

Verhaltensbasierte Validierung von Usability-Fragebögen

Frank Ollermann

Fachgebiet Arbeits- und Organisationspsychologie, Universität Osnabrück

Zusammenfassung

Es wird ein Verfahren zur Validierung eines Fragebogens zur Usability-Evaluation vorgestellt. Als Validierungskriterium dient die Anzahl der Problemnennungen, die in einer unabhängigen formativen Evaluationsuntersuchung desselben Softwareprodukts gewonnen wurden. Die Anwendung auf einen Fragebogen zur Evaluation von Hypermedia-Anwendungen deutet auf eine Eignung dieses Verfahrens zur Validierung summativer Evaluationsinstrumente hin. Alternative Anwendungen dieses Prinzips werden vorgestellt und Möglichkeiten der Weiterentwicklung beschrieben.

1 Einleitung

Trotz einiger wichtiger Einschränkungen sind Fragebögen eine sehr häufig eingesetzte Methode der Usability-Evaluation. Dies liegt vor allem am vergleichsweise geringen Durchführungs- und Auswertungsaufwand. Beispiele für Fragebögen in der Usability-Evaluation sind der „Questionnaire for User Interface Satisfaction“, QUIS (Chin et al. 1988), das „Software Usability Measurement Inventory“, SUMI (Kirakowski 1994; Kirakowski & Corbett 1993), der ISONORM 9241/10 (Prümper 1997; Prümper & Anft 1993) oder der IsoMetrics^S (Gediga & Hamborg 1999). Diese Fragebögen dienen vorwiegend der Gewinnung quantitativer Daten und können somit als psychometrische Tests aufgefasst werden.

An derartige Tests sind verschiedene Anforderungen zu stellen. Das wichtigste der so genannten Hauptgütekriterien stellt die *Validität* dar. Sie gibt an, „wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt“ (Bortz & Döring 1995, 185). Eine Möglichkeit der Validierung besteht in der Ermittlung der *Kriteriumsvalidität*. Hierbei wird der Zusammenhang zwischen den Ergebnissen des Fragebogens zur Messung eines latenten Konstrukts und den Messungen eines korrespondierenden manifesten Kriteriums ermittelt. Entscheidend ist hierbei die Wahl eines adäquaten Kriteriums. Ein Beispiel für ein häufig eingesetztes Kriterium im Bereich der Usability-Evaluation ist die Einschätzung der Gebrauchstauglichkeit durch Experten. Diese kann mit der durch den Fragebogen ermittelten

Gebrauchstauglichkeit korreliert werden. Eine andere Strategie ist die „Technik der bekannten Gruppen“, bei der zwei Gruppen von Softwareprodukten evaluiert werden, von denen erwartet werden kann, dass sie sich in Bezug auf ihre Gebrauchstauglichkeit unterscheiden. Der Fragebogen sollte in der Lage sein, zwischen diesen beiden Gruppen zu differenzieren.

Oftmals sind die in der Forschungspraxis gewählten Kriterien nicht sehr aussagekräftig. So findet der QUIS Unterschiede zwischen Programmen, die von den Benutzern gemocht werden, und solchen, die nicht gemocht werden (Chin et al. 1988). Das Validierungskriterium ist hier der Aspekt des Gefallens. Dies ist jedoch ein sehr globales Kriterium, dessen Zusammenhang mit der Gebrauchstauglichkeit von den Autoren nicht diskutiert wird.

Weiterhin vermag der QUIS zwischen kommandozeilen- und menügesteuerten Programmen zu unterscheiden. Ein Fragebogen sollte jedoch nicht nur zwischen diesen extrem unterschiedlichen Arten von Softwareprogrammen unterscheiden können, sondern auch beispielsweise zwischen zwei unterschiedlich ausgereiften Prototypen ein- und desselben Produkts.

Der SUMI vermag zwischen Textverarbeitungsprogrammen unterschiedlicher Generationen zu unterscheiden. Ein weiterer Hinweis für die Validität ergibt sich aus der Fähigkeit des Fragebogens, zwischen zwei Datenbank Anwendungen zu unterscheiden, die bewusst ergonomisch bzw. unergonomisch gestaltet wurden (Kirakowski 1994). Die beiden Gruppen von Software unterscheiden sich in ihrer angenommenen Gebrauchstauglichkeit zwar nicht so stark wie diejenigen, die für die Validierung des QUIS herangezogen werden. Dennoch können die Aussagen über die Validität des Instruments nicht als gesichert angesehen werden, da auch hier die Eignung der eingesetzten Validierungskriterien nicht in ausreichendem Maße dargestellt wird.

Beiden genannten Instrumenten gemeinsam ist die Vernachlässigung der Validierung der einzelnen Fragebogendimensionen. Da jede Dimension einen anderen Aspekt der Gebrauchstauglichkeit zu messen beansprucht, muss auch für jede Dimension getrennt die jeweilige Validität ermittelt werden.

Bei der Validierung des ISONORM 9241/10 wurden als Validierungskriterien die Gesamtscores zweier anderer Usability-Evaluationsfragebögen herangezogen. Außerdem wurde der Gesamtscore des ISONORM 9241/10 mit dem Ergebnis einer Expertenevaluation korreliert (Prümper 1997). Wiederum wurde bei keiner dieser beiden Untersuchungen die Validität der einzelnen Dimensionen berücksichtigt. Der ISONORM 9241/10 differenziert zwar zwischen Office-Software-Produkten mit und ohne grafische Benutzungsschnittstelle; wie schon beim QUIS scheint dies jedoch eher eine Mindestanforderung zu sein, die an Fragebögen dieser Art zu stellen sind.

Die Validierung des IsoMetrics^S berücksichtigt zwar ausdrücklich die unterschiedlichen Fragebogendimensionen, basiert darüber hinaus jedoch weitgehend auf Plausibilitätsannahmen. Allerdings vermag das Instrument in zwei der sieben Subskalen zwischen zwei aufeinander folgenden Prototyp-Versionen eines Systems zu unterscheiden (Gediga et al. 1999).

Eine viel versprechende Validierungsstrategie soll im Folgenden beschrieben werden. Als Validierungskriterium wird dabei die relative Anzahl von Problemnennungen aus einer un-

abhängigen formativen Evaluationsuntersuchung zum selben Softwareprodukt herangezogen. Wird zu einem bestimmten Aspekt (in Form einer Fragebogenskala) eine im Mittel besonders schlechte Bewertung ermittelt, dann sollten in einer unabhängigen formativen Untersuchung zu eben diesem Aspekt besonders viele Schwachstellen gefunden werden. Die Eignung dieses Kriteriums zur Validierung wird dadurch gestützt, dass die Häufigkeit von Problemnennungen als Indikator für den Schweregrad des betreffenden Usability-Problems herangezogen werden kann (Gediga et al. 1999; Mayhew 1999; Nielsen 1994; Rubin 1994). Durch dieses Vorgehen wird die Messung eines subjektiven Eindrucks von der Software an einem objektiveren, verhaltensbasierten Kriterium validiert. Ein Zusammenhang dieser beiden Maße würde einen deutlicheren Hinweis auf die Validität des Fragebogens liefern als die Fähigkeit zur Differenzierung zwischen extrem unterschiedlichen Softwareprodukten oder die Korrelation mit dem Gesamtergebnis eines anderen subjektiven Evaluationsinstruments.

2 Methode

In einem Usability-Test in Form einer Videokonfrontation (Hamborg & Greif 1999) wurde ein Online-Medienshop in Bezug auf seine Gebrauchstauglichkeit evaluiert. Das Ergebnis der Evaluation waren von den Versuchsteilnehmern identifizierte Schwachstellen des Online-Shops.

In einer weiteren Untersuchung wurde derselbe Online-Shop mittels eines Fragebogens zur Gebrauchstauglichkeit von Hypermedia-Anwendungen evaluiert. Der Fragebogen bestand aus 94 Items in zehn Subskalen.

Die von den Teilnehmern der Videokonfrontation gemachten Aussagen über Schwachstellen des Online-Shops wurden anschließend inhaltlich den einzelnen Subskalen des Fragebogens zugeordnet.

2.1 Fragebogen

Der Evaluationsfragebogen beruhte auf den Gestaltungsgrundsätzen der DIN EN ISO 9241-10 (DIN 1996), die um Aspekte ergänzt worden waren, die für die Evaluation von Hypermedia-Anwendungen relevant erschienen. Im Einzelnen handelte es sich um die folgenden Dimensionen:

- Aufgabenangemessenheit (9 Items)
- Selbstbeschreibungsfähigkeit (16 Items)
- Steuerbarkeit (9 Items)
- Erwartungskonformität (10 Items)
- Fehlerrobustheit (9 Items)

- Individualisierbarkeit (5 Items)
- Erlernbarkeit (7 Items)
- Informationsstruktur (15 Items)
- Medienqualität (8 Items)
- Akzeptanz (6 Items)

Die einzelnen Items waren zur Bearbeitung durch Benutzer vorgesehen und wurden auf einer fünfstufigen Ratingskala beantwortet.

2.2 Untersuchung 1: Videokonfrontation

2.2.1 Teilnehmer

Acht Personen (3 weiblich, 5 männlich) nahmen an der Untersuchung teil. Die Rekrutierung erfolgte mittels Anzeigen in der lokalen Presse. Das mittlere Alter der Teilnehmer betrug 28.1 Jahre ($SD = 8.1$). Die Teilnehmer erhielten für ihre Teilnahme eine Aufwandsentschädigung von DM 100,-.

2.2.2 Materialien

Die Untersuchung fand im Labor des Fachgebiets Arbeits- und Organisationspsychologie der Universität Osnabrück statt. Der Teilnehmer saß im Untersuchungsraum vor einem Einzelplatzrechner, der über eine Standleitung mit dem Internet verbunden war. Der angrenzende Steuerungsraum enthielt die für die Beobachtung verwendete Videotechnik und war mittels eines Einwegspiegels vom Untersuchungsraum getrennt, durch den die Nutzungssituation zusätzlich beobachtet werden konnte. Es wurden das Monitorbild mit Hilfe eines Scan Converters sowie Gesicht und Hände des Teilnehmers mit Hilfe von Videotechnik aufgenommen und mit Hilfe eines Videomischers zu einem Gesamtbild kombiniert.

Der Teilnehmer erhielt gedruckte Unterlagen, in denen jeder Schritt der Untersuchung beschrieben war. Insbesondere enthielten sie die Testaufgaben, die als Grundlage für die Exploration des Online-Shops dienten. Diese wurden nacheinander bearbeitet und enthielten Aufgaben, wie sie für die Benutzung eines solchen Online-Shops typisch waren:

- Stöbern in den Produktkategorien
- Suche nach Produkten
- Sortieren und Eingrenzen von Suchergebnissen
- Verwalten von Produkten im Warenkorb
- Bestellen mehrerer Artikel, teilweise als Geschenk
- Verwaltung eines Benutzerkontos

- Suche nach Informationen zu Artikeln, Zahlungsweisen und dem Bestellvorgang

2.2.3 Durchführung

Nach einer kurzen Begrüßung und der Vorstellung der Untersuchung wurde ein kurzer Fragebogen zu demografischen Daten und Computer- und Internetnutzung ausgefüllt.

Dann begab sich der Versuchsleiter in den Steuerungsraum und der Teilnehmer begann mit der Bearbeitung der Nutzungsszenarien. Während der Bearbeitung machte sich der Versuchsleiter Notizen zu erkennbaren Problemsituationen. Bei Nutzungsproblemen griff der Versuchsleiter nur dann helfend ein, wenn die weitere Durchführung der Untersuchung in Frage stand.

Nach der Bearbeitung der Szenarien schauten sich Versuchsleiter und Teilnehmer gemeinsam die Videoaufzeichnung der Benutzungssituation an. An den zuvor als problematisch notierten Stellen wurde das Band angehalten und der Teilnehmer gebeten zu sagen, was er in der jeweiligen Situation als problematisch oder auch hilfreich empfunden habe und welche Verbesserungsvorschläge er möglicherweise zu diesem Problem hätte. Außerdem wurde der Teilnehmer ermutigt, jederzeit Anmerkungen zu ergonomischen Schwachstellen des Online-Shops zu machen. Sämtliche Problemnennungen wurden vom Versuchsleiter notiert.

2.2.4 Auswertung

Um die Vergleichbarkeit beider Untersuchungsgruppen überprüfen zu können, wurde ein Vorerfahrungsindex gebildet, indem die Punktwerte für Fragen zu Computer- und Internetnutzung summiert wurden.

Die 336 Problemnennungen wurden den zehn Subskalen des Fragebogens (siehe Abschnitt 2.1) zugeordnet. Diese Zuordnung wurde durch den Autor vorgenommen.

2.3 Untersuchung 2: Fragebogenevaluation

2.3.1 Teilnehmer

An der Untersuchung nahmen 31 Personen (8 männlich, 23 weiblich) teil. Das Alter der Teilnehmer betrug im Mittel 28 Jahre ($SD = 7.1$). Sie wurden mittels Zeitungsannoncen und Aushängen für die Untersuchung angeworben und erhielten für ihre Teilnahme eine Aufwandsentschädigung von DM 50,-.

2.3.2 Materialien

Die Untersuchung fand in einem Computer-Gruppenraum statt. Zum Zeitpunkt der Untersuchung befanden sich in diesem Raum zehn Rechner, die über eine Standleitung mit dem Internet verbunden waren.

Die Teilnehmer erhielten ausgedruckte Unterlagen, in denen die Schritte des Untersuchungsablaufs erläutert waren. Wichtigster Teil dieser Unterlagen waren die Testaufgaben, anhand derer die Exploration des Online-Shops erfolgen sollte. Diese Aufgaben waren identisch mit denen, die bereits in der Videokonfrontation zum Einsatz gekommen waren.

2.3.3 Durchführung

Die Untersuchung fand in Gruppen statt, die jeweils aus sieben bis zehn Teilnehmern bestanden. Nach einer kurzen Vorstellung der Untersuchung folgte die Beantwortung der demografischen Fragen. Dann wurden die Testaufgaben bearbeitet und im Anschluss daran der Evaluationsfragebogen ausgefüllt.

2.3.4 Auswertung

Analog zur Videokonfrontation wurde auf die dort beschriebene Art ein Vorerfahrungsindex gebildet.

Sieben Items des Evaluationsfragebogens mussten für die weitere Analyse unberücksichtigt bleiben, da sie von nur sehr wenigen Teilnehmern ausgefüllt worden waren. Eine inhaltliche Betrachtung ergab, dass die in diesen Items enthaltenen Aussagen auf den untersuchten Online-Shop nicht anwendbar waren.

Die weitere Auswertung erfolgte analog zum IsoMetrics^S: Pro Subskala wurden die Datensätze von Teilnehmern unberücksichtigt gelassen, die weniger als 80 % der Items beantwortet haben. In den verbliebenen Daten wurden dann die fehlenden Ankreuzungen und Ankreuzungen von „Keine Angabe“ durch den mittleren Skalenwert 3 ersetzt. Dieses Verfahren hat keine bedeutsamen Auswirkungen auf Mittelwerte, Standardabweichungen und Reliabilitäten der Skalen (Gediga et al. 2000).

3 Ergebnisse

Die Teilnehmer beider Untersuchungen unterschieden sich nicht hinsichtlich ihres Alters ($t = 0.021$, $df = 37$, $p = .98$). Die Verteilung von männlichen und weiblichen Teilnehmern auf die beiden Untersuchungen zeigt keine signifikante Abweichung von der erwarteten Verteilung ($\chi^2 = 3.85$, $df = 1$, $p = .09$, Fisher's exact test). Schließlich zeigt sich auch kein signifikanter Unterschied in den Vorerfahrungs-Indizes der Teilnehmer (Mann-Whitney- $U = 84.00$, $N = 39$, $p = .17$).

3.4 Reliabilität

Die Reliabilitäten (Cronbach's alpha) der einzelnen Subskalen des Evaluationsfragebogens sind in Tabelle 1 aufgeführt.

	Reliabilität (Cronbach's alpha)
Aufgabenangemessenheit	.83
Akzeptanz	.77
Erwartungskonformität	.80
Fehlerrobustheit	.68
Informationsstruktur	.88
Medienqualität	.86
Selbstbeschreibungsfähigkeit	.89
Steuerbarkeit	.65
Erlernbarkeit	.84
Individualisierbarkeit	.65

Tabelle 1: Reliabilitäten der Subskalen des Evaluationsfragebogens

3.5 Validität

Von den 336 in der Videokonfrontation gewonnenen Problemnennungen konnten 302 den Subskalen des Fragebogens zugeordnet werden (s. Abbildung 1).

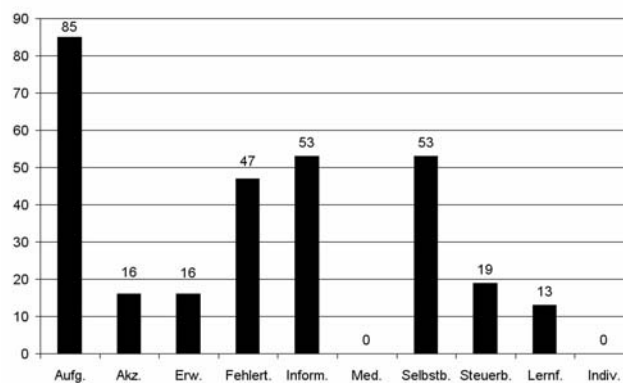


Abbildung 1: Verteilung der Problemnennungen auf die zehn Fragebogendimensionen

Die meisten Problemnennungen beziehen sich auf die Aufgabenangemessenheit. Weiterhin wurden viele Probleme zur Informationsstruktur, zur Selbstbeschreibungsfähigkeit und zur Fehlertoleranz angemerkt. Zur Medienqualität und zur Individualisierbarkeit wurden keine Anmerkungen gemacht.

Um die Ergebnisse aus der Fragebogenuntersuchung an denen der Videokonfrontation zu validieren, wurden die Skalenwerte des Fragebogens mit den Häufigkeiten der dazu gehörigen Anmerkungen korreliert. Dies ergibt einen nicht signifikanten Zusammenhang ($r = .277$, $p = .438$).

Zur Visualisierung wurden die Skalenwerte nach der Formel $x' = 6 - x$ umgepolt. Sowohl diese umgepolten Skalenwerte als auch die Häufigkeiten der Problemnennungen wurden standardisiert und zueinander in Beziehung gesetzt (Abbildung 2). Man erkennt eine weitgehend hohe Übereinstimmung. Einzige Ausnahme bildet die Subskala Individualisierbarkeit. Der Online-Shop wurde in dieser Hinsicht überdurchschnittlich schlecht bewertet, bei der Videokonfrontation wurden aber keine Problemnennungen zu diesem Aspekt gemacht. Lässt man diese Skala unberücksichtigt, so ergibt sich ein hoher Zusammenhang, der trotz der geringen Anzahl an Wertepaaren signifikant ist ($r = .756$, $p = .019$).

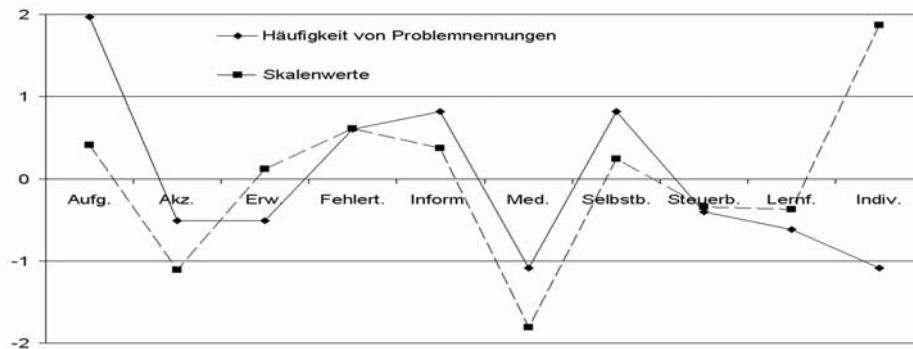


Abbildung 2: Vergleich von standardisierten Problemnennungshäufigkeiten und standardisierten, umgepolten Skalenwerten

4 Diskussion

Es wurde gezeigt, dass die Anzahl von Problemnennungen aus einer formativen Evaluationsuntersuchung als Kriterium zur Validierung eines summativen Evaluationsfragebogens dienen kann.

Die Korrelation zwischen den Skalenwerten und den Häufigkeiten der Problemnennungen stellt sich zunächst als nicht signifikant heraus. Eine nähere Betrachtung ergibt, dass die Subskala Individualisierbarkeit für die geringe Korrelation verantwortlich ist. Bleibt diese Dimension unberücksichtigt, so ergibt sich ein bedeutsamer und statistisch signifikanter Zusammenhang. Die Ursache für diese Diskrepanz ist das Fehlen von Problemnennungen zu diesem Aspekt im Usability-Test. Im Gegensatz zu anderen Aspekten scheinen Benutzer die

Individualisierbarkeit nicht von sich aus zur Bewertung der ergonomischen Qualität heranzuziehen.

Die vorgestellte Validierungsstrategie ist auch auf eindimensionale Fragebögen anwendbar. Hierzu muss der Fragebogen auf mehrere unterschiedliche Systeme angewendet werden. Eine formative Untersuchung muss dann für dieselben Systeme durchgeführt werden. Zu vergleichen wären in diesem Fall nicht Skalenwerte und Nennungshäufigkeiten verschiedener Dimensionen in Bezug auf ein System, sondern einer Dimension in Bezug auf verschiedene Systeme.

Auch muss zur Validierung der quantitativen Ergebnisse eines Fragebogens nicht auf die vergleichsweise aufwändige Methode der Videokonfrontation zurückgegriffen werden. Ebenso könnten andere formative Evaluationsuntersuchungen wie das Laute Denken oder der IsoMetrics^L (Willumeit et al. 1996) eingesetzt werden, um zu einer Menge von Problemnennungen zu kommen.

Die vorgestellte Untersuchung weist zwei methodische Einschränkungen auf. Zum einen erfolgte die Zuordnung der Problemnennungen zu den Fragebogendimensionen durch den Autor selbst. Hier wäre die Zuordnung durch mehrere Personen sinnvoll gewesen. Mit einem Übereinstimmungsmaß wie Cohen's Kappa hätte dann zudem die Güte der Zuordnungsergebnisse überprüft werden können. Optimal erscheint hier der Einsatz von mehreren unabhängigen Experten, die in Bezug auf die Untersuchungsziele und -annahmen blind sind.

Eine zweite Einschränkung betrifft den Einsatz derselben Stichprobe für die Bestimmung von Itemkennwerten, Reliabilität und Validität. Hier hätte zunächst eine Untersuchung zur Ermittlung von Itemkennwerten und Reliabilitäten durchgeführt werden sollen. Der aufgrund dieser Ergebnisse optimierte Fragebogen wäre dann im Anschluss unter Einsatz einer zweiten, unabhängigen Stichprobe validiert worden.

Ein letzter Punkt betrifft die mögliche Weiterentwicklung der vorgestellten Validierungsstrategie. Empirische Befunde deuten darauf hin, dass die reine Auftretenshäufigkeit alleine kein hinreichend genauer Indikator für den Schweregrad eines Usability-Problems darstellt (Hassenzahl et al. 1997). Vielmehr sollte zusätzlich in Betracht gezogen werden, wie viel zeitlicher und psychologischer Aufwand seitens des Benutzers zur Bewältigung des Problems eingesetzt werden muss. Ein häufig auftretendes, aber leicht zu bewältigendes Problem ist demnach möglicherweise genauso schwerwiegend einzuschätzen wie ein selten auftretendes, aber schwer zu bewältigendes Problem. In nachfolgenden Validierungsstudien der dargestellten Art sollte eine Schätzung dieses Bewältigungsaufwands erfolgen, um zu einem noch adäquateren Validierungskriterium zu gelangen.

Literaturverzeichnis

- Bortz, J.; Döring, N. (1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler. Berlin, Heidelberg, New York: Springer.
- Chin, J. P.; Diehl, V. A.; Norman, K. L. (1988): Development of an instrument measuring user satisfaction of the human-computer interface. In: Soloway, E.; Frye, D.; Sheppard, S. B. (Hrsg.): Human Factors in Computing Systems – CHI '88. New York: ACM, S. 213–218.

- Gediga, G.; Hamborg, K.-C. (1999): IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241-10. In: Holling, H.; Gediga, G. (Hrsg.): Evaluationsforschung. Göttingen: Hogrefe, S. 195–234.
- Gediga, G.; Hamborg, K.-C.; Dünsch, I. (1999): The IsoMetrics usability inventory: An operationalisation of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour and Information Technology*, Vol. 18, Nr. 3, S. 151–164.
- Hamborg, K.-C.; Greif, S. (1999): Heterarchische Aufgabenanalyse. In: Dunkel, H. (Hrsg.): *Handbuch psychologischer Arbeitsanalyseverfahren*. Zürich: vdf, S. 147–177.
- Hassenzahl, M.; Prümper, J.; Sailer, U. (1997): Die Priorisierung von Problemhinweisen in der software-ergonomischen Qualitätssicherung. In: Wüschmann, W. (Hrsg.): *Software-Ergonomie '97. Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung*. Stuttgart: Teubner, S. 191–201.
- Kirakowski, J.; Corbett, M. (1993): SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, Vol. 24, Nr. 3, S. 210–212.
- Mayhew, D. (1999): *The Usability Engineering Lifecycle. A Practitioner's Handbook for User Interface Design*. San Francisco: Morgan Kaufmann Publishers.
- Nielsen, J. (1994): Heuristic evaluation. In: Nielsen, J.; Mack, R. L. (Hrsg.): *Usability Inspections Methods*. New York: Wiley & Sons, S. 25–62.
- Prümper, J. (1997): Der Benutzerfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In: Wüschmann, W. (Hrsg.): *Software-Ergonomie '97. Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung*. Stuttgart: Teubner, S. 253–262.
- Prümper, J.; Anft, M. (1993): Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung: Ein Fallbeispiel. In: Rödiger, K.-H. (Hrsg.): *Software-Ergonomie '93 - Von der Benutzungsoberfläche zur Arbeitsgestaltung*. Stuttgart: Teubner, S. 145–156.
- Rubin, J. (1994): *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: John Wiley & Sons.
- Willumeit, H.; Gediga, G.; Hamborg, K.-C. (1996): IsoMetricsL: Ein Verfahren zur formativen Evaluation von Software nach ISO 9241/10. *Ergonomie & Informatik*, Vol. 27, S. 5–12.

Kontaktinformationen

Frank Ollermann
Universität Osnabrück
FB 08, Fachgebiet Arbeits- und Organisationspsychologie
Universität Osnabrück
Seminarstr. 20
49069 Osnabrück
E-Mail: Frank.Ollermann@Uni-Osnabrueck.de