

# Fehlerbeurteilungen in der Mensch-Computer Interaktion

## Reliabilitätsanalysen und Training einer handlungstheoretischen Fehlertaxonomie

Jochen Prümper

*In diesem Beitrag wird von einer Dissertation berichtet, die am Institut für Arbeits- und Organisationspsychologie der Justus-Liebig-Universität Gießen von Prof. Dr. Michael Frese betreut wurde.*

*Fehler in der Mensch-Computer Interaktion sind mit hohen Kosten und psychischen Stress verbunden. Zuverlässig analysiert stellen sie jedoch eine fruchtbare Möglichkeit dar, um sowohl software-ergonomische Defizite als auch organisationale Mängel zu evaluieren. Wie man reliable Fehlerbeurteilungen durchführt ist Thema dieses Buches.*

Die vorliegende Arbeit widmet sich zunächst der Frage, wie reliabel sich Probleme und Schwierigkeiten mit Computern in eine handlungsorientierte Fehlertaxonomie einordnen lassen. Die empirische Grundlage dieser Untersuchung liefern drei Feldstudien, bei denen 244 Benutzer und Benutzerinnen aus 13 Organisationen für jeweils etwa 2 Stunden bei ihrer alltäglichen Computerarbeit mit unterschiedlichster Bürosoftware beobachtet wurden.

Infolge wird mit dem „Fehlertrainingsbogen für die Mensch-Computer Interaktion“ FTB-MCI in zwei experimentellen Studien die Entwicklung eines Verfahrens vorgestellt, das als Basisinstrument im Rahmen einer Beurteiler-schulung für eine handlungsorientierte Fehlertaxonomie dienen kann.

Grundlage der Arbeit ist eine auf handlungstheoretischen Überlegungen aufgebaute Fehlertaxonomie für die Mensch-Computer Interaktion, in der Funktionsprobleme als „mismatch“ zwischen Computer und Arbeitsaufgabe und Nutzungsprobleme als „mismatch“ zwischen Benutzer und Computer unterschieden werden. Darüber hinaus wird eine Abgrenzung zwischen Nutzungsproblemen und Ineffizienzen vorgenommen. Nutzungsprobleme werden nach Regulationsebenen und nach Schritten im Handlungsprozeß aufgeschlüsselt. Zusätzlich zu den Ebenen der Handlungsregulation wird noch die Regulationsgrundlage beschrieben, auf der Fehler aufgrund unvollständigen, unrealistischen oder fehlenden Wissens auftreten. Weiterhin werden Interaktionsprobleme als „mismatch“ zwischen mehreren an der Tätigkeit beteiligten Personen vorgestellt.

Nachdem zur Heranführung an das Gütekriterium der Reliabilität eine Unterscheidung zwischen „diachronischer“ und „synchronischer“ Reliabilität getroffen wird, wird die Beziehung zwischen dem Begriff der Reliabilität und den Gütekriterien der Validität und Objektivität geklärt. Anschließend werden Probleme diskutiert, die im besonderen bei der Beobachtung von Fehlern in der Mensch-Computer Interaktion im Feld zu erwarten sind und es wird eine Übersicht über in diesem Zusammenhang auftretende Beurteilungsfehler gegeben.

Nach einer Analyse von Modellvorschlägen zur Kontrolle von methodischen Problemen der Beobachtungsmethode werden mit der „synchronischen Doppelanalyse“ und der „synchronischen Wiederholungsanalyse“ zwei Modelle zur Kontrolle der Reliabilität vorgestellt.

Ausgehend von Cohen's kappa werden in Folge diverse Varianten zur Bestimmung der Reliabilität von nominalskalierten Beobachtungsdaten dargestellt, die für die vorliegende Fragestellung bedeutsam sind. Dabei handelt es sich um kappa-Formeln für die Bestimmung der Intra- und Inter-Klassen-Reliabilitäten für ein und mehr als ein Beurteilerpaar und um ein kappa-Maß für die Berechnung der Inter-Rater-Reliabilität. Bei der Bestimmung der Intra-Klassen-Reliabilität werden die Übereinstimmungen bei gleichen Kategorien und bei der Bestimmung der Inter-Klassen-Reliabilität die Übereinstimmungen bei ungleichen Kategorien berechnet. Unter der Inter-Rater-Reliabilität wird der Vergleich von Beurteilungen mit einem Standard verstanden.

Die empirischen Untersuchungen liefern im großen und ganzen zufriedenstellende Ergebnisse. Hohe Inter-Klassen-Reliabilitäten sprechen dafür, daß es sich bei der handlungsorientierten Fehlertaxonomie um ein System von gut abgrenzbaren Ereignisklassen handelt. Allerdings weisen die nicht immer zufriedenstellenden Intra-Klassen-Reliabilitäten darauf hin, daß seitens der Untersucher noch Qualifikationsbedarf besteht.

Um dieses Defizit zu beheben, wird in zwei experimentellen Studien ein Verfahren für die Beurteilerschulung der Fehlertaxonomie (FTB-MCI) entwickelt. Den Kern dieses Trainings bildet ein schriftlicher Fehlertrainingsbogen für die Mensch-Computer Interaktion, der aus einer Sammlung von Fehlerereignissen besteht. Der FTB-MCI wird ebenfalls bezüglich seiner Intra- und Inter-Klassen-Reliabilitäten sowie auf seine Inter-Rater-Reliabilitäten hin untersucht und einer Validierung unterzogen. Bezüglich der Gesamt-Reliabilität werden für beide Teile und Zeitpunkte des FTB-MCI „gute“ Werte erzielt. Allerdings zeigt die Analyse der

Inter-Klassen-Reliabilität, daß bei der Beurteilung des FTB-MCI hauptsächlich Probleme hinsichtlich der Einschätzung des einer Handlung zugrundeliegenden Geübtheitsgrades auftreten. Bezüglich der Inter-Rater-Reliabilität zeigt sich, daß sich der FTB-MCI aufgrund einer deutlichen Streuung zwischen den kappa-Werten der einzelnen Versuchspartner als differentielles Verfahren dazu eignet, die Beurteilerleistung einzelner Personen zu bestimmen. Zur Validitätsprüfung des FTB-MCI dient als äußeres Validitätskriterium ein Fehlerbeurteilungsfilm, in dem den Versuchspartnern mehrere Fehlersequenzen präsentiert werden. Die Korrelationen zwischen den in dem FTB-MCI und den in dem Fehlerbeurteilungsfilm erzielten kappa-Werten sind hoch signifikant. Damit zeigt sich, daß der FTB-MCI einen unmittelbaren Rückschluß auf den Ausprägungsgrad des Verhaltensmerkmals „Beurteilungsfähigkeit von Fehlern in der Mensch-Computer Interaktion“ zuläßt.

In einem Exkurs wird zum Schluß noch der Frage nachgegangen, ob die Erfahrung mit Computern bei Fehlerbeurteilungen eine Rolle spielt. Es zeigt sich, daß der Einfluß der Computere Expertise auf die Güte der Fehlerbeurteilungen in dem FTB-MCI durch eine eingeschobene Diskussionsphase ausgeglichen werden kann. Allerdings wird auch deutlich, daß in dem Fehlerbeurteilungsfilm umso reliablere Fehlerbeurteilungen abgegeben werden, je länger die Untersucher bereits mit Computern arbeiten, je mehr Programme sie beherrschen und je mehr Erfahrung sie mit dem untersuchten Programm haben. Für praktische Fehlerbeobachtungen bedeutet dies, daß nur solche Untersucher zum Einsatz kommen dürfen, die sowohl über eine ausreichende Kompetenz, reliable Fehlerbeurteilungen abzugeben, als auch über eine genügend große Computere Expertise verfügen.

Prümper, Jochen  
Fehlerbeurteilungen in der Mensch-Computer Interaktion - Reliabilitätsanalysen und Training einer handlungstheoretischen Fehlertaxonomie.  
Münster/New York: Waxmann Verlag.  
1994, 160 Seiten, broschiert  
ISBN 3-89325-245-2, DM 49.90

Quelle:

Prümper, J. (1994). *Fehlerbeurteilungen in der Mensch-Computer Interaktion. Reliabilitätsanalysen und Training einer handlungstheoretischen Fehlertaxonomie*. Münster: Waxmann.

Bezugsquelle: [Waxmann Verlag GmbH](http://www.waxmann-verlag.de)