der Usability

2.1 Problematik der Definition der Usability

Über die Definition des Begriffs *Usability* herrscht keine Einigkeit in der Literatur. Obwohl im Bereich der Usability schon seit langer Zeit aktiv geforscht wird, existiert auch heute immer noch keine einheitliche Definition des Begriffs.⁶ Dies führt bei potenziellen Anwendern von Usability-Maßnahmen zu Verwirrung und Orientierungslosigkeit, da die vielen existierenden Definitionen es schwer machen, sich ein klares Bild von diesem Bereich zu verschaffen.

Der Grund für die mangelnde Konsistenz des Verständnisses des Begriffs ist die Tatsache, dass Usability ein sehr breites Spektrum an Konzepten und Methoden abdecken muss, wobei jeder Autor dieses Spektrum weiter oder enger fasst und den Fokus anders setzt. Durch die verschiedenen Möglichkeiten und Ziele der Usability-Messung ergeben sich unterschiedliche Sichtweisen des Konzepts.

Da Usability ein so weit gefasster Begriff ist, wird er meistens über konkretere Begriffe und Konzepte definiert.⁷ Hier kann man leicht erkennen, wo der Schwerpunkt der Autoren in deren Definitionen und den darauf aufbauenden Konzepten besteht. Einige Autoren verzichten aufgrund der Mehrdimensionalität des Begriffs gar ganz auf eine prägnante Definition und beschreiben Usability unmittelbar und explizit über weitere Begriffe.⁸

Eine wichtige Maßnahme zur Konsolidierung des Begriffs der Usability stellt die Entwicklung internationaler Standards dar, die (teilweise) die Usability behandeln. Die im Softwarebereich populärsten Standards sind hier sicherlich ISO 9126 und ISO 9241 (insb. Teil 11), die in Kapitel 2.2 vorgestellt werden. Auch hier zeigen sich jedoch wieder verschiedene Konzepte in den Standards, die es erschweren, sich ein konsistentes Bild der Usability zu verschaffen. Dieses Problem nahmen sich Abran, Khelifi und Surin zum Anlass, die Bedeutungen und Interpretationen der Usability in ISO-Standards zu untersuchen. Sie entwickelten ein integriertes Modell, das zur Berücksichtigung bei der weiteren Entwicklung der ISO-Standards vorgeschlagen wird.⁹

Da sich keine einheitliche Definition der Usability durchgesetzt hat, werden im nächsten Abschnitt einige populäre Modelle und mit ihnen verbundene Definitionen der Usability dargestellt.

⁶Vgl. Seffah, Metzker /Usability and SWE/ 72 sowie Bevan /Quality of Use/ 115

⁷Vgl. zu diesem Absatz Kapitel 2.2

⁸Vgl. Kapitel 2.2.2

⁹Vgl. Abran, Khelifi, Suryn /Usability Iso/ 332-337

2.2 Definitionen und Modelle

2.2.1 Shackel (1991)

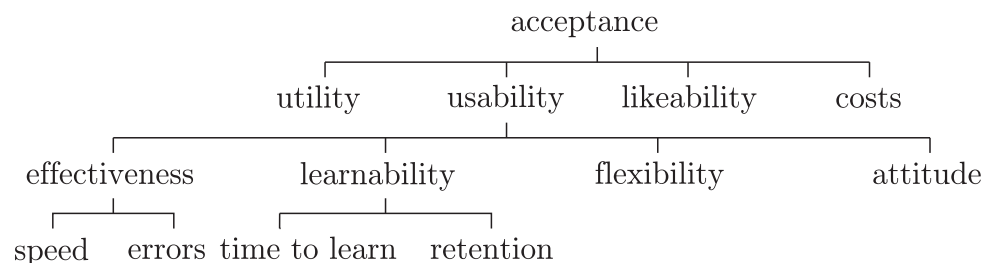
Shackel definiert Usability als

The capability in human functional terms to be used *easily* and *effectively* by the *specified range of users*, given *specified training and user support*, to fulfil the *specified range of tasks*, within the *specified range of environmental scenarios*.¹⁰

Die Bewertung der Usability eines Systems ist nach Shackel vom Kontext, insb. von den Benutzern, abhängig. Somit ist die Wahrnehmung eines Systems stets subjektiv. Er betont aber auch die Existenz objektiver Maße der Interaktion. Diese Bivalenz der Messung der Usability beschreiben Nielsen und Levy später auch ausführlich in einem Artikel, der diesen Sachverhalt zum Thema hat.¹¹

Shackel ordnet die Usability als eine Dimension der Produktakzeptanz ein, die sich relativ zu den anderen Dimensionen verhält und in gewisser Weise in Konkurrenz mit ihnen steht (insb. Kosten). Abb. 2-1 verdeutlicht den konzeptionellen Rahmen seines Verständnisses der Usability.

Abb. 2-1: Shackels Definition der Usability



Er war einer der ersten Autoren, der die Bedeutung und die Kontextabhängigkeit der Usability ausgedrückt hat. Sein Ansatz wurde breit akzeptiert und weiterentwickelt. Das zeigt sich z.B. in den Parallelen, die das im nächsten Abschnitt dargestellte Modell von Nielsen zu Shackels Definition zieht.

2.2.2 Nielsen (1993)

Auch Nielsen hebt als übergeordnetes Kriterium die Produktakzeptanz hervor und übernimmt einige Komponenten aus Shackels Definition, führt aber auch einige Änderungen ein.

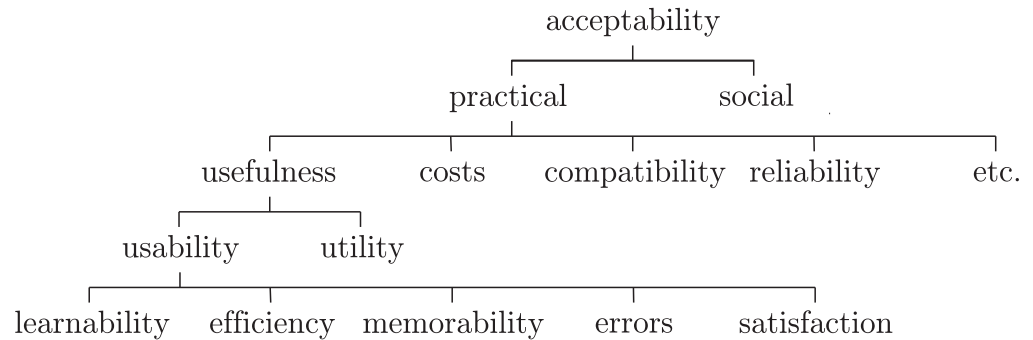
¹⁰Vgl. Shackel, Richardson /Human Factors Usability/ 24

¹¹Vgl. Nielsen, Levy /Preference vs. Performance/

Er verzichtet explizit auf eine Definitionsphrase und betont die Mehrdimensionalität der Usability.¹² Für ihn wird Usability durch die Komponenten¹³ *Learnability*, *Efficiency*, *Memorability*, *Errors* und *Satisfaction* determiniert, die er auch genau definiert.

Seine Definition der Usability wird auch heute noch als aktuell angesehen.¹⁴

Abb. 2-2: Niensens Definition der Usability



2.2.3 ISO 9241-11 (1998)

ISO 9241 ist eine Sammlung 17 internationaler Standards bzgl. der Ergonomie der Büroarbeit an Bildschirmsystemen. Teil 1 bis 9 beschreiben, neben allgemeinen Grundlagen, Empfehlungen für die physischen Schnittstellen („Hardware“) sowie Anforderungen an die Arbeitsumgebung. Die Teile 10 bis 17 legen Anforderungen an die Software fest.

Im Kontext der Usability ist insb. ISO 9241 Teil 11¹⁵ relevant. Usability wird dort definiert als:

The extent to which a product can be used by *specified users* to achieve *specified goals* with *effectiveness*, *efficiency* and *satisfaction* in a *specified context of use*.

Auch diese Definition setzt sich wesentlich aus weiteren Dimensionen zusammen, wie Abb. 2-3 verdeutlicht.

ISO 9241-11 beschreibt, wie die Informationen identifiziert werden können, die für die Spezifikation oder Beurteilung der Usability bzgl. dessen Messung nötig sind.¹⁶

¹²Vgl. Nielsen /Usability Engineering/ 26

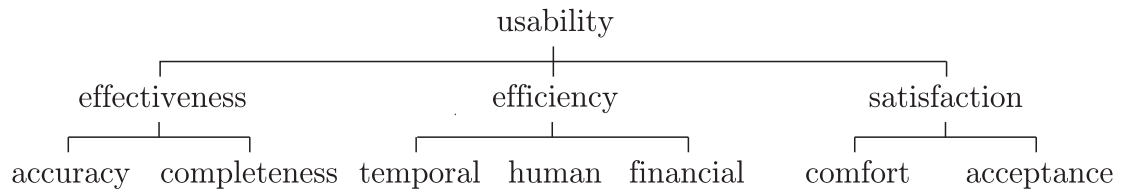
¹³Vgl. Abb. 2-2

¹⁴Vgl. Holzinger /Usability Engineering Methods/ 72

¹⁵Vgl. ISO /9241-11/

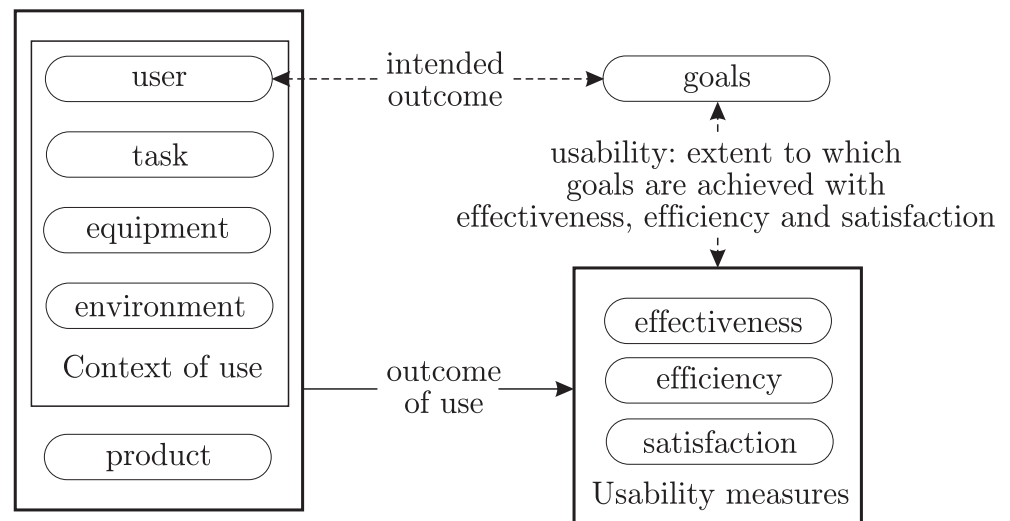
¹⁶Vgl. Bevan /International Standards/ 536

Abb. 2-3: Usability in ISO 9241-11



Die Kontextabhängigkeit der Messung der Usability wird stark betont: Es wird nicht das isolierte Produkt, sondern die Interaktion zwischen Benutzer und dem System in einem bestimmten Nutzungskontext untersucht.¹⁷ Abb. 2-4 veranschaulicht u.a. diesen Sachverhalt.

Abb. 2-4: Definitionsrahmen der Usability in ISO 9241-11



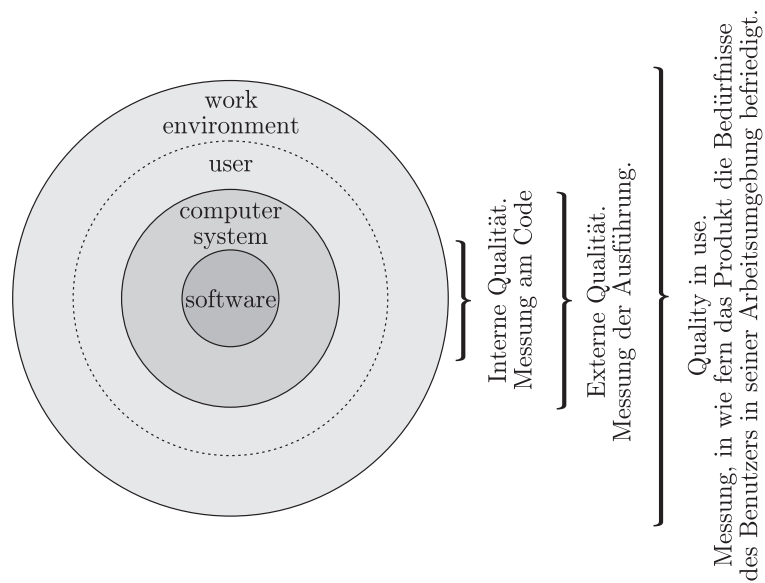
2.2.4 ISO 9126-1 (2001)

ISO 9126 ist ein aus 4 Teilen bestehender internationaler Standard zur Beurteilung der gesamten Softwarequalität, wovon die Usability nur eine Komponente darstellt. Teil 1 beschreibt das der Beurteilung zu Grunde liegende Qualitätsmodell, die Teile 2-4 stellen interne und externe Beurteilungsmaße sowie Maße zur Beurteilung der sog. *Quality in Use* dar. In ISO 9126 wird die Softwarequalität auf mehreren Ebenen beschrieben, die den Begriff enger oder weiter gefasst auslegen, wie Abb. 2-5¹⁸ verdeutlicht.

¹⁷Vgl. Avouris /Introduction/ 2

¹⁸Angelehnt an Bevan /Meeting User Needs/ 91

Abb. 2-5: Verschiedener Umfang der Softwarequalität



Usability wird im Qualitätsmodell¹⁹ nach ISO 9126-1²⁰ in einem recht schmalen Kontext als relativ unabhängige Komponente der Softwarequalität verstanden und ist nur für die interne und externe Messung der Qualität relevant. Usability wird dort als

The capability of the software product to be *understood, learned, used* and *attractive* to the user, when used *under specified conditions*.

definiert. Usability gliedert sich nach ISO 9126-1 also in die Dimensionen *Understandability, Learnability, Operability* und *Attractiveness*. Der Zusatz „*under specified conditions*“ soll betonen, dass Usability keine absolute Produkteigenschaft ist, sondern die Fähigkeit des Produktes, in einem bestimmten Kontext genutzt werden zu können.²¹

Ein weiter gefasstes Qualitätsverständnis, das den gesamten Nutzungskontext beinhaltet und das übergeordnete Ziel der Befriedigung der Benutzerbedürfnisse durch die Software verfolgt, wird in ISO 9126-1 als *Quality in Use*²² definiert:

The capability of the software product to enable *specified users* to achieve *specified goals* with *effectiveness, productivity, safety* and *satisfaction* in *specified contexts of use*.

¹⁹Vgl. Abb. 2-6

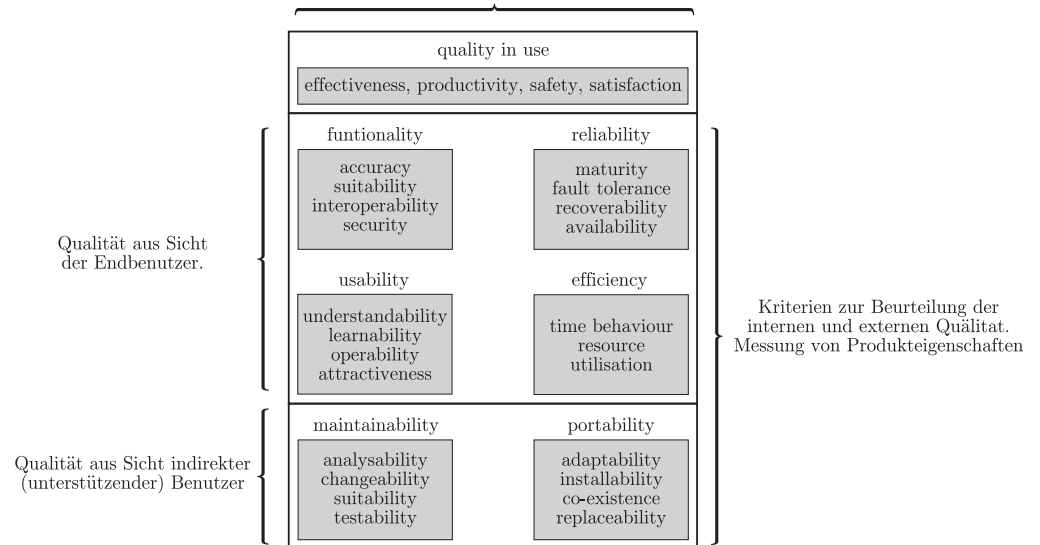
²⁰Vgl. ISO /9126-1/

²¹Vgl. Bevan /International Standards/ 537

²²Vgl. auch Bevan /Quality of Use/ 118-119

Abb. 2-6: Kommentiertes Qualitätsmodell nach ISO 9126-1

Quality in use ist die kombinierte Wirkung aller sechs Qualitätskategorien während das Produkt in Benutzung ist. Messung des Ergebnisses.



Diese Definition hat eine starke Ähnlichkeit mit der Definition der Usability in ISO 9241-11, ergänzt sie jedoch um die Komponente *Safety*, die in vielen anderen Standards und Definitionen enthalten²³ ist.

²³Vgl. Abran, Khelifi, Suryn /Usability Iso/ 333

3 Methoden zur Messung der Usability

3.1 Klassifizierung der Methoden

Man kann die Methoden zur Messung der Usability nach verschiedenen Gesichtspunkten charakterisieren und klassifizieren. Ein verbreitetes Unterscheidungskriterium ist die Teilnahme von Benutzern an der Messung.²⁴ Methoden, an deren Durchführung Benutzer teilnehmen, werden unter den Begriffen *Usability Testing* oder *Empirical Evaluation* zusammengefasst. Hierbei werden i.d.R. einige ausgewählte Benutzer bei der Ausführung bestimmter Aufgaben an Prototypen oder den tatsächlichen Anwendungen unter Laborbedingungen beobachtet, um Usability-Probleme zu entdecken. Diese Methoden haben den Vorteil, dass die Informationen ohne Umwege direkt bei der Benutzung erhoben werden, und dass sie als dynamischer Test mit Endbenutzern recht realitätsnah sind.

Methoden, die allein durch sog. Evaluatoren (üblicherweise Usability-Experten, evtl. geschulte Softwareentwickler) durchgeführt werden, werden der Klasse *Usability Inspection* oder *Analytic Evaluation* zugeordnet. Dabei wird das Design der Schnittstelle (also durchaus auch vor Implementierung) durch die Evaluatoren gegen bestimmte Standards, Richtlinien oder Checklisten geprüft, um mögliche Usability-Probleme zu identifizieren. Diese Methoden haben die positive Eigenschaft, schon in frühen Entwicklungsphasen anwendbar, relativ schnell durchführbar und somit kostengünstig zu sein.²⁵ Nachteilig ist die fehlende Involvierung von Endbenutzern. Es werden Vermutungen über die Anforderungen und Arbeitsweisen der Benutzer angestellt.²⁶

Zusätzlich wird oft²⁷ eine Klasse *Inquiry* genannt, in der die Benutzer zwar auch eingebunden werden, die sich jedoch stark vom *Testing* unterscheidet. Die Benutzer werden nicht durch Evaluatoren beobachtet, sondern meist mittels Fragebögen oder Interviews über die Anwendung befragt. Solche Methoden haben meist den Zweck, subjektive Eindrücke wie die Zufriedenheit mit dem System zu erfassen.

In Abb. 3-7 wird die beschriebene Klassifizierung verdeutlicht und es werden einige Methoden in dieses Schema eingeordnet.

3.2 Beschreibung ausgewählter Methoden

Im Folgenden wird je Klasse eine typische Methode kurz beschrieben.

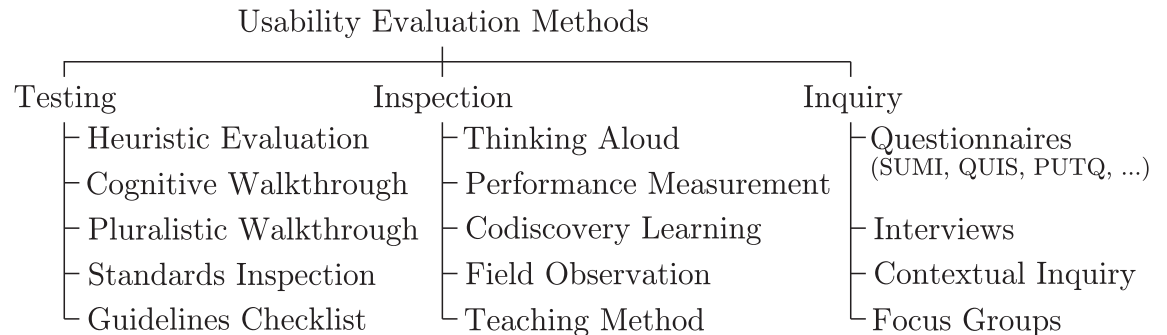
²⁴Vgl. Holzinger /Usability Engineering Methods/ 72 und Zhang /Perspective/ 43

²⁵Vgl. Mack, Nielsen /Usability Inspection/ 28-29

²⁶Vgl. Holzinger /Usability Engineering Methods/ 72-73

²⁷z.B. in Folmer, Bosch /Survey/ 69 und Avouris /Introduction/ 519

Abb. 3-7: Einordnung einiger Methoden zur Beurteilung der Usability in ein Klassifikationsschema



3.2.1 Testing: Thinking Aloud

Der Kerngedanke des *Thinking Aloud* ist die Beobachtung von Benutzern (empfohlen werden 4-5), die in Einzelsitzungen laut ihre Gedanken bei der Ausführung bestimmter Aufgaben äußern.²⁸ Das Laute Denken ermöglicht es dem Evaluator, zu verstehen, wie das System aus Sicht der Benutzer verstanden wird. Missverständnisse und Probleme können so direkt identifiziert werden. Der Beobachter erfährt nicht nur, *was* der Benutzer tut, sondern auch *warum* er es tut.

Dieser Technik wird oft die größte Bedeutung bei der Beurteilung der Usability zugesprochen.²⁹ Sie wird in der Literatur unter vielen verschiedenen Namen³⁰ beschrieben und es wurde eine Vielzahl von Varianten entwickelt³¹.

Die Varianten reichen von der aufwändigen Evaluierung durch Usability-Spezialisten in Laboren bis hin zur wesentlich einfacheren Durchführung beim Endbenutzer durch einen Software-Entwickler. Die Labore bestehen üblicherweise aus zwei durch Einwegspiegel getrennten Räumen für Evaluatoren und Benutzer. Die Handlungen werden dort auf Video aufgezeichnet und nachträglich ausgewertet. Bei der Durchführung beim Endbenutzer werden die Erkenntnisse unmittelbar notiert. Es ist kein Labor nötig und die Handlungen werden nicht weiter aufgezeichnet.

Wesentliche Vorteile sind die Unmittelbarkeit, die Tendenz, besonders schwerwie-

²⁸Vgl. zu diesem Absatz Nielsen /Usability Engineering/ 195-198; Holzinger /Usability Engineering Methods/ 73-74 sowie Hertzum, Jacobsen /Evaluator Effect/ 186-187

²⁹Vgl. Holzinger /Usability Engineering Methods/ 73 sowie Nielsen /Usability Engineering/ 195

³⁰„verbal reports, concurrent verbal protocols, retrospective verbal protocols, after think aloud and verbal protocols“ – Zitat aus Nielsen, Clemmensen, Yssing /Reflections Think-Aloud/ 102)

³¹Vgl. Hertzum, Jacobsen /Evaluator Effect/ 186 sowie Nielsen, Clemmensen, Yssing /Reflections Think-Aloud/ 102-103

gende Fehler zu finden und die Möglichkeit, objektive (z.B. Geschwindigkeit) und subjektive (z.B. Zufriedenheit) Daten gleichzeitig zu erheben.³² Von Nachteil ist es, dass den Benutzern die unnatürliche Situation, die Gedanken laut zu äußern, schwer fällt. Zudem ist diese Technik (je nach Variante) sehr zeit- und ressourcenintensiv.

3.2.2 Inspection: Heuristic Evaluation

Die *Heuristic Evaluation* (HE) wurde 1990 von Nielsen und Molich vorgestellt³³ und entspricht einer informellen Inspektion des Systems durch Usability-Spezialisten gegen eine geringe Anzahl von Usability-Prinzipien, den sogenannten *Heuristiken*.

HE ist die informellste Methode zur Beurteilung der Usability.³⁴ Zudem ist sie sehr kostengünstig³⁵. Sie ist Bestandteil des von Nielsen beschriebenen Konzepts „discount usability engineering“³⁶. Die breite Einsetzbarkeit durch die geringe Formalität sowie die hohe Effektivität machen sie zur Inspektions-Methode mit der weitesten Verbreitung.³⁷

Die HE wird auf Grundlage von UI-Attrappen oder Ausführung (eines Prototyps) des Systems durchgeführt. Dabei bewertet eine kleine Gruppe (3-5 Personen) von Evaluatoren die interaktiven Elemente der UI in Hinblick auf eine kleine Menge von Heuristiken und hält gute sowie schlechte Aspekte bzgl. dieser Heuristiken fest. Nielsen und Molich schlagen 9 grundlegende Heuristiken vor, die eine gewisse Ähnlichkeit zu den Prinzipien aus ISO 9241-10³⁸ besitzen. Die Auswahl der Heuristiken ist grundsätzlich von der zu beurteilenden Anwendung abhängig – so werden für Websites üblicherweise andere Usability-Prinzipien gelten als für klassische Anwendungssoftware. Zusätzlich zu diesen allgemeinen Heuristiken können für spezifische UI-Elemente weitere Heuristiken angewandt werden.

Um möglichst verlässliche bzw. unvoreingenommene Ergebnisse zu erhalten, wird die HE zunächst individuell durch jeden Evaluator durchgeführt. Erst im Anschluss werden die Ergebnisse gemeinsam in der Gruppe diskutiert und aggregiert.

³²Vgl. zu diesem Absatz Holzinger /Usability Engineering Methods/ 73 sowie Jeffries, Desurvire /Contest/ 40

³³Vgl. Nielsen, Molich /Heuristic Evaluation/ 25-62; Vgl. auch Nielsen /Heuristic Evaluation/ 373-380; Vgl. auch Hertzum, Jacobsen /Evaluator Effect/ 186

³⁴Vgl. Nielsen /Usability Inspection Methods/ 413

³⁵Vgl. Jeffries u.a. /Comparison/ 123

³⁶Vgl. Nielsen /Usability Engineering/ 16-20

³⁷Vgl. Holzinger /Usability Engineering Methods/ 72

³⁸Vgl. ISO /9241-10/

3.2.3 Inquiry: Questionnaires

Fragebogenbasierte Methoden dienen der indirekten Erhebung von Aussagen über bestimmte Dimensionen der Usability.³⁹ Sie sind besonders zur Erhebung der Zufriedenheit und subjektiver Einschätzungen über mögliche Problemquellen geeignet – also für Informationen, die kaum objektiv erfassbar sind. Am populärsten ist die Erhebung der Zufriedenheit mit dem betrachteten System, obwohl auch durchaus andere Kriterien wie z.B. die Vorliebe für bestimmte Features oder gar die Art der Benutzung bestimmter Funktionen untersucht werden können.

Es existiert eine Vielzahl von vorgegebenen Fragebögen, die unterschiedlichste Zwecke verfolgen. Besonders populär für die Erhebung der Zufriedenheit ist SUMI⁴⁰ – ein international standardisierter Fragebogen bestehend aus 50 Punkten.⁴¹

Vorteilhaft sind die Möglichkeiten, Statistiken über die meist quantitativen Ergebnisse zu erstellen und Vergleiche (z.B. nach Nutzergruppen oder auch im Zeitverlauf) anzustellen.⁴² Die Befragung von großen Gruppen und wiederholte Befragungen sind relativ kostengünstig durchzuführen. Nachteilig ist allerdings die Indirektheit der Methode – man erfährt, was die Leute angeben oder denken, zu tun oder zu wissen, nicht aber was sie tatsächlich tun oder wissen. Dies senkt die Validität der Methode. Außerdem ist die Erstellung der Fragebögen (falls individuell nötig) recht aufwändig, lohnt sich also nur für Umfragen mit vielen Befragten oder wiederholten Befragungen.

³⁹Vgl. zu diesem Absatz Nielsen /Usability Engineering/ 209 sowie Avouris /Introduction/ 519-520

⁴⁰Vgl. Kirakowski, Corbett /SUMI/ zitiert nach Avouris /Introduction/ 519-520

⁴¹Vgl. Avouris /Introduction/ 519-520

⁴²Vgl. zu diesem Absatz Holzinger /Usability Engineering Methods/ 74 sowie Nielsen /Usability Engineering/ 209, 211-212

4 Probleme bei der Messung der Usability

4.1 Ursachen für die Probleme

Die Ergebnisse der Usability-Messung sind stark vom Kontext der Messung abhängig. Der Kontext wird je Methodenklasse durch unterschiedliche, variable Komponenten geprägt: Beim *Testing* sind das im Wesentlichen die an der Beurteilung beteiligten Benutzer, die Evaluatoren, die Aufgaben und die Laborumgebung. Bei der *Inspection* sind es die Evaluatoren. Bei der *Inquiry* bestimmen die befragten Benutzer wesentlich den Kontext. Zusätzlich kann es in jeder Klasse zu Differenzen durch unterschiedliche Auslegungen der Methoden, über deren Definition oft keine Einigkeit herrscht⁴³, kommen.

Die Variabilität dieser Komponenten kann die Zuverlässigkeit sowie die Vollständigkeit der Ergebnisse negativ beeinflussen, wie die weiter unten erläuterten Problembeschreibungen darlegen. Diese Variabilität ist nicht zu vermeiden. Usability ist ein wesentlich vom Menschen abhängiges Konzept und Menschen unterscheiden sich untereinander. Sie haben unterschiedliche Erfahrungen und Hintergründe, was bei den Evaluatoren zu verschiedener Testausführung und Beurteilungen, sowie bei den Benutzern zu unterschiedlicher Aufgabenbewältigung und Präferenzen führt.

Subjektivität bei der Beurteilung ist, bis auf sehr wenige Ausnahmen, wie z.B. der Messung von Arbeitsgeschwindigkeiten, nicht zu vermeiden.⁴⁴ Selbst objektiv erscheinende, systematische Methoden, wie z.B. die Anwendung von Checklisten, unterliegen letztendlich der subjektiven Einschätzung der Evaluatoren, da sie über jeden Punkt der Checkliste entscheiden müssen.

Wenn man die Variabilität der Komponenten nicht ausschließen kann, muss man sie bei der Messung zumindest beachten, um zuverlässige Ergebnisse zu bekommen. Hierzu wurden in der Literatur zahlreiche zu berücksichtigende Probleme diskutiert, von denen im nächsten Abschnitt einige dargestellt werden.

Das Ausmaß der kombinierten Wirkung dieser und anderer Probleme haben Molich u.a. in einer vergleichenden Studie einer internationalen, organisationsübergreifenden Usability-Prüfung verdeutlicht.⁴⁵

⁴³Vgl. Molich u.a. /Comparative Evaluation/ 66 sowie Gray, Salzman /Damaged Merchandise/ 213

⁴⁴Vgl. zu diesem Absatz Baber /Subjective Evaluation/ 1022-1024

⁴⁵Vgl. Molich u.a. /Comparative Evaluation/ 73-74

4.2 Beschreibung ausgewählter Probleme

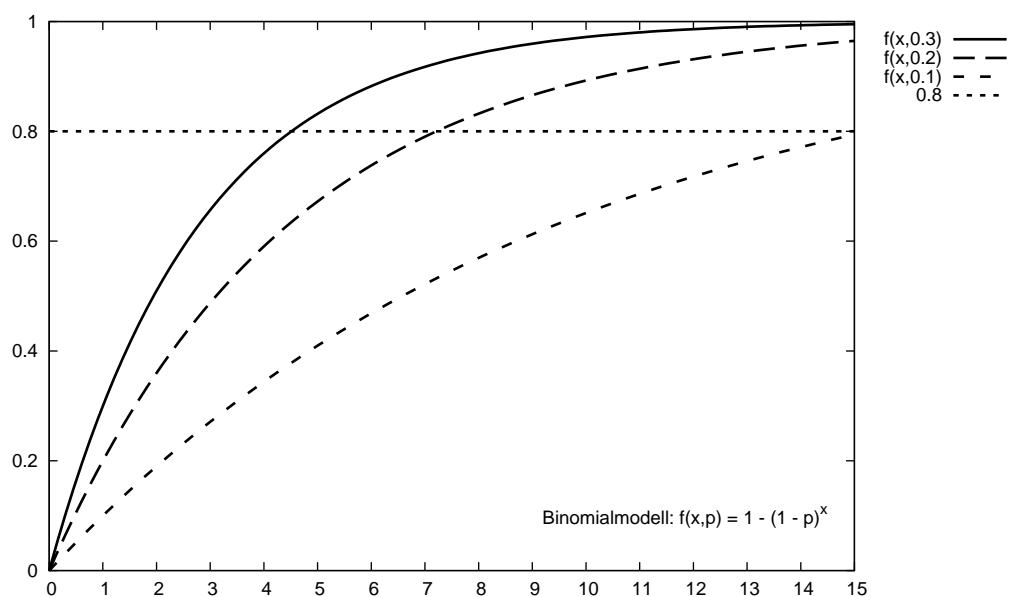
4.2.1 User Effect

Der *User Effect* beschreibt die Unterschiede bei der benutzerindividuellen Fehlerfindung beim Usability Testing.⁴⁶

Usability Testing ist immer eine Stichprobe – man versucht durch den Test mit wenigen Benutzer auf die Probleme aller Benutzer zu schließen. Eine vollständige Prüfung, d.h. das Finden *aller* potenzieller Probleme, ist praktisch nicht möglich. Jedoch nähert man sich mit zunehmender Probandenzahl asymptotisch, also mit abnehmender Anzahl neuer gefundener Probleme je zusätzlichem Probanden, dieser Vollständigkeit.⁴⁷

Aus Kostengründen ist es das Ziel, mit möglichst wenigen Probanden, möglichst viele Probleme zu entdecken. Es stellt sich also die Frage nach der Mindestanzahl an Probanden, um z.B. 80% der Probleme zu entdecken. Virzi fand heraus, dass sich die Anzahl der gefundenen Probleme gut über ein Binomialmodell⁴⁸ schätzen lässt und dass 4-5 Probanden ausreichend sind, um 80% der Probleme zu entdecken.⁴⁹

Abb. 4-8: Binomialmodell der Fehlerfindung für $p = 0.3$, $p = 0.2$ und $p = 0.1$



Problematisch ist jedoch, dass ein wesentlicher Parameter dieses Modells, die *Fehlerfindungsrate* p , nicht immer wie in seiner Untersuchung bei ungefähr 30% liegt,

⁴⁶Vgl. Law, Hvannberg /User Effect/ 9

⁴⁷Vgl. Virzi /How Many Subjects/ zitiert nach Law, Hvannberg /User Effect/ 10

⁴⁸Vgl. Abb. 4-8

⁴⁹Vgl. Virzi /How Many Subjects/ zitiert nach Law, Hvannberg /User Effect/ 10

sondern stark variieren kann. Sie entspricht dem durchschnittlichen Anteil der durch einen Benutzer gefundenen Probleme an allen gefundenen Problemen. Bei einer Verringerung von p erhöht sich die Mindestanzahl an Probanden.⁵⁰

Ein wesentlicher Grund für eine Verringerung von p ist die von Caulton beschriebene Auswirkung der Unterschiedlichkeit der Benutzergruppen.⁵¹ Je mehr unterschiedliche Benutzergruppen es gibt und je unterschiedlicher diese Gruppen sind, desto geringer fällt die Fehlerfindungsrate aus und desto höher wird die benötigte Mindestanzahl an Probanden. Dieser Effekt tritt besonders deutlich bei internationalen Benutzern auf.

Ein weiteres Problem, auf das Lewis 2001 samt Lösungsvorschlägen hinwies, ist die Überschätzung der Fehlerfindungsrate bei kleinen Probandenzahlen.⁵²

4.2.2 Evaluator Effect

Der *Evaluator Effect* beschreibt die Differenz der Ergebnisse zwischen einzelnen Evaluatoren bei Problemdetektion und Beurteilung der Problemschwere beim Einsatz der gleichen Methoden (mit gleichen Benutzern). Er wurde von Jacobsen, Hertzum und John 1998 erstmals⁵³ allgemein für das Usability Testing und 2003⁵⁴ speziell für den Cognitive Walkthrough, die Heuristic Evaluation und die Thinking Aloud-Methode nachgewiesen, sowie ausführlich beschrieben. Der Evaluator Effekt wird auch als Maß für die Zuverlässigkeit, also die Konsistenz der Ergebnisse, verstanden. Je größer der Effekt ausfällt, desto unzuverlässiger sind die Ergebnisse.

Als Ursachen wurden die Notwendigkeit der Interpretation (der Interaktion der Benutzer, der Designs, der Heuristiken etc.) durch die Evaluatoren sowie eine mangelnde Lenkung der Evaluatoren durch die Methoden (unklare Ziele, unklare Bewertungsprozeduren, unklare Problemkriterien) genannt.

Lösungsvorschläge sind die Erhöhung der Anzahl der Evaluatoren⁵⁵, sowie die Schaffung eines gemeinsamen Problemverständnisses zwischen den Evaluatoren (z.B. über die Durchführung der Methoden und die Definition eines Problems).

4.2.3 Task-Selection Bias

Dieses von Cordes beschriebene Problem tritt bei der Auswahl bzw. Definition zu testender Aufgaben beim Usability Testing auf.⁵⁶

⁵⁰Vgl. Lewis /Sample Sizes/ zitiert nach Caulton /Homogeneity Assumption/ 2

⁵¹Vgl. zu diesem Absatz Caulton /Homogeneity Assumption/ 1-7

⁵²Lewis /Problem-Discovery Rates/

⁵³Vgl. Jacobsen, Hertzum, John /Evaluator Effect/

⁵⁴Vgl. Hertzum, Jacobsen /Evaluator Effect/

⁵⁵Beim Usability Testing wird für die besten Ergebnisse eine bestimmte Anzahl an Benutzern durch eine gleich große Anzahl an Evaluatoren in unabhängigen Untersuchungen untersucht.

⁵⁶Vgl. zu diesem Abschnitt Cordes /Task-Selection Bias/ 411-418

Er identifiziert ein oft übersehenes und unterschätztes Problem: Die Neigung der Evaluatoren zur Auswahl tatsächlich erfüllbarer Aufgaben. Wenn die Benutzer wissen, dass die Aufgaben erfüllbar sein müssen, wird ihre Handlung dadurch wesentlich beeinflusst. Sie werden auch bei nicht erfüllbaren Aufgaben viel Zeit investieren, um sie zu erfüllen, was sie außerhalb des Tests nicht tun würden. Der Anteil der erfolgreich erfüllten Aufgaben wird durch die höhere Anstrengung künstlich vergrößert. Im umgekehrten Falle, also wenn man den Benutzern vor dem Test den Hinweis gibt, dass nicht alle Aufgaben erfüllbar sind, zeigte sich in seinem Experiment ein weiterer Effekt: Die Benutzer brechen den Lösungsversuch sehr früh ab, obwohl evtl. eine Lösungsmöglichkeit existiert. Der Anteil erfolgreich erfüllter Aufgaben wird hier übermäßig verringert.

Der Grund für dieses Problem ist laut Cordes, dass das System aus der Perspektive der vorhandenen Funktionen und nicht aus Sicht der Benutzer evaluiert wird. Als Lösung schlägt er vor, zusätzlich benutzerdefinierte Aufgaben zu erstellen und in Kombination mit vom Evaluator definierten Aufgaben zu testen. Dies kombiniert die Vorteile benutzerdefinierter Aufgaben, also die Nähe an den Bedürfnissen und Erwartungen der Benutzer, mit einer gewissen Vollständigkeit aus Sicht der Produktfunktionen.

5 Fazit

Die Problematik der Uneinigkeit über den Begriff *Usability* wurde hervorgehoben und begründet. Gängige Definitionen der Usability wurden dargestellt und erläutert. Methoden zur Messung der Usability wurden abgegrenzt und erklärt. Ursachen für die bei der Messung der Usability auftretenden Probleme wurden erläutert und anhand von in der Literatur dargestellten Problemen konkretisiert.

Usability ist ein in der Literatur und in Standards nicht vollkommen konsistent beschriebenes Konzept. Man muss sich über die Ziele, die man bei der Messung der Usability verfolgt, im Klaren sein. Im Kontext dieser Ziele, sowie im Kontext des zu bewertenden Systems und der damit adressierten Benutzer, muss man zur Messung geeignete Methoden auswählen und die bei der Durchführung möglichen Probleme beachten.

Für den Praktiker wäre es sicherlich hilfreich, wenn sich dieser Bereich konsolidieren ließe, und wenn eine anerkannte Definition der Usability existieren würde. Ein klares Rahmenwerk zur Beurteilung der Usability, das ihm bei der Auswahl für den individuellen Zweck geeigneter Methoden sowie bei dessen Durchführung Hilfestellung leistet, wäre wünschenswert. Ob diese Ziele, aufgrund des enormen Umfangs des Usability-Konzeptes, jedoch überhaupt sinnvoll erreichbar sind, ist fraglich.

Thesenpapier

Thema Usability

Vorgelegt im Hauptseminar Wirtschaftsinformatik

Referent Thomas Wittek

Datum (Voraussichtlich) 01.06.2006

Thesen

1. Usability hat eine große Bedeutung bei Entwicklung und Bewertung von Software.
2. *Die* Usability gibt es nicht.
3. Es gibt keine beste Methode zur Beurteilung der Usability.
4. Usability ist nicht zuverlässig messbar.

Literaturverzeichnis

Abran, Khelifi, Suryn /Usability Iso/

Alain Abran, Adel Khelifi, Witold Surin: Usability Meanings and Interpretations in ISO Standards. In: Software Quality Journal, Jg. 11, Nr. 4, 2003, S. 325–338

Avouris /Introduction/

N.M. Avouris: An Introduction to Software Usability. In: Proceedings of the 8th Panhellenic Conference on Informatics (Workshop on Software Usability), November, 2001, Nicosia. Athens, 2001, S 514–522

Baber /Subjective Evaluation/

Chris Baber: Subjective evaluation of usability. In: Ergonomics, Nr. 14, Jg. 45, 2002, S. 1021–1025

Bevan /International Standards/

Nigel Bevan: International standards for HCI and usability. In: International Journal of Human-Computer Studies, Nr. 4, Jg. 55, 2001, S. 533–552

Bevan /Meeting User Needs/

Nigel Bevan: Quality in use: Meeting user needs for quality. In: Journal of Systems and Software, Nr. 1, Jg. 49, 1999, S. 89–96

Bevan /Quality of Use/

Nigel Bevan: Measuring usability as quality of use. In: Software Quality Journal, Nr. 2, Jg. 4, 1995, S. 115–130

Caulton /Homogeneity Assumption/

David A. Caulton: Relaxing the homogeneity assumption in usability testing. In: Behaviour & Information Technology, Nr. 1, Jg. 20, 2001, S. 1–7

Cordes /Task-Selection Bias/

Richard E. Cordes: Task-Selection Bias - A Case for User-Defined Task. In: International Journal of Human-Computer Interaction, Nr. 4, Jg. 13, 2001, S. 411–419

Folmer, Bosch /Survey/

Eelke Folmer, Jan Bosch: Architecting for usability: a survey. In: Journal of Systems and Software, Nr. 1-2, Bd. 70, 2004, S. 61–78

Gray, Salzman /Damaged Merchandise/

Wayne D. Gray, Marilyn C. Salzman: Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. In: Human-Computer Interaction, Jg. 13, 1998, S. 203–261

Hertzum, Jacobsen /Evaluator Effect/

Morten Hertzum, Niels Ebbe Jacobsen: The Evaluator Effect: A Chilling Fact About Usability Evaluation Method. In: International Journal of Human-Computer Interaction, Nr. 1, Jg. 15, 2003, S. 183–204

Holzinger /Usability Engineering Methods/

Andreas Holzinger: Usability Engineering Methods for Software Developers. In: Communications of the ACM, Nr. 1, Jg. 48, 2005, S. 71–74

ISO /9126-1/

ISO/IEC (Hrsg.): Software engineering – Product quality. Part 1: Quality model. ISO/IEC 9126-1: 2001. Geneva 2001

ISO /9141-11/

ISO (Hrsg.): Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability. ISO 9241-11: 1998. Geneva 1998

Jacobsen, Hertzum, John /Evaluator Effect/

Niels Ebbe Jacobsen, Morten Hertzum, Bonnie E. John: The evaluator effect in usability tests. In: Conference on Human Factors in Computing Systems, CHI 98 conference summary on Human factors in computing systems, April 18 - 23, 1998, Los Angeles, California. New York, S. 255–256

Jeffries u.a. /Comparison/

Robin Jeffries, James R. Miller, Cathleen Wharton, Kathy M. Uyeda: User interface evaluation in the real world: a comparison of four techniques. In: Proceedings of the ACM CHI'91 Conference, April 27 - May 02, 1991, New Orleans, Louisiana. New York, S. 119–124

Jeffries, Desurvire /Contest/

Robin Jeffries, Heather Desurvire: Usability testing vs. heuristic evaluation: was there a contest? In: ACM SIGCHI Bulletin, Nr. 4, Jg. 24, 1992, S. 39–41

Kirakowski, Corbett /SUMI/

J. Kirakowski, M. Corbett: SUMI: The Software Usability Measurement Inventory. In: British Journal of Educational Technology, Nr. 3, Jg. 24, 1993, S. 210–212

Law, Hvannberg /User Effect/

Effie Lai-Chong Law, Ebba Thora Hvannberg: Analysis of Combinatorial User Effect in International Usability Tests. In: Proceedings of the SIGCHI conference on Human factors in computing systems, April 24-29, 2004, Vienna, Austria. New York, S. 9–16

Lewis /Problem-Discovery Rates/

James R. Lewis: Evaluation of Procedures for Adjusting Problem-Discovery Rates Estimated From Small Samples. In: International Journal of Human-Computer Interaction, Nr. 4, Jg. 13, 2001, S. 446–479

Lewis /Sample Sizes/

James R. Lewis: Sample sizes for usability studies: additional considerations. In: Human Factors, Nr. 2, Jg. 36, 1995, S. 368–378

Lin, Choong, Salvendy /Proposed Index/

Han X. Lin, Yee-Yin Choong, Gavriel Salvendy: A proposed index of usability: a method for comparing the relative usability of different software systems. In: Behaviour & Information Technology, Nr. 4, Jg. 16, 1997, S. 267–278

Mack, Nielsen /Usability Inspection/

Robert Mack, Jakob Nielsen: Usability inspection methods: report on a workshop held at CHI'92. In: ACM SIGCHI Bulletin. Nr. 1, Jg. 25, 1993. S 28–33

Molich u.a. /Comparative Evaluation/

Rolf Molich, Meghan R. Eed, Klaus Kaasgaard, Barbara Karyukin: Comparative usability evaluation. In: Behaviour & Information Technology, Nr. 1, Jg. 23, 2004, S. 65–74

Nielsen /Heuristic Evaluation/

Jakob Nielsen: Finding usability problems through heuristic evaluation. In: Proceedings of the ACM CHI'92 Conference, May 3-7, 1992, Monterey, California. New York, S. 373–380

Nielsen /Usability Engineering/

Jakob Nielsen: Usability Engineering. London, 1993

Nielsen /Usability Inspection Methods/

Jakob Nielsen: Usability Inspection Methods. In: Proceedings of the ACM CHI'94 Conference, April 24-28, 1994, Boston, Massachusetts. New York, S. 413–414

Nielsen, Clemmensen, Yssing /Reflections Think-Aloud/

Janni Nielsen, Torkil Clemmensen, Carsten Yssing: Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In: ACM International Conference Proceeding Series, Vol. 31; Proceedings of the second Nordic conference on Human-computer interaction (NordiCHI 2002), October 19-23, 2002, Aarhus, Denmark. New York, NY, S. 101–110

Nielsen, Levy /Preference vs. Performance/

Jakob Nielsen, Jonathan Levy: Measuring Usability – Preference vs. Performance. In: Communications of the ACM, Nr. 4, Jg. 37, 1994, S. 66–75

Nielsen, Molich /Heuristic Evaluation/

Jakob Nielsen, Robert L. Mack: Usability inspection methods. New York, 1994

Seffah, Metzker /Usability and SWE/

Ahmed Seffah and Eduard Metzker: The Obstacles and Myths of Usability and Software Engineering. In: Communications of the ACM, Nr. 12, Jg. 47, 2004, S. 71–76

Shackel, Richardson /Human Factors Usability/

Brian Shackel, Simon Richardson: Human Factors for Informatics Usability. Cambridge, 1991

Virzi /How Many Subjects/

Robert A. Virzi: Refining the test phase of usability evaluation: how many subjects is enough? In: Human Factors, Nr. 4, Jg. 34, 1992, S. 457–468

Zhang /Perspective/

Zhijun Zhang, Victor Basili, Ben Shneiderman: Perspective-based Usability Inspection: An Empirical Validation of Efficacy. In: Empirical Software Engineering, Nr. 1, Jg. 4, 1999, S. 43–69