EINDHOVEN UNIVERSITY OF TECHNOLOGY

Laesst sich die Gebrauchstauglichkeit interaktiver Software messen?

Citation for published version (APA):
Rauterberg, G. W. M. (1992). Laesst sich die Gebrauchstauglichkeit interaktiver Software messen? *Ergonomie & Informatik*, *16*, 3-18.

Document status and date:
Published: 01/01/1992

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

Lässt sich die Gebrauchstauglichkeit interaktiver Software messen? Und wenn ja, wie?

Matthias Rauterberg

Institut für Arbeitspsychologie (IfAP)
Eidgenössische Technische Hochschule (ETH)
Nelkenstr. 11, CH-8092 Zürich

Der vorliegende Beitrag entstand im Rahmen des Forschungsprojekt BOSS - Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (Förderkennzeichen 01 HK 706-0), das vom BMFT (AuT-Programm) gefördert wird, und ist eine überarbeitete Fassung des Beitrages auf der 22.GI-Jahrestagung 1992.

1. Einleitung

Die Ergebnisse des Forschungsgebietes zur Software-Ergonomie sind zum Teil in nationale Normen (zB. DIN 66 234; Dzida, Herda & Itzfeldt 1978), bzw. neuerdings in internationale Normungsaktivitäten (zB. ISO 9241) eingegangen (Jüptner 1991). Eines der Hauptprobleme bei der Anwendung der einzelnen genormten Vorschriften, Leitsätze, bzw. Gestaltungskriterien liegt jedoch darin begründet, daß "es derzeit noch nicht möglich ist, die Erfüllung einzelner der ... genannten Leitsätze objektiv zu überprüfen, da geeignete Überprüfungsverfahren noch nicht bekannt sind. Wenn Prüfverfahren bekannt sind, bedarf es noch einer Weiterentwicklung dieser oder weiterer Normen, z.B. hinsichtlich quantifizierbarer Größen und anwendungsspezifischer Anforderungen." (DIN 66 234, Teil 8, 1988:1). Es wird im folgenden aufgezeigt und diskutiert, welche sinnvollen und praktikablen Möglichkeiten zur Zeit bestehen, die Gebrauchstauglichkeit interaktiver Softwareprodukte zu messen und zu beurteilen (Kirakowski & Corbett 1990).

2. Gebrauchstauglichkeit = Benutzbarkeit + Benutzungsfreundlichkeit

Softwareprodukte müssen zuallererst einmal ihre Gebrauchstauglichkeit unter Beweis stellen. Gebrauchstauglichkeit setzt sich zusammen aus Benutzbarkeit und Benutzungsfreundlichkeit. Benutzbar ist interaktive Software, wenn sie mit aufgabenangemessener Funktionalität ausgestattet ist. Mit Arbeitsaufgabe ist der 'Schnittpunkt' zwischen Organisation und Benutzer gemeint (Volpert 1987:14) und muß die folgenden Eigenschaften aufweisen (Ulich 1991:157):

- *Ganzheitlichkeit*: Benutzer erkennen die Bedeutung und den Stellenwert ihrer Tätigkeit; sie erhalten entsprechende Rückmeldung über den eigenen Arbeitsfortschritt aus der Tätigkeit selbst.
- *Anforderungsvielfalt*: Unterschiedliche Fähigkeiten, Kenntnisse und Fertigkeiten können von dem Benutzer eingesetzt werden; einseitige Beanspruchungen können vermieden werden.
- *Möglichkeiten der sozialen Interaktion*: Schwierigkeiten können gemeinsam zwi-

erschienen als:

Rauterberg, M. (1992) Lässt sich die Gebrauchstauglichkeit interaktiver Software messen? Und wenn ja, wie?. Ergonomie & Informatik 16:3-18.

schen den Benutzern bewältigt werden; gegenseitige Unterstützung hilft Belastungen besser ertragen.

- *Autonomie*: das Selbstwertgefühl und die Bereitschaft zur Übernahme von Verantwortung wird durch Autonomie gestärkt; ebenso wird die Erfahrung vermittelt, nicht einfluss- und bedeutungslos zu sein.
- *Lern- und Entwicklungsmöglichkeiten*: die allgemeine geistige Flexibilität bleibt erhalten; berufliche Qualifikationen werden erhalten und weiterentwickelt.

Diese Anforderungen lassen sich am besten im Rahmen partizipativer Softwareentwicklung verwirklichen (Rauterberg 1992) und betreffen die Spezifikation der Organisations- und Werkzeugschnittstelle (Dzida 1987).

2.1. Benutzbarkeit

Um die Benutzbarkeit eines Softwareproduktes sicherzustellen, muß in den frühen Phasen der Softwareentwicklung eine Analyse der Arbeitsaufträge und eine Analyse der Arbeitstätigkeiten durchgeführt werden (Ulich 1991):

Analyse der Arbeits-Aufträge: Bei der Auftrags- und Bedingungsanalyse wird schrittweise eine vertiefende Analyse von Arbeitsaufträgen durchgeführt (Hacker & Matern 1980). Die Auftragsanalyse dient der Gewinnung von organisationalen Gestaltungsvorschlägen (siehe auch Macaulay et al. 1990).

Analyse der Arbeits-Tätigkeiten: Für die Erarbeitung von arbeitsplatzbezogenen Gestaltungsvorschlägen ist es oft unumgänglich, eine Tätigkeitsanalyse durchzuführen (Ulich 1991:72ff). Sie liefert Kenntnisse über Abläufe, Auftrittshäufigkeiten und Zeitanteile der einzelnen Teiltätigkeiten. Während die globale Analyse der Arbeitsaufträge in dem zu analysierenden Arbeitssystem im Rahmen traditioneller Softwareentwicklung teilweise zum Tragen kommt, wird die Analyse der Arbeitstätigkeiten und der Auswirkungen dieser Tätigkeiten weitgehend außer acht gelassen. Einen systematischen Überblick über mögliche Analyse- und Gestaltungsmaßnahmen gibt Upmann (1989:114) (siehe auch Baitsch et al. 1989; Dzida et al. 1990; Lim, Long & Silcock 1990).

Die Benutzbarkeit interaktiver Software entscheidet darüber, ob das Softwareprodukt im geplanten Aufgabenkontext von den Endbenutzern überhaupt sinnvoll verwendbar ist (Carroll 1988). Die Messung der Benutzbarkeit kann über die Messung der Benutzungsfreundlichkeit erfolgen (Spencer 1985; Brooke et al. 1990; Kirakowski & Corbett 1990).

2.2. Benutzungsfreundlichkeit

Unter Benutzungsfreundlichkeit ("usability") versteht man ganz allgemein

"the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment; where 'ease of use' affects user performance and satisfaction, and 'acceptability' affects whether or not the product is used" (Bevan, Kirakowski & Maissel 1991:652).

Wenn die Benutzbarkeit gewährleistet ist, bleibt noch offen, wie gut die aufgabenbezogene Funktionalität von dem jeweiligen Benutzer im Rahmen seiner Tätigkeit genutzt werden kann. Zur Messung der Benutzungsfreundlichkeit lassen sich vier Meßansätze unterscheiden (siehe auch Bevan, Kirakowski & Maissel 1991:651):

- ***der interaktions-zentrierte Meßansatz*** **(IM); Benutzungsfreundlichkeit läßt sich über Eigenschaften der Interaktion zwischen Benutzer und System messen (Performanz, psychomentale Leistungen, etc.);**
- ***der benutzer-zentrierte Meßansatz*** **(BM); subjektive Beurteilungen des Benutzers lassen sich im Rahmen dieses Ansatzes erfassen (subjektive Ratings, etc.); das System ist "nur" als mental antizipiertes Modell und/oder als Spezifikation gegeben;**
- ***der produkt-zentrierte Meßansatz*** **(PM); die Benutzungsfreundlichkeit des Systems wird in ergonomischen Eigenschaften des Produktes selbst bestimmt (Maskenaufbau,**

Dialogtechnik, etc.); der Benutzer wird über die impliziten Modellannahmen, welche sich in der "Ergonomie der Eigenschaften" niederschlagen, berücksichtigt;

- ***der formal-analytische "Meß"-ansatz* (FM); Benutzungsfreundlichkeit läßt sich anhand der Eigenschaften formaler Modelle über den Benutzer (das mentale Modell) und über das interaktive System (die Spezifikation) messen (Performanz, psycho-mentale Leistungen, Funktionalität, etc.) (siehe Ziegler 1988:231ff).**

Jede Messung setzt sich aus dem Inhalt der Messung und der Form, unter der die Messung verläuft, zusammen (Zülch & Englisch 1991). Die Inhalte der Messung von Benutzungsfreundlichkeit sind Kriterien zur benutzer-orientierten Dialoggestaltung. In wie weit das jeweilige Kriterium erfüllt sein sollte, wird im Rahmen der Messung normativ vorgegeben. Um den Ausprägungsgrad eines Kriteriums bestimmen zu können, muß eine Meßvorschrift erstellt werden; dieser Vorgang wird 'Operationalisierung' genannt (Sarris 1990:142). Zum Zwecke einer empirischen Messung müssen also Zuordnungen ("Operationalisierungen") zwischen theoretischen Konstrukten (z.B. "Flexibilität") und messbaren Phänomenen (z.B. Bearbeitungszeit) getroffen werden. Diese Zuordnung von Meßskalen zu Gestaltungsprinzipien ist ein bis heute noch weitgehend ungelöstes Problem.

2.3. Prinzipien zur Gestaltung gebrauchstauglicher Software

Aus der Fülle möglicher Gestaltungsansätze wird hier das empirisch am besten abgesicherte Konzept von Ulich (1991; Ulich et al. 1991) vorgestellt. Ulich (1991:256ff) unterscheidet basierend auf handlungspsychologischen Überlegungen drei Bereiche: "Aufgabenorientierung", "Kalkulierbarkeit als Voraussetzung für Kontrolle" und "Kontrolle".

Der Bereich "Aufgabenorientierung" wurde bereits oben unter dem Aspekt der Benutzbarkeit besprochen. Der Bereich "Kalkulierbarkeit..." umfaßt die folgenden Kriterien (Ulich 1991:258-259):

"Transparenz: Benutzer/innen sollten erkennen können, ob ein *eingegebener Befehl* behandelt wird oder ob das System auf weitere Eingaben wartet. Bei längeren Vorgängen sollte das System *Zwischenzustandsmeldungen* abgeben können.

Konsistenz: Die *Antwortzeiten* des Systems sollten wenig variieren; wichtiger als kurze Antwortzeiten sind regelmässige und damit kalkulierbare Intervalle. Das System sowie dessen *Antwortverhalten* sollten für Benutzer/innen transparent und konsistent sein; ähnliche Aktionen sollten ähnliche Ausführungen bewirken, andernfalls muss dies durchschaubar gemacht werden. ...

Kompatibilität: Bei der *Darstellungsform* für Einzelinformationen sollte ebenso wie für ganze Bilder ggf. auf Übereinstimmung mit entsprechenden gedruckten Vorlagen oder Unterlagen geachtet werden. *Sprache* und begriffliche Komplexität des Dialoges sollten an den Gepflogenheiten und Kenntnissen des spezifischen Benutzerkreises orientiert sein; anstelle von EDV-Kürzeln sollte mit den jeweils fachspezifischen Begriffen der Benutzer/innen gearbeitet werden können.

Unterstützung: *Dialoghilfen* sowohl zu inhaltsbezogenen wie zu vorhergehensbezogenen Aspekten sollte von den Benutzer/innen während des Dialogs jederzeit abgerufen werden können; das Betätigen einer allfälligen Help-Taste sollte gegenüber anderen Befehlen einen Sonderstatus einnehmen. Das System sollte eine *Rückfragemöglichkeit* derart bereitstellen, dass auf eine Aufforderung durch die Benutzer/innen hin ggf. ausführlichere Antworten abgegeben werden".

Der Bereich "Kontrolle" wird durch die folgenden Kriterien beschrieben:

Flexibilität ist die "Summe objektiv vorhandener Freiheitsgrade zur selbständigen Setzung und Erreichung von (Teil-) Zielen durch variable Abfolge von (Teil-) Schritten" (Spinas 1987:146).

Individuelle Auswahlmöglichkeiten: der Benutzer kann das Systemverhalten durch die Einstellung von Systemparametern auf seine individuellen und aufgabenbezogenen Bedürfnisse abstimmen.

Individuelle Anpassungsmöglichkeiten ist "die Möglichkeit der eigenständigen Gestaltung und dementsprechend auch der Erweiterung objektiver Tätigkeitsspielräume" (Spinas 1987:177).

Partizipation beinhaltet die verschiedenen Formen und Grade der Benutzerbeteiligung bei der Systementwicklung (Spinas, Waeber & Strohm 1990; Rauterberg 1992).

Um nun diese inhaltlichen Gestaltungsprinzipien messen zu können, müssen sie 'operationalisiert' werden (Peercy 1981; Whitefield, Wilson & Dowell 1991: 72). Die am häufigsten verwendeten Meßskalen für bestimmte Eigenschaften des Interaktionsprozesses werden nur 'lose' - wenn überhaupt - inhaltlichen Kriterien zugeordnet. Rengger (1991: 658) konnte zeigen, daß in ca. 500 Veröffentlichungen gefundene Operationalisierungen in folgende vier Bereiche aufgeteilt werden können:

- *Benutzbarkeits-Indikatoren*;
- *Leistungs-Indikatoren*;
- *Handhabungs-Indikatoren*;
- *Qualifizierungs-Indikatoren*;

Der folgende Bereich muß unbedingt noch hinzugenommen werden (Boucsein 1987; Kishi & Kinoe 1991):

- *Belastungs-Indikatoren*.

Ein gutes Meßverfahren sollte sich nicht nur durch die Eigenschaften der "Objektivität", der "Reliabilität" und der "Validität" (siehe unten), sondern sich auch durch einen *minimalen Meßaufwand* auszeichnen. Dies ist einer der Gründe, warum ein großes Interesse daran besteht, den aufwendigen interaktions- und benutzer-zentrierten Meßansatz durch den weniger aufwendigen produkt-zentrierten, bzw. formal-analytischen Meßansatz zu ergänzen. Es gibt jedoch Kriterien wie z.B. "Transparenz", welche sich nur im Bezug auf die Wahrnehmungs- und Interpretationsleistung des Benutzers (seine mentalen Interpretationsschemata) messen lassen (siehe Weidenmann 1988:95), sodaß der interaktions-, bzw. benutzer-zentrierte Meßansatz sich nicht vollständig durch den produkt-zentrierten Meßansatz ersetzen läßt. Dieser Umstand wird oft auch als Kritik an dem formal-analytischen Meßansatz aufgeführt (z.B. Ziegler 1988: 250).

3. Die verschiedenen Ansätze zur Messung von Gebrauchstauglichkeit

Durch die Modellierung, bzw. Simulierung des Benutzers oder des interaktiven Systems versucht man, den Meßaufwand beim Einsatz empirischer Tests zu reduzieren. Wenn sowohl die Software, als auch der Benutzer nur als Modell gegeben ist, handelt es sich um den formal-analytischen Ansatz (z.B. "keystroke level model": Card, Moran & Newell 1983; "cognitive complexity theory": Kieras & Polson 1985; etc.). Da die Modellannahmen dieser formal-analytischen Methoden aufgrund fehlender empirischer Forschungsergebnisse noch nicht ausgereift sind, werden wir auf diesen Ansatz nicht weiter eingehen (zur einschlägigen Kritik siehe Greif & Gediga 1987, bzw. Karat & Bennett 1991). Der formal-analytische Meßansatz (FM) wird an anderer Stelle ausführlicher diskutiert (Ziegler 1988).

Wir werden zuerst den am häufigsten eingesetzten interaktions-zentrierten Meßansatz (IM), dann den benutzer-zentrierten Meßansatz (BM) und zum Schluß die bisher möglichen Meßkriterien im Rahmen des produkt-zentrierten Meßansatzes (PM) vorstellen. Die Zuordnung zu einem dieser Meßansätze bestimmt die Quelle, für welche die jeweilige Meßskala definiert ist. Oft werden IM und BM im Rahmen von empirischen Tests gemeinsam eingesetzt. Eine empirische Messung setzt sich aus einer Datenerhebungs- und einer Datenaufzeichnungsmethode, sowie einer Auswertungs-, bzw. Meßvorschrift zusammen.

3.1. Der interaktions-zentrierte Meßansatz (IM)

Wenn die Meßskala Eigenschaften des Interaktionsprozesses zwischen Benutzer und System mißt, handelt es sich um den interaktions-zentrierten Meßansatz (IM). Es lassen sich verschiedene Aspekte des Interaktionsprozesses messen (siehe Tabelle 1).

Leistungsfähige Datenaufzeichnungsmethoden sind: Testleiterprotokollierung, Video-Aufzeichnung des Bildschirminhaltes ("screen-recording"), des Benutzers, sowie der Eingabeschnittstelle, automatische Aufzeichnung der benutzten Dialogoperatoren ("logfile-recording") (Crellin, Horn & Preece 1990). Als brauchbarer Kompromiß hat sich eine Kombination zwischen "logfile-recording" und unmittelbarer Testleiterprotokollierung ergeben (Müller-Holz et al. 1991: 418).

Die Datenaufzeichnung auf Video oder Tonband ist zwar sehr praktisch, benötigt aber bei der Auswertung einen drei- bis zehnfachen zeitlichen Auswertungsaufwand. Um diesen Auswertungsaufwand zu minimieren, empfiehlt es sich möglichst viele Daten während der Testung mitzuerheben (Vossen 1991). Die wichtigsten Daten lassen sich jedoch oftmals problemlos auf dem Testleiterprotokollbogen vermerken.

Um die vom Benutzer jeweils angestrebten Bearbeitungsziele messen zu können, wird er gebeten, seine kognitiven Ziele während der Aufgabenbearbeitung laut auszusprechen ("lautes Denken"). Problematisch ist diese Datenerhebungsmethode, wenn der Benutzer ungeübt im Verbalisieren oder sehr intensiv mit der zu bearbeitenden Aufgabenstellung beschäftigt ist. Der Benutzer neigt dann dazu, mit dem Aussprechen aufzuhören. Die Videokonfrontationsmethode (Neal & Simons 1984; Moll 1987) kann die Schwächen der Datenerhebungsmethode des "lauten Denkens" zum Teil ausgleichen.

Tabelle 1: Zuordnung interaktionsbezogener Meßskalen zu den fünf Meßbereichen.

Mess-bereich	MessKriterium	Daten-Erhebungs-methode	Daten-Auf-zeichnungs-methode	Daten-Aus-wertungs-methode
Benutz-barkeit	Art und Anzahl benutzter Funktionen (Moll 1987)	aufgabenbezogene und/oder explorative Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokollierung	Häufigkeiten
	Produktgüte, Mängelrate, Verhältnis von Produktgüte zu Mängel, Genauigkeit, Effektivität (Rengger 1991)	aufgabenbezogene Rechnerbenutzung	Backup des Anwendungsobjektes (z.B. Textdokument, Datenbank; Datei; etc.)	Ratingverfahren anhand eines Kategoriensystems

Mess-bereich	MessKriterium	Daten-Erhebungs-methode	Daten-Auf-zeichnungs-methode	Daten-Aus-wertungs-methode
Lei-stung	Aufgabenbearbeitungsgeschwin-digkeit	aufgabenbezogene Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokol-lierung	Bearbeitungszeit
	Anzahl Dialogoperatoren	aufgabenbezogene und/ oder explorative Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording	Häufigkeiten
	durchschnittliche Bearbeitungs-zeit pro Dialogoperator	aufgabenbezogene Rechnerbenutzung	"logfile"-Recording	Mittelwertberech-nung
Hand-habung	durchschnittliche Dauer der Pausen zwischen zwei Dialog-operatoren	aufgabenbezogene Rechnerbenutzung	"logfile"-Recording	Mittelwertberech-nung
	Übergangswahrscheinlichkeiten zwischen verschiedenen Dialog-operatoren (Schmid & Meseke 1991)	aufgabenbezogene Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokol-lierung	Ratingverfahren an-hand eines Katego-riensystems
	Art und Anzahl benutzter Dialogoperatoren	aufgabenbezogene und/ oder explorative Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokol-lierung	Häufigkeiten
	Art und Anzahl Fehler; Art und Anzahl interaktiver Probleme ("interaktive Deadlocks")	aufgabenbezogene Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokol-lierung	Häufigkeiten auf-grund eines Rating-verfahren anhand eines Kategorien-systems
	Art der Problemlöse-Strategie	aufgabenbezogene Rechnerbenutzung mit "lautem Denken"	Videoaufnahmen mit Tonaufzeichnung	Häufigkeiten anhand eines Kategorien-systems
	subjektive Beurteilung der Dialogschnittstelle	aufgabenbezogene und/ oder explorative Rechnerbenutzung	Handhabungsbogen (Spinas 1987; Rauterberg 1991)	bipolare Rating-skalen

Mess-bereich	MessKriterium	Daten-Erhebungs-methode	Daten-Auf-zeichnungs-methode	Daten-Aus-wertungs-methode
Qualifi-zierung	angestrebte Bearbeitungsziele	aufgabenbezogene Rechnerbenutzung mit "lautem Denken"	Videoaufnahmen mit Tonaufzeichnung	Häufigkeiten anhand eines Kategorien-systems
	Blickbewegungen (Fleischer et al. 1984)	aufgabenbezogene und/ oder explorative Rechnerbenutzung mit "lautem Denken"	Blickregistrierung mittels einer Blickbe-wegungskamera	Häufigkeiten anhand eines Kategorien-systems
	Art und Anzahl von Problem-löse-Strategiewechsel	aufgabenbezogene Rechnerbenutzung mit "lautem Denken"	Videoaufnahmen mit Tonaufzeichnung	Häufigkeiten anhand eines Kategorien-systems
	Zeit für die Benutzung des Hilfe-system; Zeit für die Benutzung der Dokumentation	aufgabenbezogene und/ oder explorative Rechnerbenutzung	Videoaufnahmen; "logfile"-Recording; rechnerunterstützte Verhaltensprotokol-lierung	Summenbildung
	Überlegungszeitschwellen (Ackermann & Greutmann 1987)	aufgabenbezogene Rechnerbenutzung	"logfile"-Recording	Mittelwertberech-nung
Be-lastung	Herzrate (Wiethoff, Arnold & Houwig 1991)	aufgabenbezogene Rechnerbenutzung	papier- und/oder digi-tale Aufzeichnung mittels Elektrokardio-gramm (EKG)	spezielle Auswert-ungsprogramme
	Atmung (Wiethoff, Arnold & Houwig 1991)	aufgabenbezogene Rechnerbenutzung	papier- und/oder digi-tale Aufzeichnung mittels Atemfrequenz-messgerät (Termistor)	spezielle Auswert-ungsprogramme
	Blutdruck	aufgabenbezogene Rechnerbenutzung	papier- und/oder digi-tale Aufzeichnung mittels Blutdruck-messgerät	spezielle Auswert-ungsprogramme
	Hautleitfähigkeit (Wiethoff, Arnold & Houwig 1991)	aufgabenbezogene Rechnerbenutzung	papier- und/oder digi-tale Aufzeichnung mittels Elektromyo-gramm (EMG), bzw. Psychogalvanischer Reflex	spezielle Auswert-ungsprogramme
	Hirnstromfrequenzen (Wiethoff, Arnold & Houwig 1991)	aufgabenbezogene Rechnerbenutzung	papier- und/oder digi-tale Aufzeichnung mittels Elektroence-phalogramm (EEG)	spezielle Auswert-ungsprogramme
	bipolare Videoratingskalen (Rauterberg 1988)	aufgabenbezogene Rechnerbenutzung	Videoaufzeichnung	Ratingverfahren
	subjektive Befindlichkeit (Rauterberg 1991)	aufgabenbezogene Rechnerbenutzung	z.B. Eigen-Zustands-fragebogen (Apenburg 1986)	Schablonen

3.2. Der benutzer-zentrierte Meßansatz (BM)

Alle Meßwerte, die ausschließlich über Eigenschaften des Benutzers erhoben werden, gehören zum benutzer-zentrierte Meßansatz (BM). Es lassen sich zwei Meßwertbereiche ausmachen: Eigenschaften des Benutzers selbst und/oder Eigenschaften des simulierten Systems gemessen über die Einschätzungen seitens des Benutzers (siehe Tabelle 2). Bei diesem Meßansatz liegt das System nur als Modell vor: a) als mentales Modell in der Vorstellung des Benutzers und/oder b) als formales Modell in Form einer Spezifikation.

Tabelle 2: Zuordnung benutzerbezogener Meßskalen zu den fünf Meßbereichen.

Mess-bereich	MessKriterium	Daten-Erhebungs-methode	Daten-Auf-zeichnungs-methode	Daten-Aus-wertungs-methode
Benutz-barkeit	Aufgabeneigenschaften (Rudolph, Schönfelder & Hacker 1987)	arbeitswissenschaft-liche Aufgabenanalyse	Fragebogen	Häufigkeiten mittels Rating-verfahren anhand eines Kategorien-systems
	Beurteilungsskalen (Shneiderman 1987:402-407, Spinas 1987)	Fragebogen, Interview	Fragebogen, Tonband	Mittel-, bzw. Summenwerte
Lei-stung	Fragebögen zur Messung der In-telligenz, - der Leistungsmotiva-tion, der Aufmerksamkeits-spanne, etc.	Fragebogen	Fragebogen	Mittel-, bzw. Summenwerte
Hand-habung	Handhabungsbogen (Spinas 1987, Rauterberg 1991), "Questionnaire for User Interface Satisfaction (QUIS)" (Chin, Diehl & Norman 1988)	Fragebogen, Interview	Fragebogen, Tonband	Mittel-, bzw. Summenwerte
Qualifi-zierung	Zeit zur Bewältigung eines Trainingprogramms	Trainingsprogramm	Fragebogen	Mittel-, bzw. Summenwerte
	"Questionnaire for User Interface Satisfaction (QUIS)", Wissens-fragebogen (Dutke 1988), Vorer-fahrungsfragebogen (Rauterberg 1991)	Fragebogen, Interview	Fragebogen, Tonband	Mittel-, bzw. Summenwerte
Be-lastung	Fragebogen zur Messung psy-chomentaler Belastungen (Apenburg 1986)	Fragebogen, Interview	Fragebogen, Tonband	Mittel-, bzw. Summenwerte

3.3. Der produkt-zentrierte Meßansatz (PM)

Bei dem produkt-zentrierten Meßansatz (PM) werden Eigenschaften des Softwareproduktes direkt am Produkt selbst gemessen. Das Benutzermodell ist eingebettet in die Operationalisierungen der verwendeten Meßwertskalen. Es gibt drei mögliche Zugangsweisen: Kriterien (z.B. DIN 66 234), Sichtprüfung mittels Checklisten (z.B. EVADIS, Oppermann et al. 1988) und quantitative Maße (Gunsthövel & Bösser 1991).

Grundsätzlich ist für diesen produkt-zentrierten Meßansatz eine Beschreibungssprache für Eigenschaften von Benutzungsoberflächen notwendig, welche nicht zu allgemein ist, aber auch nicht zu spezifisch am technischen Detail hängen bleibt. Der "Granulationsgrad" dieser Beschreibungssprache sollte so gewählt sein, daß die verwendeten Beschreibungskonstrukte die spezifischen Eigenschaften der verschiedenen Oberflächentypen hinreichend genau differenzieren können, aber dennoch auf möglichst viele Oberflächentypen einheitlich anwendbar sind. Das EVADIS-Verfahren stellt daher den Prüffragen eine Erläuterung der technischen Komponenten der Benutzungsschnittstelle voran (Oppermann et al. 1988:21-23).

	E/A-Schnittstelle	Dialog-Komponente	Anwendungs-komponente
Individuelle Anpassung ("eins zur Zeit, nach eigener Gestaltung")	freie Namensgebung, bzw. Icon-Gestaltung	freie Änderung der Interaktions-Art	
	freie Gestaltung der Bildschirmlayouts	freie Änderung der Dialogstruktur	
	freie Tastenbelegung	Makrodefinition	Eigenprogrammierung
Individuelle Auswahl ("eins zur Zeit, nach Auswahl aus dem Angebot")	Auswahl von Namen, bzw. Icons		
	Auswahl des Bildschirmlayouts	Wahl zwischen Interaktions-Arten	
	Auswahl von Tastenbelegungen	Wahl zwischen Makros	Auswahl des Funktionsangebotes
Flexibilität ("mehrere Möglichkeiten gleichzeitig")	verschiedene Namen, bzw. Icons	verschiedene Interaktions-Arten	
	[verschiedene Arten des Bildschirmlayouts]*	"modeless states"	
	[verschiedene Arten an Tastaturbelegungen]**	verschiedene Makros	vollständiges Funktionsangebot

Legende:
* setzt mehrere Bildschirme nebeneinander voraus
** setzt mehrere Tastaturen nebeneinander voraus

Abbildung 1: Übersicht über die ergonomischen Eigenschaften für den Gestaltungsbereich der Kontrolle nach Ulich (1991).

Das Beschreibungskonzept der "interaktiven Aufsetzpunkte" (Rauterberg, in Vorbereitung) erlaubt es nun, nicht nur verschiedenste Arten von Benutzungsoberflächen einheitlich zu beschreiben, sondern auch wesentliche Unterschiede zwischen diesen Oberflächen einfach darzustellen. Es lassen sich in Abhängigkeit von der jeweiligen interaktiven Bedeutung verschiedene Mengen von Aufsetzpunkten unterscheiden: repräsentationale Interaktionspunkte (RIPe) und funktionale Interaktionspunkte (FIPe); unterscheidet man nun die Funk

tionalität in Dialogfunktionen und Anwendungsfunktionen, so erhält man dialog-funktionale Interaktionspunkte (DFIPe) und anwendungs-funktionale Interaktionspunkte (AFIPe); sind diesen beiden Typen von FIPen jeweils wahrnehmbare Repräsentationen auf der Ein/Ausgabeschnittstelle zugeordnet, so ergeben sich repräsentierte dialog-funktionale Interaktionspunkte (RDFIPe) und repräsentierte anwendungs-funktionale Interaktionspunkte (RAFIPe). Aufbauend auf diesen Beschreibungskonstrukten lassen sich nun Kriterien wie "Flexibilität", "individuelle Auswahl" und "individuelle Anpassung" in quantifizierbare Formeln überführen.

Wie sich der Gestaltungsbereich der Kontrolle gegeben über die drei Gestaltungsprinzipien der "individuellen Anpassung", der "individuellen Auswahl" und der "Flexibilität" auf konkrete Produkteigenschaften beziehen läßt, zeigt Abbildung 1. Es läßt sich nun unmittelbar an dem Produkt das Ausmaß der erreichten "Kontrolle" mittels dieser Operationalisierungen abzählen.

4. Die Güte der Messung

Messen ist die Zuordnung von Zahlen (numerisches Relativ) zu Objekten und deren Eigenschaften (empirisches Relativ) mit dem Ziel einer isomorphen oder homomorphen Abbildung (Kirakowski & Corbett 1988:156). Damit verbunden sind die Objektivität, die Reliabilität und die Validität einer Messung, weil jede Messung durch systematische und/oder zufällige Meßfehler beeinflußt wird. Das Ergebnis einer Messung wird Skala genannt, die durch das geordnete Tripel (A, Z, Ω) mit A als empirischen Relativ, Z als numerischen Relativ und Ω als Zuordnungsfunktion definiert ist. Zu jedem empirischen Relativ gibt es eine Menge numerischer Relative, die durch alle diejenigen Transformationen gegeben sind, welche die Isomorphie- oder Homomorphiebedingungen der Abbildung bewahren. Skalen mit gleichen Transformationseigenschaften haben dasselbe Meß- bzw. Skalenniveau.

4.1. Das Meßniveau

Es lassen sich vier verschiedene Meßniveaus unterscheiden: die Nominal-Skala, die Ordinal-Skala, die Intervall-Skala und die Verhältnis-Skala. Die folgende Tabelle zeigt die Eigenschaften dieser vier Meßniveaus:

Tabelle 3: Übersicht über die vier verschiedenen Meßskalen und ihre Eigenschaften.

	Nominal-Skala	**Ordinal-Skala**	**Intervall-Skala**	**Rational-Skala**
andere Bezeichnungen	topologische Skalen "qualitative" Skalen		metrische Skalen "quantitative" Skalen	
definierte Relationen	= ≠	= ≠ < >	= ≠ < > + -	= ≠ < > + - * /
zulässige Transformation	alle eindeutigen	alle monotonen	x'= bx + a mit b≠0	x'= bx mit b≠0
Interpretation	gleich / ungleich Relationen im numerischen Relativ geben entsprechende Eigenschaften des empirischen Relativs wieder	kleiner / größer Relationen im numerischen Relativ geben entsprechende Eigenschaften des empirischen Relativs wieder	Differenz-Relationen im numerischen Relativ geben entsprechende Eigenschaften des empirischen Relativs wieder	Verhältnis-Relationen im numerischen Relativ geben entsprechende Eigenschaften des empirischen Relativs wieder
Erwartungswert	Modalwert	Median	arithmetisches Mittel	geometrisches Mittel

4.2. Die "klassische" Testtheorie

Die "klassische" Testtheorie der Sozialwissenschaften ist aus der Ausgleichs- und Fehlerrechnung der Physik entstanden, d.h. jeder konkret gewonnene Meßwert X im Rahmen eines Tests wird aus zwei additiven Komponenten zusammengesetzt gedacht: dem "wahren" Wert W und einem Fehleranteil F, d.h. also X = W + F. Über den Meßfehler F werden nun spezifische Annahmen getroffen: der Erwartungswert ist 0; W und F sind unabhängig voneinander ("unkorreliert"); F_1 und F_2 sind unabhängig voneinander, d.h. die Fehleranteile unterschiedlicher Skalen sind unkorreliert. Insgesamt müssen die Meßwerte voneinander stochastisch unabhängig sein. Vor dem Hintergrund dieser allgemeinen Annahmen werden die folgenden Gütekriterien formuliert:

4.3. Objektivität einer Messung

Die Ergebnisse eines Tests sollen ausschließlich die zu messende Eigenschaften des empirischen Relativs repräsentieren und insbesondere unabhängig von Testleitereffekten, unterschiedlichen Interpretationen und sonstigen situativen Bedingungen sein. Eine hohe Objektivität kann durch klar geregelte Test- und Meßvorschriften erreicht werden (siehe als Beispiel Dutke 1988:215).

4.4. Reliabilität ("Zuverlässigkeit") der Messung

Das Gütekriterium der Reliabilität (Landauer 1988) ist ein Maß für den Anteil der "wahren" Varianz $\sigma(W)$ an der Gesamtvarianz $\sigma(X)$ (siehe Peercy 1981: 347-348). Geschätzt wird dieses Verhältnis durch die Berechnung von Zusammenhangsmaßen ("Korrelationen") zwischen den Meßwerten bei...

Testwiederholungen: ein und derselbe Test wird mit demselben System in einem bestimmten Zeitabstand wiederholt; man spricht hier daher auch von der Test-Retest-Reliabilität oder Stabilität; problematisch sind jedoch alle zwischenzeitlichen Veränderungseffekte bei dem getesteten System;

Parallel-Tests: zu einem Test wird eine möglichst "gleichwertige" Parallelform entwickelt und beide Formen zur Testung von zwei identischen Systemen eingesetzt; praktisch kann es jedoch recht schwierig sein, inhaltlich und formal zufriedenstellende Parallelformen zu entwickeln.

4.5. Validität ("Gültigkeit") der Messung

Die Validität ist ein Maß dafür, inwieweit ein Test inhaltlich das mißt, was er zu messen vorgibt (Lienert 1989). Gemeint ist die Übereinstimmung der Ergebnisse eines Tests mit den durch die Testung vorgegebenen theoretisch-begrifflich zu erfassenden Sachverhalt. Dies setzt voraus, daß die Abbildung des empirischen Relativs auf das numerische Relativ semantik-bewahrend ist. Durch eine hohe Validität wird gewährleistet, daß die Testergebnisse eine sinnvolle Grundlage für eine Beurteilung darstellen.

Man unterscheidet verschiedene Validitäts-Konzepte, die jeweils unterschiedliche Aspekte von Validität betonen...

augenscheinliche Gültigkeit ("face-validity"): diese Validitätsbestimmung beschränkt sich auf die Frage der Plausibilität von Testergebnissen; die erhaltenen Testergebnisse dürfen dem akzeptierten Wissensstand nicht widersprechen; man spricht auch von inhaltlicher Validität ("content validity");

interne, logische oder Eigenschaftsvalidität: unter dieser Validitätsbestimmung versteht man die Eindeutigkeit einer Messung, d.h. ob unkontrollierte Variablen eine Messung beeinflussen; dieser Validitätsaspekt ist eine notwendige Voraussetzung für die externe oder Kriteriumsvalidität, die sich auf die Generalisierbarkeit der Testergebnisse bezieht;

Parallelen-Validität: diese Validitätsbestimmung beruht auf dem strukturellen

Zusammenhang mit einem Kriterium, von dem man weiß, daß es eng mit dem gemessenen Merkmal zusammenhängt (Außenkriterium);

Konstruktvalidität: die Abschätzung der Konsistenz mit den entsprechenden theoretischen Konstrukten liefert Werte für die Konstruktvalidität; dabei werden die Meßergebnisse durch vorhandene oder neu formulierte Hypothesen einer Theorie erklärt und anhand weiterer Messungen überprüft.

5. Testen, Messen und Beurteilen

Objektivität, Reliabilität und Validität haben die vor allem pragmatische Funktion, daß unterschiedliche Tester hinsichtlich einer bestimmten Testung zu vergleichbaren Ergebnissen gelangen (Landauer 1988). Eine hohe Reliabilität garantiert die intersubjektive Erfahrbarkeit im Gegensatz zu raum-zeitlich singulärer und individueller Erfahrung. Aus diesem Grund ist eine hohe Reliabilität eine notwendige, wenn gleich keine hinreichende, Voraussetzung für die Validität einer Messung. Die Forderung nach der Reproduzierbarkeit von Testergebnissen erfordert ein Konstanthalten aller "relevanten" Testbedingungen. Je besser diese Forderung erfüllbar ist (z.B. in einigen Bereichen der Naturwissenschaften), desto leichter ist die Beurteilbarkeit der Meßergebnisse und damit die Erkenntnisgewinnung.

Die Meßergebnisse eines Test, die ausschließlich unter künstliche Laboratoriumsbedingungen gewonnen wurden, haben jedoch eine geringe "ökologische" Validität (Benda 1983). Die Minimierung der großen Variabilität der zu messenden Phänomene in ihrem "natürlichen" Entstehungskontext, um eine möglichst hohe Reliabilität zu erreichen, steht der Forderung nach "ökologischer" Validität entgegen. Da aber zugleich eine hohe Reliabilität eine notwendige Voraussetzung für Validität ist, befindet sich der Tester in einem Dilemma, das nur zu lösen ist, indem man entweder hofft, daß auf der Grundlage einer Vielzahl von artifiziellen Testergebnissen sich eine umfassende und alltagsrelevante Beurteilung abgeben läßt, oder indem man multivariate Testdesigns wählt, um so der Vielfalt aller "relevanten" Einflußgrößen auf das Meßergebnis zumindest einigermaßen gerecht werden kann (Tabachnik & Fidell 1989).

Mögliche Zielkonflikte (Greutmann & Ackermann 1989) zwischen verschiedenen Kriterien lassen sich konstruktiv durch die Abhängigkeitsmatrix (Evans & Marciniak 1987: 180) und die Paarvergleichsmethode (Sherwood-Smith 1989: 87) lösen.

5.1 Test auf Übereinstimmung ("conformance")

Der "Test auf Übereinstimmung" (Dzida 1992) mit ausgewählten Kriterien und normativ vorgegebenen Soll-Werten dient der Überprüfung, in wie weit ein Softwareprodukt dem jeweils ausgewählten Kriterium genügt. Um die Meßbarkeit zu ermöglichen, muß ein Meßprotokoll mit einem normativ gesetzten Soll-Wert und dem un-, bzw. günstigsten Fall vorgegeben werden (siehe Tabelle 4).

Tabelle 4: Meßprotokoll z.B. für den Test der Installationssoftware einer Workstation (aus Whiteside, Bennett & Holtzblatt 1988:795).

Produkteigenschaft	Testaufgabe	Meßskala	im ungünstigsten Fall	im günstigsten Fall	SOLL-Wert	IST-Wert
Installierbarkeit	Installieren einer Workstation	Bearbeitungszeit	ein Tag mit Hilfestellung	10 Minuten ohne Hilfestellung	eine Stunde ohne Hilfestellung	viele können überhaupt nicht installieren

5.2. Benutzungsorientierte Benchmarktests ("benchmarking")

Benutzungs-orientierte Benchmark-Tests (bBTs) lassen sich in zwei Arten unterteilen: *induktive* und *deduktive bBTs* (Rauterberg 1991). Die induktiven bBTs sind bei der Evaluation eines (z.B. vertikalen) Prototypen, oder einer (Vor)-Version zur Gewinnung von Gestaltungs- und Verbesserungsvorschlägen, bzw. zur Analyse von Schwachstellen in der Gebrauchstauglichkeit einsetzbar. Induktive bBTs können immer dann zum Einsatz kommen, wenn nur *ein* Prototyp, bzw. *eine* Version der zu testenden Software vorliegt. Demgegenüber verfolgen deduktive bBTs primär den Zweck, zwischen mehreren Alternativen (mindestens zwei Prototypen, bzw. Versionen) zu entscheiden. Zusätzlich lassen sich jedoch auch mit deduktiven bBTs Gestaltungs- und Verbesserungsvorschläge gewinnen.

5.3. Fehler und Fallen beim Testen ("pitfalls")

Da interaktions- und benutzer-zentrierte Meßansätze zur Zeit am besten entwickelt sind, müssen bei dieser Messung unbedingt folgende Punkte beachtet werden (Holleran 1991):

- die Testaufgabe ist dem Aufgabenkontext des potentiellen Benutzerkreises zu entnehmen;
- die Auswahl der Benutzer muß repräsentativ sein (Bortz 1984:239-347);
- die Testung ist mit mindestens sechs verschiedenen Benutzern durchzuführen, um die Auswertung mit inferenzstatistischen Methoden zu ermöglich, damit eine möglichst gute Generalisierbarkeit der Meßergebnisse gewährleistet wird (Bortz 1989);
- manchmal ist es wichtig, daß der Testleiter nicht auch gleichzeitig Entwickler der zu testenden Software ist (Bortz 1984:61-62);
- es sollten möglichst alle relevanten Einflußgrößen gleichzeitig gemessen werden;
- die Testung sollte in dem "natürlichen" Arbeitskontext der potentiellen Benutzer erfolgen;
- die Meßmethode sollte so objektiv, reliabel und valide wie möglich sein.

6. Fazit

Um eine gebrauchstaugliche Software zu gewährleisten, muß die Gebrauchstauglichkeit bestimmt, d.h. gemessen werden können. Gebrauchstauglichkeit läßt sich am prägnantesten durch Kriterien inhaltlich ausfüllen, welche auf der Grundlage arbeitswissenschaftlicher Forschungen entwickelt wurden. Diese inhaltlichen Kriterien lassen sich zur Zeit am besten im Rahmen von drei verschiedenen Meßansätzen (interaktions-, benutzer- und produkt-zentriert) messen. Die Messung kann auf vier unterschiedlichen Meßniveaus erfolgen, wobei das Meßniveau weitgehend durch die Art der zu messenden empirisch beobachtbaren Eigenschaft festgelegt wird. Eine Messung sollte objektiv, reliabel und valide sein. Die Erhebung und Aufzeichnung der Daten, sowie die Auswertung zu Meßwerten sollte mit einem möglichst geringen Aufwand erfolgen können.

Der Aufwand für die Messung ließe sich dadurch verringern, möglichst viele Eigenschaften, welche zur Zeit nur interaktions- oder benutzer-zentriert messbar sind, produkt-zentriert zu messen. Jede Messung setzt eine Zuordnung von inhaltlichen Kriterien zu messbaren Eigenschaften der Realität voraus. Dazu stehen arbeitswissenschaftlich abgesicherte Kriterien, sowie das ausgereifte Methodenspektrum der Testtheorie, der Testplanung, der Skalierungstheorie und der angewandten Statistik zur Verfügung. Es ist zur Zeit also sehr gut möglich, die Gebrauchstauglichkeit von Software in einem hinreichenden Ausmaß messen und beurteilen zu können. Man muß es nur tun.

7. Literaturverzeichnis

Ackermann, D / Greutmann, T, 1987: Interaktionsgrammatik und kognitiver Aufwand. In: Schönpflug, W / Wittstock, M (Hrsg.) Software-Ergonomie ' 87 "Nützen Informationssysteme dem Benutzer?". Stuttgart: Teubner. 262-270

Apenburg, E, 1986: Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung. Zeitschrift für Arbeits- und Organisationspsychologie 30(1):3-14

Baitsch, C / Katz, C / Spinas, P / Ulich E, 1989: Computerunterstützte Büroarbeit - Ein Leitfaden für Organisation und Gestaltung. Zürich: Verlag der Fachvereine.

Benda, von, H, 1983: Feldversuch zur Gestaltung des Mensch-Maschine-Dialogs im Verwaltungsbereich. In: Balzert, H (ed.) Software-Ergonomie. Stuttgart: Teubner. 266-277

Bevan, N / Kirakowski, J / Maissel, J, 1991: What is usability? In: Bullinger, H-J (ed.) Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam London New York: Elsevier. 651-655

Bortz, J, 1984: Lehrbuch der empirischen Forschung. Berlin Heidelberg New York: Springer.

Bortz, J, 1989: Statistik. Berlin Heidelberg New York Tokyo: Springer.

Boucsein, W, 1987: Psychophysiological investigation of stress induced by temporal factors in human-computer interaction. In: Frese, M / Ulich, E / Dzida, W (eds.) Human Computer Interaction in the Work Place. Amsterdam: Elsevier (North-Holland). 163-181

Brooke, J / Bevan, N / Brigham, F / Harker, S / Youmans, D, 1990: Usability statements and standardisation - work in progress in ISO. In: Diaper, D / Gilmore, D / Cockton, G / Shackel, B (eds.) Human-Computer Interaction INTERACT'90. Amsterdam New York: North-Holland. 357-361

Card, S K / Moran, T P / Newell, A, 1983: The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum.

Carroll, J, 1988: Integrating Human Factors and Software Development. In: Soloway, E / Frye, D / Sheppard, S B (eds.) Proceedings of CHI'88 "Human Factors in Computing System". New York: ACM. 157-159

Chin, J P / Diehl, V A / Norman, K L, 1988: Development of an instrument measuring user satisfaction on the human-computer interface. In: Soloway, E / Frye, D / Sheppard, S B (eds.) Proceedings of CHI'88 "Human Factors in Computing System". New York: ACM. 213-218

Crellin, J / Horn, T / Preece, J, 1990: Evaluating Evaluation: A Case Study of the Use of Novel and Conventional Evaluation Techniques in a Small Company. In: Diaper D et al. (eds.) Human-Computer Interaction - INTERACT '90. Amsterdam: Elsevier Science. 329-335

DIN 66 234, Teil 8, 1988: Bildschirmarbeitsplätze - Grundsätze ergonomischer Dialoggestaltung. Beuth-Verlag GmbH, Burggrafenstarße 6, D-1000 Berlin 30.

Dutke, S, 1988: Lernvorgänge bei der Bedienung eines Textkommunikationssystems. Frankfurt Bern New York Paris: Peter Lang.

Dzida, W, 1987: On tools and interfaces. In: Frese, M / Ulich, E / Dzida, W (eds.) Psychological Issues of Human Computer Interaction in the Work Place. Amsterdam: Elsevier (North-Holland). 339-355

Dzida, W, 1992 (in press): A methodological framework for software-ergonomic evaluation. In: van der Veer, G / Arnold, A G (eds.) Proceedings of "Interacting with Computers - Preparing for the Nineties". Amsterdam: Elsevier.

Dzida, W / Freitag, R / Hoffmann, R / Valder, W, 1990: Bridging the gap between task design and interface design In: Diaper, D / Gilmore, D / Cockton, G / Shackel, B (eds.) Human-Computer Interaction INTERACT'90. Amsterdam New York: North-Holland. 239-245

Dzida, W / Herda, S / Itzfeldt, W D, 1978: User-perceived quality of interactive systems. IEEE Transactions on Software Engineering SE-4(4):270-276

Eason, K, 1981: An annoted bibliography of user-friendly systems. In: Murray, G (ed.) User-friendly systems. Maidenhead, UK: Infotech International.

Evans, M W / Marciniak, J J, 1987: Software Quality Assurance and Management. New York: Wiley.

Fleischer, A G / Becker, G / Knabe, K P / Rademacher, U, 1984: Analyse der Augen- und Kopfbewegungen bei der Textverarbeitung. Zeitschrift für Arbeitswissenschaft 38(3):156-160

Greif, S / Gediga, G, 1987: A critique and empirical investigation of the "one-best-way-models" in human-computer interaction. In: Frese, M / Ulich, E / Dzida, W (eds.) Psychological Issues of Human Computer Interaction in the Work Place. Amsterdam: Elsevier (North-Holland). 357-377

Greutmann, T / Ackermann, D, 1989: Zielkonflikte bei Software-Gestaltungskriterien. In: Maaß, S / Oberquelle, H (Hrsg.) Software-Ergonomie '89 "Aufgabenorientierte Systemgestaltung und Funktionalität". Stuttgart: Teubner. 144-152

Gunsthövel, D / Bösser, T, 1991: Predictive metrics for usability. In: Bullinger, H-J (ed.) Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam London New York: Elsevier. 666-670

Hacker, W / Matern, B, 1980: Methoden zum Ermitteln tätigkeitsregulierender kognitiver Prozesse und Repräsentationen bei industriellen Arbeitstätigkeiten. In: Volpert, W (ed.) Beiträge zur psychologischen Handlungstheorie. (Schriften zur Arbeitspsychologie, Vol. 28; Ed.: E Ulich). Bern: Huber. 29-49

Holleran, P A, 1991: A methodological note on pitfalls in usability testing. Behaviour and Information Technology 10(5):345-357

Jüptner H, 1991: Stand der europäischen und internationalen Normung im Bereich der Ergonomie. Zeitschrift für Arbeitswissenschaft, 45(4):213-215

Karat, J / Bennett, J, 1991: Modelling the user interaction methods imposed by designs. In: Tauber, M J / Ackermann, D (eds.) Mental Models and Human Computer Interaction 2. Amsterdam: Elsevier (North-Holland). 257-269

Kieras, D / Polson, D, 1985: An approach to the formal analysis of user complexity. International Journal of Man-Machine Studies 22:365-394

Kirakowski, J / Corbett, M, 1988: Usability and measurement. In: Bullinger, H-J et al. (eds.) Information Technology for Organizational Systems. Amsterdam: Elsevier. 153-157

Kirakowski, J / Corbett, M, 1990: Effective methodoloy for the study of HCI. Amsterdam: North-Holland.

Kishi, N / Kinoe, Y, 1991: Assessing usability evaluation methods in a software development process. In: Bullinger, H-J (ed.) Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam London New York: Elsevier. 597-601

Landauer, T K, 1988: Research methods in human-computer interaction. In: Helander, M (ed.) Handbook of Human-Computer Interaction. Amsterdam: Elsevier. 905-928

Lienert, G A, 1989: Testaufbau und Testanalyse. München Weinheim: Psychologie Verlagsunion.

Lim, K Y / Long, J B / Silcock, N, 1990: Integrating human factors with structered analysis and design methods: an enhanced conception of the extended Jackson System Development Method. In: Diaper, D / Gilmore, D / Cockton, G / Shackel, B (eds.) Human-Computer Interaction INTERACT'90. Amsterdam New York: North-Holland. 225-230

Macaulay, L / Fowler, C / Kirby, M / Hutt, A, 1990: USTM: a new approach to requirements specification. Interacting with Computers 2(1):92-118

Martin, T M, 1973: Design of man-computer dialogues. Englewood Cliffs, N.J.: Prentice Hall.

Moll, T, 1987: Über Methoden zur Analyse und Evaluation interaktiver Computersysteme. In: Fähnrich, K-P (ed.) Software-Ergonomie. München Wien: Oldenbourg. 179-190

Müller-Holz, B / Aschersleben, G / Hacker, S / Bartsch, T, 1991: Methoden zur empirischen Bewertung der Benutzerfreundlichkeit von Bürosoftware im Rahmen von Prototyping. In: Frese, M / Kasten, C / Skarpelis, C / Zang-Scheucher, B (Hrsg.) Software für die Arbeit von morgen. Berlin Heidelberg New York: Springer. 409-420

Neal, A S / Simons, R M, 1984: Playback - a method for evaluating the usability of software and its documentation. In: Janda, A (ed.) Proceedings of CHI'83 "Human Factors in Computing Systems". Amsterdam New York Oxford: North-Holland. 78-82

Oppermann, R / Murchner, B / Paetau, M / Pieper, M / Simm, H / Stellmacher, I, 1988: Evaluation von Dialogsystemen - der software-ergonomische Leitfaden EVADIS. Berlin New York: de Gruyter.

Peercy, D E, 1981: A software maintainability evaluation methodology. IEEE Transactions on Software Engineering SE-7(4):343-351

Rauterberg, M, 1988: Video-rating - a reliable and valide evaluation method for the man-computer interaction (MCI). In: Adams, A S / Hall, R R / McPhee, B J / Oxenburgh, M S (eds.) Proceedings of 10th Congress of the International Ergonomics Association "Designing a better World" vol. II. 633-635

Rauterberg, M, 1991: Benutzungsorientierte Benchmark-Tests: eine Methode zur Benutzerbeteiligung bei Standardsoftwareentwicklungen. In: Spinas, P / Rauterberg, M / Strohm, O / Waeber, D / Ulich, E (Hrsg.) Projektbericht Nr. 3 zum Forschungsprojekt "Benutzerorientierte Softwareentwicklung und Schnittstellengestaltung (BOSS)". Institut für Arbeitspsychologie, Zürich: Eidgenössische Technische Hochschule.

Rauterberg, M, 1992: Partizipative Modellbildung zur Optimierung der Softwareentwicklung. In: Studer, R (ed.) "Informationssysteme und Künstliche Intelligenz", Informatik Fachbericht Nr. 303. Berlin Heidelberg New York: Springer. 113-128

Rauterberg, M, (in Vorbereitung): Spezifikation und Entwurf von Benutzungsoberflächen - das Konzept der interaktiven Aufsetzpunkte.

Rengger, R, 1991: Indicators of usability based on performance. In: Bullinger, H-J (ed.) Human-Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam: Elsevier. 656-660

Rudolph, E / Schönfelder, E / Hacker, W, 1987: Tätigkeits-Bewertungs-System für Geistige Arbeit. Psychodiagnostisches Zentrum, Humboldt-Universität, Oranienburger Strasse 18, D-O-1020 Berlin.

Sarris, V, 1990: Methodologische Grundlagen der Experimentalpsychologie. Band 1: Erkenntnisgewinnung und Methodik. Müchen Basel: Ernst Reinhard.

Schmid, U / Meseke, B, 1991: Deskription und Analyse komplexer Verhaltenssequenzen - Benutzerstrategien beim Arbeiten mit CAD-Systemen. Zeitschrift für experimentelle und angewandte Psychologie 38(2):307-320

Sherwood-Smith, M, 1989: The evaluation of computer-based office systems. unpublished dissertation. Department of Computer Science, University College Dublin, Belfield, Dublin 4 (Ireland).

Shneiderman, B, 1987: Designing the User Interface. Reading: Addison-Wesley.

Spencer, R H, 1985: Computer usability testing and evaluation. Englewood Cliffs: Prentice Hall.

Spinas, P, 1987: Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen. Dissertation. Institut für Arbeitspsychologie, Zürich: Eidgenössische Technische Hochschule.

Spinas, P / Waeber, D / Strohm, O, 1990: Kriterien benutzerorientierter Dialoggestaltung und partizipative Softwareentwicklung - eine Literaturaufarbeitung. In: Spinas, P / Rauterberg, M / Strohm, O / Waeber, D / Ulich, E (Hrsg.) Projektbericht Nr. 1 zum Forschungsprojekt "Benutzerorientierte Softwareentwicklung und Schnittstellengestaltung (BOSS)". Institut für Arbeitspsychologie, Zürich: Eidgenössische Technische Hochschule.

Tabachnik, B G / Fidell, L S, 1989: Using multivariate statistics. New York: HarperCollins.

Ulich, E, 1991: Arbeitspsychologie. Stuttgart: Poeschel.

Ulich, E / Rauterberg, M / Moll, T / Greutmann, T / Strohm, O, 1991: Task Orientation and User-Oriented Dialogue Design. International Journal of Human Computer Interaction 3(2):117-144

Upmann, R, 1989: Aufgaben- und nutzerorientierte Gestaltung rechnergestützter, kooperativer Arbeitssysteme in den indirekten Produktionsbereichen mittelständischer Maschinenbauunternehmen. In: Maass, S / Oberquelle, H (eds.) Software-Ergonomie '89. (Berichte des German Chapter of the ACM, Vol. 32). Stuttgart: Teubner. 110-122

Volpert, W, 1987: Psychische Regulation von Arbeitstätigkeiten. In: Kleinbeck, U / Rutenfranz, J (eds.) Arbeitspsychologie. (Enzyklopädie der Psychologie, Themenbereich D, Serie III, Vol. I). Göttingen: Hogrefe. 1-42

Vossen, P, 1991: Rechnerunterstützte Verhaltensprotokollierung und Protokollanalyse. In: Rauterberg, M / Ulich, E (Hrsg.) Posterband zur Software-Ergonomie '91. Institut für Arbeitspsychologie, Zürich: Eidgenössische Technische Hochschule. 181-188

Weidenmann, B, 1988: Psychische Prozesse beim Verstehen von Bildern. Bern Stuttgart Toronto: Huber.

Wiethoff, M / Arnold, A G / Houwig, E M, 1991: The value of psychophysiological measures in human-computer interaction. In: Bullinger, H-J (ed.) Human-Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam: Elsevier. 661-665

Whiteside, J / Bennett, J / Holtzblatt, K, 1988: Usability engineering - our experience and evolution. In: Helander, M (ed.) Handbook of Human-Computer Interaction. Amsterdam: Elsevier Science. 791-817

Whitefield, A / Wilson, F / Dowell, J, 1991: A framework for human factors evaluation. Behaviour and Information Technology 10(1):65-79

Ziegler, J, 1988: Aufgabenanalyse und Funktionsentwurf. In: Balzert, H. et al. (Hrsg.) Einführung in die Software-Ergonomie. Berlin New York: de Gruyter. 231-252

Zülch, G / Englisch, J, 1991: Procedures to evaluate the usability of software products. In: Bullinger, H-J (ed.) Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam London New York: Elsevier. 614-620