Methoden zur Konformitaetspruefung im Rahmen der ISO 9241
Rauterberg, G.W.M.

Published in:
Zukunftorientiert handeln! : die EU-Bildschirmrichtlinie in der Praxis - Ergebnisse aus dem SANUS-Projekt

Published: 01/01/1997

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 author's 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

Citation for published version (APA):

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?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Methoden zur Normenkonformitätsprüfung im Rahmen von ISO 9241

Matthias Rauterberg

Forschungsgruppe "Mensch-Maschine Interaktion"
Institut für Hygiene und Arbeitsphysiologie
Eidgenössische Technische Hochschule Zürich

Anschrift:
Dr. M. Rauterberg
IHA-ETHZ
Clausiusstrasse 25
CH-8092 Zürich
Schweiz
Tel: 0041-1-632 7082
Fax: 0041-1-632 1173
Email: rauterberg@iha.bepr.ethz.ch
Internet: http://www.iha.bepr.ethz.ch/pages/forschung/MMI/MMI.HTM
Usability Methods

Usability inspection is the name of a set of highly cost-effective methods for finding usability problems and improving the usability of a user interface design by inspection.

Topics to be covered include...
  • **Definition of usability inspection,**
  • the heuristic evaluation method,
  • other inspection methods.
  • Relation between usability inspection methods and user testing.
  • Severity of usability problems found by usability inspection.
  • Cost-benefit characteristics of usability inspection methods.
  • Positioning inspection in the usability engineering lifecycle.

Evaluation
Assessing the usability of an existing design
  • finding usability problems (to fix them)
  • formative evaluation: improve interface, find good/bad parts
  • summative evaluation: are goals met?

Only one part of the usability engineering lifecycle
(task analysis, goal setting, design, prototyping, iteration, field studies, etc.)
Inspection methods

• pluralistic walkthrough [Bias 1991]
  – define a scenario (linear path through interface)
  – get users, designers/developers, usability specialists in one room
  – show user interface one screen at a time (e.g., overheads)
  – have participants write down problems before discussion
  – discuss the screen (let users speak first)
    {may use designer/developer as 'living manual' for early help}

• standards inspection [Wixon, Jones, Tse & Casaday 1994]
  – have a standard expert inspect interface for compliance
    {may cover most of standards without much task knowledge}

• consistency inspection [Wixon, Jones, Tse & Casaday 1994]
  – team of designers/developers (one from each project) inspects a set of interfaces
    for more than one system/application, one at a time

• feature inspection [Bell 1992]
  – imagine typical user task
  – list sequence of features used to accomplish the task
  – check for long sequences, cumbersome steps, additional knowledge, etc.

• cognitive walkthrough [Polson, Lewis, Rieman & Wharton 1992]
  – imagine typical user task
  – use the system to perform the task, 'defining' the correct solution sequence
  – hand-simulate user's problem solving process at each step
  – check if user's goal/memory leads to the defined solution sequence

• quantitative metrics [Rauterberg 1994]
Evaluation methods

- **highly informal evaluation: heuristic evaluation**
  
  Look at interface and make lists of its problems [Nielsen and Molich 1990]:
  - according to checklist of established usability heuristics
  - may also apply any additional usability knowledge
  
  Two or more passes through interface:
  - inspect flow of interface
  - inspect each screen (dialog box, system message, etc.), one at a time
  
  Typical session length: 1–2 hours.
  
  May use observer to help evaluator and note problems mentioned.
  
  Afterwards: aggregate lists of problems from multiple evaluators

- **informal evaluation: usability inspection**
  
  Goals to be met in a somewhat systematic way:
  - generate list of usability problems (main goal)
  - contribute to building design rationale (artifact inspection)
  - provide feedback in design courses [Nielsen et al. 1992]
  - evolve a parallel design [Nielsen 1993]
  
  Tools support for inspection:
  - mostly none
  - online forms for cognitive walkthroughs [Lewis et al. 1992]
  - online/hypertext guidelines/standards documents
  - CSCW tools for team heuristic evaluations
  
  (show panel to be discussed for annotation/drawing/pointing)

- **structured evaluation: usability tests**
Frage:
Gibt es Methoden zur Normenkonformitätsprüfung im Rahmen der ISO 9241?

Antwort:
Nein!

Begründung:
Die ISO 9241 enthält keine ge-"normten" Metriken mit zugehörigen Grenzwerten.

Was nun?
## Methoden zur Qualitätssicherung

### Benutzer vorhanden...

<table>
<thead>
<tr>
<th>virtuell</th>
<th>real</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>formaler Ansatz:</strong> formale Theorie</td>
<td><strong>benutzer-zentriert:</strong> Fragebogen, Interview, Mock-ups</td>
</tr>
<tr>
<td><strong>produkt-zentriert:</strong> Experten Evaluation</td>
<td><strong>interaktions-zentriert:</strong> Usability-Test</td>
</tr>
</tbody>
</table>

Aufwand und Kosten
der produkt-zentrierte Meß-Ansatz

- Checklisten
- Experten-Evaluation
der benutzer-zentrierte Meß-Ansatz

- mündliche Befragung (Interview)
- schriftliche Befragung (Umfragen)
- Diskussionen (zB. in Workshops)
interaktions-zentrierter Meß-Ansatz

He!
Ich Chef - du Werkzeug!
Begreifen?

- aufgaben-orientierte Usability-Tests
- induktive Usability-Tests (formative evaluation)
- deduktive Usability-Tests (summative evaluation)
Modell-1

Auftraggeber  Software-Entwickler  BenutzerIn

Modell-2

Auftraggeber  Software-Entwickler  BenutzerIn

Usability-Experten

Modell-3

Auftraggeber  Software-Entwickler  BenutzerIn
Das Quadranten-Modell

[BOSS-Projekt, Rauterberg 1991]
ein formale Gestaltungs-Theorie

• Metriken
An abstract concept to describe usability aspects
quantitative measure of "feedback":

\[
FB = \frac{1}{D} \sum_{d=1}^{D} \left( \frac{\#PF_d}{\#HF_d} \right) \times 100\%
\]

quantitative measure of "interactive directness":

\[
ID = \left\{ \frac{1}{P} \sum_{p=1}^{P} \min[\ln\text{(PATH}_p)] \right\}^{-1} \times 100\%
\]

[visual] feedback (FB)

<table>
<thead>
<tr>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>menu interface</td>
</tr>
<tr>
<td>MI</td>
<td></td>
</tr>
<tr>
<td>command language</td>
<td>desktop style</td>
</tr>
<tr>
<td>CI</td>
<td>direct manipulation</td>
</tr>
<tr>
<td>DI</td>
<td></td>
</tr>
</tbody>
</table>
The outcomes of nine (9) different comparison studies between command (CI) and menu (MI) interfaces.

"CI < MI" means that the average usage/preference with/for MI is better than with/for CI; "CI = MI" means that there are no published data to decide; "CI > MI" means that the average usage/preference with/for CI is better than with/for MI; "sig." means that p ≤ 0.05; "not sig." means that p > 0.05

<table>
<thead>
<tr>
<th>Reference</th>
<th>Interface</th>
<th>Skill level</th>
<th>Usability metric</th>
<th>Outcome</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streitz et al. (1987)</td>
<td>CI, MI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Chin et al. (1988)</td>
<td>CI, MI</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Ogden &amp; Boyle (1982)</td>
<td>CI, MI, HY</td>
<td>beginner</td>
<td>preferences</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Roy (1992)</td>
<td>CI, MI</td>
<td>advanced</td>
<td>error rate</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Roberts &amp; Moran (1983)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>task solving time</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Chin et al. (1988)</td>
<td>CI, MI</td>
<td>experts</td>
<td>subjective rating</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>slips</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>recognition errors</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>efficiency</td>
<td>CI &lt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Ogden &amp; Boyle (1982)</td>
<td>CI, MI, HY</td>
<td>beginner</td>
<td>task time</td>
<td>CI &lt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Roy (1992)</td>
<td>CI, MI</td>
<td>advanced</td>
<td>task solving time</td>
<td>CI &lt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Antin (1988)</td>
<td>CI, MI</td>
<td>advanced</td>
<td>subjective rating</td>
<td>CI &lt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Hauptmann &amp; Green (1983)</td>
<td>CI, MI, NO</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI = MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Hauptmann &amp; Green (1983)</td>
<td>CI, MI, NO</td>
<td>beginner</td>
<td>number of errors</td>
<td>CI = MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Hauptmann &amp; Green (1983)</td>
<td>CI, MI, NO</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI = MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Whiteside et al. (1985)</td>
<td>CI, MI, IO</td>
<td>beginner</td>
<td>task completion rate</td>
<td>CI &gt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Antin (1988)</td>
<td>CI, MI</td>
<td>KMI advanced</td>
<td>preferences</td>
<td>CI &gt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Roberts &amp; Moran (1983)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>error-free task time</td>
<td>CI &gt; MI</td>
<td>not sig</td>
</tr>
<tr>
<td>Whiteside et al. (1985)</td>
<td>CI, MI, IO</td>
<td>advanced</td>
<td>task completion rate</td>
<td>CI &gt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Streitz et al. (1987)</td>
<td>CI, MI</td>
<td>advanced</td>
<td>task solving time</td>
<td>CI &gt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Antin (1988)</td>
<td>CI, MI</td>
<td>KMI advanced</td>
<td>task completion rate</td>
<td>CI &gt; MI</td>
<td>sig.</td>
</tr>
<tr>
<td>Whiteside et al. (1985)</td>
<td>CI, MI, IO</td>
<td>experts</td>
<td>task completion rate</td>
<td>CI &gt; MI</td>
<td>sig.</td>
</tr>
</tbody>
</table>
The outcomes of twelve (12) different comparison studies between command (CI) and direct manipulative (DI) interfaces.

"CI < DI" means that the average usage/preference with/for DI is better than with/for CI;
"CI = DI" means that there are no published data to decide;
"CI > DI" means that the average usage/preference with/for CI is better than with/for DI;
"sig." means that $p \leq 0.05$; "not sig." means that $p > 0.05$

<table>
<thead>
<tr>
<th>Reference</th>
<th>Interface</th>
<th>Skill level</th>
<th>Usability metric</th>
<th>Outcome</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altmann (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Karat et al. (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Streitz et al. (1989)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Sengupta &amp; Te'eni (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Margono et al. (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>number of errors</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Morgan et al. (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>number of errors</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Morgan et al. (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>time between errors</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Karat et al. (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>error correction time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Morgan et al. (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>error-free time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Margono et al. (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Morgan et al. (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Torres-Chazaro et al. (1992)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Sengupta &amp; Te'eni (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>efficient usage</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Tombaugh et al. (1989)</td>
<td>CI, DI</td>
<td>advanced</td>
<td>subjective rating</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Torres-Chazaro et al. (1992)</td>
<td>CI, DI</td>
<td>advanced</td>
<td>subjective rating</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Roberts &amp; Moran (1983)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>oblivion's errors</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>recognition error</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Peters et al. (1990)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>efficiency</td>
<td>CI &lt; DI</td>
<td>sig.</td>
</tr>
<tr>
<td>Margono et al. (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>not sig.</td>
</tr>
<tr>
<td>Morgan et al. (1991)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>not sig.</td>
</tr>
<tr>
<td>Tombaugh et al. (1989)</td>
<td>CI, DI</td>
<td>advanced</td>
<td>task solving time</td>
<td>CI &lt; DI</td>
<td>not sig.</td>
</tr>
<tr>
<td>Roberts &amp; Moran (1983)</td>
<td>CI, MI, DI</td>
<td>experts</td>
<td>error correction time</td>
<td>CI &lt; DI</td>
<td>not sig.</td>
</tr>
<tr>
<td>Altmann (1987)</td>
<td>CI, DI</td>
<td>beginner</td>
<td>subjective rating</td>
<td>CI &gt; DI</td>
<td>not sig.</td>
</tr>
<tr>
<td>Masson et al. (1988)</td>
<td>CI, DI</td>
<td>advanced</td>
<td>task solving time</td>
<td>CI &gt; DI</td>
<td>sig.</td>
</tr>
</tbody>
</table>
Contingency tables of a meta-analysis for all data

[Cell Content: observed frequency  (expected frequency)]

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>DI</th>
<th>outcome of this meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI better as</td>
<td>7</td>
<td>2</td>
<td>Chi** = 5.52, df = 1</td>
</tr>
<tr>
<td>CI worse as</td>
<td>12</td>
<td>23</td>
<td>p ≤ .019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>DI</th>
<th>outcome of this meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI better as MI,DI</td>
<td>2</td>
<td>7</td>
<td>Chi** = 2.95, df = 1</td>
</tr>
<tr>
<td>CI worse as MI,DI</td>
<td>19</td>
<td>16</td>
<td>p ≤ .086</td>
</tr>
</tbody>
</table>

Contingency tables only for significant differences
(SELECTION for "result" = "sig.").

[Cell Content: observed frequency  (expected frequency)]

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>DI</th>
<th>outcome of this meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI better as</td>
<td>4</td>
<td>1</td>
<td>Chi** = 4.07, df = 1</td>
</tr>
<tr>
<td>CI worse as</td>
<td>9</td>
<td>19</td>
<td>p ≤ .044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>DI</th>
<th>outcome of this meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI better as MI,DI</td>
<td>0</td>
<td>5</td>
<td>Chi** = 5.55, df = 1</td>
</tr>
<tr>
<td>CI worse as MI,DI</td>
<td>16</td>
<td>12</td>
<td>p ≤ .018</td>
</tr>
</tbody>
</table>
quantitative measure of "dialog flexibility":

\[ DFD = \frac{1}{K} \sum_{d=1}^{K} \#HDFIP_d \]

quantitative measure of "application flexibility":

\[ DFA = \frac{1}{K} \sum_{d=1}^{K} \#HAFIP_d \]

performance

\[ P \]

flexibility

\[ Fl = DFA + DFD \]
mentales Modell

Syntax und Semantik

Dialog-Struktur

Sicht der BenutzerIn ('von aussen')

Operator

Operation

Funktion

Sicht der SoftwareentwicklerIn ('von innen')
Zukunftssorientiert handeln!

Die EU-Bildschirmrichtlinie in der Praxis

Ergebnisse aus dem SANUS-Projekt

2. SANUS - Kongress in Bad Honnef

- Tagungsunterlagen -

Herausgeber: Prof. Dr. Ing. Peter Kern
© by Institut für Arbeitswissenschaft und Technologiemanagement IAT,