

# Zulassung und Gebrauchstauglichkeit technischer Medizinprodukte

(Ergonomie und Sicherheit von Medizinprodukten)

*Usability-Engineering II*  
*Verifizierung und Validierung*

**K. Radermacher, A. Janß**

**Lehrstuhl für Medizintechnik**

Helmholtz-Institut für Biomedizinische Technik der RWTH Aachen  
Pauwelsstraße 20, 52074 Aachen  
[www.meditec.hia.rwth-aachen.de](http://www.meditec.hia.rwth-aachen.de)

# Nutzungshinweis

---

Die im Folgenden bereitgestellten Folien stellen einen Auszug aus dem Foliensatz der Vorlesung bzw. Übung dar und erheben keinen Anspruch auf Vollständigkeit. Der regelmäßige Besuch der Veranstaltung bleibt daher weiterhin dringend empfohlen.

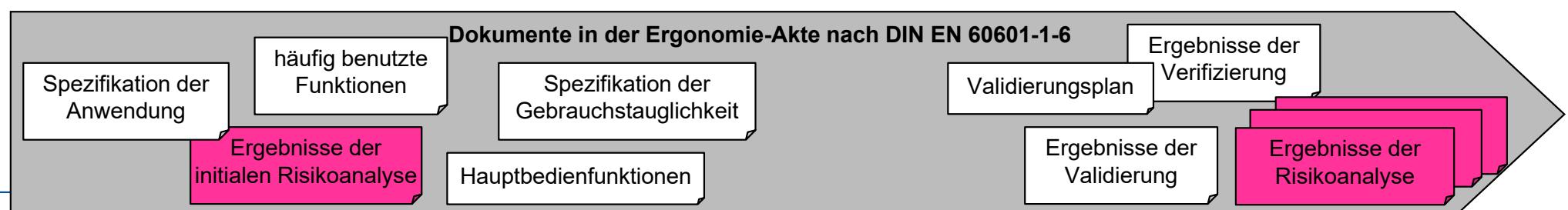
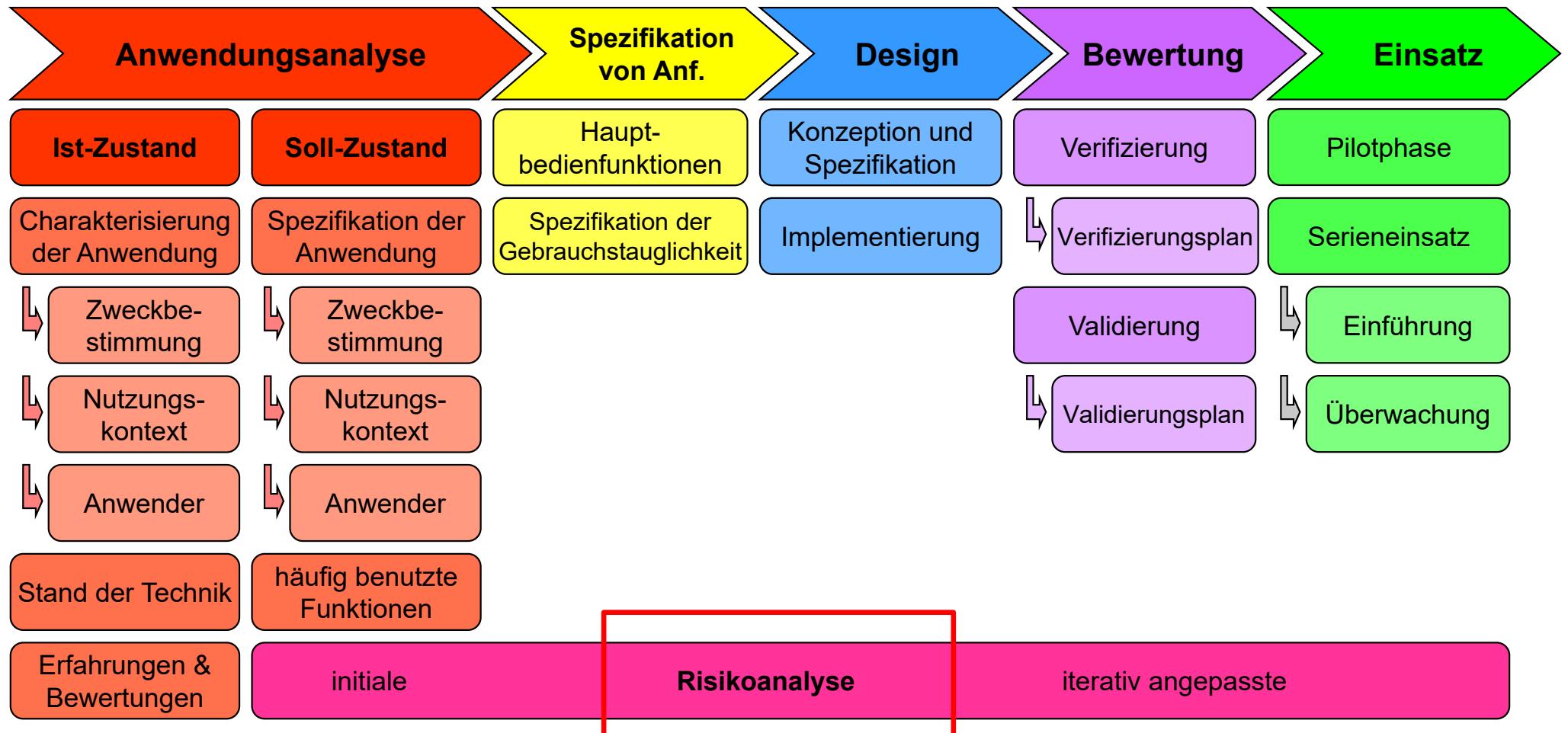
Die Folien und das auf ihnen dargestellte Material sind ausschließlich zur persönlichen Nutzung im Rahmen der Lehrveranstaltung bestimmt.

Eine Weitergabe oder Veröffentlichung – auch auszugsweise – ist nicht gestattet.

Die Folien sind mit größtmöglicher Sorgfalt erstellt worden.  
Für etwaige Fehler wird keine Haftung übernommen.

© mediTEC

# Usability-Engineering-Prozess nach DIN EN 60601-1-6



# Exkurs/Ergänzung: Human-Risiko-Analyse

## - Ziele (nach VDI 4006) -

---

Ziel ist eine Einschätzung der menschlichen Zuverlässigkeit und Verbesserung der Voraussetzung für zuverlässiges Handeln des Menschen im technischen System zur Steigerung der Sicherheit, Leistungsfähigkeit und Zuverlässigkeit des Gesamtsystems.

Hierzu werden...

- menschliche Handlungen qualitativ analysiert und mögliche Fehlhandlungen identifiziert
- Schwachstellen identifiziert und damit die Voraussetzung für geeignete Gegenmaßnahmen geschaffen
- die Zuverlässigkeit der menschlichen Handlung quantifiziert, um so u.a. die Wirksamkeit von Gegenmaßnahmen (z.B. einen Sicherheitsgewinn) innerhalb des Gesamtsystems einschätzen zu können

Bei der Analyse menschlicher Zuverlässigkeit muss zwischen fertigkeitsbasiertem, regelbasiertem und wissensbasiertem Handlungen unterschieden werden.

# Human-Risiko-Analyse

## - Generisches Vorgehen (nach VDI 4006) -

z.B.  
maixuse

?



Verfahrensschritte	Inhalt
<b>Festlegen des zu bewertenden Systems (Abschnitt 3)</b>	
<ul style="list-style-type: none"> <li>Identifikation und Festlegung von Aufgaben für die Systemkomponente „Mensch“</li> </ul>	<p>Beschreibung der zu untersuchenden und zu bewertenden menschlichen Handlungskomplexe; festgelegt werden:</p> <ul style="list-style-type: none"> <li>• Randbedingungen der Analyse</li> <li>• Qualitätsparameter der Analyse</li> <li>• Geltungsbereich der Analyse</li> <li>• Analysetiefe</li> </ul>
<b>Qualitative Analyse der Aufgaben (Abschnitt 4)</b>	
<ul style="list-style-type: none"> <li>Informationssammlung (Abschnitt 4.1)</li> <li>Analyse der zu bewertenden Aufgabe (Abschnitt 4.2)</li> <li>Identifikation möglicher Fehlhandlungen (Abschnitt 4.3)</li> <li>Identifikation der wirksamen leistungsbeeinflussenden Faktoren (PSF) (Abschnitt 4.4)</li> <li>Identifikation möglicher Fehlerkorrekturen (Abschnitt 4.5)</li> <li>Zusammenfassung der Ergebnisse der Aufgabenanalyse (Abschnitt 4.6)</li> </ul>	<p>Über die Aufgaben in der zu bewertenden Situation sind Informationen zu sammeln.</p> <p>Auf der Basis der Informationen sind Handlungsabläufe zu Tätigkeitelementen zu zerlegen, die einer qualitativen und quantitativen Bewertung und Repräsentation nach dem gewählten Modellansatz zugänglich sind.</p> <p>Die bei der Aufgabenanalyse ermittelten Einzelschritte sind hinsichtlich potenzieller menschlicher Fehlhandlungen auszuwerten.</p> <p>Die PSF bilden einen Satz von Modellparametern, die eine graduierter Anpassung der Zuverlässigkeitssparameter der Systemkomponente „Mensch“ an die aktuelle Charakteristik des MMS gestatten.</p> <p>Die im vorangegangenen Schritt unterstellten Fehlhandlungen können z.B. durch redundante oder diversitäre Maßnahmen oder die System- bzw. Prozesscharakteristik kompensiert werden.</p> <p>Die Aufgabenanalyse ist Basis für die nachfolgende quantitative Bewertung. Entsprechend wird das Ergebnis der Aufgabenanalyse in einem Handlungsmodell dokumentiert.</p>
<b>Quantitative Vorhersage (Abschnitt 5)</b>	
<ul style="list-style-type: none"> <li>Quantifizierung der Erfolgs- bzw. Fehlerwahrscheinlichkeiten</li> </ul>	<p>Quantifizierung der Nichtverfügbarkeiten der Einzelschritte unter Berücksichtigung der PSF auf Grundlage von modellspezifischen, generischen, empirischen Zuverlässigkeitssdaten oder Expertenschätzwerten; modellgemäße Berechnung der Erfolgs- bzw. Versagenswahrscheinlichkeit der Handlungskomplexe. Festgelegt werden</p> <ul style="list-style-type: none"> <li>• Versagensereignisse (Abschnitt 5.1)</li> <li>• Kenngrößen für menschliche Zuverlässigkeit (Abschnitt 5.2)</li> <li>• Prädiktive Quantifizierungsmethoden (Abschnitt 5.3)</li> <li>• Quantifizierungsprinzipien (Abschnitt 5.4)</li> <li>• Auswahl von Quantifizierungsverfahren (Abschnitt 5.5)</li> </ul>
<b>Wertung und Dokumentation der Ergebnisse einer HRA (Abschnitt 6)</b>	
<ul style="list-style-type: none"> <li>Wertung und Dokumentation der Ergebnisse</li> </ul>	<p>Wichtung der HRA-Ergebnisse in Bezug auf die Funktion des MMS; in Abhängigkeit von der Zielstellung der Analyse iterative Wiederholung der Analyse mit geänderten Parametern</p>

Quelle: VDI 4006 - Blatt 2 (2003): Menschliche Zuverlässigkeit – Methoden zur quantitativen Bewertung menschlicher Zuverlässigkeit

# Leitfaden zur Aufgabenanalyse nach VDI 4006

Aspekt des MMS	Beschreibung und beispielhafte Fragen
Situation	Organisatorische Faktoren wie Zeitangabe, Betriebszustand, mögliche Konsequenzen <i>Wie groß ist das Zeitfenster zur Bewältigung der Aufgabe?      In welchem Zustand befindet sich das technische System?      Welche Konsequenzen sind zu erwarten, falls der Mensch nicht oder fehlerhaft eingreift?      Innerhalb welcher Zeit darf ein bestimmter Eingriff nicht erfolgen?</i>
Aufgabe	Beschreibung der Aufforderungen an die Arbeitsperson(en), eine Arbeit nach einem vorgegebenen Arbeitsverfahren auszuführen (Soll-Zustand bzw. Soll-Funktion) <i>Welche Aufgaben haben die Personen, um die Störung des technischen Systems zu bewältigen?</i>
Auftragserteilung	Organisatorische und administrative Maßnahmen, die zur Einleitung der erforderlichen Aufgabe dienen (z.B. Prozeduren, Checklisten, mündliche Aufträge) <i>Welche organisatorischen Maßnahmen haben die Personen durchzuführen?      In welcher Form wird die Aufgabe erteilt (mündlich, schriftlich, geplant usw.)?</i>
Operateur	Personen, die an der Aufgabenerfüllung beteiligt sind, die Art der erforderlichen kognitiven Verarbeitung sowie deren Fähigkeiten und Fertigkeiten <i>Welche Personen sind an den Aufgaben beteiligt?      Wie vertraut sind sie mit den Aufgaben (Wissen, Fertigkeit, Übung)?</i>
Tätigkeit	Tätigkeiten, die bei der Durchführung der Handlungen erforderlich sind <i>Was ist unter den gegebenen Umständen zu tun (Kompromisse zwischen Ist- und Soll-Zustand)?      Wie lange dauern die Tätigkeiten?</i>
Auftragserledigung	Organisatorische Maßnahmen, um die Aufgabenerledigung zu protokollieren oder anderen Personen bekannt zu machen <i>Welche organisatorischen Aufgaben haben die Personen zu erledigen?      Wie sind die aufbauorganisatorischen Maßnahmen umgesetzt (Aufgaben und Funktionen der handelnden Personen)?      Wie ist die Ablauforganisation aufgebaut (wem muss wie über die Handlungen berichtet werden, mündliche, schriftliche Weitergabe der Informationen)?</i>
Rückmeldung	Informationen, an denen die Personen die Durchführung der Handlungen kontrollieren können <i>Welche Informationsmittel haben die Personen zur Verfügung, um die Notwendigkeit von Aufgaben (z.B. eine Störung des technischen Systems) zu erkennen oder deren Erfolg zu kontrollieren?</i>
Technisches System	Systeme oder Teilsysteme, welche die Tätigkeiten des Operateurs verarbeiten <i>Mit welchem technischen System oder Teilsystem haben die Personen zu tun?      Welche ergonomisch relevanten Eigenschaften hat das System (z.B. Dynamik, Vermaschung von Redundanzen)?</i>
Ergebnis	Realisierung einer Aufgabe (Ist-Zustand bzw. Ist-Funktion) <i>Zu welchen Änderungen an den oder Zuständen der Systemgrößen führen die Tätigkeiten (z.B. Druckanstieg, System nicht funktionsbereit)?</i>
Umgebung	Einflussfaktoren in der Umgebung bzw. den Räumlichkeiten (z.B. Beleuchtung) und deren Auswirkung auf die menschliche Zuverlässigkeit <i>An welchen Orten sind Handlungen erforderlich und welche Bedingungen herrschen dort vor?</i>

Quelle: VDI 4006 - Blatt 2 (2003): Menschliche Zuverlässigkeit – Methoden zur quantitativen Bewertung menschlicher Zuverlässigkeit

# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierung -

---

- **Menschliche Zuverlässigkeit:**  
„Fähigkeit des Menschen, eine Aufgabe unter vorgegebenen Bedingungen für ein gegebenes Zeitintervall im Akzeptanzbereich durchzuführen.“
- **Wahrscheinlichkeit einer menschlichen Fehlhandlung (Human Error Probability - HEP):**

$$HEP = \frac{\text{Zahl der beobachteten Fehler}}{\text{Zahl der Möglichkeiten für einen Fehler}}$$

- **HRP (Human Reliability Probability)** als Kennzahl für die menschliche Zuverlässigkeit:

$$HRP = 1 - HEP$$

Zuverlässigkeitsskenngrößen zur quantitativen Behandlung der „Systemkomponente Mensch“ weist wesentlich größere Unsicherheiten auf als Kenngrößen technischer Komponenten und hängt von Faktoren (PSFs) ab, deren Auswirkungen häufig nicht quantitativ bestimmt werden können.

# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierung -

- Bei der Bewertung der menschlichen Zuverlässigkeit wird eine logarithmische Normalverteilung angenommen. Sie ist durch Median ( $M$ ) und Streufaktor ( $K$ ) bestimmt. Für das 5. und 95. Percentil besteht folgender Zusammenhang:
  - 5. Percentil = Untere Grenze =  $M_{\text{HEP}} / K$
  - 95. Percentil = Obere Grenze =  $M_{\text{HEP}} \cdot K$
- Üblicherweise werden HEP in HRA-Verfahren als Median-Werte angegeben. Für den Zusammenhang von Median ( $M$ ) und Erwartungswert ( $EW$ ) gilt für logarithmische Normalverteilungen:

$$M_{\text{HEP}} = \frac{EW_{\text{HEP}}}{e^{(\ln K)^2 / 5,411}}$$

„Bei fehlender Datenbasis kann der Streufaktor ( $K$ ) aus den in: *Swain, A.D., H.E. Guttmann: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final Report. NUREG/CR-1278, Washington, DC (USA), 1983* angegebenen Richtlinien ermittelt werden.“

Quelle: VDI-Richtlinie VDI 4006: Menschliche Zuverlässigkeit - Ergonomische Forderungen und Methoden der Bewertung; Blatt 2: Methoden zur quantitativen Bewertung menschlicher Zuverlässigkeit

# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierung -

- Datenquellen für Angaben zur Fehlhandlungswahrscheinlichkeit:
  - Literaturdaten für vergleichbare Handlungen (generische Daten); Problem i.a.: Vergleichbarkeit / Validität
  - Beobachtungen von Fehlhandlungen in dem zu analysierenden bzw. in einem vergleichbaren (Labor-)Mensch-Maschine-System (systemspezifische Daten); Problem i.a.: statistische Basis wg. geringer Anzahl Fehlhandlungen, Validität
- Generische Wahrscheinlichkeiten für menschliche Fehler bei unterschiedlichen Aufgaben:

Aufgabenbeschreibung in Abhängigkeit von den situativen Anforderungen und der kognitiven Belastung	Fehlerwahrscheinlichkeit
Einfache und häufig durchgeführte Aufgaben bei geringem Stress und genügend zur Verfügung stehender Zeit in gewohnten Situationen (z.B. ohne ablenkende oder störende zusätzliche Einflüsse, gute Rückmeldung)	1*10E-3
Komplexe und häufig durchgeführte Aufgaben in gewohnten Situationen mit geringem Stress und genügend zur Verfügung stehender Zeit, wobei eine gewisse Sorgfalt bei der Durchführung notwendig ist	1*10E-2
Komplexere und regelmäßig durchgeführte Aufgaben in ungewohnten Situationen (z.B. ablenkende oder störende Einflüsse, unzureichende Rückmeldung), bei hohem Stress oder geringer zur Verfügung stehender Zeit	1*10E-1
Komplexere und selten durchgeführte Aufgaben in ungewohnten Situationen (z.B. ablenkende oder störende Einflüsse, unzureichende Rückmeldung), bei hohem Stress oder geringer zur Verfügung stehender Zeit	3*10E-1
Hochkomplexe oder sehr selten durchgeführte Aufgaben in ungewohnten Situationen (z.B. ablenkende oder störende Einflüsse, unzureichende Rückmeldung), bei sehr hohem Stress oder geringer zur Verfügung stehender Zeit	1*10E-0

Quelle: VDI-Richtlinie VDI 4006: Menschliche Zuverlässigkeit - Ergonomische Forderungen und Methode der Bewertung; Blatt 2: Methoden zur quantitativen Bewertung menschlicher Zuverlässigkeit

# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierungsmethoden -

- Methoden zur probabilistischen Risiko-oder Sicherheitsanalyse (Übersicht in [Hennings95]; alphabetische Reihenfolge):
  - ASEP (Accident Sequence Evaluation Program) [Swain 87]
  - ESAT (Expertensystem zur Aufgaben-Taxonomie) [Brauser 92]
  - HCR (Human Cognitive Reliability model) [Hannaman 84b]
  - HCR/ORE (HCR/ Operator Reliability Experiments) [Moieni 94]
  - HEART (Human Error Assessment and Reduction Technique) [Williams 88]
  - SLIM (Success Likelihood Index Methodology) [Embrey 84]
  - THERP (Technique of Human Error Rate Prediction) [Swain 83]

### Literatur:

- Swain, A.D.: Accident Sequence Evaluation Program Human Reliability Analysis Procedure. NUREG/CR-4772, Washington, DC (USA), 1987
- Brauser, K.: ESAT – ein neues Verfahren zur Abschätzung der menschlichen Zuverlässigkeit. In: K.P. Gärtner (Hg.), Menschliche Zuverlässigkeit: Maßnahmen zur Eingrenzung menschlichen „Versagens“, DGLR-Bericht 92-04, Bonn (D), 1992
- Hannaman, G.W. & Spurgin, A.J.: Human Cognitive Reliability Model for PRA Analysis. NUS-4531. NUS-Corp. San Diego, 1984
- Hennings, W., J. Mertens, B. Reer: Methodik der Risikoanalyse für Kernkraftwerke. Eine bewertende Bestandsaufnahme mit Blick auf regionale Sicherheitsplanung. vdf Verlag der Fachvereine, Zürich (CH), 1995
- Moieni, P., A.J. Spurgin, A. Singh: Advances in Human Reliability Analysis Methodology. Part I: Frameworks, Models and Data. Reliability Engineering and System Safety 44 (1994), pp. 27–55
- Williams, J.C.: A Data-Based Method for Assessing and Reducing Human Error to Improve Operational Performance. Proceedings of the IEEE 4th Conference on Human Factors in Power Plants, Monterey, California, June 6–9, 1988, Institute of Electronic and Electrical Engineers, New York (USA), 1988
- Embrey, D.E., P. Humphreys, E.A. Rosa, B. Kirwan, K. Rea: SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgement, Vol I: Overview of SLIM-MAUD, Vol. II: Detailed Analyses of the Technical Issues. NUREG/CR-3518, Washington, DC (USA), 1984
- Swain, A.D., H.E. Guttmann: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final Report. NUREG/CR-1278, Washington, DC (USA), 1983

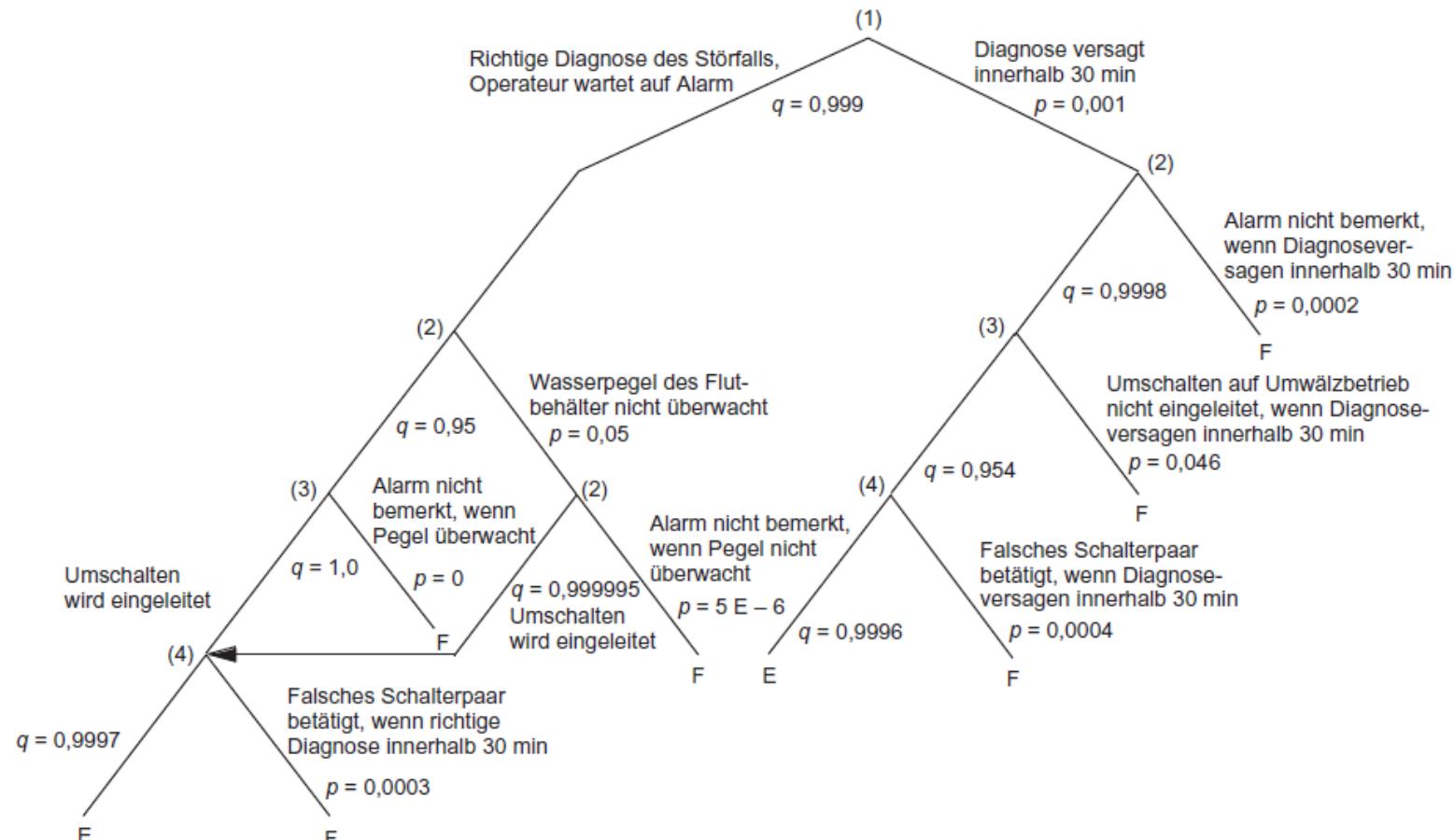
# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierungsmethoden -

### ■ Beispiel Ereignisbaummodell THERP

q: Erfolgswahrscheinlichkeit, p: Fehlerwahrscheinlichkeit, E: Erfolg, F: Fehler (Misserfolg). Übersetzt aus [Swain 83] (p. 21/22) und mit zusätzlichen Erläuterungen versehen, welche die von Swain und Guttmann beschriebene Analyse widerspiegeln.

Die in Klammern genannten Zahlen beziehen sich auf Beispiel eines hypothetischer Kühlmittelverlust-Störfall an einem Druckwasserreaktor bei dem ein Flutbehälter nach 32 Minuten soweit entleert ist, dass als Folge die für den Nachkühlbetrieb benötigten Pumpen ausfallen würden (s. VDI 4006)



# Menschliche Zuverlässigkeit

## - Prädiktive Quantifizierungsmethoden -

- Auswahl der Quantifizierungsprinzipien sollte aufgrund des Kenntnisstandes über die zu bewertende Situation sowie aufgrund einer Einschätzung des führenden Parameters für die menschliche Zuverlässigkeit geschehen.
  - **Aufgabenbezogene Quantifizierung** durch Aufgabenzerlegung und Vergleich mit bekannten Fehlerwahrscheinlichkeiten für Teilaufgaben, z.B. bei THERP (detailliert), ASEP; z.T. auch HEART
  - **Zeitbezogene Quantifizierung**, z.B. bei HCR, HCR/ORE (normierte Zeitskala), z.T. auch THERP (absolute Zeitskala)
  - **PSF-bezogene Quantifizierung**, z.B. bei HEART, SLIM, ESAT, z.T. auch HCR/ORE

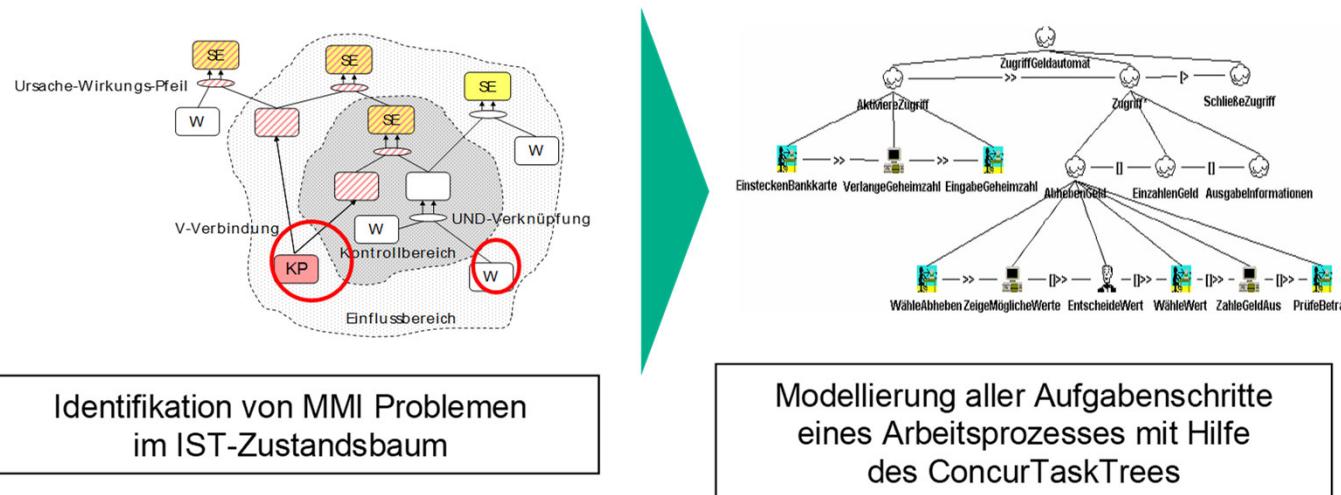
*Quantitative Analysemethoden aufgrund hoher Variabilität der Arbeitsprozesse und Randbedingungen in der Medizin meist nicht sinnvoll einsetzbar!*

**Weitere Informationen zu quantitativen HRA-Methoden:**

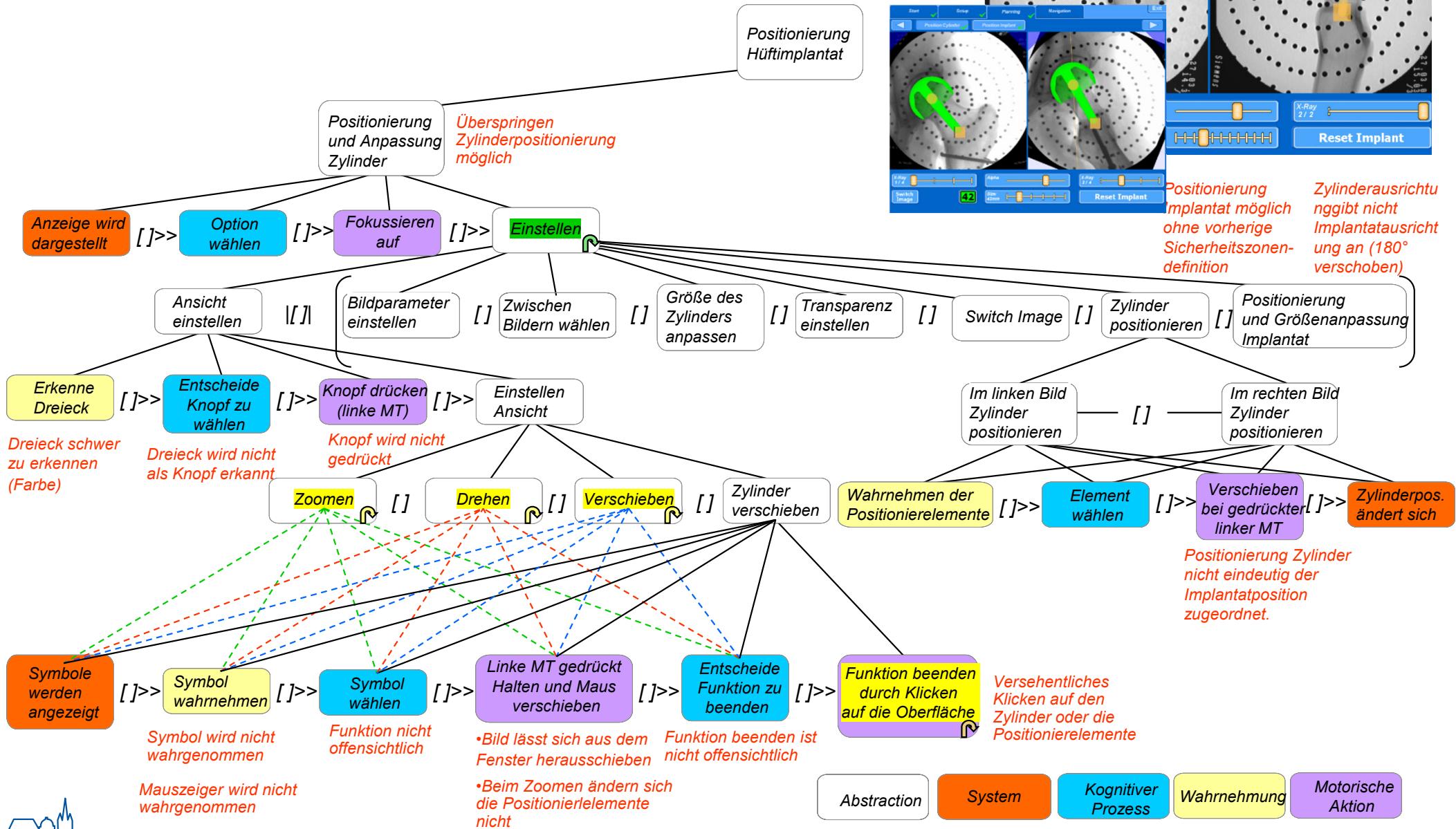
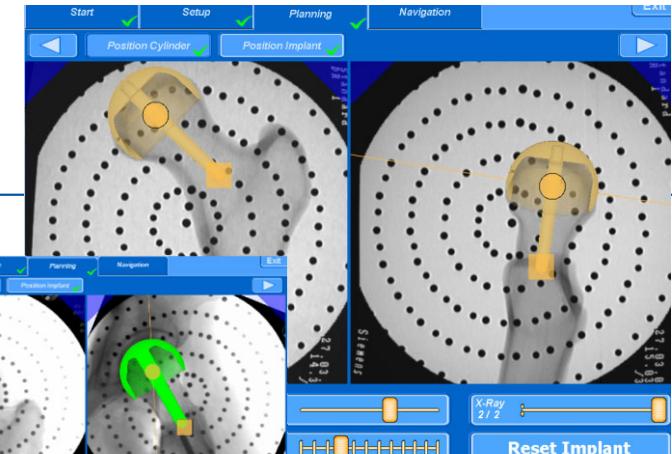
*VDI-Richtlinie VDI 4006: Menschliche Zuverlässigkeit - Ergonomische Forderungen und Methoden der Bewertung; Blatt 2: Methoden zur quantitativen Bewertung menschlicher Zuverlässigkeit*

# mAIxuse: Modellbasierte qualitative / semi-quantitative (Risiko-)Analyse der Mensch-Maschine-Interaktion

- softwareunterstützte, modellbasierte Analyse von Mensch-Maschine-Interaktionsprozessen und –Schnittstellen
- Identifikation und Bewertung von Fehlerpotentialen

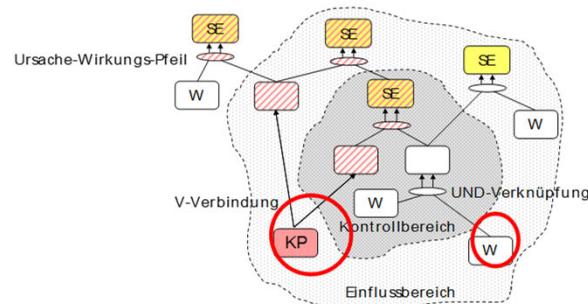


# Beispiel Operationsplanungssystem - qualitative HRA -

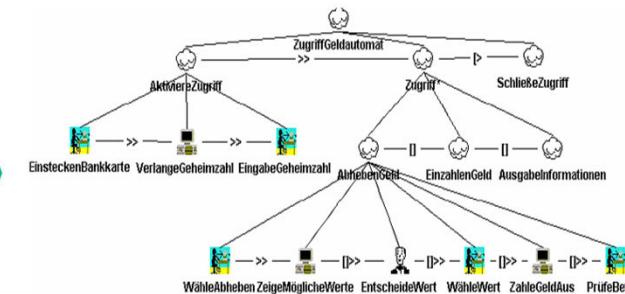


# mAIxuse: Modellbasierte qualitative / semi-quantitative (Risiko-)Analyse der Mensch-Maschine-Interaktion

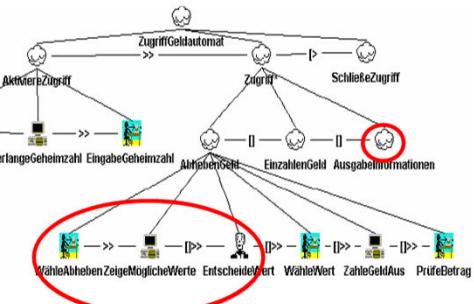
- softwareunterstützte, modellbasierte Analyse von Mensch-Maschine-Interaktionsprozessen und –Schnittstellen
- Identifikation und Bewertung von Fehlerpotentialen



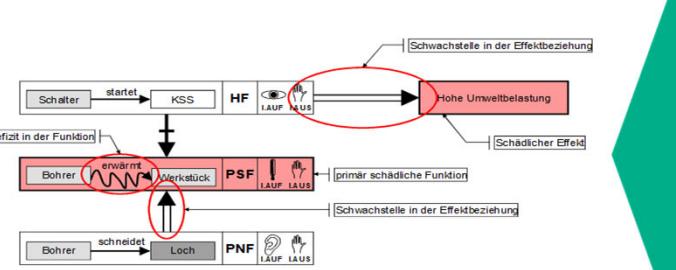
Identifikation von MMI Problemen im IST-Zustandsbaum



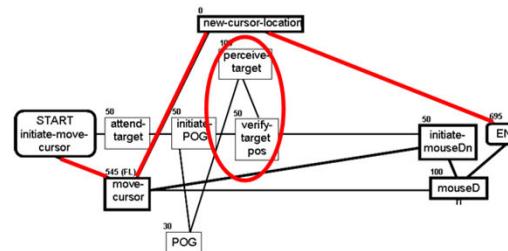
Modellierung aller Aufgabenschritte eines Arbeitsprozesses mit Hilfe des ConcurTaskTrees



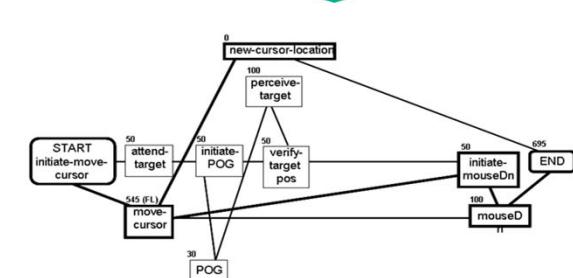
Analyse des ConcurTaskTrees auf potenzielle Problembereiche bei der MMI



Progressive Gebrauchstauglichkeitsprüfung durch spezifische Ausrichtung experimenteller Usability-Tests

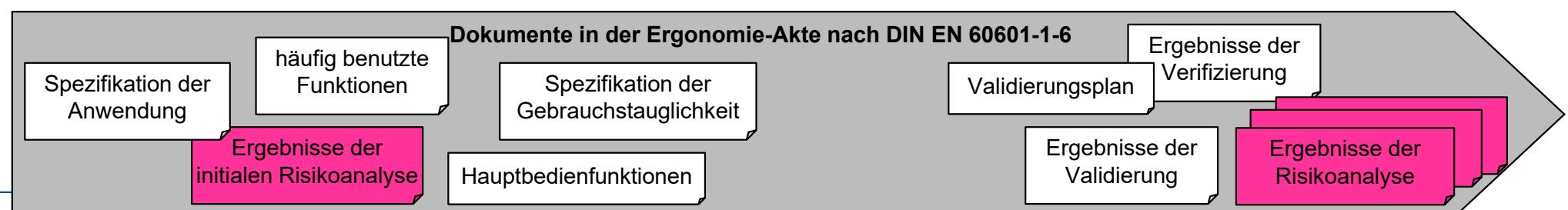
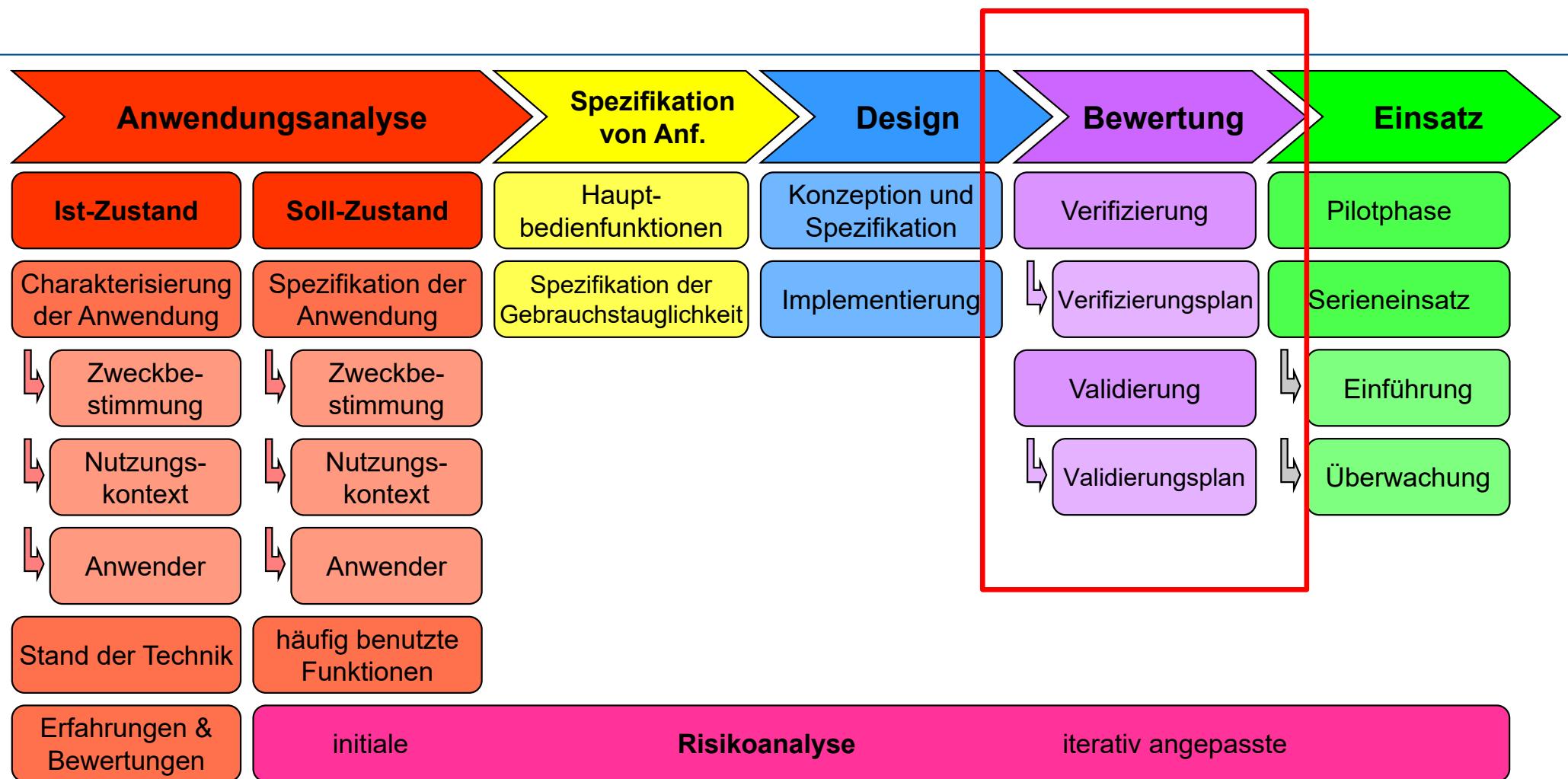


Ableitung der dem Modell innewohnenden Widersprüche und Konflikte bei der kognitiven Ressourcennutzung



Modellierung der Problembereiche mit Hilfe des CPM-GOMS Modells

# Usability-Engineering-Prozess nach DIN EN 60601-1-6



# Gliederung

---

- Verifikation und Validierung – Begriffsklärung
- Validierungsplanung
- Methoden der Gebrauchstauglichkeits-Evaluierung
  
- Beispiele interaktions- und benutzerzentrierter Evaluierungen

# Begriffsabgrenzung

## Verifizieren

- Duden: “durch Überprüfung **Richtigkeit** bestätigen”
  - DIN EN ISO 14971: „Bestätigung aufgrund einer Untersuchung und durch Bereitstellen eines Nachweises, dass **festgelegte Forderungen erfüllt** sind“.
  - DIN EN 60601-1-6: “Nachweis der Erfüllung der **Anforderungen** an die **Benutzer-Geräte-Schnittstelle**”  
=> z.B. *Erfüllung der Design-Prinzipien für Eingabedialoge*

Ziel der Verifizierung der Gebrauchstauglichkeit ist es demnach, sicherzustellen, dass das Produkt **die Spezifikation der Gebrauchstauglichkeit erfüllt**.

Beispiel: Führt das Betätigen des „Drucken“ Buttons führt zu einem formatierten Ausdruck einer Informationsseite?



# Begriffsabgrenzung

## Validieren

- Duden: “**gültig** machen”
- DIN EN 60601-1-6: “...Bestätigung durch Bereitstellung eines objektiven Nachweises, daß die Anforderungen **für einen spezifischen beabsichtigten Gebrauch oder eine spezifische Anwendung** erfüllt worden sind”  
=> *Erfüllung der Anwenderbedürfnisse & der Zweckbestimmung*

Die Validierung soll als **abschließende Bewertung** der Gebrauchstauglichkeit sicherstellen, dass **das „richtige Produkt“** gebaut wird und überprüfen, ob das Produkt die **Anwenderbedürfnisse** und die **Zweckbestimmung** erfüllt.

Beispiel: Ist eine Druckfunktion im intendierten Nutzungskontext sinnvoll und zielführend nutzbar?

“**Doing the right thing in the right way**”



# Gliederung

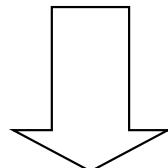
---

- Verifikation und Validierung – Begriffsklärung
- **Validierungsplanung**
- Methoden der Gebrauchstauglichkeits-Evaluierung  
(im Kontext der Entwicklungsphasen)
- Beispiele interaktions- und benutzerzentrierter Evaluierungen

# Validierungsplanung

**DIN EN 60601-1-6: “Validierung der Gebrauchstauglichkeit innerhalb der Zweckbestimmung”**

- Zweckbestimmung
  - Nutzungskontext
  - Anwendercharakterisierung
- } **Spezifikation der Gebrauchstauglichkeit**



## Szenariendefinition

aus Anwendungsspezifikation und Ergebnissen der Risikoanalyse

=> besonders häufige / Routineszenarien

=> worst case-Situationen / seltene aber wichtige/risikobehaftete Funktionen /  
erwartete Fehlerfälle

# Gliederung

---

- Verifikation und Validierung – Begriffsklärung
- Validierungsplanung
- Methoden der Gebrauchstauglichkeits-Evaluierung
  
- Beispiele interaktions- und benutzerzentrierter Evaluierungen

# Nutzung von Usability Methoden

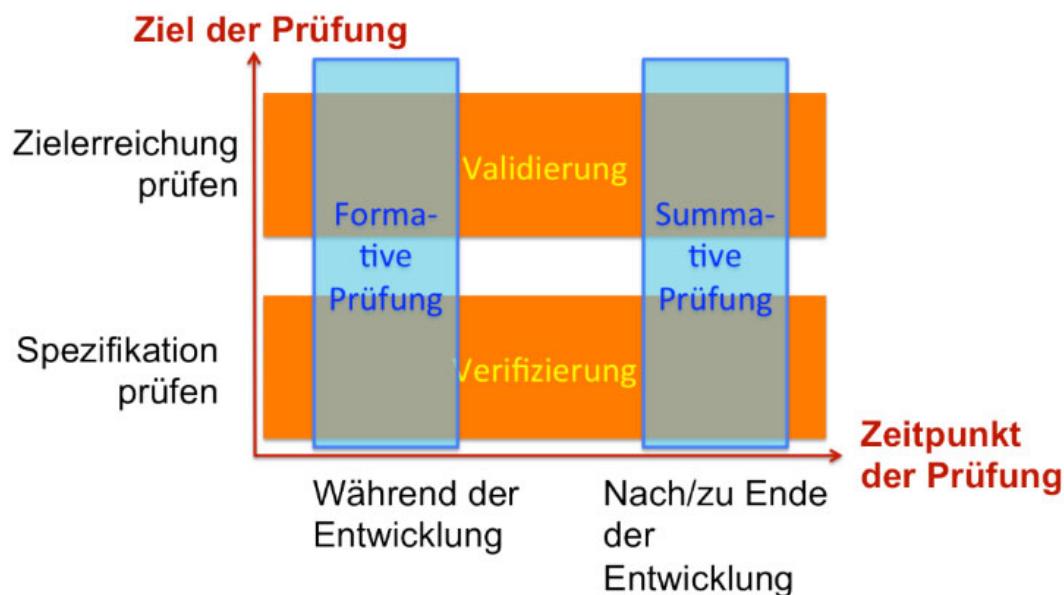
(nach Technical Report ISO/TR 16982:2002 - Ergonomics of human-system interaction — Usability methods supporting human-centred design)

## Die Faktoren, die die Wahl der Methoden beeinflussen:

- Stadium im Produktlebenszyklus (Konzeption, Entwicklung, Evaluation, Markt)  
(formative (ggf. fokussiert) vs. summativ)
- Eigenschaften d. Benutzer (u.a. Verfügbarkeit repräsentativer Nutzer u. Modelle)
- Nutzungsszenario (Realitätsnähe/Validität von vereinfachten Labor-Szenarien und Komplexität der Abbildung relevanter ePSFs)
- Merkmale der zu erfüllenden Aufgabe (Komplexität, Standardisierbarkeit,...)
- Merkmale des Produktes bzw. Systems (Komplexität, Software, Hardware,...)
- Zeitliche und ökonomische Randbedingungen des (Bewertungs-)Projektes
- Maß an Ergonomie-Fachwissen im Entwicklungs- oder Bewertungsteam

# Usability Verifizierung und Validierung vs. formative und summative Bewertung

- Die IEC 62366-1:2017 fordert statt einer **Usability Verifizierung und Validierung** (FDA) eine **formative (entwicklungsbegleitende) und summative (abschließende) Bewertung**.



## Stellenwert der formativen Bewertung

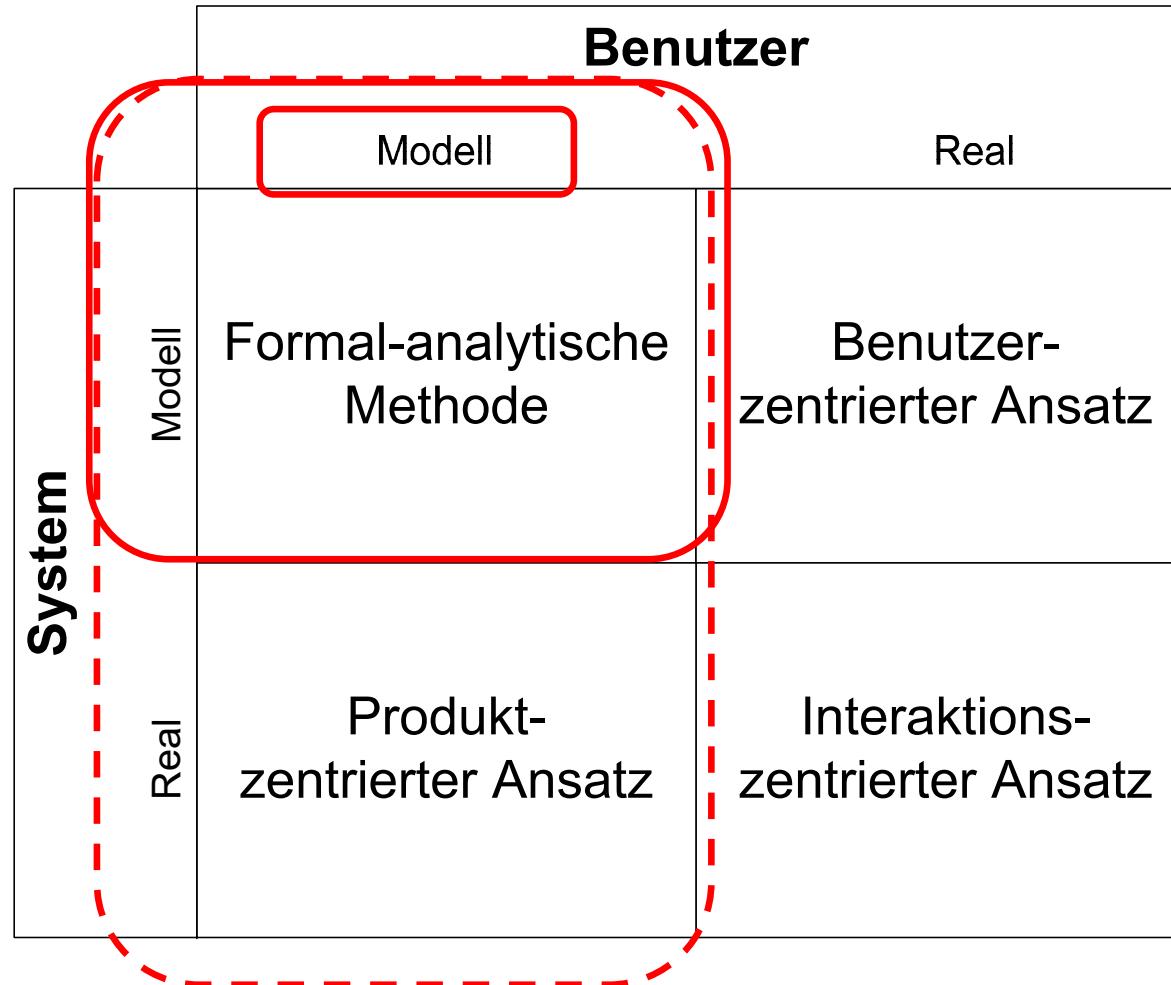
(nach FDA Guidance Dokument „Applying Human Factors and Usability Engineering to Medical Devices“)

- unbekannte benutzungsbezogene Gefährdungen, Benutzungsfehler und kritische Aufgaben identifizieren,
- Änderungen an der Benutzer-Produkt-Schnittstelle zeitnah umsetzen,
- die Wirksamkeit von risikominimierenden Maßnahmen an der UI überprüfen,
- die Notwendigkeit von Trainingsmaßnahmen bestimmen
- Inhalte für den Validierungsplan (für die summative Evaluierung) ableiten**

# Einordnung der Methoden zur Bewertung der Gebrauchstauglichkeit

		Benutzer	
		Modell	Real
System	Modell	Formal-analytische Methode	Benutzer-zentrierter Ansatz
	Real	Produkt-zentrierter Ansatz	Interaktions-zentrierter Ansatz

# Einordnung der Methoden zur Bewertung der Gebrauchstauglichkeit



# Mockups beim Prototyping von Medizinprodukten

---

## ■ **Mockups:**

meist einfache (UI-) Prototypen, mit denen Hersteller insbesondere die Gebrauchstauglichkeit von Medizinprodukten früh im Entwicklungsprozess – beim Prototyping – prüfen können

- Papier-Prototyp (ggf. auch mit Powerpoint oder CAD gezeichnet)
- Interaktive Prototypen z.B. mit Power-Point (ggf. mit <https://www.powermockup.com/>) oder in HTML; <https://balsamiq.com/index.html>; <http://www.flairbuilder.com/>; <http://www.foreui.com/>; ...)
- 3D Prototypen (z.B. 3D Druck) und/oder VR-Prototypen (ggf. auch Hybrid)

## ■ **Formative Evaluation:**

„User interface evaluation conducted with the intent to explore user interface design strengths, weaknesses, and unanticipated use errors.“

(IEC 62366-1:2015 - „Medical devices – Part 1: Application of usability engineering to medical devices“)

„Process of assessing, at one or more stages during the device development process, a user interface or user interactions with the user interface to identify the interface's strengths and weaknesses and to identify potential use errors that would or could result in harm to the patient or user.“ (FDA)

# Evaluierung der Gebrauchstauglichkeit

---

- **Plan für die (formative) Bewertung (FDA / IEC 62366-1:2017)**
  - Ziele und Prioritäten der Bewertung
  - zu bewertende Elemente der Benutzer-Produkt-Schnittstelle
  - Benutzungsszenarien und Aufgaben
  - Teilnehmer / Benutzer (nur für die Usability Validierung (summative Evaluierung)\* sind zwingend repräsentative Nutzer und repräsentativer Nutzungskontext nötig)
  - Methoden zur Datenerhebung und zur Datenauswerten
  - Verwendung der Ergebnisse
- **Methoden (Beispiele):**
  - Cognitive Walkthrough
  - Richtlinienbasierte Evaluierung z.B. vs. Style-Guides, Heuristiken, Design Pattern
  - Usability Test (benutzer-/interaktionszentriert) mit Prototypen unterschiedlicher Güte (low, high fidelity)
  - Interviews
  - Fragebögen

## Cognitive Walkthrough

- Inspektionsmethode, die von Usability-Experten durchgeführt wird  
=> Experten führen alle Handlungsschritte durch und protokollieren, ob der Anwender im gegebenen Kontext exakt diese Handlung ausführen würde.
- Voraussetzung: der Usability-Experte muss den Anwender richtig einschätzen können
- Erfassung dreier Problembereiche:
  - Punkte, an denen die Konzepte der Anwender und Entwickler über die Aufgaben nicht übereinstimmen
  - ungünstige Benennung von Bedienelementen
  - inadäquates Feedback
- Vor- und Nachteile:
  - + früh einsetzbar
  - + kostengünstig, da nicht unbedingt an Prototypen gearbeitet werden muss, sondern auch mit Bleistift und Papier gearbeitet werden kann
  - Eine Papiervariante erlaubt nicht die Möglichkeiten eines Prototyps, also sind solche Erkenntnisse mit Vorsicht zu behandeln.
  - es kann nur eine im Vorfeld definierte Bedienreihenfolge überprüft werden, nicht aber die beste Bedienreihenfolge ermittelt werden

# Methoden der Usability-Evaluierung

## Heuristische Evaluierung

- Inspektionsmethode, bei der Usability-Experten überprüfen, ob eine Bedienoberfläche bzw. ein Produkt bestimmten Usability-Prinzipien, so genannten Heuristiken, entspricht  
=> Experten nehmen Sicht des späteren Anwenders ein (**anhand Modell den Benutzers**)
- Heuristiken: allgemein anerkannte Prinzipien bzw. Kriterien und Regelsätze,
- Jeder Verstoß gegen ein Kriterium wird als potentielles Usability-Problem gewertet, wenngleich es keinen zwingenden kausalen Zusammenhang zwischen Verstoß und Problem gibt
- Vor- und Nachteile:
  - + früh einsetzbar ohne Prototyp
  - + schnell und kostengünstig
  - + qualitativ gute Ergebnisse
  - misst nicht in der wirklichen Gebrauchsumgebung
  - typische Anwender sind an der Bewertung nicht beteiligt
  - oft sind existierende Heuristiken zu allgemein
  - die evtl. notwendige Entwicklung eigener Heuristiken ist sehr zeitintensiv
  - subjektiv

# Beispiel für Heuristiken

## Zehn generelle Heuristiken nach Nielsen (1994)

1. **Sichtbarkeit des Systemstatus:** Das System sollte den Anwender durch angemessene Rückmeldungen zu jeder Zeit darüber informieren, in welchem Zustand es sich befindet.
2. **Übereinstimmung zwischen Systemen und realer Welt:** Das System sollte die Sprache des Anwenders verwenden. Darin verwendete Konzepte, Worte und Phrasen sollten dem Anwender aus seiner Umgebung vertraut sein.
3. **Benutzerkontrolle und Freiheit:** Anwender verwenden Funktionen aus Versehen oder benötigen einen Notausgang, um auf einfache Art und Weise einen ungewollten Systemzustand zu beenden.
4. **Konsistenz und Standards:** Anwender sollten nicht dadurch verwirrt werden, dass verschiedenen Bezeichnungen für das gleiche Konzept stehen bzw. verschiedene Befehlschritte zum gleichen Ergebnis führen. Bei der Gestaltung der Bedienoberfläche sollten die Gestaltungsregeln des jeweiligen Betriebssystems beachtet werden.
5. **Fehler vermeiden:** Besser als gute Fehlermeldungen ist eine Gestaltung, welche Fehler von vornherein verhindert. Mögliche Fehlerquellen sollten beseitigt werden oder durch entsprechende Dialoge abgefangen werden.

# Beispiel für Heuristiken

## Zehn generelle Heuristiken nach Nielsen (1994)

6. **Erkennen vor Erinnern:** Objekte, Aktionen oder Optionen sollten leicht erkennbar und erinnerbar sein. Es sollte nicht notwendig sein, dass sich Anwender bestimmte Dinge von einem Dialog zum anderen merken müssen. Hilfeinformationen sollten sichtbar oder im Bedarfsfall einfach zugänglich sein.
7. **Flexibilität und effiziente Nutzung:** Ein System sollte sowohl unerfahrene als auch erfahrene Anwender unterstützen und für beide Gruppen geeignete Bedienmöglichkeiten bieten. Wiederkehrende Aktionen sollten automatisierbar sein.
8. **Ästhetisches und minimalistisches Design:** Dialoge sollten nur für die Bedienung notwendige Informationen beinhalten. Zusätzliche bzw. unnötige Informationseinheiten in einem Dialog vermindern die Sichtbarkeit der wichtigen Informationen.
9. **Unterstützung beim Erkennen, Verstehen und Bearbeiten von Fehlern:** Fehlermeldungen sollten allgemein verständlich formuliert sein, das Problem klar beschreiben und eine mögliche Lösung aufzeigen.
10. **Hilfe und Dokumentationen:** Sollte das System nicht ohne eine Hilfe oder Dokumentation bedient werden können, sollte diese einfach zu finden und kontextbezogen sein. Hilfeinformationen sollten konkrete Bedienschritte und keine unnötigen Informationen beinhalten.

# Methoden der Usability-Evaluierung

## Richtlinienbasierte Evaluierung

- Gestaltungsrichtlinien (engl. Guidelines) bestehen aus einer Liste von Prinzipien, die zu gebrauchstauglichen Produkten führen sollen, wenn sie befolgt werden. Sie gehen im Gegensatz zu den Heuristiken auf sehr spezifische Details ein und geben dem Entwickler damit klare Gestaltungsanforderungen oder -regeln vor. Die entsprechende Umsetzung lässt sich anschließend von Usability-Experten objektiv bewerten.
- Ziele: Qualitative und quantitative Überprüfung eines Produktes auf die Einhaltung geltender Prinzipien bzgl. der Gebrauchstauglichkeit
- Vor- und Nachteile:
  - + Eigenschaften des Produktes können konkret anhand der Prinzipien überprüft werden
  - + relativ einfache Anwendung
  - keine umfassende Berücksichtigung des Nutzungskontextes und der Zielgruppe
  - relativ aufwändig

aus „Research-based Webdesign and Usability-Guidelines“ [Koyani et al., 2001]

### 5.3 Create a Positive First Impression of Your Site

**Guideline:** Treat your homepage as the key to conveying the quality of your site.

**Comments:** In terms of conveying quality, the homepage is probably the most important page on a Web site. One study found that when asked to find high quality Web sites, about half of the time participants looked only at the homepage. You will not get a second chance to make a good first impression on a user.

**Sources:** Amento, et al., 1999; Coney and Stehouder, 2000; Mahlke, 2002; Nielsen and Tahir, 2002.

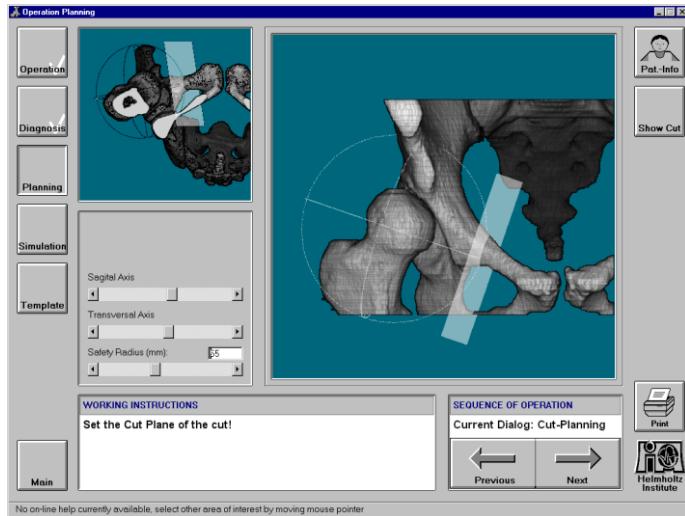
**Example:** This homepage creates a positive first impression:

- Tag line increases users' understanding of site;
- Key topic areas are presented in order of importance and are easy to scan; and
- Up-to-date news stories are available.



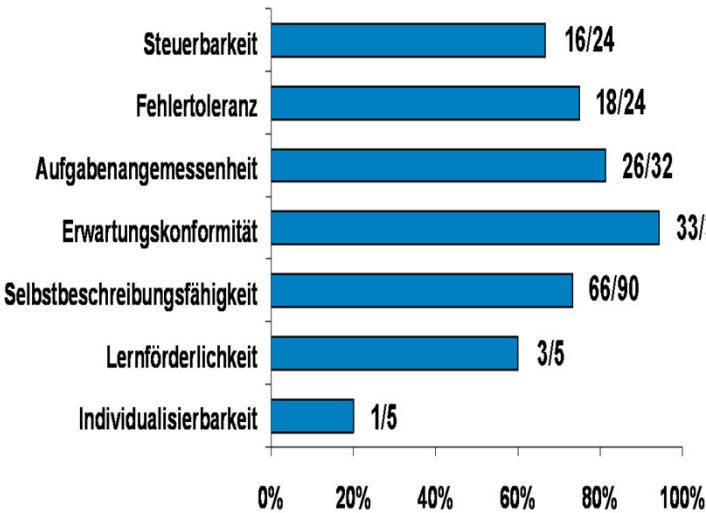
# Richtlinienbasierte Evaluierung

## Beispiel Operationsplanungssystem

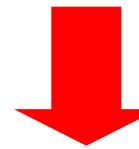


=> CeMPEG-EVAS

Datenbankgestützte richtlinienbasierte Evaluierung  
anhand von über 500 Kriterien aus Normen,  
Styleguides etc...



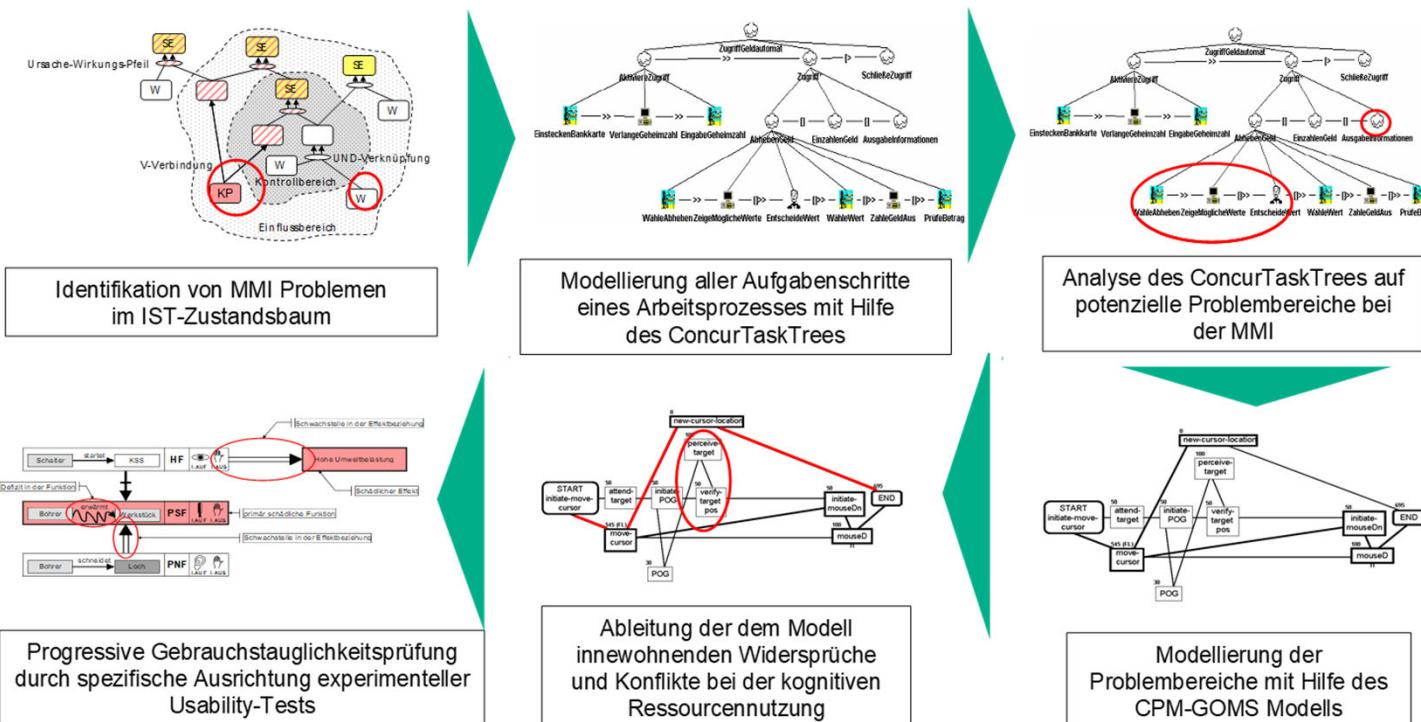
- Eindeutigkeit
- Komplexität
- Aufgabenangemessenheit
- Umweltgrenzwerte
- Anthropometrische Grenzwerte
- Kompatibilität
- Informationsdarstellung
- **Fehlertoleranz**
- Selbstbeschreibungsfähigkeit
- ...



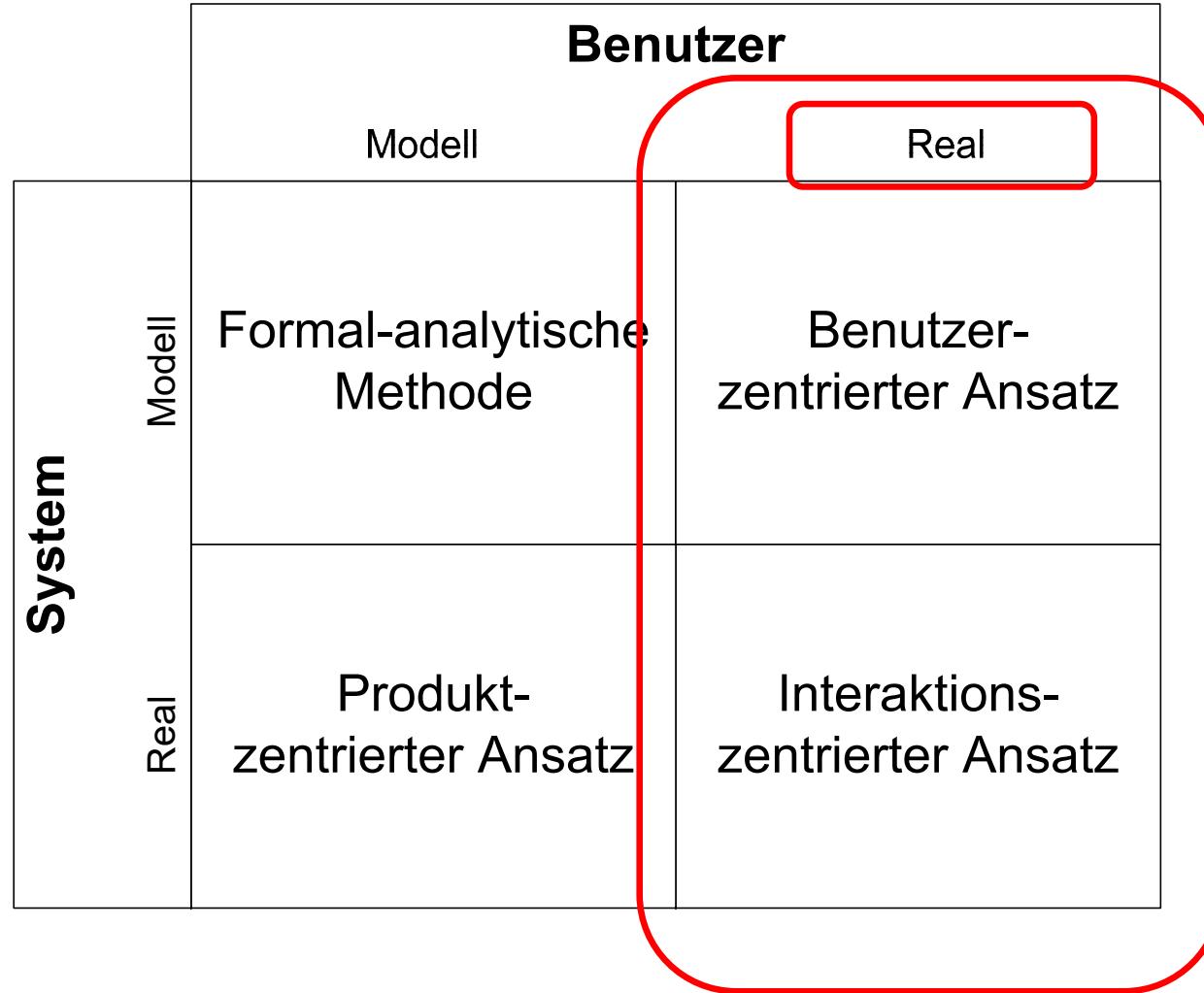
- Beschränkung und Überprüfung von Eingaben
- Unterstützung bei der Fehlersuche
- Anzeigen von Fehlern
- Hinweise zur Fehlerbehebung
- doppelte Sicherheit bei kritischen Eingaben

Kriterien aus ISO 6385, ISO 10075-2, ISO 9241, VDI/VDE 3850

- softwareunterstützte, modellbasierte Analyse von Mensch-Maschine-Interaktionsprozessen und –Schnittstellen
  - + Identifikation und qualitative / semi-quantitative Bewertung von Fehlerpotentialen
  - Validität der Modelle im spezifischen Nutzungskontext vs. realen benutzer- bzw. interaktionszentrierten Tests ggf. eingeschränkt



# Einordnung der Methoden zur Bewertung der Gebrauchstauglichkeit



*Für summative Evaluierung interaktionszentrierter Ansatz zwingend erforderlich!*

# Benutzer- und Interaktionszentrierte Evaluierung

## Validierungsumfeld

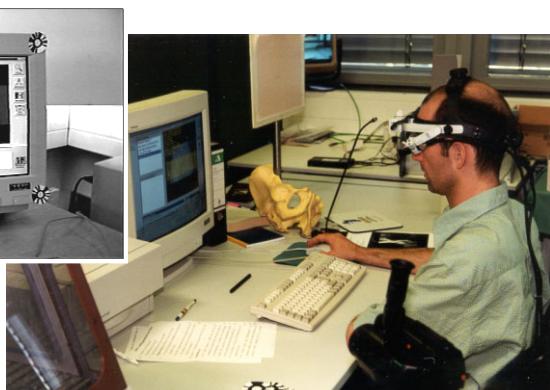
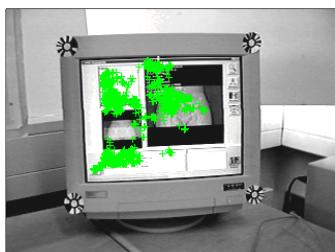
**DIN EN 60601-1-6: “Durchführung in Laboreinrichtung oder simulierter bzw. wirklicher Gebrauchsumgebung”**

- ? umfassende Einbeziehung des Nutzungskontextes
- ? Evaluierung einzelner Nutzungs-Aspekte oder integrierte Nutzung
- ? Abbildbarkeit komplexer Einflusszusammenhänge (Dienstplan, Alltag, Räume...)

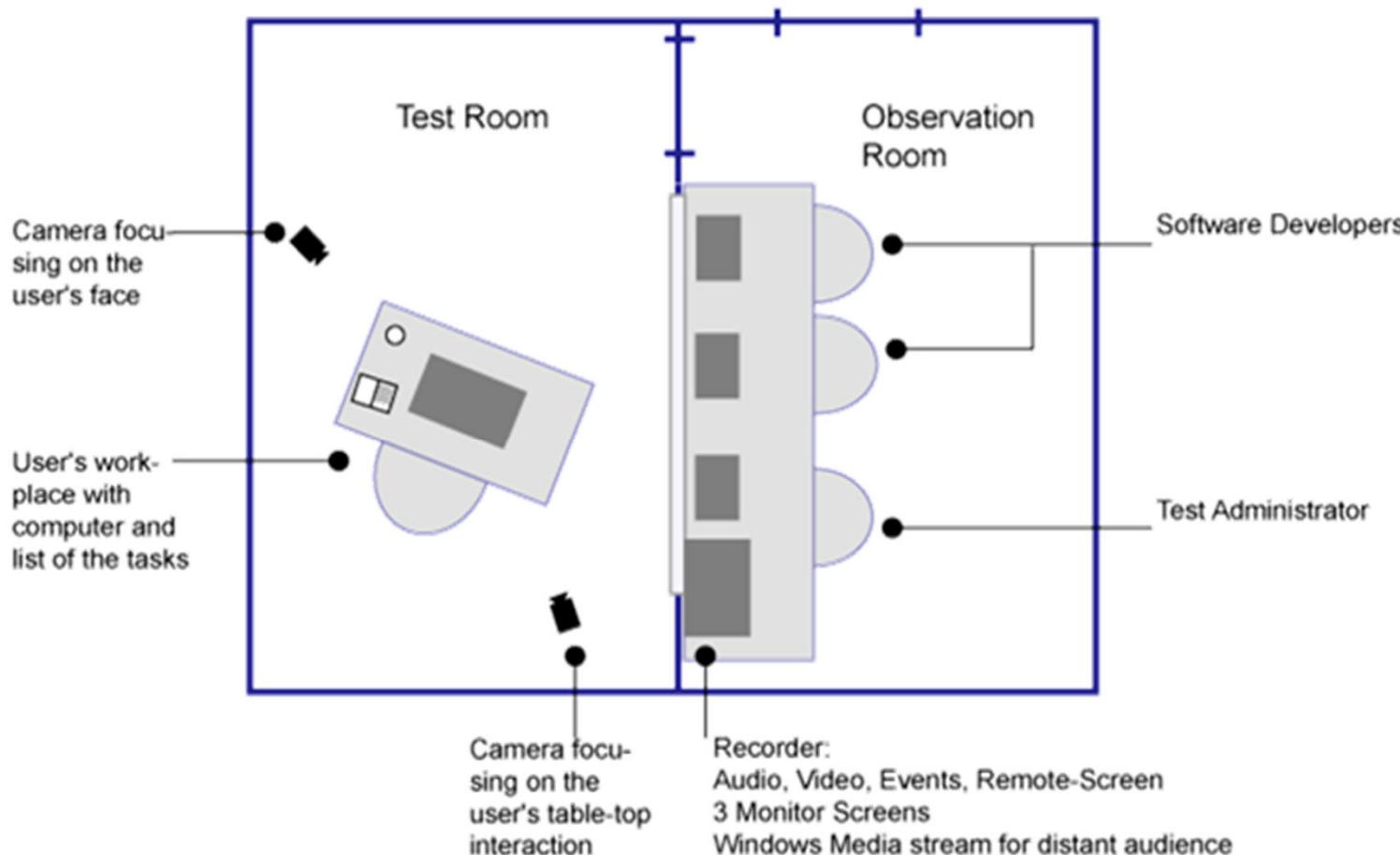
**Vorgehen am Centrum für Medizinprodukt ergonomie und –gebrauchstauglichkeit (CeMPEG) Aachen:**

=> Usability-Labor (Zweiraum-Konzept, Videobeobachtung...)

=> Evalierungs-OP Universitätsklinikum Aachen/Feldstudien

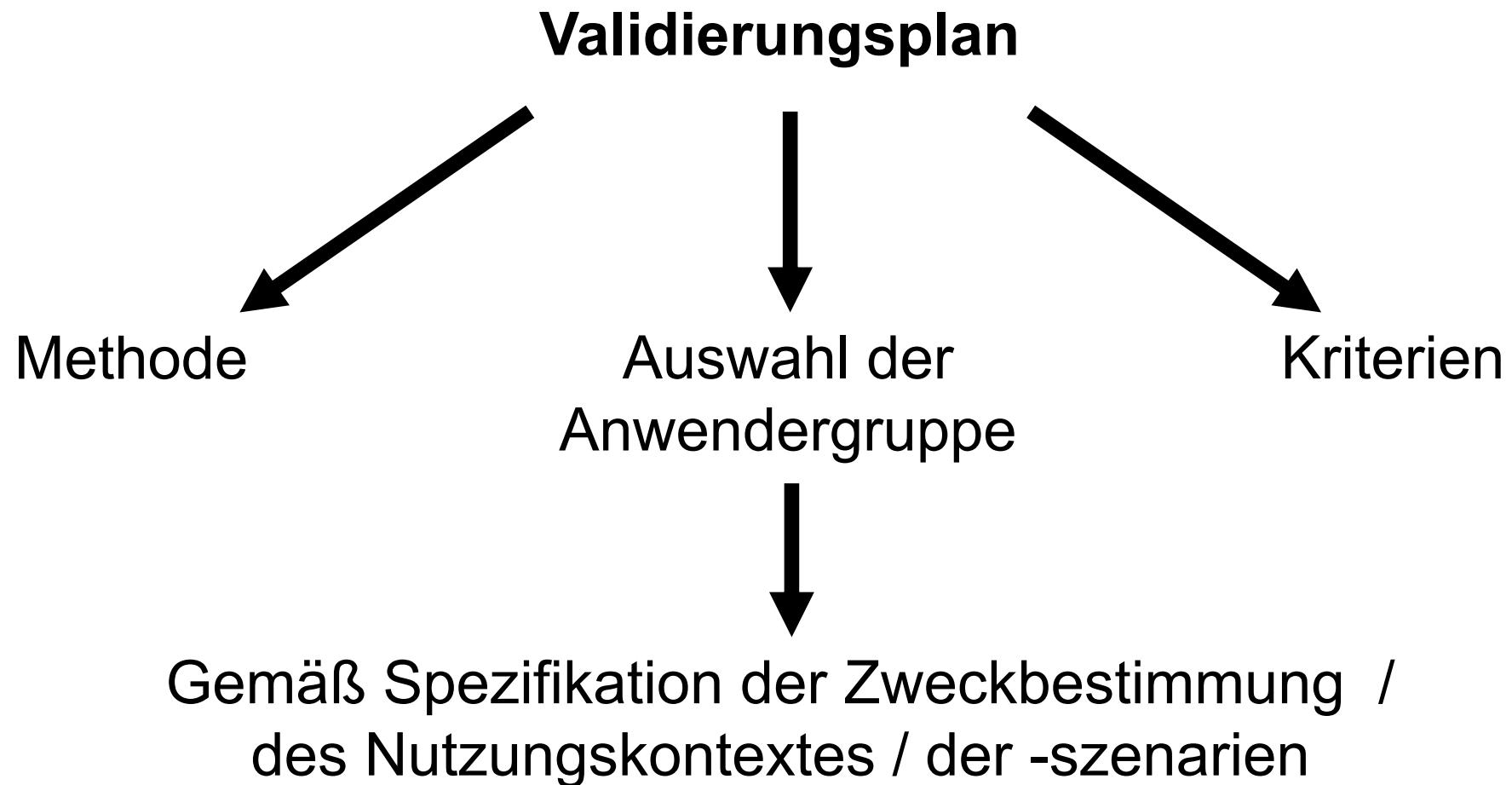


# Usability-Labor



# Validierungsplan

---



# Vorüberlegungen für Benutzer-/Interaktionszentrierte Usability-Tests

---

- Quantitative Usability Evaluierung
  - > Methodenauswahl & messbaren Kriterien
- Grundlegende Schritte für eine quantitative Evaluierung
  - 1) Festlegung welche Teile und Funktionen der Applikation/ der MMI getestet werden (Hauptbedienfunktionen !)
  - 2) Zieldefinition der Tests:
    - a) Entdecken/Beheben möglichst vieler Usability Probleme (Formative Eval.)
    - b) Usability Evaluierung im Vergleich mit Anforderungen (Summative Eval.)
    - c) Usability Vergleichs-/Konkurrenzprodukte
  - 3) Beschreibung von mindestens 3, i.a. 5-15...Aufgaben in **realistischen Szenarien** (unter Berücksichtigung der relevanten (e/i) **PSFs im Nutzungskontext!**)
  - 4) Ablaufplanung
  - 5) Akquisition von Testpersonen
    - 5-10 für iterative formative Tests entspr. 2a)
    - Für summative Tests bzw. 2b+c) typischerweise 10-20 (s.u.)
  - 6) Durchführung
  - 7) Auswertung und Dokumentation

# Anzahl an Probanden bei der Usability Validierung

- Anzahl hängt u.a. ab von: (s. J. Sauro: A Practical Guide to Measuring Usability, 2010)
  - Formative (iterativ Evaluierung; i.a. weniger Probanden) vs. summative Bewertung
  - Homogenität der Nutzergruppen
  - bei Produkt-/Anforderungsvergleich:
    - Sinnvolle Effektgröße (welcher Unterschied ist sinnvoll)
    - Konfidenzniveau: Wie sicher soll ein gemessener Unterschied tatsächlich vorhanden sein
    - Güte (Power): Wie sicher soll ein vorhandener Unterschied tatsächlich erkannt werden
- (Weitere Quelle u.a.: Jeff Sauro, James R. Lewis: What sample sizes do we need? (Part 1: summative studies; Part 2: formative studies; <https://www.sciencedirect.com/topics/computer-science/sample-size-estimation>), in Jeff Sauro, James R. Lewis: Quantifying the User Experience (Second Edition), 2016 (<https://www.sciencedirect.com/book/9780128023082/quantifying-the-user-experience>))
- Die IEC 62366 verlangt bei der Validierung der Gebrauchstauglichkeit repräsentative Anwender ohne eine konkrete Zahl zu nennen
- Die **FDA** empfiehlt im Human Factors Engineering Guide **5-10 für formative und 15 Teilnehmer pro Benutzergruppe für summative Evaluierung.**

(referenziert in ihrem Guidance Dokument „Applying Human Factors and Usability Engineering to Optimize Medical Device Design“ eine Studie von Faulkner, 2003 (vereinfachte (!) Kernaussage: mit fünf repräsentativen Nutzern werden im Mittel 85% der Nutzungsprobleme, mit zehn bereits ca. 95% gefunden)

Table B-1. Percentage of Total Known Usability Problems Found in 100 Analysis Samples  
(Faulkner, 2003).

No. users	Min. % Found	Mean % Found	SD	SE
5	55	85.55	9.2957	.9295
10	82	94.69	3.2187	.3218
15	90	97.05	2.1207	.2121
20	95	98.4	1.6080	.1608
30	97	99.0	1.1343	.1051

# Vorüberlegungen zu Ablauf von Usability Tests

- Testdauer pro Proband max. 1- max. 2 h für 5-15 Aufgaben (je nach Komplexität)
- Aufklärung der Probanden (ggf. Ziele der Untersuchung/des Projektes, Ergebnisverwertung etc.); ggf. Unterzeichnung Einwilligungserklärung
- Standardisierte Einweisung in die Aufgaben! (ggf. über Video bzw. Tutorial; z.B. wenn Selbstbeschreibungsfähigkeit getestet werden soll die Bedienung dieser Funktionen z.B. ggf. nur generisch, d.h. nicht per Clicks darstellen; notwendige Aufgaben möglichst konkret vorgeben; ggf. in Ausgangssituation des Szenarios einführen („stellen Sie sich vor Sie hätten Herzrasen und wollten Ihren Blutdruck messen,...“; Abhängigkeiten von Folgeaufgaben (wenn möglich) vermeiden,...))
- Erfassung der Messparameter (papierbasiert bzw. ggf. automatisiert) , u.a.:
  - Usability Probleme, Fehlhandlungen, Workarounds, ...
  - Zeiten pro Aufgabe, Fertigstellungsralten,
  - ggf. (Blick-)Bewegungsanalysen und Eventlogging (Clicks,...)
  - ggf. physiologische Beanspruchungsverlaufsmessungen: Hautleitwert, Herzrate,...)
  - Subjektive Belastungs-/Beanspruchungsmessung vor vs. nach Aufgabe (Skalen) bzw. Test (ggf. Fragebögen; s.u.)
  - Benutzerzufriedenheitsmessung (vor/)nach jeder Aufgabe (kurz!; i.a. Skalen; s.u.)
  - Benutzerzufriedenheitsmessung (vor/)nach Test (i.a. Fragebögen; s.u.)
  - Kommentare (ggf. begleitend (think-aloud,...) oder nach Aufgabe/Test
- Ggf. Durchführung von Pilottests um Probleme im Testablauf zu erkennen !

# Vorüberlegungen zu Ablauf von Usability Tests

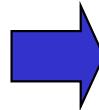
---

- Abhängigkeiten von Folgeaufgaben (wenn möglich) vermeiden (s.u.)
- Aufgabenreihenfolge (A, B, C, D,...) variieren um Lerneffekte auszugleichen (z.B.: Proband 1: A>B>C>D...; Proband 2: B>C>D>A...; Proband 3: C>D>A>B; ...oder Randomisierung,...); (vgl. J. Sauro: A Practical Guide to Measuring Usability, 2010)
- direkte Wiederholungen (A,A,...) vermeiden: A,B,C,D,A,B,C,D
- Keine Erläuterungen /Unterstützungen während Testablauf!
- wenn Unterstützung bei einer Aufgabe zum weiteren Testablauf unumgänglich ist, sollte dies (und die Häufigkeit) im Protokoll notiert werden (i.a. sollte diese Aufgabe in der Auswertung als nicht fertiggestellt gewertet werden!)

# Benutzerbasierte Evaluierung - Bewertungskriterien

## => Effektivität

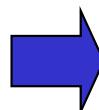
Wie gut kann die Aufgabe erfüllt werden?



- Messung der Bearbeitungsqualität und Bestimmung des Grades der Aufgabenerfüllung (Kriterien!!)

## => Effizienz

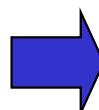
Mit welchen Ressourcenaufwand kann die Aufgabe erfüllt werden?



- Kennzahlbildung (z.B. effektive zu uneffektiven Schritten, Zeit / Schritte zur Aufgabenerfüllung)
- Dokumentation von „workarounds“

## => Erlernbarkeit

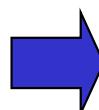
Wie gut und schnell können Anfänger den Umgang mit dem System lernen?



- Vergleich Zeiten & Probleme von Novizen zu Experten
- Anteile exploratives Erlernen / Nutzung von Tutorials und Hilfe
- Benutzerbefragung

## => Benutzerzufriedenheit

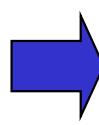
Wie zufrieden sind die Benutzer mit dem System?



- Skalen, Interviews & Fragebögen zu Zufriedenheit bzw. Diskomfort
- Methode des lauten Denkens

## => Systemzuverlässigkeit

Wie zuverlässig kann die Aufgabe erfüllt werden?



- Beobachtung, Erfassung und Klassifikation von Störungen
- Variabilität v. Effektivität und Effizienz
- Benutzerbefragung

# Messparameter für Effektivität und Effizienz (weitere Beispiele)

---

- Zeitaufwand für die Erledigung einer Aufgabe (Feldbeobachtungen auch ggf. Multimomentstudien zu beobachtbaren Zeitanteilen/Häufigkeiten)
- Erledigte Aufgaben pro vorgegebener Zeit
- Anzahl der Fehler
- Zeit für die Behebung von Fehlern
- Anzahl der genutzten Befehle
- Zeit für das Auffinden und Interpretieren von Informationen im Benutzerhandbuch
- Häufigkeit der Nutzung von Hilfsmaterialien (Dokumentation, Hilfesystem usw.)
- Anzahl der Fälle, in denen die Benutzeraufgabe abgebrochen wurde
- Anzahl der Abschweifungen
- Anzahl der Leerlaufzeiten (Unterscheidung von systembedingten Verzögerungen, Denkzeiten und Verzögerungen durch externe Faktoren!)
- Anzahl der gesamten Tastenanschläge/Klicks
- (Blickbewegungen: Fixationsdauer/-häufigkeit für Informationsaufnahme ?!?)
- Linkanalyse/Stringdiagramm/Harmonographie\*: Häufigkeit, Zeiten/Strecken für (räumliche) Ortswechsel f. Informationsaufnahme (?s.o.?) u. –ausgabe/Aktionen
- **Erlernbarkeit: Effektivität und Effizienz ...im zeitlichen Verlauf bei wiederholter Messung**
- **Zuverlässigkeit u.a.: Variation von Effektivität und Effizienz ...bei wiederholter Messung/Variation der Randbedingungen/...**

# Methoden der benutzer- und interaktionszentrierten Usability-Evaluierung

---

## Methode des lauten Denkens („think aloud“)

- Variante des Usability-Tests
- Das Produkt wird von Anwendern anhand realer oder realistischer Aufgaben erprobt. Der Anwender erhält eine Aufgabe und wird angehalten, seine Gedanken bzgl. der gerade ablaufenden Vorgänge, eigener Ziele mit einem Handlungsschritt und der Gründe für eine Handlung während der Bearbeitung laut zu äußern. Ein Usability-Experte beobachtet den Probanden und fragt für den Fall, dass der Proband in seinen Äußerungen stockt, gezielt nach.
- Vor- und Nachteile:
  - + es wird sofort ersichtlich, an welcher Stelle und warum Probleme auftauchen
  - + erkennbar, in welcher Art und Weise ein Anwender das Produkt / den Prototyp nutzt
  - verträgt sich nicht mit Performance-Messungen
  - Doppelbelastung des Anwenders
  - Senkung der Bearbeitungsgeschwindigkeit
  - Anwender fühlen sich bei der Äußerung unwohl und verstummen immer wieder
  - Vorgehen des Anwenders wird beeinflusst

# Methoden der benutzer- und interaktionszentrierten Usability-Evaluierung

---

## Konstruktive Interaction (Constructive Interaction)

- Variante des Usability-Tests
- Produkt wird von Anwendern anhand realer oder realistischer Aufgaben erprobt
- Zwei Probanden arbeiten gemeinsam an einer Aufgabe. Sie äußern – ähnlich der Methode des Lauten Denkens – ihre Erwartungen, Vorstellungen und Gedanken im Gespräch mit dem Partner. So ergibt sich eine Art Gespräch, die Probanden ergänzen sich. Ein Usability-Experte beobachtet beide Probanden
- Vor- und Nachteile:
  - + komfortabler als die Methode des Lauten Denkens
  - + natürliche Gesprächssituation führt zu gesteigerter Redebereitschaft beim Probanden
  - Senkung der Bearbeitungsgeschwindigkeit
  - Probanden haben häufig unterschiedliche Herangehensweisen und Lösungsstrategien
  - Beeinflussung der Aufgabenerfüllung durch Zusammenarbeit und Kommentierung
  - Auswertung ist zeitintensiv

# Methoden der benutzer- und interaktionszentrierten Usability-Evaluierung

---

## Retrospektives Testen (Retrospective Testing)

- Variante des Usability-Tests
- Das Produkt von Anwendern anhand realer oder realistischer Aufgaben erprobt, die Tests werden auf Video aufgezeichnet. Anschließend wird der Proband von einem Usability-Experten anhand des Videobandes zu seinem Test befragt bzw. soll seine Handlungen kommentieren.
- Vor- und Nachteile:
  - + Proband wird während der Bearbeitung der Aufgabe nicht doppelt belastet
  - + der Experte kann zusätzliche Fragen stellen, die nicht Gegenstand des Tests waren
  - + liefert häufig konkrete Verbesserungsvorschläge
  - + da die Befragung direkt im Anschluss an den Test durchgeführt wird und der Proband sich selber beobachten kann, ergeben sich keine falschen Erinnerungen oder Fehldeutungen durch den Experten
  - für den Probanden zeitintensiv
  - Notwendigkeit der Auswahl geeigneter Videosequenzen ist aufwändig

# Methoden der benutzer- und interaktionszentrierten Usability-Evaluierung

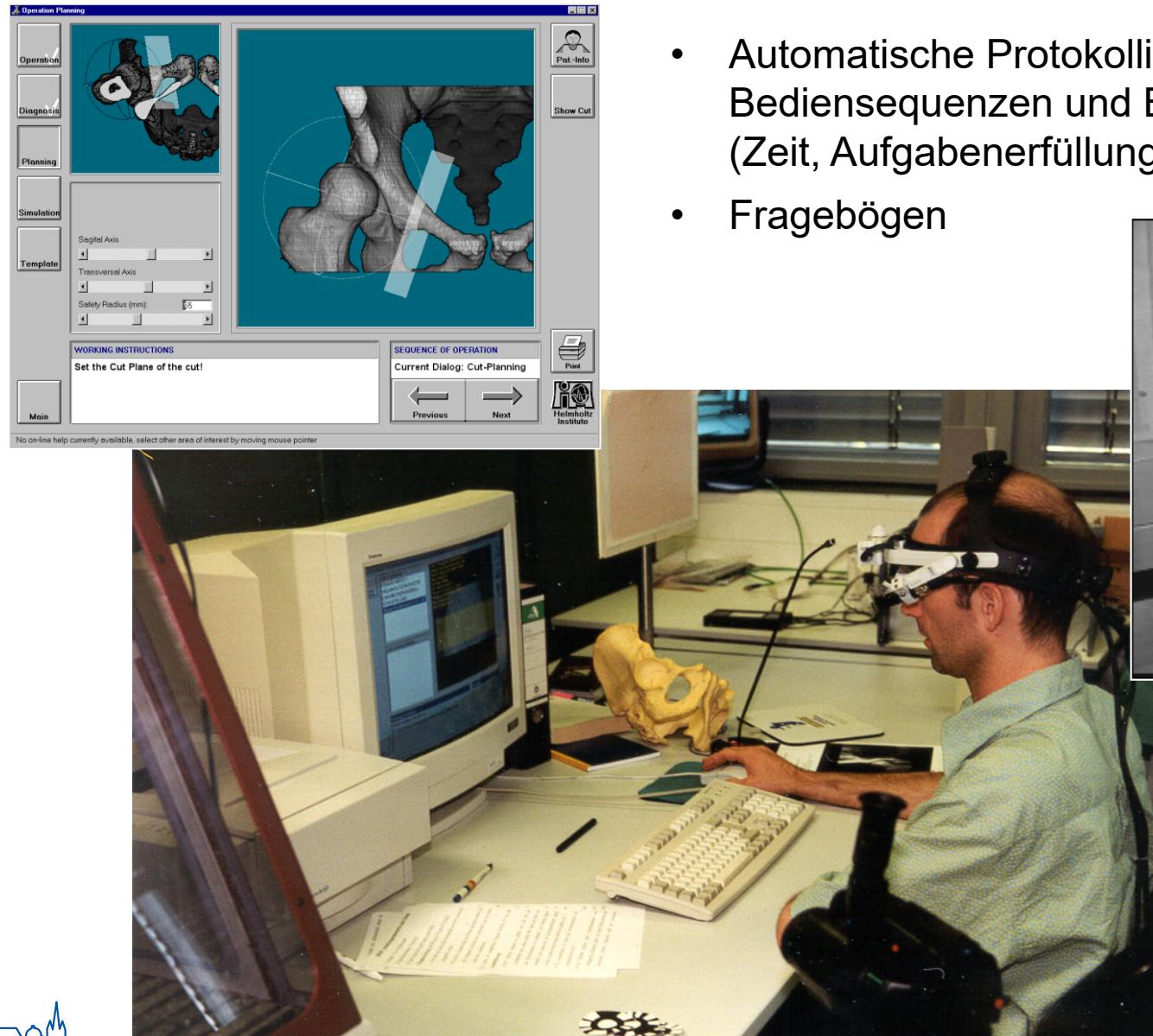
---

## Aufmerksamkeits-Analyse (Eye- / Attention Tracking)

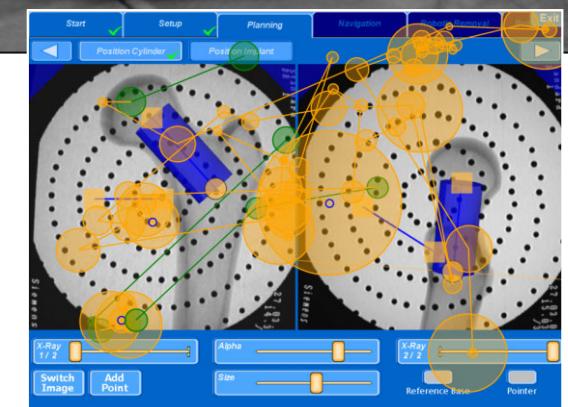
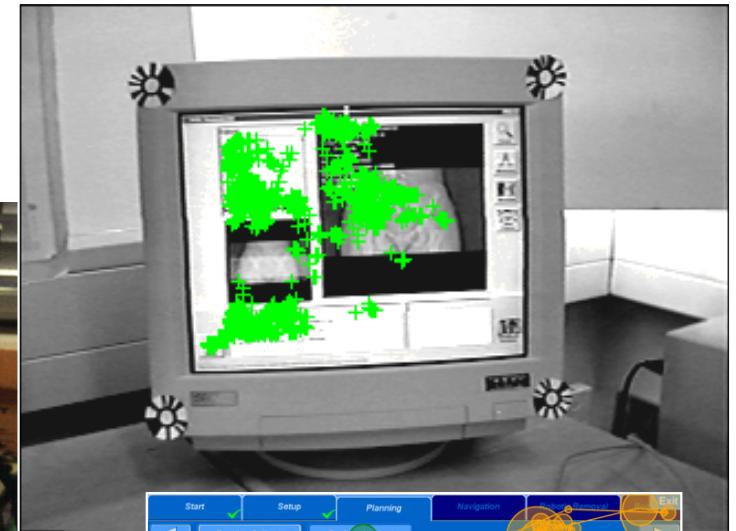
- Variante des Usability-Tests, bei der Augen- oder Mauszeigerbewegungen gemessen werden. Über Blickbewegungen oder den Weg des Mauszeigers kann auf Verschiebungen der Aufmerksamkeit geschlossen und somit analysiert werden, welchen Teilen einer Benutzeroberfläche der Anwender besonderes „Augenmerk“ schenkt.
- Messung von Augenbewegungen: Eye-Tracking,  
Analyse von Mauszeigerbewegungen: Attention-Tracking
- Vor- und Nachteile:
  - + Möglichkeit des Vergleichs verschiedener Designs
  - + Daten über den Verlauf der Aufmerksamkeit
  - + Interpretation von Zeiten möglich, in denen der Anwender keine Aktion ausführt
  - hoher technischer, finanzieller und methodischer Aufwand
  - fehlende Standards
  - Aussagekraft fraglich

# Eye-Tracking

## Beispiel Operationsplanungssystem



- Automatische Protokollierung und Auswertung der Bediensequenzen und Blickbewegungen (Zeit, Aufgabenerfüllung, Störungen, Hilfenutzung...)
- Fragebögen



# Eyetracking

The house was approached by gravelled driveways which wound about through wide-spreading lawns and under the interlacing boughs of tall poplars. At the rear things were on even a more spacious scale than at the front. There were great stables, where a dozen grooms and boys held forth, rows of vine-clad servants' cottages, an endless and orderly array of outhouses, long grape arbors, green pastures, orchards, and berry patches. Then there was the pumping plant for the artesian well, and the big cement tank where Judge Miller's boys took their morning plunge and kept cool in the hot afternoon.

And over this great demesne Buck ruled. Here he was born, and here he had lived the four years of his life. It was true, there were other dogs, There could not but be other dogs on so vast a place, but they did not count



# Messung von subjektivem Belastungsempfinden und Beanspruchung

---

- Zur Erinnerung: **Beanspruchung** (muskulär, physisch, emotional, informatorisch-mental) ist die Folge von externen **Belastungsprofilen** (ePSF) und individuellen Vorbedingungen (iPSF)  
-> **Beanspruchung ist individuell ausgeprägt (d.h. relativ)!**
- Beanspruchung (und Ermüdung) ist nicht direkt messbar
- **Beanspruchung ist nur indirekt messbar** (d.h. relativ über Indikatoren)
  - **Objektiv:** Herzrate, Blutdruck, Hautleitwert, Augen- und Lidschlussbewegungen, bewegungen (Videookulografie oder EOG), Muskelaktivität (EMG; z.B. Nackenmuskulatur), ... (oft werden – je nach Einsatzbedingungen - mehrere komplementäre Parameter erfaßt...)
  - **Subjektiv:** Skalen zur Eigenzustandsskalierung (Borg\*, Shackel\*, Nitsch\*\*), sowie Fragebögen\*\*,...

\* Nennung von Einzel-Skalenwerten –zeitlich kompakt – daher vor/nach einzelnen Aufgaben möglich

\*\*zeitaufwändiger, daher prä-/post-Test ...; Nitschtest differenziert nach Motivation und

Beanspruchung ! (J. Nitsch: Eigenzustandsskala (EZ-Skala): ein Verfahren zur hierarchisch-mehrdimensionalen Befindlichkeitsskalierung. Beanspruchung im Sport, Limpert-Verlag, Bad Homburg, 1976)

# Beispiel: RPE Skala nach BORG (subjektives Belastungsempfinden)

Der bei der Belastung angegebene RPE-Wert (Received Perception of Exertion) basiert auf der Annahme, dass das Belastungsempfinden mit der **Herzfrequenz** zusammenhängt:  
**RPE = Herzfrequenz x 0,1**

Da in der Regel die Ruheherzschlagfrequenz bei ca. 60 Schlägen pro Minute liegt, beginnt die Borg Skala bei 6, was den 60 Schlägen pro Minute entspricht. Die maximale Herzschlagfrequenz liegt bei gesunden Menschen meist bei ca. 200 Schlägen pro Minute - daher endet die Borg Skala bei 20.

Im Test nennt die Testperson z.B. **jeweils VOR und NACH einer Aufgabe einen Zahlenwert auf der im Versuchsraum aushängenden Skala.**

RPE-Wert	Belastungsempfinden
6	
7	sehr, sehr leicht
8	
9	sehr leicht
10	
11	leicht
12	
13	etwas anstrengend
14	
15	anstrengend
16	
17	sehr schwer
18	
19	sehr, sehr schwer
20	

# Beispiel: Diskomfortskala nach BORG

(Borg, G., A Category Scale With Radio properties for Intermodal and Interindividual Comparision. In: Geissler, H.G.; Petzold, P. (eds.) Psychological Judgement and the Process of Perception. Amsterdam, New, York, Oxford: North-Holland Publ., 1982) (dt. Übersetzung von Hahner)

- subjektives Belastungs- (Beanspruchungs-)empfinden
- Skala sehr einfach gestaltet und wird von Borg als Alternative zur RPE-Skala (s.o.) empfohlen
- beansprucht Intervallskalenniveau, so daß nicht nur Höhen des Diskomforts, sondern auch Verhältnisse messbar sind.
- für unterschiedliche Modalitäten geeignet (Geschmack, Temperatur, Druck,...)

> Versuchsperson nennt z.B. jeweils

- **VOR und NACH einer Aufgabe**  
...oder...
- zu definierten/protokollierten Zeitpunkten (> Verlauf)

<b>0</b>	<b>keine</b>
<b>0,5</b>	<b>äußerst schwach/gerade wahrnehmbar</b>
<b>1</b>	<b>sehr schwach</b>
<b>2</b>	<b>schwach/leicht</b>
<b>3</b>	<b>mäßig</b>
<b>4</b>	<b>etwas stark</b>
<b>5</b>	<b>stark/schwer</b>
<b>6</b>	<b>stark/schwer</b>
<b>7</b>	<b>sehr stark</b>
<b>8</b>	<b>sehr stark</b>
<b>9</b>	<b>sehr stark</b>
<b>10</b>	<b>äußerst stark/beinahe maximal</b>
<b>*</b>	<b>maximal</b>

...einen Zahlenwert der Skala.

# Bespiel: Diskomfortskala nach Shackel

(Shackel, B. et al. The Assessment of Chair Comfort. In: Grandjean, E. (ed.): Sitting postures. Taylor&Francis, London, 1969)

- *subjektives Belastungs- (Beanspruchungs-)empfinden*
- Ordinalskalenniveau
- Ursprünglich für Sitzkomfortbewertung entwickelt
- Sinnvolle Aussagen über den Gradient des Diskomforts über die Zeit
- i.a. Anwendung auf Empfinden bezogen auf **Mechanorezeptoren** (von Haut, Gelenken, Sehnen, Muskeln)

> Versuchsperson nennt z.B. jeweils

- **VOR und NACH einer Aufgabe**

...oder...

- zu definierten/protokollierten Zeitpunkten (> Verlauf)

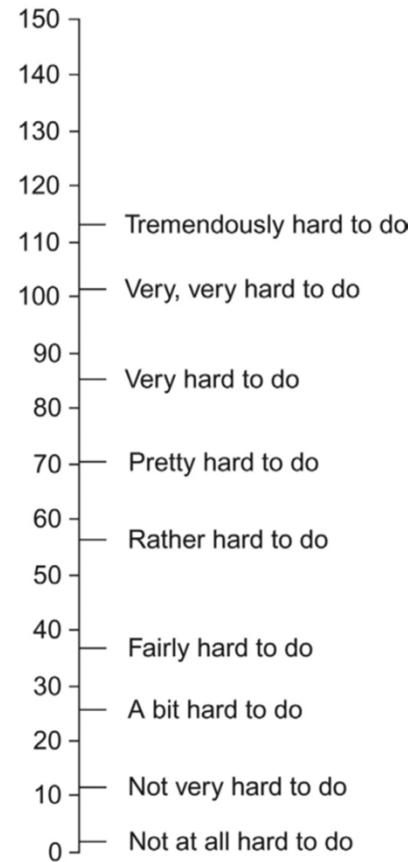
...einen Zahlenwert der Skala.

0	total entspannt
1	
2	sehr wohl
3	
4	ganz gut
5	
6	noch akzeptabel
7	
8	unbequem
9	
10	sehr unbequem
11	
12	verkrampft
13	
14	steif
15	
16	empfindungslos
17	
18	zermürbt
19	
20	unerträglich belastet

# Weitere Beispiele

## Post-Aufgaben-Usability-Einzelfrage-Tests

**Subjective Mental Effort Questionnaire (SMEQ):** Benutzer markiert mit einer Linie auf Skala wie hoch die mentale Beanspruchung war (alternativ: nennt Zahl)



**Single Question Likert:** „Overall the Task was:“

Very difficult                                Very easy

**Anmerkung zu Skalenformat (Likert unipolar vs. bipolar) und Art und Zahl von Ankern/Stufen:**

- **Likert-Skala** für Grad der Zustimmung könnte unipolar (eine Ausprägung...wird zwischen Extremen eher als gleichabständig angesehen) z.B. „nicht so einfach“ bis „sehr einfach“ alternativ als **Bipolar-Skala mit zwei Endpunkten ausgeführt sein** (schlecht/sehr gut....Endpole werden eher interpretiert und damit Skala wird eher als nicht gleichabständig interpretiert)

Beispiel Bipolar: Messages that appear on the screen are...

Confusing	1	2	3	4	5	6	7	8	9	NA
	<input type="radio"/>									

- diverse Empfehlungen zu Zahl d. Anker (i.a. 3-9), Verwendung eines neutralen Ankers; mit/ohne Zahlen / verbale Anker,...
- **Skalenformat eher zweitrangig (vgl. Quelle unten); Wichtig ist die konsistente Nutzung i.R. der Evaluierung**

# Fragebögen

---

- Fragebögen haben ihre Wurzeln in den Sozialwissenschaften, um z.B. Meinungen zu erfassen
- Beurteilung (von **Benutzerzufriedenheit/subjektiver Gebrauchstauglichkeit** oder **subjektiver Beanspruchung**) aus Sicht der Benutzer im Anwendungskontext
- bestehen in der Regel aus einer Anleitung und einzelnen Items: Fragen oder Aussagen mit dazugehörigen Antwortmöglichkeiten, häufig als Likert-Skala oder in Form eines semantischen Differenzials
- Da Fragebögen eher zeitaufwändiger sind, finden diese i.a. NICHT nach Einzelaufgaben sondern eher **vor und nach einem Test** statt

(Anmerkung: der **Vorher-Nachher-Vergleich im Kontext der subjektiven Bewertungen** ist (soweit möglich) allgemein wichtig um **interindividuelle (persönlichkeitsspezifische) Unterschiede in der Bewertung auszugleichen** und (möglichst nur) den **Einfluss der Aufgabe bzw. des Tests (in der Veränderung) zu erfassen.**)

# Methoden der benutzerzentrierten Usability-Evaluierung

---

## Fragebögen

- Items zur Erhebung subjektiver Urteile zu einzelnen Aspekten
- Arten: Multiple Choice, abgestufte bipolare Einschätzungen in Form einer Skala („stimme gar nicht zu“ bis „stimme voll zu“), Freitextfragebogen
- Gütekriterien für Fragebögen:
  - Objektivität: Sind die Ergebnisse unabhängig vom Versuchsleiter?
  - Reliabilität: Produziert der Fragebogen bei einer Wiederholung der Erhebung annähernd die gleichen Ergebnisse?
  - Validität: Misst der Fragebogen, was er messen soll?
  - Vergleichsdatenbasis (um Usability Qualität im Vergleich zu anderen Systemen einschätzen zu können; z.B. kostenpflichtiger SUMI Questionnaire - <https://sumi.uxp.ie/>)
- Vor- und Nachteile:
  - + es können leicht und kostengünstig Daten von vielen Anwendern erhoben werden
  - + kaum Beeinflussung durch Dritte während der Evaluation
  - + es liegt eine Vielzahl von standardisierten Fragebögen vor
  - + subjektive Daten können kategorisiert und je nach Antwortformat statistisch analysiert werden
  - je strukturierter das Verfahren, umso weniger Freiraum bleibt dem Anwender
  - Entwicklung eigener, spezifischer Fragebögen sehr aufwändig

# Software Usability Measurement Inventory - SUMI

- Kostenpflichtiger softwaregestützter Fragebogen zur Messung der Benutzerzufriedenheit (user satisfaction) bzw. „User Experience“ (<https://sumi.uxp.ie/>)
- Vergleichsdatenbasis aus 25 Jahren industrieller Anwendung
- 50 Aussagen (ggf. zusätzliche kunden- bzw. anwendungsspezifisch angepaßte Aussagen) zu denen der Benutzer mit „agree/don't know/disagree“ antworten muss (in über 20 Sprachen verfügbar)
- Softwaregestützte Online-Auswertung
- Auszüge aus Report:

## Contents of this Report

[Graphical summaries of SUMI scales.](#)  
Three graphs output.

[Strengths and weaknesses analysis.](#)  
There are 31 items above 60th percentile.

[User records.](#)  
There are 15 user records, including 1 user with potential outliers.

[Summary statistics.](#)  
Average Global SUMI score is 55.53.

[Comments by users to: What, in general, do you use this software for?](#)  
15 comments listed.

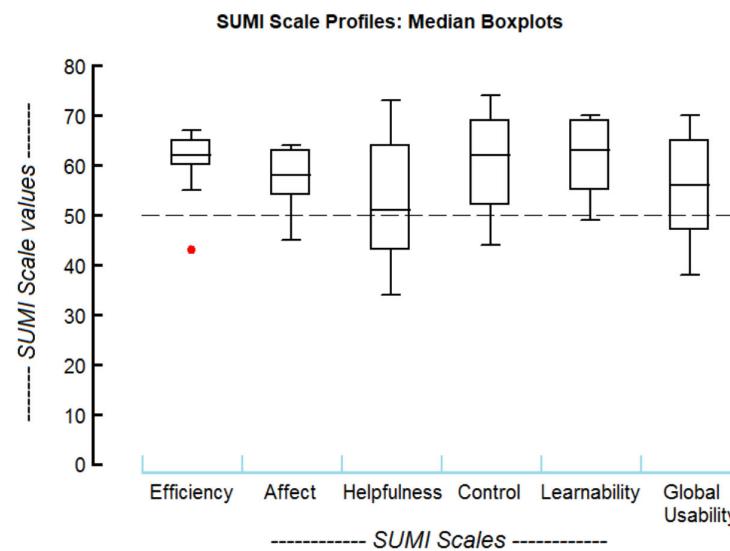
[SUMI scores by answers to: How important for you is the kind of software you have just been rating? \(radio\)](#)  
8 respondents said: 'Extremely important'.

[SUMI scores by answers to: How would you rate your software skills and knowledge? \(radio\)](#)  
5 respondents said: 'Very experienced and technical'.

[Comments by users to: What do you think is the best aspect of this software, and why?](#)  
15 comments listed.

[Comments by users to: What do you think needs most improvement, and why?](#)  
15 comments listed.

[CSV format data file](#)  
15 records analysed. There are no ignored records in this data set.



# System Usability Scale – SUS

## (Subjektive Usability Bewertung)

(Brooke, J., 1996. SUS: A 'quick and dirty' usability scale. In: Jordan, P., Thomas, B., Weerdmeester, B. (Eds.), *Usability Evaluation in Industry*. Taylor & Francis, London, UK, pp. 189–194.)

The System Usability Scale Standard Version		Strongly disagree	Strongly agree			
		1	2	3	4	5
1	I think that I would like to use this system frequently.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	I found the system unnecessarily complex.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	I thought the system was easy to use.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	I think that I would need the support of a technical person to be able to use this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	I found the various functions in the system were well integrated.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	I thought there was too much inconsistency in this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	I would imagine that most people would learn to use this system very quickly.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	I found the system very awkward to use.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	I felt very confident using the system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	I needed to learn a lot of things before I could get going with this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rechts: Relative Einordnung der SUS-Score-Ergebnisse in umfangreiche Vergleichsdaten aus Usability Tests: Beispiel SUS Score 81 bedeutet, daß Usability (Grade A) besser als 90-95% der in Datenbasis getesteten Systeme ist.

(Quelle: Jeff Sauro, James R. Lewis: Standardized Usability Questionnaires. in Jeff Sauro, James R. Lewis: Quantifying the User Experience (Second Edition), 2016 (<https://www.sciencedirect.com/book/9780128023082/quantifying-the-user-experience>))

grobe Einschätzung der Usability eines interaktiven Systems, z.B. für **Versionsvergleiche**

- + extrem kurze Bearbeitungszeit
- + dadurch geeignet zur Verwendung in Benutzertests
- grobe Einschätzung (vergleichbar ASQ)
- eigene Vergleichsdatenbasis nötig (alternativ näherungsw.:)

SUS Score Range	Grade	Percentile Range
84.1–100	A+	96–100
80.8–84.0	A	90–95
78.9–80.7	A–	85–89
77.2–78.8	B+	80–84
74.1–77.1	B	70–79
72.6–74.0	B–	65–69
71.1–72.5	C+	60–64
65.0–71.0	C	41–59
62.7–64.9	C–	35–40
51.7–62.6	D	15–34
0.0–51.6	F	0–14

# IBM Usability Satisfaction Questionnaires

Alle Fragebögen basieren auf 7 Punkt Skala mit Endpunktankern „strongly agree (1)/strongly disagree(7) oder not applicable (N/A). \*ASQ u.U. als Post-Aufgaben (s.o.)-Fragebogen nutzbar)

- ASQ – **After-Scenario\***-Questionnaire: 3 Fragen (und jeweils Kommentarmöglichkeit)

The After-Scenario Questionnaire Version 1		Strongly agree		Strongly disagree		NA			
		1	2	3	4	5	6	7	
1	Overall, I am satisfied with the ease of completing the tasks in this scenario.		o	o	o	o	o	o	o
2	Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.		o	o	o	o	o	o	o
3	Overall, I am satisfied with the support information (online help, messages, documentation) when completing the tasks.		o	o	o	o	o	o	o

- PSSUQ – **Post-Study**-System Usability Questionnaire (V3, Lewis 2012: 16 Fragen; Auswertung in 4 Kategorien (overall satisfaction score (1-16), system usefulness (1-6), information quality (7-12), interface quality (13-15))
- CSUQ - Computer System Usability Questionnaire (wie PSSUQ – aber allgemeinere Formulierungen, die sich nicht auf Testsituation beziehen)

(Quellen und weitere Informationen zu diesen Fragebögen inkl. Originalfragebögen:

- James R. Lewis (1995) IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, International Journal of Human-Computer Interaction, 7:1, 57-78, DOI: 10.1080/10447319509526110
- Jeff Sauro, James R. Lewis: Standardized Usability Questionnaires. in Jeff Sauro, James R. Lewis: Quantifying the User Experience (Second Edition), 2016 (<https://www.sciencedirect.com/book/9780128023082/quantifying-the-user-experience>)

# Post-Study System Usability Questionnaire (PSSUQ)

The Post-Study System Usability Questionnaire Version 3		Strongly agree	Strongly disagree							
			1	2	3	4	5	6	7	NA
1	Overall, I am satisfied with how easy it is to use this system.		0	0	0	0	0	0	0	0
2	It was simple to use this system.		0	0	0	0	0	0	0	0
3	I was able to complete the tasks and scenarios quickly using this system.		0	0	0	0	0	0	0	0
4	I felt comfortable using this system.		0	0	0	0	0	0	0	0
5	It was easy to learn to use this system.		0	0	0	0	0	0	0	0
6	I believe I could become productive quickly using this system.		0	0	0	0	0	0	0	0
7	The system gave error messages that clearly told me how to fix problems.		0	0	0	0	0	0	0	0
8	Whenever I made a mistake using the system, I could recover easily and quickly.		0	0	0	0	0	0	0	0
9	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.		0	0	0	0	0	0	0	0
10	It was easy to find the information I needed.		0	0	0	0	0	0	0	0
11	The information was effective in helping me complete the tasks and scenarios.		0	0	0	0	0	0	0	0
12	The organization of information on the system screens was clear.		0	0	0	0	0	0	0	0
13	The interface* of this system was pleasant.		0	0	0	0	0	0	0	0
14	I liked using the interface of this system.		0	0	0	0	0	0	0	0
15	This system has all the functions and capabilities I expect it to have.		0	0	0	0	0	0	0	0
16	Overall, I am satisfied with this system.		0	0	0	0	0	0	0	0

# Computer System Usability Questionnaire (CSUQ)

The Computer System Usability Questionnaire Version 3		Strongly agree	Strongly disagree								
		1 2 3 4 5 6 7	NA								
1	Overall, I am satisfied with how easy it is to use this system.	<input type="radio"/>									
2	It is simple to use this system.	<input type="radio"/>									
3	I am able to complete my work quickly using this system.	<input type="radio"/>									
4	I feel comfortable using this system.	<input type="radio"/>									
5	It was easy to learn to use this system.	<input type="radio"/>									
6	I believe I became productive quickly using this system.	<input type="radio"/>									
7	The system gives error messages that clearly tell me how to fix problems.	<input type="radio"/>									
8	Whenever I make a mistake using the system, I recover easily and quickly.	<input type="radio"/>									
9	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear.	<input type="radio"/>									
10	It is easy to find the information I needed.	<input type="radio"/>									
11	The information provided with the system is effective in helping me complete my work.	<input type="radio"/>									
12	The organization of information on the system screens is clear.	<input type="radio"/>									
13	The interface* of this system is pleasant.	<input type="radio"/>									
14	I like using the interface of this system.	<input type="radio"/>									
15	This system has all the functions and capabilities I expect it to have.	<input type="radio"/>									
16	Overall, I am satisfied with this system.	<input type="radio"/>									

# NASA-TLX Fragebogen

(TLX = Task-Load-Index > misst subjektive Belastung / Beanspruchung)

Beispiel: Touchpad

## 1. Absolute Bewertung

Bitte bewerten Sie den Beitrag der jeweiligen Faktoren zur Arbeitsbelastung durch das Markieren einer Position auf der zugehörigen Skala

### Geistige Anforderungen

Wie hoch waren die geistigen Anforderungen der Aufgabe?



### Körperliche Anforderungen

Wie hoch waren die körperlichen Anforderungen der Aufgabe?



### Zeitliche Anforderungen

Wie hoch war das Tempo, mit dem die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten?



### Leistung

Wie erfolgreich haben sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?



### Anstrengung

Wie sehr mußten Sie sich anstrengen, um Ihre Leistung zu erreichen?



### Frustration

Wie verunsichert, entmutigt, gereizt und verärgert waren Sie?



# NASA-TLX Fragebogen

(TLX = Task-Load-Index > misst subjektive Belastung / Beanspruchung)

Beispiel: Maus vs. Touchpad

## 2. Relative Bewertung

Bitte bewerten Sie den Beitrag der jeweiligen Faktoren zur Arbeitsbelastung durch das markieren einer Position auf der zugehörigen Skala im Vergleich zu ..... :

### Geistige Anforderungen

Wie hoch waren die geistigen Anforderungen der Aufgabe im Vergleich zur Mausbenutzung?



### Körperliche Anforderungen

Wie hoch waren die körperlichen Anforderungen der Aufgabe im Vergleich zur Mausbenutzung?



### Zeitliche Anforderungen

Wie hoch war das Tempo, mit dem die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten im Vergleich zur Mausbenutzung?



### Leistung

Wie erfolgreich haben sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt im Vergleich zur Mausbenutzung?



### Anstrengung

Wie sehr mußten Sie sich anstrengen, um Ihre Leistung zu erreichen im Vergleich zur Mausbenutzung ?



# IsoMetrics

---

- Messung und Vergleich der Usability verschiedener interaktiver Systeme oder Versionen eines interaktiven Systems
- wird unterteilt in eine kurze Version IsoMetrics<sup>S</sup> und eine lange Version IsoMetrics<sup>L</sup>  
Unterschied: im Fragebogen IsoMetrics<sup>L</sup> kann zu jeder Frage noch einmal ausführlich und frei handschriftlich eine Stellungnahme genommen werden
  - + starker Bezug zur DIN EN ISO 9241-110
  - allerdings sehr zeitaufwendig (eher Experten-Evaluierung (s.o.))

## Über den Fragebogen 'IsoMetrics<sup>S</sup>'

Liebe Untersuchungsteilnehmerin, lieber Untersuchungsteilnehmer

Der Ihnen vorliegende Fragebogen dient zur Einschätzung der Benutzbarkeit von Anwendungsprogrammen, die mit graphisch gestalteten Benutzungsschnittstellen ausgestattet sind. Durch das Ausfüllen des Fragebogens helfen Sie uns, die Schwächen und Stärken des in Frage stehenden Produktes festzustellen.

Der Fragebogen enthält Aussagen zur Benutzungsfreundlichkeit von Software. Bitte schätzen Sie Ihre Zustimmung zu jeder Aussage auf der unter der Frage befindlichen Skala ein. Hierzu ein Beispiel:

		stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr	
Index		1	2	3	4	5	Keine Angabe
D	Gestaltungsgrundsatz					X	
D	Diese Software ist für mich ein nützliches Arbeitsmittel.						

Wenn Sie der Meinung sind, daß diese Aussage für Sie zutrifft, sollte Ihr Kreuz bei "5" für "Stimmt sehr" gesetzt sein. Falls Sie dieser Aussage nicht zustimmen können, sollte Ihre Kreuz entsprechend bei "1" für "Stimmt nicht" gesetzt sein. Angekreuzte Zahlen zwischen diesen Polen bedeuten eine graduelle Zustimmung oder Ablehnung. Für den Fall, daß Sie sich aus irgendwelchen Gründen dazu nicht äußern wollen oder können, sollten Sie "keine Angabe" ankreuzen.

Vielen Dank für Ihre Mitarbeit

## IsoMetrics<sup>S</sup>

Index	Aufgabenangemessenheit	stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr	Keine Angabe
A.1	Die Software zwingt mich, überflüssige Arbeitsschritte durchzuführen.						
A.3	Mit der Software kann ich zusammenhängende Arbeitsabläufe vollständig bearbeiten.						
A.4	Die Software bietet mir alle Möglichkeiten, die ich für die Bearbeitung meiner Aufgaben benötige.						
A.6	Die Software ermöglicht es mir, Daten so einzugeben, wie es von der Aufgabenstellung gefordert wird.						
A.7	Die für die Aufgabenbearbeitung notwendigen Informationen befinden sich immer am richtigen Platz auf dem Bildschirm.						
A.8	Es müssen zuviele Eingabeschritte für die Bearbeitung mancher Aufgaben durchgeführt werden.						
A.9	Die vom Programm erzeugten Ausgaben passen zu meinen Aufgabenstellungen, d.h. sie erhalten keine überflüssigen, zu knappen oder unverständlich formulierten Informationen.						
A.10	Die Software ist auf die von mir zu bearbeitenden Aufgaben zugeschnitten.						
A.11	Auf dem Bildschirm finde ich alle Informationen, die ich gerade benötige.						
A.12	Die in der Software verwendeten Begriffe und Bezeichnungen entsprechen denen meiner Arbeitstätigkeit.						
A.14	Die Software bietet mir eine Wiederhol-Funktion für wiederkehrende Arbeitsschritte.						

# IsoMetrics<sup>L</sup>

IsoMetrics<sup>L</sup>

- dabei sind die Fragen identisch zu denen im IsoMetrics<sup>s</sup>

	stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr	Keine Angabe
A.1 Die Software zwingt mich, überflüssige Arbeitsschritte durchzuführen.	1	2	3	4	5	

	nicht wichtig	wenig wichtig	mittelmaßig wichtig	ziemlich wichtig	sehr wichtig	Keine Angabe
Wie wichtig ist dieser Aspekt für Ihren Gesamteindruck von der Software?	1	2	3	4	5	

# ISONORM 9241

## Noch ein Hinweis zur Beantwortung des Beurteilungsbogens:

Die einzelnen Normen werden über Beschreibungen konkretisiert. Diese Beschreibungen weisen immer folgende Form auf.

### Beispiel Nr.1:

Die Software ...	---	--	-	-/+	+	++	+++	Die Software ...
ist schlecht.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ist gut.				

Beispiel 1

Im ersten Beispiel wird danach gefragt, wie gut, bzw. wie schlecht die Software ist. Der Benutzer beurteilt in diesem Fall die Software zwar als gut, sieht jedoch noch Verbesserungsmöglichkeiten.

### Beispiel Nr.2:

Die Software ...	---	--	-	-/+	+	++	+++	Die Software ...
ist langsam.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ist schnell.				

Beispiel 2

Im zweiten Beispiel beurteilt der Benutzer die Software als ziemlich langsam.

Füllen Sie bitte den Beurteilungsbogen äußerst sorgfältig aus und lassen Sie keine der Fragen aus!

Die Auswertung der Daten erfolgt anonym.

## Selbstbeschreibungsfähigkeit

Gibt Ihnen die Software genügend Erläuterungen und ist sie in ausreichendem Maße verständlich?

Die Software ...	---	--	-	-/+	+	++	+++	Die Software ...
bietet einen schlechten Überblick über ihr Funktionsangebot.	<input type="checkbox"/>	bietet einen guten Überblick über ihr Funktionsangebot.						
verwendet schlecht verständliche Begriffe, Bezeichnungen, Abkürzungen oder Symbole in Masken und Menüs.	<input type="checkbox"/>	verwendet gut verständliche Begriffe, Bezeichnungen, Abkürzungen oder Symbole in Masken und Menüs.						
liefert in unzureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.	<input type="checkbox"/>	liefert in zureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.						
bietet auf Verlangen keine situationspezifischen Erklärungen, die konkret weiterhelfen.	<input type="checkbox"/>	bietet auf Verlangen situationspezifische Erklärungen, die konkret weiterhelfen.						
bietet von sich aus keine situationspezifischen Erklärungen, die konkret weiterhelfen.	<input type="checkbox"/>	bietet von sich aus situationspezifische Erklärungen, die konkret weiterhelfen.						

# ISONORM 9241

---

- Messung der Usability von interaktiven Systemen und Prototypen, Identifizierung von Schwachstellen
  - + sehr starker Bezug zur DIN EN ISO 9241-110 aus diesem Grund auch den IsoMetrics Fragebögen sehr ähnlich
  - + Vergleichbarkeit mit anderen Softwareprodukten
  - keine Hinweise zur Mängelbildung
  - sehr zeitaufwändig (eher Experten-Evaluierung (s.o.))

# UEQ – Usability Evaluation Questionnaire

Bewertung einer Nutzungssituation durch Messung des Gesamteindrucks eines Nutzers in Bezug auf ein interaktives System

- + gute Ergänzung zu anderen Methoden
- allerdings nicht normgerecht in Bezug auf die 9241-210, da eine andere Definition von User Experience zugrunde liegt

Bitte geben Sie Ihre Beurteilung ab.

Um das Produkt zu bewerten, füllen Sie bitte den nachfolgenden Fragebogen aus. Er besteht aus Gegensatzpaaren von Eigenschaften, die das Produkt haben kann. Abstufungen zwischen den Gegensätzen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Zustimmung zu einem Begriff äußern.

Beispiel:

attraktiv	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	unattraktiv				
-----------	-----------------------	----------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-------------

Mit dieser Beurteilung sagen Sie aus, dass Sie das Produkt eher attraktiv als unattraktiv einschätzen.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Begriffe nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einem Begriffspaar unsicher sind oder finden, dass es nicht so gut zum Produkt passt.

Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

Bitte geben Sie nun Ihre Einschätzung des Produkts ab. Kreuzen Sie bitte nur einen Kreis pro Zeile an.

	1	2	3	4	5	6	7		
unerfreulich	<input type="radio"/>	erfreulich	1						
unverständlich	<input type="radio"/>	verständlich	2						
kreativ	<input type="radio"/>	phantasielos	3						
leicht zu lernen	<input type="radio"/>	schwer zu lernen	4						
wertvoll	<input type="radio"/>	minderwertig	5						
langweilig	<input type="radio"/>	spannend	6						
uninteressant	<input type="radio"/>	interessant	7						
unberechenbar	<input type="radio"/>	voraussagbar	8						
schnell	<input type="radio"/>	langsam	9						
originell	<input type="radio"/>	konventionell	10						
behindernd	<input type="radio"/>	unterstützend	11						
gut	<input type="radio"/>	schlecht	12						
kompliziert	<input type="radio"/>	einfach	13						
abstoßend	<input type="radio"/>	anziehend	14						
herkömmlich	<input type="radio"/>	neuartig	15						
unangenehm	<input type="radio"/>	angenehm	16						
sicher	<input type="radio"/>	unsicher	17						
aktivierend	<input type="radio"/>	einschläfernd	18						
erwartungskonform	<input type="radio"/>	nicht erwartungskonform	19						
ineffizient	<input type="radio"/>	effizient	20						
übersichtlich	<input type="radio"/>	verwirrend	21						
unpragmatisch	<input type="radio"/>	pragmatisch	22						
aufgeräumt	<input type="radio"/>	überladen	23						
attraktiv	<input type="radio"/>	unattraktiv	24						
sympathisch	<input type="radio"/>	unsympathisch	25						
konservativ	<input type="radio"/>	innovativ	26						

# Methoden der benutzerzentrierten Usability-Evaluierung

---

## Interviews und Fokusgruppen

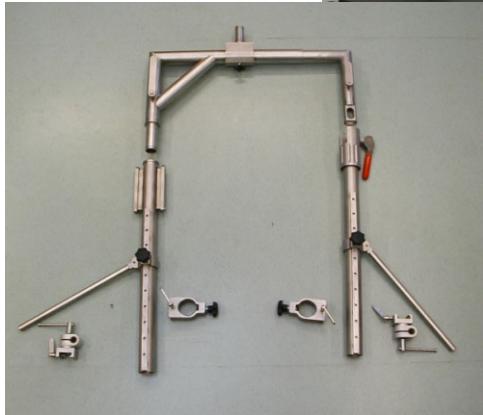
- Erhebung subjektiver Eindrücke und Meinungen von Anwendern einzuholen, die auf deren Erfahrungen im Umgang mit einem Produkt basieren.
- Interviews: (Anwender-) Einzelgespräche  
Fokusgruppen: moderierte und fokussierte Diskussionen in der (Anwender-) Gruppe
- Interview-Arten:
  - **Strukturierte Interviews:** Formulierung der Fragen, ihre Reihenfolge, die Antwortmöglichkeiten und das Interviewverhalten genau festgelegt sind.  
=> gezielt Antworten auf bestimmte Fragen bekommen
  - **Semi-strukturierte Interviews:** es existiert meist nur Fragenkatalog oder ein Gesprächsleitfaden, der eine gewisse Struktur in das Gespräch bringen soll
  - **Unstrukturierte Interviews:** nützlich, um z.B. zu Beginn eines Design-Prozesses differenzierte Informationen zu kommen
- Vor- und Nachteile:
  - + liefert im Designprozess evtl. Anhaltspunkte für die Ursache von Problemen
  - + früh einsetzbar, wenn die Anforderungen an das Produkt noch veränderlich sind
  - mangelhafte Fähigkeiten des Interviewers/Moderators können den Erfolg gefährden
  - möglicherweise antwortet der Befragte nicht wahrheitsgemäß oder unvollständig

# Gliederung

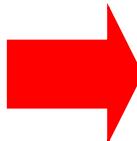
---

- Verifikation und Validierung – Begriffsklärung
- Validierungsplanung
- Methoden der Gebrauchstauglichkeits-Evaluierung  
(im Kontext der Entwicklungsphasen)
- Beispiele interaktions- und benutzerzentrierter Evaluierungen

# Beispiel benutzer- und interaktionszentrierte Evaluierung: Multifunktionaler Anästhesiebogen



- **Aufbau, Abdeckung und Notfallentfernung**  
=> mit und ohne schriftliche Anleitung
- **Zeitmessung und Videobeobachtung**
- **Fragebögen / Interview**



ibm  
RWTH AACHEN

MOP<sub>2</sub>

Fragebogen

	sehr schlecht	schlecht	0	schön	+	sehr gut			
1	2	3	4	5	6	7	8	9	10

Evaluation

Fragebogen

	sehr schlecht	schlecht	0	schön	+	sehr gut			
1	2	3	4	5	6	7	8	9	10

Fragebogen

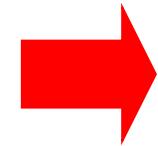
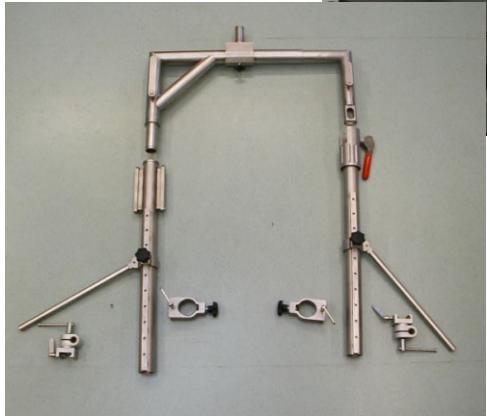
	sehr schlecht	schlecht	0	schön	+	sehr gut			
1	2	3	4	5	6	7	8	9	10

Fragebogen

	sehr schlecht	schlecht	0	schön	+	sehr gut			
1	2	3	4	5	6	7	8	9	10

Kommentare

# Beispiel benutzer- und interaktionszentrierte Evaluierung: Multifunktionaler Anästhesiebogen



- **Aufbau, Abdeckung und Notfallentfernung**  
=> mit und ohne schriftliche Anleitung
- **Zeitmessung und Videobeobachtung**



# Beispiel benutzerzentrierte Evaluierung zur koordinierten Sprach- / Kopfsteuerung eines Mikroskopmanipulators



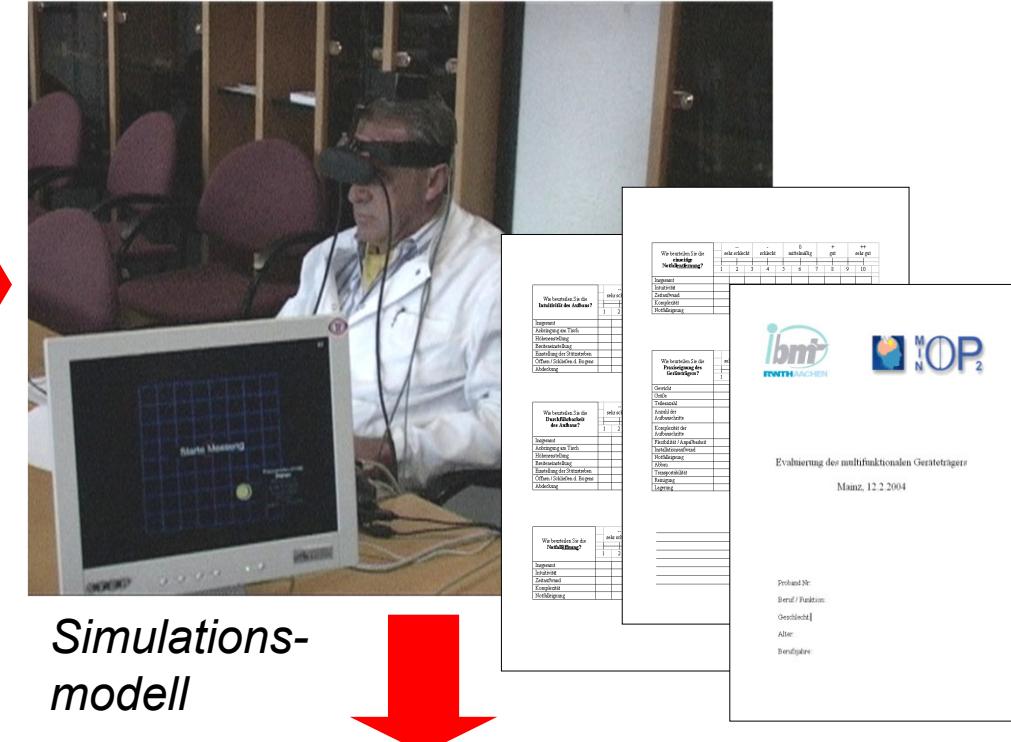
Prototyp real

Modellbasierte benutzerzentrierte Untersuchung der kombinierten Kopf- und Sprachsteuerung sowie der Bewegungsmodi am Simulationsmodell

=> virtuelle Testumgebung  
(Zeit, Aufgabenerfüllung...)

=> Fragebögen

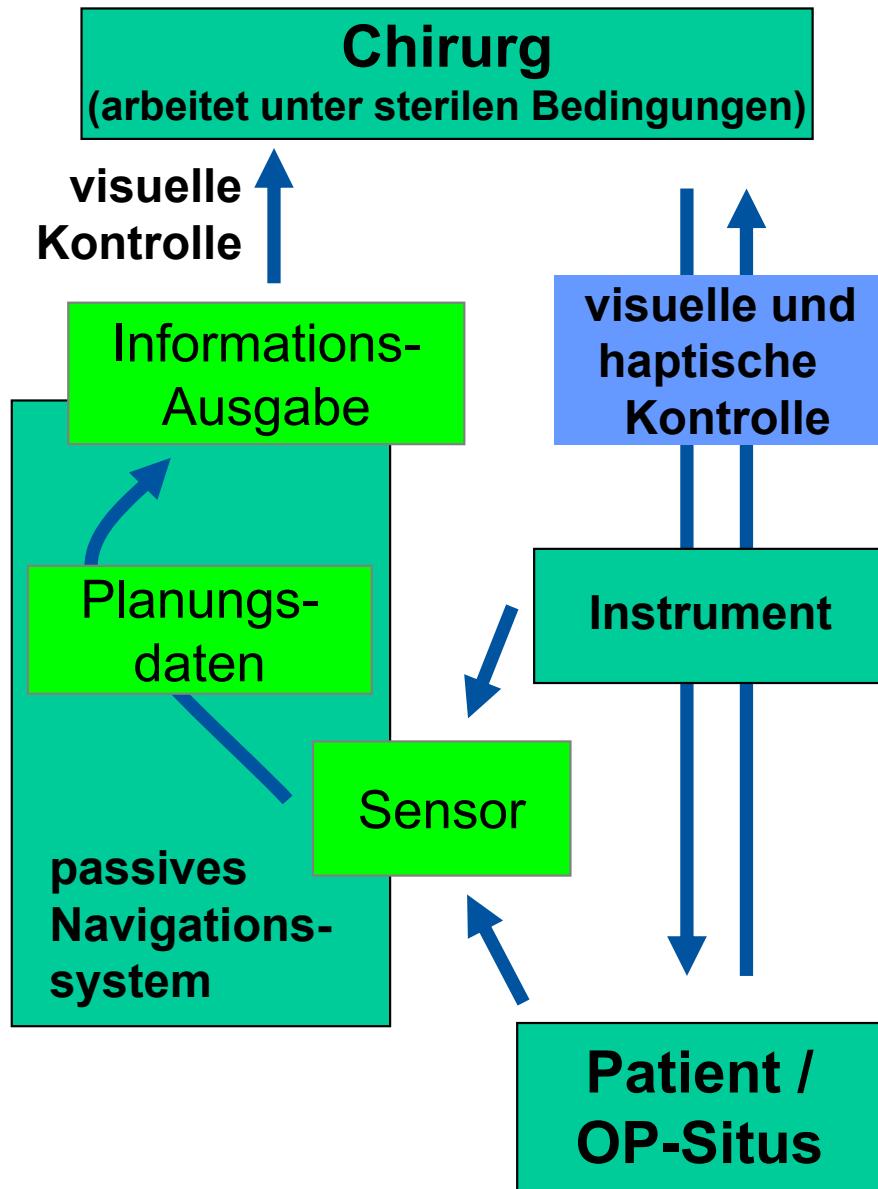
dann: => Test-Setup im OP  
(Beobachtung, Benutzerbefragung)



Modus	Effizienz	Intuitivität	Komfort	Kontrollierbarkeit
Turn	+ 36 %	+ 16 %	+ 16 %	+ 24 %
Pivot	+ 20 %	+ 13 %	+ 18 %	+ 22 %
Lean	- 29 %	+ 7 %	+ 2 %	+ 16 %
Shift	- 10 %	+ 11 %	+ 11 %	+ 16 %
Zoom	- 6 %	+ 7 %	+ 7 %	+ 4 %

# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Mensch-Maschine-Interaktion



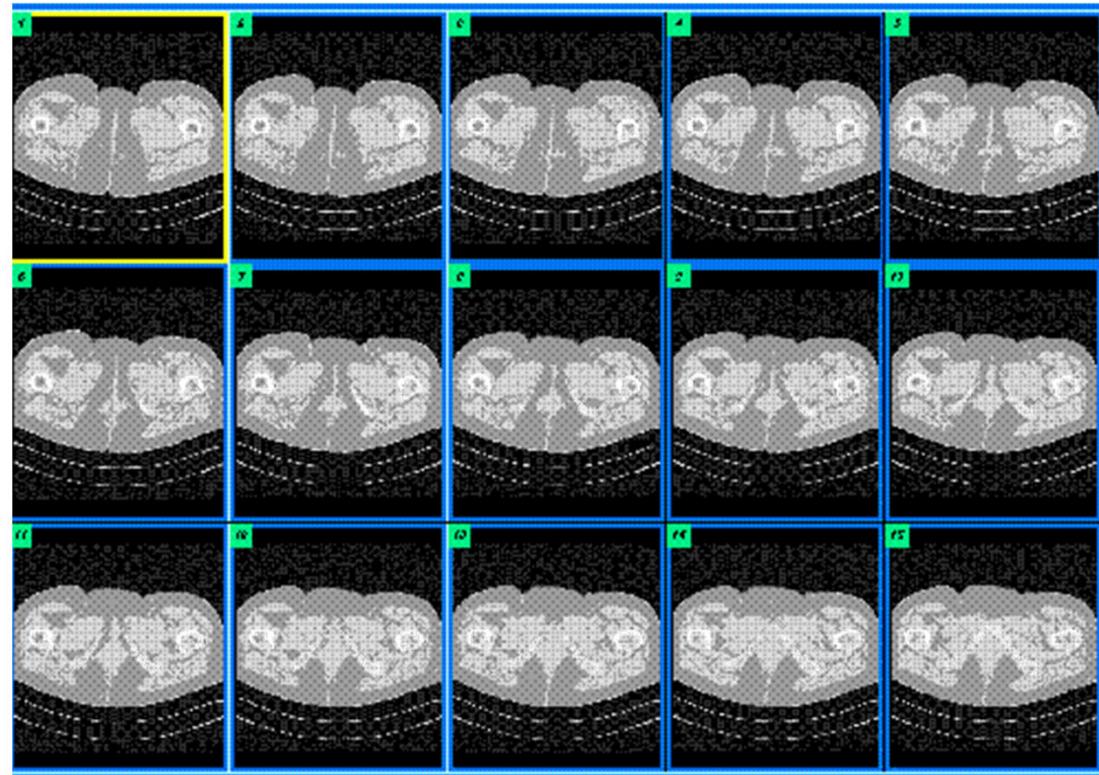
Wirbelsäulenchirurgie mit StealthStation  
(KKMH Würselen, 1999)

# Beispiel: Usability-Evaluierung CAS Navigationssystem

## OP-Ablauf SurgiGATE Pfannennavigation



1) Informations-  
akquisition

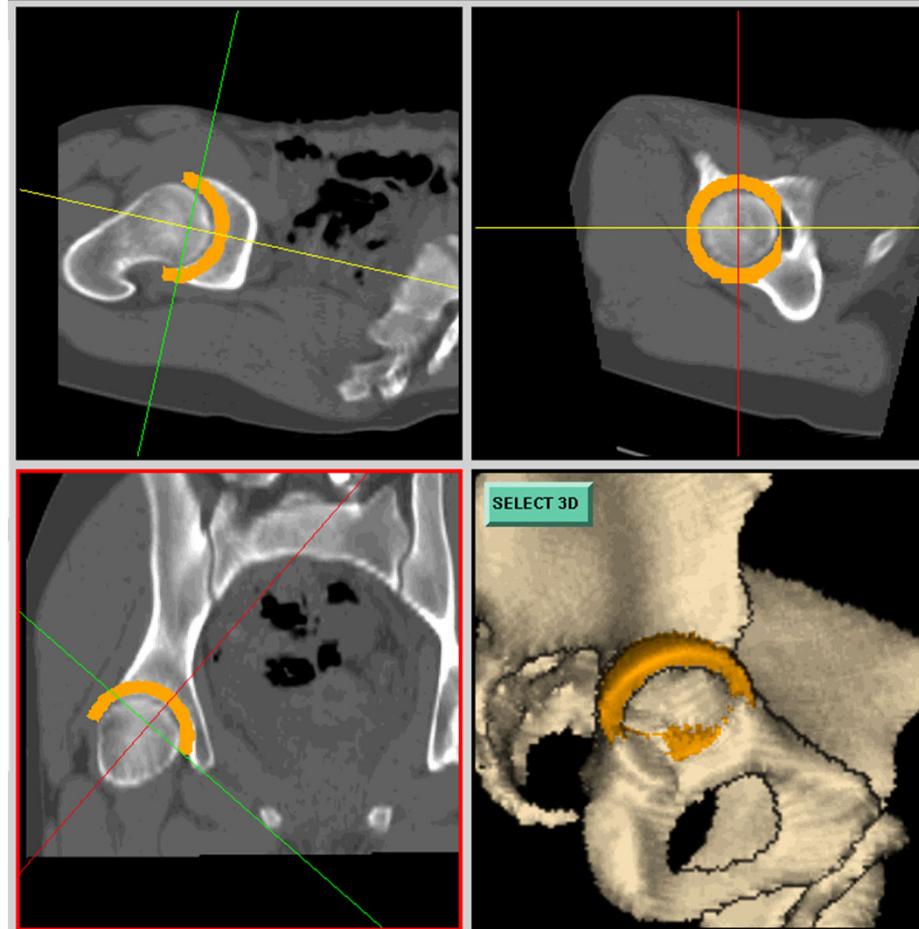


# Beispiel: Usability-Evaluierung CAS Navigationssystem

## OP-Ablauf SurgiGATE Pfannennavigation

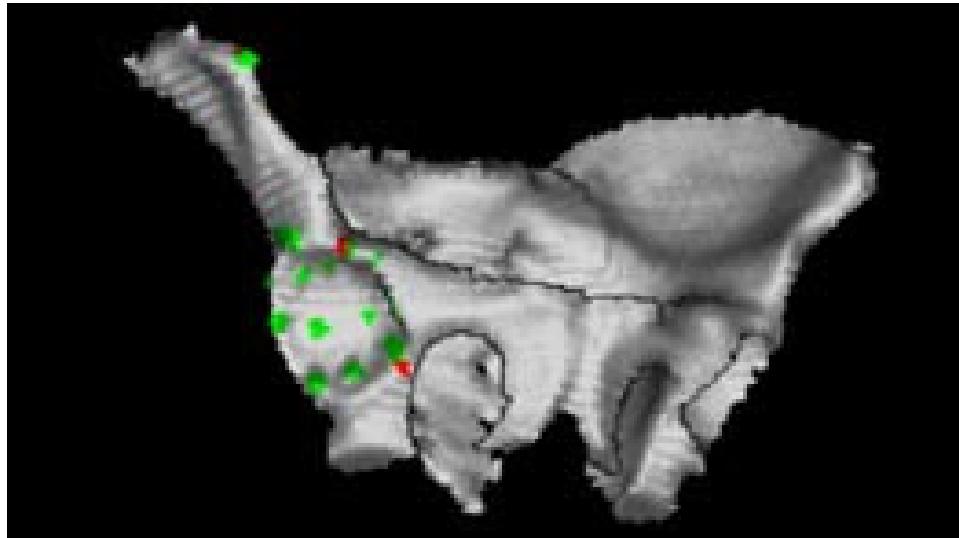


2) Operations-  
planung



# Beispiel: Usability-Evaluierung CAS Navigationssystem

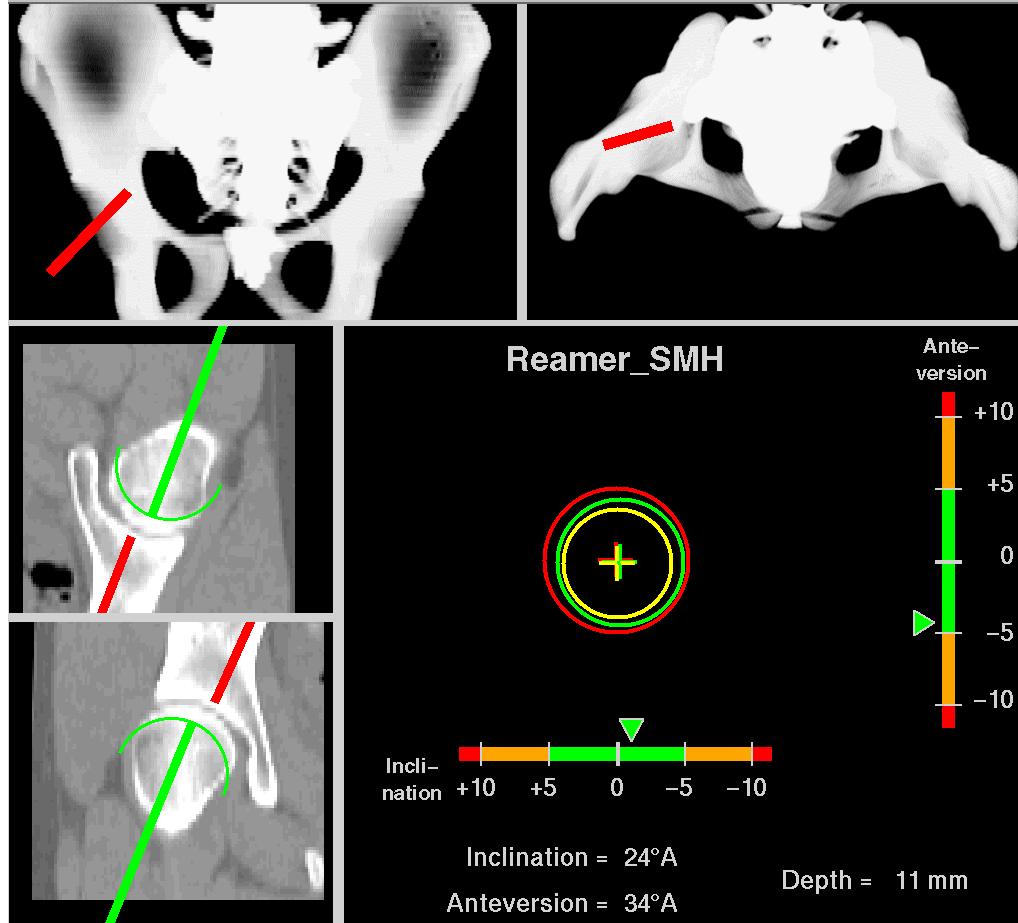
## OP-Ablauf SurgiGATE Pfannennavigation



3) Registrierung der  
Planungsdaten

# Beispiel: Usability-Evaluierung CAS Navigationssystem

## OP-Ablauf SurgiGATE Pfannennavigation



4) Durchführung  
des Eingriffs

# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Bewertungskriterien

### Bewertungskriterien :

- Wie zuverlässig kann die Aufgabe erfüllt werden?
- **Systemzuverlässigkeit**
- Wie gut (Schnelligkeit & Qualität) kann die Aufgabe erfüllt werden?
- **Effektivität / Effizienz**
- Wie gut können Anfänger den Umgang mit dem System lernen?
- **Erlernbarkeit**
- Wie zufrieden sind die Benutzer mit dem System?
- **Benutzerzufriedenheit**



In der Praxis?



# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Bewertungskriterien

### Bewertungskriterien:

- **Systemzuverlässigkeit** → Beobachtung, Erfassung und Klassifikation von Gestaltungsdefiziten, Benutzerbefragung
- **Effektivität / Effizienz** → Messung Bearbeitungszeit und Bestimmung des Grads der Aufgabenerfüllung
- **Erlernbarkeit** → Vergleich Zeiten & Probleme von Novizen zu Experten, Benutzerbefragung
- **Benutzerzufriedenheit** → Fragebogen

### Operationalisierung:

# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Ablauf der Evaluierung

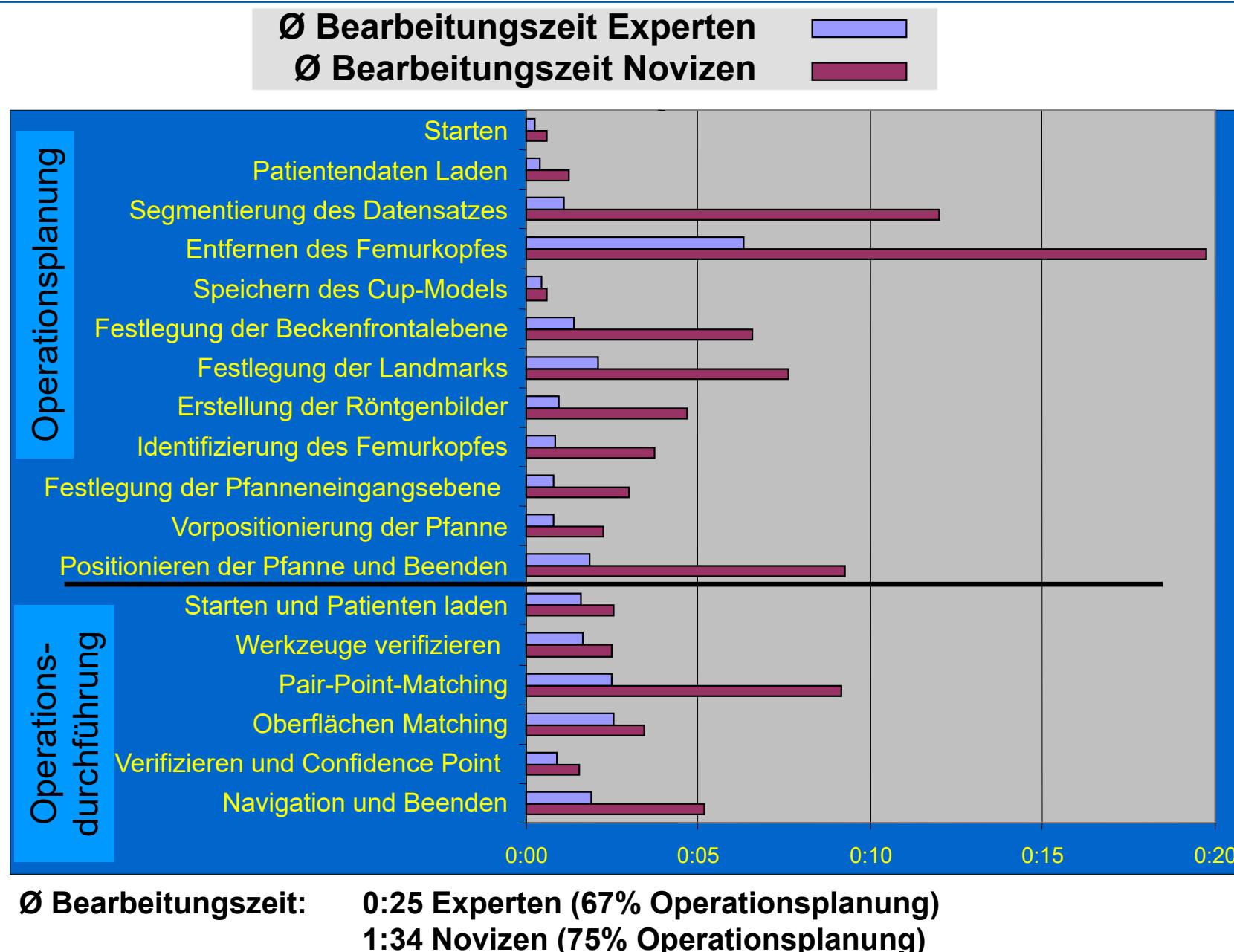
---

- Einführung in das System (möglichst standardisiert z.B. per Video)
- Erklärung der Aufgabe - Handhabung des Tutorials
- Durchführung der Aufgabe Teil A
- Interview
- Durchführung der Aufgabe Teil B
- Interview
- Durchführung der Aufgabe Teil C
- Interview
- Fragebogen



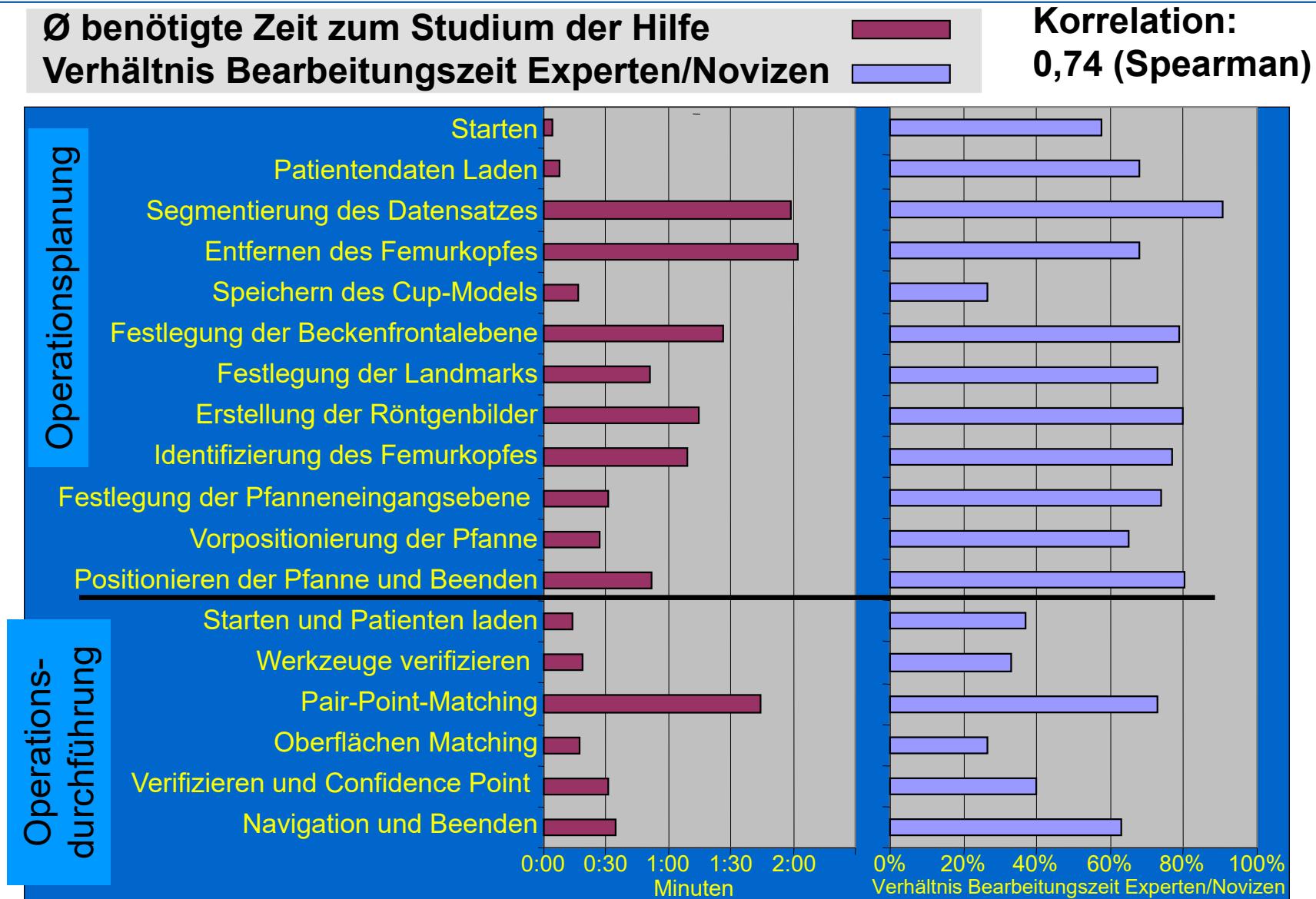
# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Bearbeitungszeiten (reine Interaktion)



# Beispiel: Usability-Evaluierung CAS Navigationssystem

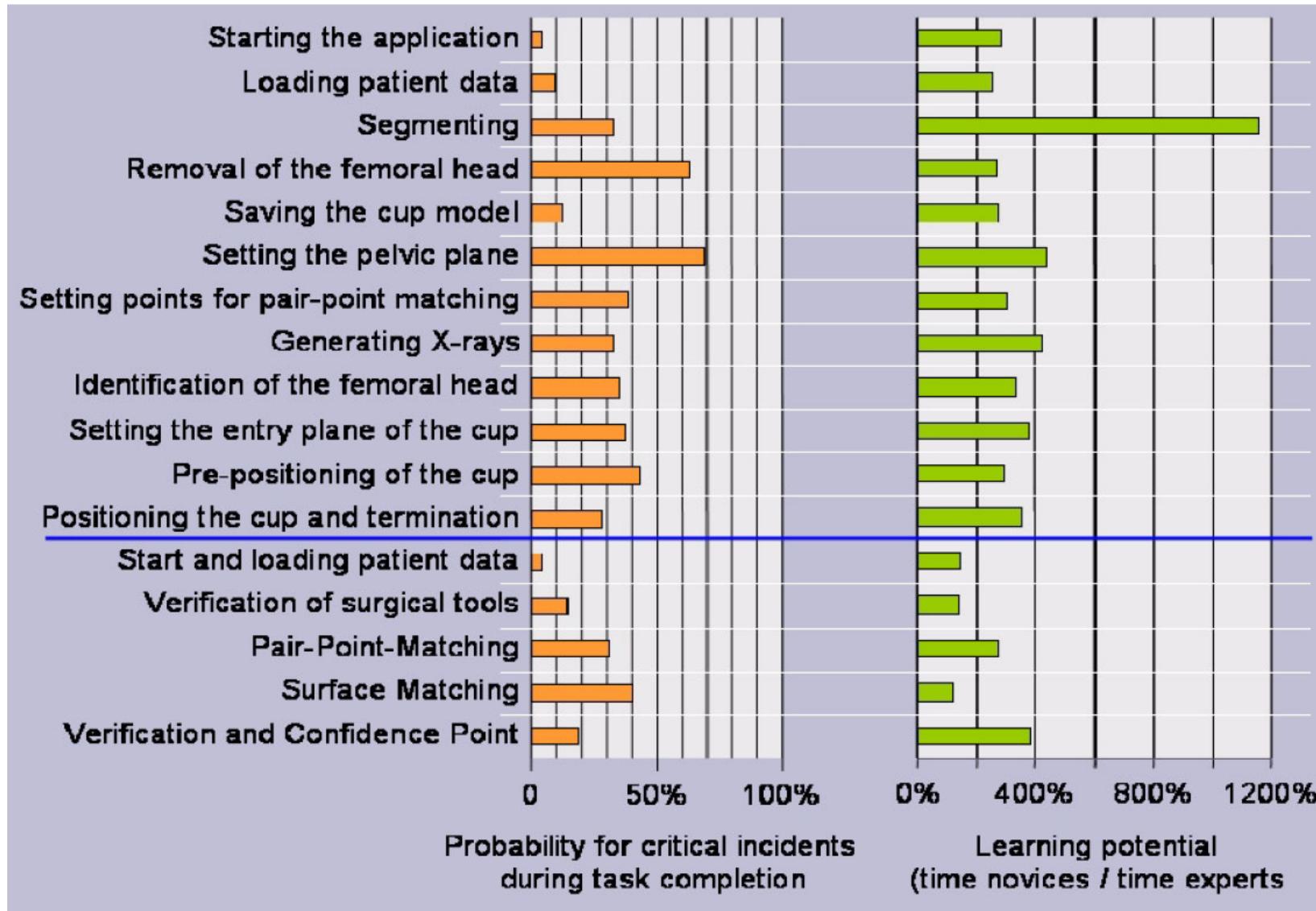
## Erlernbarkeit – komplizierte Arbeitsschritte



Ø benötigte Zeit zum Studium der Hilfe: 14 Minuten (76% Operationsplanung)  
Ø Verhältnis Bearbeitungszeit Experten/Novizen: 73%

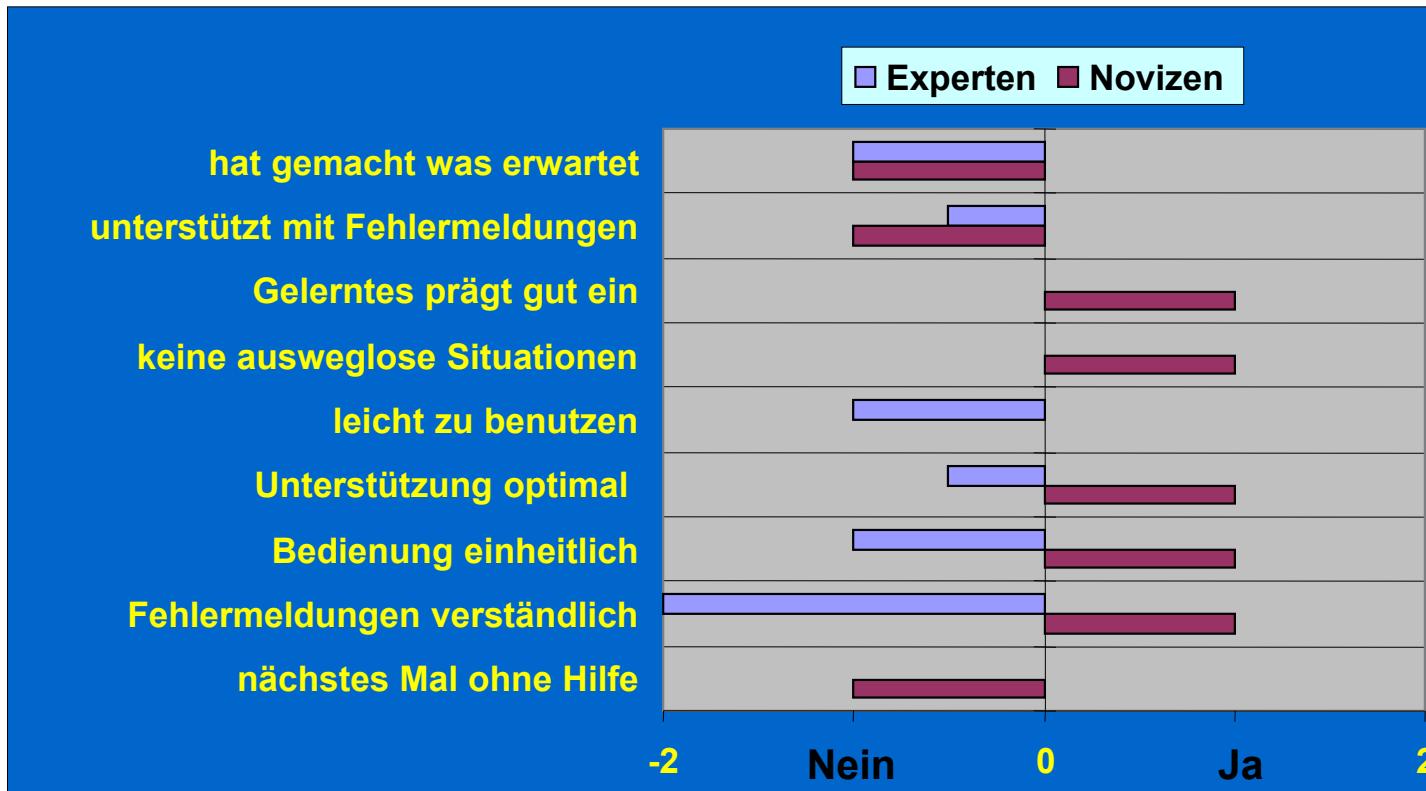
# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Wahrscheinlichkeit kritischer Schritte vs. Lernpotential



# Beispiel: Usability-Evaluierung CAS Navigationssystem

## Erlernbarkeit - Fragebogen



Novizen bewerten signifikant besser als Experten ( $p<0,05$  Wilcoxon)

# Gliederung

---

- Verifikation und Validierung – Begriffsklärung
- Validierungsplanung
- Methoden der Gebrauchstauglichkeits-Evaluierung  
(im Kontext der Entwicklungsphasen)
- Beispiele interaktions- und benutzerzentrierter Evaluierungen