

Schriftenreihe der
Bundesanstalt für
Arbeitsschutz und
Arbeitsmedizin

Forschung F 1693

W. Dzida
B. Hofmann
R. Freitag
W. Redtenbacher
R. Baggen
T. Geis
J. Beimel
C. Zurheiden
W. Hampe-Neteler
R. Hartwig
H. Peters

Gebrauchstauglichkeit von Software

**ErgoNorm:
Ein Verfahren zur Konformitätsprüfung
von Software auf der Grundlage von
DIN EN ISO 9241 Teile 10 und 11**

Arbeitsschutz

- Forschung -
F 1693

W. Dzida
B. Hofmann
R. Freitag
W. Redtenbacher
R. Baggen
T. Geis
J. Beimel
C. Zurheiden
W. Hampe-Neteler
R. Hartwig
H. Peters

Gebrauchstauglichkeit von Software

**ErgoNorm:
Ein Verfahren zur Konformitätsprüfung
von Software auf der Grundlage von
DIN EN ISO 9241 Teile 10 und 11**

Dortmund 2000

Bearbeiter:

Dr. Wolfgang Dzida
 Dipl.-Psych. Britta Hofmann
 Dipl.-Inform. Regine Freitag
 GMD, Institut für Autonome intelligente Systeme
 53754 Sankt Augustin

Dipl.-Math. Wolfgang Redtenbacher
 Redtenbacher Software
 71272 Renningen

Dipl.-Psych. Robert Baggen
 Dr. Dipl.-Inform. Wolfgang Hampe-Neteler
 Dipl.-Psych. Christoph Zurheiden
 TÜV Informationstechnik Essen GmbH
 Prüfstelle für IT-Ergonomie
 45307 Essen

Dipl.-Ing. Thomas Geis
 Dipl.-Psych. Jürgen Beimel
 Dipl.-Inform. Ronald Hartwig
 TÜV Secure iT GmbH / TÜV Rheinland Product Safety GmbH
 Prüflabor Software Usability
 51101 Köln

Dr. Helmut Peters
 TÜV Product Service GmbH
 Ergonomics
 80339 München

Kurzreferat

Konformität und Non-Konformität von Software mit ergonomischen Normen (DIN EN ISO 9241 Teile 10 und 11)

Zur Prüfung der Gebrauchstauglichkeit von Softwareprodukten auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11 wird hiermit das ErgoNorm-Prüfverfahren vorgelegt. Ein Teil des Verfahrens dient Benutzern zur subjektiven Bewertung von Effektivität und Effizienz der Softwarenutzung, um Hinweise auf Nutzungsprobleme und vermutete Normabweichungen zu geben. Der zweite Teil des Verfahrens ist für die Hand des Experten bestimmt. Softwareprüfer werden angeleitet, den Prüfaufwand auf die Anforderungen zu konzentrieren, die sich objektiv und valide ableiten lassen. Eine Aussage über Normkonformität ist somit stets auf die Liste der untersuchten Anforderungen beschränkt. Diese Prüfgrundlage ermöglicht die Reproduzierbarkeit der Prüfergebnisse. Eine vollständige Produktprüfung, die durch den Verifikationsansatz der ISO 9241 nahegelegt wird, ist kaum kostengünstig umsetzbar. Statt dessen wird einem Softwareprodukt Normkonformität bis zum Nachweis des Gegenteils unterstellt (Falsifikationsansatz). Dieser Nachweis ist stets auf objektive Erfordernisse des Nutzungskontexts und der Arbeitstätigkeit bezogen. Nur unter Anwendung dieses Ansatzes darf für ein Produkt in Anspruch genommen werden, "gebrauchstauglich" zu sein. Beim Verdacht auf eine Normverletzung besteht dieser Anspruch so lange, bis dieser Verdacht durch einen gesonderten Prüfschritt erhärtet wurde. Hierbei wird untersucht, welche Wirkung eine vermutete Normabweichung auf die Arbeitsausführung und den Benutzer hat. Der Stellenwert des Verfahrens liegt nicht allein in der Feststellung der Mängel und seiner Wirkungen, sondern in seinem konstruktiven Beitrag zur Qualitätsverbesserung. Jede ergonomische Verbesserung der Softwarenutzung hilft Nutzungskosten sparen. Heute sind es nicht mehr die Anschaffungskosten, sondern die meist unterschätzten Nutzungskosten, die den Anwendern "Kopfschmerzen machen". Die Anwendung des Prüfverfahrens ist somit ein Beitrag zur Kostensenkung bei gleichzeitiger Verbesserung der Arbeitsqualität.

Schlagwörter: Software-Ergonomie, Software, Normen, Normkonformität, Qualität, Prüfung, Gebrauchstauglichkeit, Usability, Bildschirmarbeitsplatz

Summary

Conformity and non-conformity of software with usability standards (EN ISO 9241 Parts 10 and 11)

This booklet introduces ErgoNorm, a two-part method for testing the usability of software for conformity with ISO 9241 Parts 10 and 11. One part of the method is intended for the users to enable them to subjectively assess the effectiveness and efficiency of software use, thereby encouraging them to point out problems of use and supposed deviations from the standard. The second part of the method is intended for experts. The professional tester of software is guided to restrict the test effort to those requirements that can be derived as being objective and valid. A statement on conformity is thus confined to the list of such requirements. On this basis the results achieved with a test are reproducible. A complete product test, as suggested by the verification approach of ISO 9241, can hardly be put into practice cost-effectively. Therefore in our approach a product's conformity with standards is being presumed until the reverse has been verified (so-called falsification approach). Such a proof is always linked to objective requirements derived from the context of use and user performance. Adhering to this test approach a product can be claimed as usable. If a deviation from the standard is suspected it has to be proven by means of a separate procedure. In this case the impact of a suspected deviation on user performance is looked into. The virtue of this approach is not only seen in the stated defects and related impacts but also in its constructive contribution to quality improvement. Every ergonomic improvement helps save usage costs. Nowadays, it is not the purchase costs that "gives the user a headache" but the mostly underestimated usage costs. The application of the test method can therefore contribute to reduce these costs and simultaneously improve the quality of work.

Key words: software ergonomics, software, standards, conformity with standards, quality, test, usability, office work with visual display terminals

Résumé

Conformité et non-conformité de logiciel avec des normes ergonomiques (EN ISO 9241 Parties 10 et 11)

Il s'agit ici d'une méthode en deux parties pour tester de l'aptitude à l'usage du logiciel pour la conformité avec les parties 10 et 11 d'EN DIN 9241. La première partie de cette méthode s'est dédiée aux utilisateurs pour évaluer subjectivement l'efficacité et l'efficience de l'utilisation de logiciel, afin de pouvoir signaler des problèmes de l'utilisation et des déviations supposées de la norme. La deuxième partie de la méthode est destinée à l'experte. L'appareil de contrôle professionnel du logiciel est guidé pour limiter l'effort d'essai à ces conditions qui peuvent être dérivées en tant qu'étant objectives et valides. Un rapport sur la conformité est ainsi confiné à la liste de telles conditions. Sur cette base les résultats réalisés avec un essai sont reproductibles. Un essai complet de produit, comme suggéré par l'approche de vérification d'EN DIN 9241, peut à peine être mis en pratique de manière rentable. Au lieu de cela, un produit est supposé norm-conforme jusqu'à ce que le contraire soit prouvé (nommé ici: approche de falsification). Cette preuve se réfère toujours aux exigences objectives du contexte d'utilisation et du travail effectué. Adhérer à cette approche d'essai un produit peut être réclamé comme utilisable. Si on suspecte une déviation de la norme, cette réclamation peut être mise à jour jusqu'à ce que l'opposé ait été prouvé au moyen d'un procédé séparé. Dans ce cas-ci l'impact d'une déviation suspectée sur l'exécution d'utilisateur est énoncé. Le mérite de la méthode sera non seulement dans la constatation des défauts et de leurs conséquences, mais également dans sa contribution constructive à une véritable amélioration de la qualité. Les aides ergonomiques de chaque amélioration sauvegardent des coûts d'utilisation. De nos jours, ce ne sont plus les coûts d'achat que „font des maux de tête“ aux utilisateurs, mais les frais d'utilisation, le plus souvent sous-estimés. L'application de la méthode d'essai peut donc contribuer pour réduire ces coûts et pour améliorer simultanément la qualité du travail.

Mots clés: ergonomie de logiciel, logiciel, normes, conformité avec des normes, qualité, travail de bureau avec des terminaux à écrans de visualisation

Inhaltsverzeichnis

1 Einleitung.....	1
2 Prüf- und Beurteilungsgrundlagen.....	8
2.1 Übereinstimmung der Beurteilungsgrundlagen	9
2.2 Verifikation oder Falsifikation	12
2.3 Unzureichende Grundlagen der bisherigen Prüfpraxis	13
2.4 Erfordernisse, Anforderungen und Prüfkriterien	15
2.5 Durch Benutzerbefragung ermittelte Beurteilungsgrundlagen.....	19
2.6 Normen als Prüfgrundlage.....	22
3 Prüfverfahren für Experten.....	26
3.1 Vorbereitung der Normkonformitätsprüfung	27
3.1.1 Vorbereitung mittels Benutzerfragebogen.....	28
3.1.2 Prüfungsvorbereitung.....	29
3.2 Durchführung der Normkonformitätsprüfung	33
3.3 Bewertung von Abweichungen: Erhärtungsprüfung.....	37
3.4 Gütekriterien der Konformitätsprüfung	45
4 Analyse des Nutzungskontexts und der Nutzung.....	47
4.1 Erhebung von Kontext-Szenarien	47
4.1.1 Instruktion zur Erhebung von Kontext-Szenarien.....	52
4.1.2 Trainingsanleitung zur Erhebung von Kontext-Szenarien.....	53
4.1.3 Auswertung	56
4.1.4 Validität von Kontext-Szenarien	60
4.2 Erhebung von Use-Szenarien	60
4.2.1 Use-Case angereichert mit Use-Szenarien.....	61
4.2.2 Erhebung und Prüftätigkeit in einem Arbeitsgang.....	63
4.2.3 Validität von Use-Szenarien.....	63
4.2.4 Instruktion zur Erhebung von Use-Szenarien.....	64
4.3 Erhebungs- und Auswertungsobjektivität	65
4.3.1 Versuchsplan für Objektivitätsuntersuchungen.....	68
4.3.2 Durchführung der Untersuchung: Erhebungsobjektivität.....	70
4.3.3 Erhebungsobjektivität: Objektivitätsmessungen und Ergebnisse.....	73

4.3.4 Auswertungsobjektivität.....	75
4.4 Handhabbarkeitsprüfung	78
4.4.1 Integration in das DEKITZ-Prüfverfahren.....	78
4.4.2 Erprobung in einem Großprojekt	79
5 Bewertung durch Benutzer	81
5.1 Anforderungen an den Benutzerfragebogen	83
5.2 Anforderungen an die Handhabbarkeit.....	85
5.3 Anforderungen an die Ergebnisqualität	86
5.4 Entwicklungskonzept	90
5.5 Itemgenerierung und Itemselektion	91
5.6 Vergleich der Softwareprodukte anhand von Skalenwerten.....	102
5.7 Abschließende Itemselektion.....	104
5.8 Beanspruchungsmessung	106
5.9 Verwertung der Fragebogenergebnisse	107
5.10 Auswertungsleitfaden	110
6 Existierende Prüf- und Bewertungsverfahren	117
6.1 Literatúrauswertung	117
6.2 Beschreibung der Bewertungsverfahren	119
6.3 Zusammenfassende Beurteilung.....	146
7 Literatur	147
Anhang A: Methoden zum Prüfverfahren	153
Anhang B: Erhebung und Auswertung von Kontext-Szenarien.....	168
Anhang C: Erhebung und Auswertung von Use-Szenarien	191
Anhang D: ErgoNorm Benutzerfragebogen zu „Arbeit & Software“	197
Anhang E: Glossar	216

1 Einleitung

Die Nutzungsqualität von Software durch Anforderungsentwicklung zu sichern und auf Konformität mit ergonomischen Normen zu prüfen, ist Thema dieses Berichts. Konformität mit ergonomischen Normen (DIN EN ISO 9241) wird zur Zeit international zu wenig beachtet. Kaum ein Softwareprodukt würde deshalb den Test gegen die Normen bestehen. Wo dies versucht wurde, hat sich diese Vermutung bestätigt. Das akkreditierte Prüflaboratorium für Usability der TÜV-Informationstechnik in Essen hat z.B. im Auftrag der Zeitschrift ComputerBILD Windows 98 (CB 19/1998) und Office 2000 (CB17/1999) von Microsoft für einen bestimmten, typischen Nutzungskontext¹ geprüft. Beide Produkte haben sich als nicht normkonform herausgestellt. Ausländische Hersteller verteidigen sich manchmal mit dem Argument, ihre Produkte seien nicht nur für den deutschen Markt entwickelt worden, deshalb seien die deutschen Normen nicht allein Maßstab für Qualität. Dabei handelt es sich jedoch nicht um deutsche Normen, sondern um ISO-Normen, die ohne Änderungen in Europa übernommen wurden (erkennbar an dem Kürzel EN für Europäische Norm) und somit unverändert in Deutschland gelten (erkennbar an dem Kürzel DIN). Aber auch in Deutschland kann man in manchen Herstellerkreisen die abwehrende oder abwertende Bemerkung hören, dass Normen eine "typisch deutsche Marotte" seien. Die Autoren dieses Berichts hätten das Thema Normen und Normenkonformität gern zu den Akten gelegt, wenn es sich in der Weise von selber erledigen würde, dass Softwareprodukte in einer ergonomischen Qualität ausgeliefert würden, die das von den Normen geforderte Mindestniveau übertrifft.

Der für Anwender und Benutzer von Softwareprodukten am meisten interessierende Aspekt ergonomischer Nutzungsqualität ist die Gebrauchstauglichkeit (DIN EN ISO 9241-11, 1998). Hiermit ist die Tauglichkeit eines Softwareprodukts für den Gebrauch am Arbeitsplatz gemeint. Wie aber können Anwender und Benutzer feststellen (lassen), ob ein Produkt, das ein Hersteller für die Nutzung an bestimmten Arbeitsplätzen anbietet, auch tauglich dafür ist? Ein gründlicher und umfassender Test gegen

¹ Siehe Glossar

sämtliche software-ergonomischen Normen (DIN EN ISO 9241 Teile 10 bis 17) wäre möglich, wäre aber aufwendig und teuer. Mit diesem Aufwand und Anspruch sollte der Hersteller die Normkonformitätstests in der eigenen Werkstatt durchführen. Den Anwendern oder Benutzern ist dies nicht zuzumuten. Die Prüfpraxis von akkreditierten Prüflaboratorien hat gezeigt, dass bereits ein Konformitätstest gegen Teil 10 der Norm ausreicht, um die Erfüllung ergonomischer Mindestqualität feststellen zu können. Weil DIN EN ISO 9241-10 (1995) keine Konformitätsklausel enthält, ist die Anwendung einer Prüfmethodik grundsätzlich Vereinbarungssache zwischen Partnern (Hersteller und Anwender). Ein Vorteil solcher Vereinbarungen kann darin liegen, dass man sich auf eine kostengünstige Methodik einigt. Viele Methoden nehmen in Anspruch, Normkonformität mit ISO 9241-10 zu prüfen und kostengünstig zu sein. Aber über kaum etwas lässt sich trefflicher streiten, als über die passende Methode.

Mit dem hier vorgelegten Verfahren sollte der Methodenstreit bis auf weiteres ein Ende haben; denn parallel zu dem Projekt, in dem das Verfahren entwickelt wurde, gab es über die Anwendung des Verfahrens einen Konsensprozess in der DEKITZ (Deutsche Akkreditierungsstelle für Informations- und Kommunikationstechnik). Außerdem wurde das Verfahren während der Projektlaufzeit in vielen großen und kleinen Prüfprojekten erprobt, wobei die praktischen Erfahrungen immer wieder zur Verbesserung der Handhabung geführt haben. So wie die Normen unter anderem auf dem Konsens der Fachleute beruhen, so hat sich das Prüfverfahren als konsensfähig und ökonomisch durchführbar erwiesen. Auf diese Weise soll gewährleistet sein, dass Prüflaboratorien in methodisch einheitlicher Weise Software prüfen. Feststellungen über Konformität oder Non-Konformität sollen künftig nicht mehr strittig sein, nur weil Prüfer sich auf ihre jeweils eigene methodische Vorgehensweise berufen.

Darüber hinaus wird angestrebt, die korrekte Anwendung des Prüfverfahrens bei der Begutachtung aller Prüflaboratorien zu überwachen, die von der DEKITZ akkreditiert werden. Eine ständige Arbeitsgruppe der DEKITZ wertet auch weiterhin die Erfahrungen der Prüflaboratorien aus, und beschließt ggf. nötige methodische Verbesserungen. Mit Verbesserungen ist zu rechnen, wenn Teile des Prüfverfahrens künftig nicht nur für die Konformitätsprüfung fertiger Softwareprodukte eingesetzt

werden, sondern auch in Entwicklungsprojekten, in denen beispielsweise Anwender und Hersteller die gewünschten Anforderungen an die Gebrauchstauglichkeit einvernehmlich entwickeln und abschließend auf Normkonformität prüfen lassen. Viele Prüfvoraussetzungen, die mit dem vorliegenden Verfahren geschaffen werden, sind überaus nützlich für die Vorbereitung eines Entwicklungsprojekts und beim Usability-Prototyping.

Zweiteiliges Verfahren	Methoden
ErgoNorm - Benutzerfragebogen Untersuchung von Nutzungsproblemen	Benutzerbefragung
ErgoNorm - Prüfverfahren Anforderungsentwicklung und Normkonformitätstest sowie Erhärtungsprüfung für vermutete Normabweichung	Aufgabenanalyse mittels Szenarien Teilnehmende Beobachtung Inspektion Benutzerbefragung Dokumentenanalyse

Tabelle 1-1: ErgoNorm: Aufbau des zweiteiligen Verfahrens und Methodenübersicht

Gebrauchstauglichkeit kennzeichnet die Nutzungsqualität von Software am Benutzerarbeitsplatz. Funktionale und software-technische Merkmale sind in diese Bewertung eingeschlossen, soweit sie Auswirkungen auf die Softwarenutzung haben. Es wird jedoch nicht die technische Güte dieser Merkmale bewertet, sondern ihr Beitrag zur Erzielung der gewünschten Nutzungsqualität. Anhand der festgestellten Wirkungen der Produktmerkmale auf die Tätigkeit am Arbeitsplatz und auf die Benutzer wird beurteilt, ob Gebrauchstauglichkeit erreicht ist. Das Produkt wird vor allem danach beurteilt, ob und inwieweit es am Arbeitsplatz des Benutzers anforderungsgemäß und zuverlässig funktioniert.

Softwareprüfer werden durch das Prüfverfahren angeleitet, die passenden Methoden anzuwenden, um die Prüfung einer Software auf Gebrauchstauglichkeit vorzubereiten und durchzuführen. Anhand von Gütekriterien können Prüfer eine methodisch seriöse

Vorgehensweise absichern. Durch geeignete Dokumentation sorgt der Prüfer dafür, dass Dritte die Einhaltung der geforderten Gütekriterien nachvollziehen und die erzielten Prüfergebnisse reproduzieren können.

Das Prüfverfahren setzt in der Regel keine technisch aufwendige Laboreinrichtung voraus. Anders als bei einer rein software-technischen Produktprüfung wird die Prüfung aus der Sicht der tatsächlichen Nutzung im Nutzungskontext angegangen, so dass selbst Funktions- und Zuverlässigkeitsprüfungen keine aufwendige Testumgebung erfordern. Der Aufwand der Prüfung wird auch durch den im Prüfverfahren gewählten Falsifikationsansatz stark reduziert. Im Unterschied zu einer vollständigen Prüfung nach den Normen DIN EN ISO 9241, Teile 10-17, die eine vollständige Verifikation für den Nachweis der Konformität eines Produkts mit den anwendbaren Forderungen einer Norm nahe legen, sieht das Prüfverfahren vor, dass lediglich die aus dem Teil 10 (Grundsätze der Dialoggestaltung) der Norm abgeleiteten Prüfkriterien anzuwenden sind. Für diese Prüfkriterien wird jeweils unterstellt, dass sie durch die Qualität der Merkmale des Dialogsystems oder durch die am Dialogsystem tatsächlich ausführbaren Tätigkeiten erfüllt sind. Prüfen heißt, zu versuchen, diese Konformitätsannahme für jedes Kriterium zu falsifizieren. Nach dem Vorbild der wissenschaftlichen Hypothesenprüfung wird die Konformitätsannahme so lange aufrecht erhalten, bis sie durch den Nachweis des Gegenteils aufgegeben werden muss. Dann liegt eine vermutete Normabweichung vor, deren Auswirkung durch eine "Erhärtpungsprüfung" gesondert beurteilt wird. Um bedeutsame Normabweichungen konstruktiv zu beurteilen, werden die übrigen Normen von ISO 9241 (Teile 12 bis 17) herangezogen. Aus diesen Normen können dem Designer Vorschläge für die normkonforme Verbesserung des Produkts gemacht werden. Für den Prüfer ist es daher wichtig, die Rolle der einzelnen Teile der Norm zweckdienlich einzuschätzen, d.h., Teil 10 (Dialogprinzipien) als Maßstab für den Nachweis von Non-Konformität zu verwenden und ggf. die übrigen Normen zu konsultieren, damit eine festgestellte Non-Konformität dem Anwender oder dem Software-Designer (Usability-Engineer) gegenüber konstruktiv begründet werden kann.

In der Praxis der Prüflaboratorien hat sich der in den Normen nahe gelegte Verifikationsansatz als zu aufwendig herausgestellt. Softwareprüfer standen stets vor

der nicht entscheidbaren Frage, wo eine Prüfung anfangen und aufhören soll. Außerdem war die Aufmerksamkeit der Prüfer zu stark auf die gegebenen Produktmerkmale gelenkt und weniger auf die erforderlichen Arbeitstätigkeiten. Im Unterschied zur aufwendigen Verifikation aller geforderten Merkmale und Arbeitstätigkeiten wird mit dem vorgelegten Verfahren der Prüfaufwand auf eine Liste abgeleiteter und validierter Prüfkriterien beschränkt. Der Schwerpunkt wird aus der Merkmalsprüfung in die Anforderungsdefinition vorverlegt. Mit einer validierten Anforderungsdefinition in der Hand reduziert sich der Prüfaufwand auf das Wesentliche, das heißt auf jene Merkmale des Produkts, die sich durch Forderungen des Benutzers (Anwenders) begründen lassen.

Die Methoden des Prüfverfahrens können zu verschiedenen Anlässen angewendet werden.

- Am häufigsten ist die Anwendung im Nutzungskontext eines Softwareprodukts. In Kenntnis der aus dem Kontext abgeleiteten Anforderungen sowie der Norm DIN EN ISO 9241-10 kann die Konformität des Produkts mit den Anforderungen geprüft werden. Zunächst steht die Prüfung gegen die kontext-bezogenen Anforderungen an. Darüber hinaus ist gegen Anforderungen zu prüfen, die kontext-neutral sind.
- Eine eingeschränkte Prüfung ist in der Werkstatt des Herstellers möglich (sog. Benutzbarkeits-Prüfung im Sinne von ISO/IEC 12119, 1994), beispielsweise vor der Auslieferung des Produkts. Dies ist eine eingeschränkte Prüfung, weil der reale Nutzungskontext im Sinne von DIN EN ISO 9241-11 in dieser Prüfsituation nicht gegeben ist.
- Am wichtigsten wird künftig der Einsatz des Prüfverfahrens während des Software-Entwicklungsprozesses werden. Dies ist ein geeigneter Weg, Möglichkeiten der Produktverbesserung rechtzeitig zu erkennen, wenn die in den Normen geforderte Mindestqualität (noch) nicht erreicht wurde. Die Vorbereitungen für die Prüfung setzen keinen Aufwand voraus, der nicht ohnehin für eine angemessene Software-Entwicklung und projektbegleitende Qualitätssicherung erforderlich ist.
- Man kann den Prüfumfang auch auf die am Arbeitsplatz des Benutzers entdeckten Nutzungsprobleme beschränken. Aus der Kenntnis des Nutzungskontexts und

seiner Mängel können sich dabei für den Prüfer wertvolle Hinweise auf mögliche Ursachen für mangelnde Zufriedenstellung der Benutzer ergeben.

Soweit hier von Prüfverfahren gesprochen wird, ist immer eine Zusammenstellung von Methoden gemeint, um objektive und valide Prüfkriterien definieren und anwenden zu können (vgl. Tab. 1). Diese Methoden sind für die Hand des Experten bestimmt. Softwareprüfer sind ohne Zusatzausbildung nicht genügend vorbereitet, diese Methoden anzuwenden. In der Regel ist daher eine Schulung erforderlich. Die mit Hilfe der Methoden erzielten Ergebnisse dürfen nicht absolut gesetzt werden. Auch Experten können irren. In den meisten Prüfsituationen sind deshalb die Benutzer des zu prüfenden Produkts beteiligt. Vor jeder Konformitätsaussage über die Gebrauchstauglichkeit eines Produkts sind stets die Benutzer auf ihre Zufriedenstellung hin zu befragen (DIN EN ISO 9241-11, 1998).

Ein besonderer Teil des hier vorgelegten Verfahrens dient der subjektiven Bewertung der Softwarequalität (siehe Fragebogen im Anhang). Der Fragebogen allein ist jedoch kein ausreichendes Instrument der Normkonformitätsprüfung. Er ist vielmehr gedacht als Einstieg in eine Prüfsituation. Als Initialverfahren kann er helfen, den Blick des Prüfers auf Nutzungsprobleme zu lenken, um z.B. den Prüfumfang besser einschätzen zu können. Der besondere Wert des Fragebogens besteht darin, dass Benutzer reflektieren lernen, welche Qualitätseigenschaften der Dialog mit der Software mindestens haben muss, um anschließend besser in der Lage zu sein, Nutzungsprobleme aufzuschreiben und zu bewerten. Der Fragebogen ist für die Hand des Betreibers (Anwenders) einer Software bestimmt, soweit er in der Rolle als Arbeitgeber seine Pflichten im Rahmen des vorbeugenden Arbeitsschutzes wahrnehmen will (vgl. Arbeitsschutzgesetz, 1996 und Bildschirmarbeits-Verordnung, 1996). Aus den im Gesetz und in der Verordnung genannten Anlässen kann der Arbeitgeber den Fragebogen an Bildschirmarbeitsplätzen verteilen und die Benutzer bitten, diesen in Ruhe zu lesen und später mit Blick auf eine konkrete Arbeitsaufgabe auszufüllen. Ausgewertete Fragebögen können sehr nützliche Hinweise für den Betreiber der Software enthalten und dazu veranlassen, einen ausgewiesenen Softwareprüfer zu rufen, um vermutete Normabweichungen auf ihre Wirkungen hin zu untersuchen. Es wird empfohlen, dass Anwender eines Softwareprodukts mit dem Hersteller verein-

baren, den Fragebogen als Instrument der Qualitätssicherung im Nutzungskontext des Produkts einzusetzen und u.U. ein Prüflabor zu beauftragen, um aufgetretene Nutzungsprobleme zu bewerten und in Kooperation mit dem Hersteller beseitigen zu helfen. Einem ISO 9000-zertifizierten Hersteller wird empfohlen, den Fragebogen als Instrument der Feldvalidierung seines Produkts zu verwenden.

2 Prüf- und Beurteilungsgrundlagen

Kern des zu entwickelnden Prüfverfahrens ist ein Beurteilungsmodell. Es soll den Prüfer anleiten, die Prüf- und Beurteilungsgrundlagen zu objektivieren, um jeden Konformitätstest nachvollziehbar und reproduzierbar dokumentieren zu können. Hierbei soll gesichert werden, dass die Beurteilungsgrundlagen für Prüfer und Benutzer gleich sind (siehe Bild 2.1-1). Beurteilungsgrundlage ist u.a. der Nutzungskontext (DIN EN ISO 9241-11, 1998), in dem das zu beurteilende System eingesetzt wird oder in dem das zu entwickelnde System eingesetzt werden wird. Im wesentlichen umfasst der Nutzungskontext die Arbeitstätigkeiten der Benutzer und die Bedingungen zu ihrer Durchführung. Im Kontext-Szenario werden vom Benutzer die bedeutsamen Tätigkeiten beschrieben und die damit verbundenen Probleme, die möglicherweise mittels Software behoben werden können oder die durch existierende Software bedingt sind. Ferner sind aus dem Nutzungskontext Anforderungen an den Dialog mit der Software abzuleiten, oder es werden mit Blick auf den Nutzungskontext Abweichungen von einer dem Kontext angemessenen Gestaltung der Software festgestellt. Die im Beurteilungsmodell beschriebenen Beurteilungsgrundlagen für die Qualität von Software umfassen also den Nutzungskontext, die daraus abgeleiteten Anforderungen sowie die ggf. aufgedeckten Nutzungsprobleme. Für die Beurteilungsgrundlagen ist Übereinstimmung gegeben, wenn Benutzer und Prüfer hierin inhaltlich übereinstimmen. Für übereinstimmende Beurteilungsgrundlagen könnte man auch sagen, dass sie valide sind.

Benutzern ist es im allgemeinen nicht zuzumuten, ihre Beurteilungen anhand von Forderungen abzugeben, die in den Normen stehen, da Benutzer mit den Fachbegriffen nicht vertraut sind und sich schwer tun, Normen zu interpretieren. Experten (Prüfer und Usability-Designer) dagegen müssen auf die Normen zurückgreifen. Normen sind somit als einheitliche Beurteilungsgrundlage für Benutzer und Experten nicht geeignet. Beide Gruppen können sich aber über die gegebenen oder künftigen Nutzungsbedingungen von Software verständigen sowie über die darin enthaltenen Aufgabenerfordernisse und daraus abgeleitete Tätigkeiten oder Merkmale des Dialogsystems. Während die Experten diese Forderungen im Lichte der Normen interpretieren müssen und dann mit

den Gegebenheiten am Produkt vergleichen, liegt es nahe, die Benutzer nur über den Kontext der Arbeitstätigkeiten sowie ihre subjektive Einschätzung der mit dem Produkt ausführbaren Tätigkeiten zu befragen.

In diesem Kapitel wird zunächst ein Modell vorgestellt, das beschreibt, welche Beurteilungsgrundlagen gegeben sein müssen, bevor Normkonformität überhaupt geprüft werden kann (Abschnitt 2.1). Da die Einhaltung von Normen sowohl während des Entwicklungsprozesses als auch nach Installation von Produkten an Arbeitsplätzen geprüft wird, werden zwei unterschiedlich aufwendige Konformitätstests gegenübergestellt. Es wird begründet, warum für die Prüfung der Gebrauchstauglichkeit an den Benutzerarbeitsplätzen der Falsifikationsansatz ausreicht (Abschnitt 2.2). Verglichen mit dem hier entwickelten Beurteilungsmodell erscheint die bisherige Prüfpraxis als unzureichend (Abschnitt 2.3). Deshalb werden Konzepte der Qualitätssicherung eingeführt, die dafür sorgen helfen, dass vor der Prüfung objektive Beurteilungsgrundlagen zustande kommen und eine zuverlässige Anwendung der Norm DIN EN ISO 9241-10 (1995) möglich wird (Abschnitt 2.4). Um die Benutzer in die Prüfungsvorbereitung einzubeziehen, wird eine Beurteilungsgrundlage geschaffen, die mit der des Prüfers korrespondiert (Abschnitt 2.5). Abschließend wird dargestellt, wie die genormte Prüfgrundlage systematisch aufgebaut ist und warum es nicht nur aus Kostengründen ausreicht, sich bei der Produktprüfung am Benutzerarbeitsplatz auf die Anwendung des Teils 10 von DIN EN ISO 9241 zu beschränken (Abschnitt 2.6).

2.1 Übereinstimmung der Beurteilungsgrundlagen

In einem Modell (Abbildung 2.1-1) wird die geforderte Übereinstimmung der Beurteilungsgrundlagen für Experten und Benutzer beschrieben. Mit Übereinstimmung ist nicht gemeint, dass die Qualitätsbewertungen der Experten mit den durch die Benutzer subjektiv bewerteten Tätigkeiten übereinstimmen; es geht vielmehr um die Übereinstimmung im Sachverhalt, der beiden Beurteilungen zugrunde liegt. Benutzer sollen sich unabhängig von Experten auf denselben Sachverhalt beziehen. Experten dürfen zwar bei der Beurteilung zu anderen Schlüssen gelangen als Benutzer; wenn jedoch Benutzer in einer Nutzungssituation ein Problem sehen, so muss das Prüf-

verfahren gewährleisten, dass auch Experten dieselbe Situation als Nutzungsproblem bewerten. Um den zu beurteilenden Sachverhalt zu ermitteln, ist es zweckmäßig, dem Benutzer standardisierte Fragen zu stellen (siehe Leitfragen im Abschnitt 4.1). Der Benutzer kann den erfragten Sachverhalt wegen seiner Sachkenntnis zutreffend (valide) beschreiben, etwa die bedeutsamen Arbeitstätigkeiten und die dabei festgestellten Probleme. Auf der Grundlage der kurzen Darstellung dieses Sachverhalts durch den Benutzer soll der Experte in der Lage sein, zusammen mit dem Benutzer ein ausführlicheres Kontext-Szenario seiner Arbeitstätigkeiten zu erarbeiten. Um das Kontext-Szenario gut strukturiert und objektiv zu beschreiben, wird dem Experten ein Leitfaden mit einem ausführlichen Beispiel sowie einem Erhebungs- und Auswertungsrahmen an die Hand gegeben (siehe Anhang B). Die erhobenen Sachverhalte des Kontexts werden von der befragten Arbeitsperson auf zutreffende Beschreibung geprüft. Es entsteht somit eine validierte Grundlage zur Ableitung von Anforderungen an die Software, bevor die Software geprüft wird.

Kontext-Szenarien sind als Umsetzung der Norm DIN EN ISO 9241-11 (1998) gedacht. Die Norm fordert, die Merkmale des Nutzungskontexts zu beschreiben; dies geschieht mittels Kontext-Szenarien in episodischer und erzählender Form (siehe Anhang B).

Ein methodischer Fehler der meisten in der Literatur beschriebenen und heute in der Praxis verwendeten Evaluierungsverfahren beruht darin, dass die Beurteilungsgrundlagen für Experten- und Benutzerurteile nicht eindeutig festgestellt werden. Mag der Experte sich über seine eigene Beurteilungsgrundlage noch im Klaren sein, so ist meist nicht nachvollziehbar, ob dieser Sachverhalt mit dem übereinstimmt, den die befragten Benutzer bewertet haben. Erst aus einer validierten Beschreibung des Nutzungskontexts können valide Anforderungen abgeleitet werden.

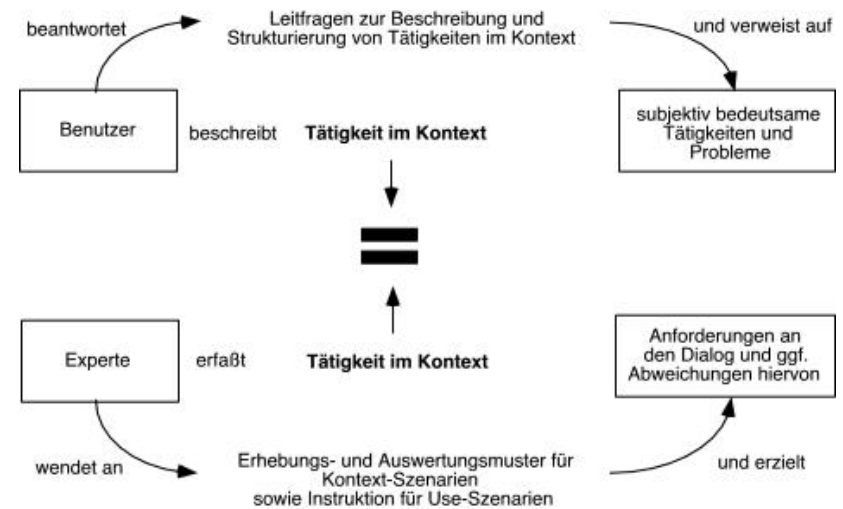


Abb. 2.1-1: Modell zur Erhebung der Beurteilungsgrundlagen als Voraussetzung für die Prüfung von Software

Eine Softwareprüfung kann ohne Beteiligung der Benutzer bei der Anforderungsdefinition methodisch gar nicht sauber vorbereitet werden. Experten mögen in der ergonomischen Beurteilung eines Sachverhalts zu anderen Ergebnissen gelangen als die Benutzer, an der zutreffenden Erfassung des Sachverhalts darf dies jedoch nicht gelegen haben. Um dies zu vermeiden, ist die Übereinstimmung der Beurteilungsgrundlagen zwischen den Benutzern, die den Sachverhalt kennen, und den Experten, die ihn kennen lernen müssen, um Normkonformität zu prüfen, zu sichern. Wir sprechen deshalb von Übereinstimmung, wenn die Prüf- und Beurteilungsgrundlagen validiert wurden, und von Normkonformität (oder Konformität), wenn die geforderten Tätigkeits- oder Produktmerkmale zu dieser Grundlage passen. Die Essenz des Beurteilungsmodells lässt sich wie folgt zusammenfassen: Normkonformität setzt Übereinstimmung der Beurteilungsgrundlagen von Softwareprüfer und Benutzer voraus.

2.2 Verifikation oder Falsifikation

Das Beurteilungsmodell ist so allgemein formuliert, dass es als Voraussetzung für Normkonformitätsprüfungen mit allen software-ergonomischen Normen anwendbar ist; denn alle Normen der Reihe DIN EN ISO 9241, Teile 10 bis 17, gehen davon aus, dass dem Prüfer der Nutzungskontext als Beurteilungsgrundlage bekannt ist. Dem gemäß könnte man verlangen, dass alle Normen zu berücksichtigen sind, wenn ein Produkt auf Normkonformität geprüft wird. Mehr noch, man könnte verlangen, dass alle am Produkt ausführbaren Tätigkeiten und die gegebenen Produktmerkmale gegen die anwendbaren Forderungen in den genannten Normen zu prüfen sind. Tatsächlich wird dieser Prüfumfang in den Normen nahe gelegt. Wir bezeichnen eine Normkonformitätsprüfung dieses Umfangs als *Verifikation* der Erfüllung aller anwendbaren Normen durch ein Produkt. Dieser mit dem Prüfumfang verbundene Aufwand ist eigentlich nur während der Software-Entwicklung notwendig, weil der Usability-Designer jede Entscheidung über Dialogschritte und Produktmerkmale mit Blick auf die Normen treffen muss. In der Designpraxis hat sich gezeigt, dass die Teile 13 bis 17 konstruktive Entscheidungshilfen bieten, während die Teile 10 und 11 für diesen Zweck zu allgemein formulierte Anforderungen enthalten. Hingegen hat sich bei der Evaluierung des Software-Entwurfs oder des Produkts herausgestellt, dass die Teile 10 bis 12 nützlich sind, weil es ausreicht, die gegebene Interaktion des Benutzers mit dem System in Bezug auf die Erfüllung von Designprinzipien zu beurteilen. Sogar Abweichungen von den Teilen 13 bis 17 können akzeptiert werden, wenn diese mit den Prinzipien vereinbar sind. Erst wenn Abweichungen nicht akzeptabel sind, wird die konstruktive Qualitätssicherung wieder aufgenommen, und die Teile 13 bis 17 sind anzuwenden.

Es wird unterstellt, dass in einem Entwicklungsprojekt alles getan wird, um für ein Produkt die Erfüllung von genormten Forderungen zu verifizieren, und zwar in der zuvor beschriebenen Weise. Einem ausgelieferten Produkt wird somit Normkonformität unterstellt. Ein Hersteller, der einen Usability-Engineering-Prozess von hoher Qualität implementiert hat, wird dies für sein Produkt ebenfalls unterstellen und ggf. bereit sein,

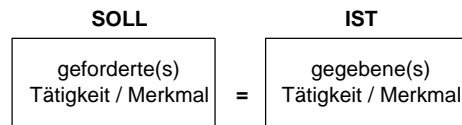
eine Hersteller-Erklärung (gem. DIN EN 45014, Allgemeine Kriterien für Konformitätserklärungen von Anbietern, 1990) abzugeben.

Bei der Normkonformitäts-Prüfung eines installierten Produkts ist es nicht mehr notwendig, alle Detailprüfungen zu wiederholen, die eigentlich während der analytischen Qualitätssicherung im Entwurfsprozess durchzuführen waren. Vielmehr kommt es darauf an, die eben formulierte Vermutung zu widerlegen, nach der ein ausgeliefertes Produkt als normkonform angesehen wird. Wir nennen dies - in Analogie zur wissenschaftlichen Hypothesenprüfung - den Falsifikationsansatz der Produktprüfung. Dieser Ansatz verträgt sich mit der Absicht, dass es unter ökonomischem Gesichtspunkt darauf ankommt, die bedeutsamen Mängel aufzudecken und deren Wirkungen zu untersuchen. Unverzichtbar ist für die Umsetzung dieses Konzepts, dass Fehler- oder Mängelnachweise einem Wartungs- oder Pflegeprozess übergeben werden. Konformitäts-Prüfungen nach dem Falsifikationsansatz sind deshalb kein Selbstzweck, sondern sie weisen stets einen konstruktiven Ausweg.

2.3 Unzureichende Grundlagen der bisherigen Prüfpraxis

Die tägliche Prüfpraxis stimmt leider nicht immer mit dem hier vorgestellten Beurteilungsmodell überein. Oft wird von Herstellern behauptet, der Nutzungskontext eines zu entwickelnden Produkts sei gar nicht genau bekannt, z.B. die Aufgabenmerkmale, die Benutzerzielgruppen, die Probleme, die Software lösen helfen soll usw. Es fehlt folglich eine valide Beurteilungsgrundlage für eine Normkonformitätsprüfung. Mit dem hier vorgelegten Prüfverfahren soll es Herstellern erleichtert werden, den potentiellen Nutzungskontext für ein zu entwickelndes Produkt besser kennen zu lernen. Darüber hinaus wird dazu angeleitet, aus dem Nutzungskontext ergonomische Anforderungen abzuleiten, die den Entwickler in die Lage versetzen, die Normen zutreffend zu interpretieren.

Genormte Forderungen enthalten entweder eine geforderte Tätigkeit oder ein gefordertes Produktmerkmal. Die durchzuführende Prüfung beruht deshalb stets auf folgendem Vergleich:



Entscheidende Grundlage dieses Vergleichs ist die Definition des Geforderten. Diese Festlegung ist leider ein in seiner Schwierigkeit unterschätzter Prozess, ein Grund, warum er oft gar nicht erst stattfindet. Für den Prüfer ist schnell erkennbar, ob die ergonomische Anforderungsdefinition unterlassen wurde, denn dann ist sie weder Teil des Lastenhefts noch des Pflichtenhefts. In dieser Prüfsituation ist das eingangs beschriebene Beurteilungsmodell nicht anwendbar, das ja die Voraussetzung für Normkonformitätsprüfungen ist. Unerfahrene Prüfer ignorieren leider diese Voraussetzung. Sie übersehen, dass ihnen die Beurteilungsgrundlage für die Prüfung fehlt. Statt dessen sehen sie die Beurteilungsgrundlage in den am Produkt ausführbaren Tätigkeiten und in den gegebenen Produktmerkmalen. Dabei wird während der Inspektion versucht, für jedes untersuchte Merkmal (und für jeden ausführbaren Dialogschritt) in den Normen passende Forderungen (SOLL) zu finden. Bei dieser Prüfpraxis ist es schwierig zu begründen, warum nur bestimmte Merkmale oder Tätigkeiten gegen die Norm geprüft werden, denn alles kann ohnehin nicht geprüft werden. Oft wird auch enttäuscht festgestellt, dass die Normen viel zu allgemein formuliert sind, und behauptet, diese seien nicht anwendbar, ja sogar unbrauchbar für den Entwurf und die Prüfung von Produkten. Dies ist ein Grund, warum sich viele Prüfer auf die in den Normen enthaltenen Merkmalsforderungen beschränken und die viel wichtigeren Tätigkeitsanforderungen kaum beachtet werden.

Es stimmt, dass ergonomische Forderungen (auch in den Normen) oft nur als Grundsätze formuliert werden. Dieser Umstand ist für die ergonomische Anforderungs-Entwicklung eher nützlich als schädlich. Denn er zwingt den Usability-Designer und Prüfer dazu, sich Klarheit über die Beurteilungsgrundlage solcher Forderungen zu

verschaffen. In Kenntnis der Bedingungen, unter denen Benutzer arbeiten, ist es möglich, die vage erscheinenden Forderungen zu konkretisieren. Das hier beschriebene Prüfverfahren bietet einen Leitfaden zur Erarbeitung einer validen Beurteilungsgrundlage, zur Ableitung von Anforderungen, zur Interpretation genormter Forderungen und deren Transformation in Prüfkriterien (SOLL). Darüber hinaus wird in die Nutzung eines Entscheidungsverfahrens eingeführt, das bei vermutlichen SOLL-IST-Differenzen aufzuklären hilft, ob eine Normabweichung bedeutsam ist.

Wir sehen die software-ergonomische Produktprüfung in der Tradition der software-technischen Qualitätssicherung und empfehlen dem *Designer* die Anwendung des Verifikationsansatzes während der Herstellung von Software und dem *Prüfer* die Anwendung des Falsifikationsansatzes bei der Prüfung installierter Produkte. Hierbei wird die Software nur stichprobenartig geprüft (z.B. für die Kern-Aufgaben), und es wird versucht, die zunächst unterstellte Normkonformität zu falsifizieren. Diese Strategie hat den Vorteil, dass man die Softwareprüfung zunächst auf die Erfüllung der Dialogprinzipien der DIN EN ISO 9241-10 (1995) beschränken kann, womit Softwareprüfung ökonomisch handhabbarer wird.

Die Umsetzung des Beurteilungsmodells empfehlen wir sowohl dem Designer wie dem Prüfer, denn ohne validierte Daten des Nutzungskontexts eines (noch zu entwickelnden) Produkts kann das SOLL nicht festgelegt werden, das für Produktentwurf und Prüfung unersetzlich ist. In Kapitel 3 werden die Maßnahmen beschrieben, die zur Umsetzung des Beurteilungsmodells notwendig sind. Alle Vorbereitungen dienen der Beschreibung einer validen Beurteilungsgrundlage und der Ableitung von Anforderungen. Die Prüfpraxis verlagert somit ihren Schwerpunkt von der Merkmalsinspektion (IST) in die Anforderungsdefinition (SOLL).

2.4 Erfordernisse, Anforderungen und Prüfkriterien

Das in Abschnitt 2.1 beschriebene Beurteilungsmodell ist Grundlage für die Ableitung von Anforderungen und die Anforderungsdefinition. In Kapitel 3 wird ausführlich beschrieben, wie das Beurteilungsmodell für diese Zwecke umgesetzt wird. Alle

Vorbereitungen dienen letztlich der Festlegung eines SOLLs, gegen das ein gegebenes Produktmerkmal oder eine am Softwareprodukt ausführbare Tätigkeit getestet wird. Wie im vorigen Abschnitt dargelegt, wird die Festlegung des SOLLs meist vernachlässigt, weil man die abstrakt und unscharf formulierten Tätigkeitsanforderungen in der Norm selbst für Prüfkriterien nimmt, was jedoch in der Regel nicht möglich ist, da diese Anforderungen erst durch Interpretation präzisiert werden müssen. Ein Schlüssel für diese Interpretation ist der Begriff *Erfordernis* (engl. "implied need", vgl. ISO 8402, 1994). Im Unterschied zu einer Anforderung, die ausdrücklich festgelegt wird, ist ein Erfordernis implizit in einem Sachverhalt enthalten. Manchmal ist ein Erfordernis so selbstverständlich, dass man darüber keine Worte verliert. Wenn man z.B. einen Autoreifen kauft, so ist es selbstverständlich, dass er mit Luft aufgepumpt ist, sonst ist es schwierig, das Rad zu montieren. Oder, wenn man in einem Dienstleistungsbüro von einem Kunden angerufen wird, so ist es *erforderlich*, das Telefonat sofort zu beantworten. Ein zutreffendes Erfordernis erkennt man daran, dass niemand über seine Berechtigung streitet.

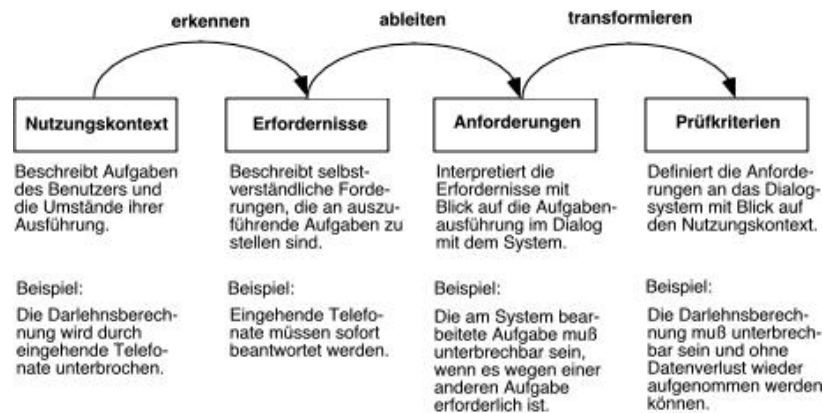


Bild 2.4.-1: Definition von Prüfkriterien anhand von Anforderungen und Erfordernissen des Nutzungskontexts.

Bild 2.4-1 beschreibt den Interpretationsprozess, ausgehend vom Nutzungskontext (DIN EN ISO 9241-11) über das Erkennen von Erfordernissen, die Ableitung von Anforderungen (unter Anwendung der Norm DIN EN ISO 9241-10) bis hin zur Transformation der Anforderungen in Prüfkriterien, und zwar mit Blick auf den Nutzungskontext. Ein Prüfkriterium ist somit eine durch Interpretation der Erfordernisse des Nutzungskontexts gewonnene Forderung.

Der zur Zeit ergiebigste Ansatz für eine Analyse des Nutzungskontexts ist die Beschreibung der Aufgaben des Benutzers mittels Kontext-Szenarien. Ein Kontext-Szenario ist die episodische Beschreibung von Aufgaben in ihrem situativen Kontext, um Motive, Ziele, Tätigkeiten und Probleme des Benutzers zu verstehen, das darin Erforderliche zu erkennen und daraus Anforderungen abzuleiten (vgl. Anhang B). Mittels Kontext-Szenario werden nur die Aufgaben und ihre Durchführungsbedingungen beschrieben. Die tatsächliche Ausführung der Aufgaben am Bildschirmarbeitsplatz ist nicht Gegenstand des Kontext-Szenarios. Der Inhalt der Kontextbeschreibungen soll nicht von den Eigenschaften eines bestimmten Softwareprodukts beeinflusst sein. Für die Beschreibung der ausgeführten Tätigkeit am Bildschirmarbeitsplatz ist eine andere Form von Szenarien angemessen: Use-Szenario. In seiner Kurzform ist das Use-Szenario dem Use-Case (Jacobson et al., 1992) ähnlich. Hierin wird die Interaktion des Benutzers mit dem Programm sequentiell beschrieben. Lediglich die aufgetauchten Nutzungsprobleme (genannt "critical incidents") werden in episodischer Breite beschrieben, um die Problembeschreibung leicht verständlich zu machen (vgl. Anhang C). Auch aus den Use-Szenarien werden Anforderungen abgeleitet. Ein ausdrückliches Formulieren von Erfordernissen ist meist nicht notwendig, es sei denn, der Prüfer stellt fest, dass die Umstände im Nutzungskontext bei der am Bildschirmarbeitsplatz auszuführenden Tätigkeit nicht angemessen berücksichtigt werden.

Der in Bild 2.4-1 beschriebene Prozess hilft, valide Prüfkriterien zu definieren. Es reicht nicht aus, lediglich die Prüfkriterien in eine Liste zu schreiben, sondern man muss deren Herkunft nachvollziehbar dokumentiert haben. In der Praxis geschieht dies durch Validierung der Anforderungsdefinition. Benutzer und Anwender werden zu jedem

Prüfkriterium gefragt, ob die im obigen Ableitungsmodell beschriebenen Schritte schlüssig sind. Validierte Prüfkriterien werden im Lasten- oder Pflichtenheft dokumentiert.

Für die Bestimmung eines Prüfkriteriums ist es notwendig, die abstrakten Tätigkeitsanforderungen, die in einer genormten Forderung stehen, bezüglich der gegebenen Aufgabenerfordernisse oder Benutzerbelange² zu interpretieren. Soll beispielsweise gemäß DIN EN ISO 9241-10 der Dialog unterbrechbar sein, so ist die Unterbrechbarkeit auf die Aufgabenerfordernisse des Nutzungskontexts zu beziehen und aus diesem Sachverhalt das Prüfkriterium abzuleiten.

Um diese Interpretationsprozesse zu vereinheitlichen, werden in dem vorliegenden Prüfverfahren standardisierte Leitfragen formuliert, die den Prüfer anleiten sollen, zu jedem im Nutzungskontext erfassten Sachverhalt die passenden Erfordernisse und Anforderungen abzuleiten (Abschnitt 4.1.3 sowie Anhang B). Damit soll erreicht werden, dass verschiedene Prüfer unabhängig voneinander dasselbe Prüfkriterium definieren, wenn sie denselben Sachverhalt eines Nutzungskontexts analysieren und dabei dieselben Leitfragen stellen. Die Objektivität der Auswertung von Daten des Nutzungskontexts soll durch die Leitfragen gesichert werden. Diese Fragen sollen helfen, folgende Problembereiche eines Nutzungskontexts zu analysieren:

- Was ist das zu erfüllende Aufgabenerfordernis? Stelle sicher, dass dies unmissverständlich und offensichtlich ist und vom Benutzer bestätigt ist.
- Welches ist der zu erfüllende Benutzerbelang? Finde das begründete subjektive Bedürfnis des Benutzers heraus.
- Inwieweit werden Aufgabenerfordernisse und Benutzerbelange von den gegebenen (sozialen, organisatorischen und technischen) Umständen im Nutzungskontext beeinflusst? Finde heraus, ob und inwieweit diese Umstände Aufgabenerfordernisse prägen oder Benutzerbelange mit sich bringen.

Der mit Hilfe der Leitfragen durchgeführte Interpretationsprozess muss folgende methodische Gütekriterien erfüllen:

² Siehe Glossar

- **Validität**, d.h. die Erfordernisse des Nutzungskontexts sind so in Tätigkeiten oder Merkmale zu transformieren, dass Benutzer oder Anwender diesen Anforderungen sachlich zustimmen. (Diese Definition der Validität einer Forderung ergibt sich aus der Definition von Anforderung gemäß ISO 8402, Abschnitt 2.3)
- **Objektivität**, d.h. die Definition des Prüfkriteriums ist unabhängig vom Prüfer. (Gegebenenfalls muss untersucht werden, warum Prüfer nicht zu derselben Definition des Kriteriums gelangt sind.)
- **Reliabilität**, d.h. ein Prüfkriterium muss von demselben Analytiker oder Prüfer bei Wiederholung des Interpretationsprozesses in gleicher Weise definiert werden.

Im Abschnitt 4.3 dieses Berichts wird dargestellt, welche Schwierigkeiten es mit der Erfüllung dieser Gütekriterien gibt und wie mit Hilfe des vorliegenden Prüfverfahrens diese Schwierigkeiten bewältigt werden können.

2.5 Durch Benutzerbefragung ermittelte Beurteilungsgrundlagen

Die bisher behandelten Grundlagen für Normkonformitätsprüfungen sind nicht vollständig, ohne dass auch die subjektiven Beurteilungen der Benutzer berücksichtigt werden. Grundlage für diese Beurteilungen sind die von den Benutzern als bedeutsam angesehenen Tätigkeiten und die bei deren Ausführung aufgedeckten Nutzungsprobleme. Benutzer sind selten in der Lage, anhand eines Nutzungsproblems die Non-Konformität einer Software festzustellen. Es kommt z.B. vor, dass Benutzer beklagen, ein bestimmtes Produktmerkmal sei an ihrem System nicht vorhanden, das sie an einem anderen System gesehen haben, z.B. auf einer Messe. Fraglich ist, ob der Hinweis auf ein fehlendes Produktmerkmal eine geeignete Beurteilungsgrundlage ist. Benutzer sind aber durchaus in der Lage, Mängel in der Ausführung ihrer Arbeitstätigkeiten zu benennen, die einem Prüfer Hinweise auf mögliche Normabweichungen liefern.

Benutzer bewerten die Software ebenfalls nach dem Falsifikationsansatz. Sie vertrauen der Tauglichkeit des Produkts solange, bis sie auf Nutzungsprobleme stoßen. An den Auswirkungen der Nutzungsprobleme baut sich der Unmut der Benutzer auf. Benutzer

bewerten die Qualität des Produkts anhand der erfahrenen Wirkungen, die das Produkt auf sie selbst oder ihre Arbeit hat. Wegen der Beurteilung dieser Wirkungen sind die Benutzer stets bei der Normkonformitätsprüfung einzubeziehen. Gemäß ISO 9241-11 ist eine von Experten festgestellte vermutete Normabweichung immer auch mit den subjektiven Bewertungen der Benutzer (Zufriedenstellung³) zu vergleichen, um die Bedeutsamkeit der Abweichung beurteilen zu können (siehe Erhärtungsprüfung, Abschnitt 3.3). Denn es geht bei Normabweichungen nicht nur um die Feststellung eines Mangels, sondern vor allem um die Bewertung der Wirkungen. Geringe Zufriedenstellung der Benutzer ist immer eine ernstzunehmende Folge, bei der man sich nicht mit einer Feststellung der Beschwerden begnügen darf. Vielmehr sind in jedem Fall die Ursachen zu untersuchen, die übrigens nicht immer allein in den Mängeln der Software liegen müssen.

Weil sich Experten irren können, sind gemäß ISO 9241-11 stets auch die Benutzer zur Bewertung der Zufriedenstellung aufzufordern. Es reicht nicht aus, aus Expertensicht die Konformität der Produkt- und Tätigkeitsmerkmale mit ISO 9241-10 festzustellen, um ein Produkt hinsichtlich der geprüften Merkmale als gebrauchstauglich einzustufen.

In Ergänzung zu den von Prüfern zu erhebenden Szenarien (Kontext- und Use-Szenarien) wird in dem hier vorliegenden Prüfverfahren ein Fragebogen für Benutzer angeboten (Anhang D). Die Angaben, die Benutzer in dem Fragebogen machen, sind eine unverzichtbare Beurteilungsgrundlage für Normkonformitätsprüfungen. Benutzer beurteilen die Zufriedenstellung, indem sie die Effektivität⁴ und Effizienz der Nutzung eines Programms bei einer gegebenen Aufgabe einschätzen. Zusätzlich machen die Benutzer Angaben über Mängel und kennzeichnen besonders störende. Sämtliche Angaben werden als Indikatoren für das Ausmaß der Zufriedenstellung angesehen.

Urteile der Benutzer über die Zufriedenstellung müssen methodischen Gütekriterien genügen. Allgemeine Voraussetzung für die Ermittlung der nachstehenden Gütekriterien ist, dass die Benutzer der zu erhebenden Stichprobe bereits in die

³ Siehe Glossar

⁴ Siehe Glossar

Nutzung des Produkts eingearbeitet sind. Gütekriterien für Zufriedenstellungsurteile sind:

- **Übereinstimmung**, d.h. der Grad der Übereinstimmung der Beurteilungssachverhalte, die der Expertenbeurteilung und der Beurteilung durch Benutzer zugrunde gelegen haben (siehe Beurteilungsmodell, Abschnitt 2.1).
- **Objektivität**, d.h. der Grad der Übereinstimmung in einer Stichprobe von Benutzern derselben Zielgruppe bezüglich der Zufriedenstellung. Unter dieses Gütekriterium fällt auch der Grad der Verallgemeinerbarkeit eines subjektiv festgestellten Mangels über alle Benutzer einer Benutzerzielgruppe.
- **Validität**, d.h. die Beschreibung einer Nutzungssituation als Nutzungsproblem durch einen Benutzer in Übereinstimmung mit der Problembeschreibung eines Experten.
- **Reliabilität**⁵, d.h. die Wiederholbarkeit des Zufriedenstellungsurteils bei denselben Benutzern nach einer mehrwöchigen Nutzungszeit ohne Änderung des Nutzungskontexts (z.B. Schulung).

Das Beurteilungsmodell (Abschnitt 2.1) impliziert, dass ein Nutzungsproblem immer eine ernst zu nehmende Beurteilungsgrundlage ist, insbesondere, wenn es ein Benutzer ist, der darauf hinweist. Denn jeder Hinweis wird als mögliche Normabweichung angesehen, die die unterstellte Normkonformität in Frage stellt. Nur wenn die mit einem Nutzungsproblem verbundenen Wirkungen keine bedeutsame Minderung der Zufriedenstellung verursachen, gilt eine Normabweichung als tolerierbar. Die Wirkungen, die ein Mangel auf die Zufriedenstellung der Benutzer hat, ist die ausschlaggebende Beurteilungsgrundlage bei Normkonformitätsprüfungen (siehe Erhärtungstest, Abschnitt 3.3). Hiermit wird dem Umstand Rechnung getragen, dass Benutzer manche Mängel gar nicht bemerken, weil sie z.B. leicht umgehbar sind oder typische Einarbeitungsprobleme nach der Einarbeitungsphase überwunden sind und kein weiteres Nutzungsproblem darstellen.

⁵ Diese Eigenschaft der Messung ist schwer zu realisieren. Statt dessen wird die Paralleltestreliabilität als Indikator für Reliabilität gewählt (siehe Abschnitt 5.1.2).

Die Messung der Zufriedenstellung mittels Benutzerfragebogen (siehe Anhang D) kann vor einer Konformitätsprüfung stattfinden oder am Ende einer durch Experten durchgeführten Prüfung, etwa im Zusammenhang mit der Erhärungsprüfung (Abschnitt 3.3). In der Erhärungsprüfung kann man zwar auf frühere Fragebogenergebnisse zurückgreifen, aber es ist besser, noch einmal zu fragen, ob ein Benutzer mit einer gegebenen Umgehbarkeit der Nutzungsbarriere zufrieden ist. Zwischen zwei Befragungen kann sich das „Messobjekt“ Benutzer verändert haben, z.B. durch Einstellungsänderung, so dass eine subjektive Minderung der Zufriedenstellung nicht mehr empfunden wird. In diesem Fall ist das frühere negative Zufriedenstellungsurteil des Benutzers eine wenig reliable Beurteilungsgrundlage und sollte in der Erhärungsprüfung nicht den Ausschlag für die Feststellung einer bedeutsamen Normabweichung geben.

2.6 Normen als Prüfgrundlage

Wie in den vorangegangenen Abschnitten schon mehrfach dargelegt, sind ergonomische Normen (insbesondere DIN EN ISO 9241 Teile 10 bis 17) nicht ohne weiteres als Prüfgrundlage geeignet, weil die in den Normen enthaltenen Tätigkeitsanforderungen erst im Zusammenhang mit den Erfordernissen des Nutzungskontexts interpretiert werden müssen. Die Autoren der Normen haben dies im Text mancher Anforderungen ausdrücklich kenntlich gemacht, indem die Erfüllung einer Anforderung davon abhängig gemacht wird, dass es die Arbeitsaufgabe erfordert. Eigentlich ist es überflüssig, solch eine Bedingung in den Text der Anforderungen aufzunehmen, da diese Bedingung generell für genormte Anforderungen gilt.

Bild 2.6-1 bietet einen Überblick über die software-ergonomischen Normen. Eine ähnliche Struktur wird auch im Anhang des Teil 1 der Norm DIN EN ISO 9141-1 beschrieben. Um die Grundzüge der Struktur zu verstehen, ist Bild 2.6-1 einfacher und spiegelt die ursprünglich beabsichtigte Struktur der Normen wider (Çakir und Dzida, 1997; Dzida 1997).

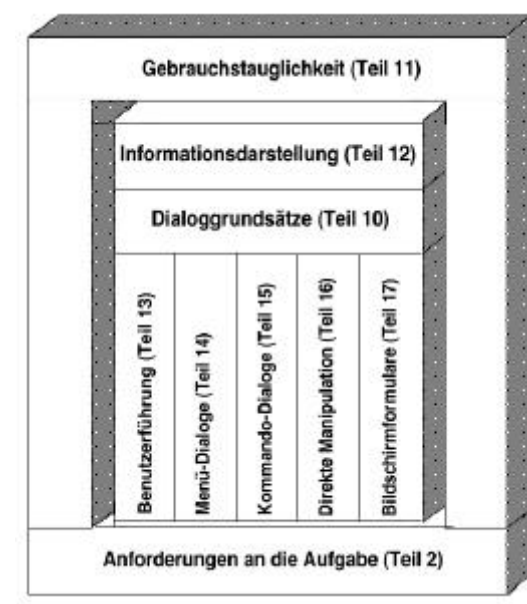


Bild 2.6-1: Struktur der software-ergonomischen Normen DIN EN ISO 9241 Teil 2 und Teile 10 bis 17

Grundlage für die ergonomische Gestaltung von Bildschirmarbeit ist DIN EN ISO 9241 Teil 2. Diese Norm definiert Anforderungen an die Qualität von Arbeitsaufgaben. Wenn schon diese Anforderungen nicht erfüllt sind, beispielsweise bei monotonen, überdauernden Eingabetätigkeiten, kann auch die beste Software keine zufriedenstellende Arbeitssituation schaffen helfen. Den software-ergonomischen Rahmen bildet DIN EN ISO 9241-11. Hier wird definiert, was Gebrauchstauglichkeit ist, nämlich die Effektivität, Effizienz und Zufriedenstellung einer Arbeitstätigkeit im Nutzungskontext eines Softwareprodukts. Effizienz setzt Effektivität voraus, d.h., dass die gewünschten Arbeitsergebnisse korrekt und vollständig mit Hilfe der Software erzielt werden können. Effizienz ist bestimmt durch den im Verhältnis zum Ergebnis notwendigen Aufwand des Benutzers. Eine Beurteilung des Aufwands kann überflüssig sein, wenn schon das

Ergebnis gar nicht oder nicht vollständig oder nicht korrekt erzielt wurde. Zur Optimierung des Aufwands sind die übrigen Teile der Norm anzuwenden, Teil 10 sowie Teile 12 bis 17. Zwei der Normen (Teile 10 und 12) enthalten Designgrundsätze, und zwar für die Darstellung der Information auf der Oberfläche (Teil 12) sowie für den Dialog mit dem Programm (Teil 10). Diese Normen sind abstrakt formuliert, da sie neutral gegenüber den Gestaltungsmerkmalen spezieller Dialogtechniken sind (siehe die Teile 14 bis 17). Eine Sonderstellung unter den speziellen Normen hat Teil 13, da Benutzerführung wohl nicht als Dialogtechnik bezeichnet werden kann.

Wie bereits oben erwähnt, hat der Designer sämtliche speziellen Normen (Teile 13 bis 17) auf die Entwurfsentscheidungen anzuwenden (Verifikationsansatz). Der Prüfer kann sich auf die Teile 10 und 11 beschränken. Teil 12 der Norm kann vernachlässigt werden, wenn dem Entwurf der Benutzungsschnittstelle ein normkonformer Styleguide zugrunde liegt. Allerdings ist von den verbreitet angewendeten Styleguides (z.B. Microsoft Windows, 1995; Sun Java, 1999; The Open Group Motif, 1997) nicht bekannt, ob sie auf Normkonformität geprüft und an ISO 9241 angepasst worden sind (Petrash, 2000). Außerdem sollte der Prüfer nicht unterstellen, dass für ein Produkt ein Style-Guide-Compliance-Test stattgefunden hat, es sei denn, ein entsprechendes Produkt- oder Projekt-Dokument existiert.

Für die Vernachlässigung des Teils 12 der Norm als Prüfgrundlage und für die Beschränkung auf Teil 10 gibt es noch einen anderen Grund. Wenn der Prüfer auf der Grundlage von Szenarien Prüfkriterien definiert, so ist die Prüfung primär am Tätigkeits-(Dialog-)ablauf orientiert. Die Dialogprinzipien des Teils 10 sind für die Gestaltung und Bewertung des Dialogablaufs gedacht. Wenn am Benutzerarbeitsplatz Nutzungsprobleme im Ablauf aufgedeckt werden, so fällt der Blick automatisch auch auf die statischen Merkmale des Anzeigenbereichs eines Bildschirms. Es kommt vor, dass es an der Anordnung dieser Merkmale oder am Bildschirmaufbau liegt, wenn ein Dialogablauf gestört ist. Der Prüfer stellt dann die Normabweichung fest und verweist ggf. auf die Verletzung des Teils 12, indem konkrete Hinweise zur Darstellung von Oberflächenmerkmalen empfohlen werden. Sollte der Prüfer auf eine Störung des Dialogablaufs stoßen, die auf unangemessene Oberflächengestaltung rückführbar ist,

so kann der Prüfer noch einen weiteren Hinweis an den Designer geben. Im allgemeinen beginnen nämlich die Designer mit der Oberflächengestaltung, bevor sie den Dialogablauf entwerfen. Diese Vorgehensweise wird durch die Styleguides nahe gelegt. Eigentlich ist diese Vorgehensweise aber untauglich; denn die Gestaltung der Oberfläche gelingt besser und bereitet weniger Aufwand, wenn der Designer erst den Tätigkeitsablauf einer Aufgabe entworfen hat und sich anschließend überlegt, wie dieser an der Oberfläche darzustellen ist. Diese Empfehlung sollte an Designer weitergegeben werden. Jedenfalls ist mit dem vorliegenden Prüfverfahren die Prüfung so aufgebaut, dass nicht der zweite Schritt vor dem ersten gemacht wird: Es werden auf der Grundlage des Teils 10 die Anforderungen an das Dialogsystem abgeleitet (SOLL) und mit den gegebenen Tätigkeitsabläufen ausführbarer Dialoge (IST) verglichen. Sollte anhand eines aufgedeckten Nutzungsproblems eine Normabweichung festgestellt werden, so wird nach den Ursachen gesucht und mit Hinweis auf konkrete Empfehlungen in den Teilen 12-17 eine Verbesserung vorgeschlagen.

Die Beschränkung der Prüfgrundlagen auf den Teil 10 der Norm ist zweckmäßig, weil sich der Prüfer gar nicht erst in Hunderte von Merkmalen der Benutzungsschnittstelle verzetteln muss, sondern den Blick auf das Prinzipielle richtet. Wenn keine bedeutsamen Abweichungen von den Prinzipien der Dialoggestaltung festgestellt werden, so können zwar bei einzelnen Oberflächenmerkmalen Mängel bestehen; diese stören aber den Benutzer oft gar nicht, oder sie sind von einem Usability-Designer leichter zu ändern, als wenn ein komplizierter oder fehlerträchtiger Dialogablauf implementiert wurde.

3 Prüfverfahren für Experten

Das Prüfverfahren ist für die Feststellung und Bewertung von Normabweichungen entwickelt worden. Hierbei wird die Abweichung eines Produktmerkmals oder einer Benutzertätigkeit von einer aus der Norm (DIN EN ISO 9241-10, 1995) abgeleiteten Mindestanforderung (Prüfkriterium) untersucht. Dieser Test ist der wichtigste Anwendungsfall des Prüfverfahrens. Hier stellt sich die Frage, ob das Verfahren Normkonformität oder Non-Konformität prüft. Da die vollständige Prüfung der Normkonformität zu aufwendig wäre (siehe Argumentation in Abschnitt 2.2), wird einem Produkt Normkonformität unterstellt, dies wird aber zunächst nicht geprüft. Erst wenn es einen Anlass gibt, der den Vertrauensvorschuss in die Qualität eines Produkts in Frage stellt, wird das Produkt bezüglich kritischer Merkmale oder Dialogschritte auf Normkonformität geprüft. Diese Prüfung ist der Versuch, die Vermutung einer Normabweichung zu bestätigen oder zu erhärten (Falsifikationsansatz). Sollte nun wegen dieser empfohlenen Prüfpraxis das vorliegende Prüfverfahren besser Non-Konformitätstest genannt werden? Wir halten dies für eine rhetorische Frage. Denn das Prüfverfahren ist so aufgebaut, dass es selbstverständlich für eine vollständige Normkonformitätsprüfung im Sinne des Verifikationsansatzes eingesetzt werden könnte. Es wird lediglich wegen der damit verbundenen Kosten von diesem Vorhaben abgeraten. Nach dem Motto, "Wo kein Kläger, da kein Richter", wird die Prüfung nicht auf Eigenschaften eines Softwareprodukts angesetzt, über deren Qualität sich niemand beschwert. Es ist ökonomischer, den mit einer Prüfung verbundenen Aufwand auf kritische Fälle zu beschränken und die Ressourcen auf konstruktive Vorschläge für die Verbesserung des Produkts zu verwenden.

Darüber hinaus kann das Prüfverfahren auch für die Entwicklung und Bewertung von Anforderungen bei der Vorbereitung eines Software-Entwicklungsprojekts angewendet werden, z.B. beim Prototyping. Hierbei haben wir es allerdings mit einer Prüfung im Sinne des Verifikationsansatzes zu tun, denn wir befinden uns in einer Entwicklungssituation. Sollte es versäumt werden, das Prüfverfahren in Entwicklungsprojekten einzusetzen, so ist der Vertrauensvorschuss in die Gebrauchstauglichkeit des Produkts nicht gerechtfertigt. Wir überlassen es der weiteren Entwicklung im Qualitäts-

management von Softwareprojekten, ob und inwieweit Softwarehersteller bereit sind, das Prüfverfahren einzusetzen, oder ob Anwender die Hersteller eines Tages vertraglich verpflichten, dies zu tun. Erste Schritte in diese Richtung hat beispielsweise ein öffentlicher Auftraggeber in einer Ausschreibung getan, indem er folgende Anforderungen an die Auftragsvergabe knüpfte:

"Für die zu entwickelnden Produkte sind folgende Richtlinien und Quasi-Standards einzuhalten:"

Es folgte eine Liste fast sämtlicher ergonomischer Normen; dann weiter:

"Für diese projektbegleitenden Arbeiten sind ca. 10% des Gesamtaufwandes einzuplanen. Das notwendige Fachwissen und die Kompetenz zur Bewertung und Zertifizierung der entwickelten Software sind bei der Analyse des Angebots zu dokumentieren. Diese Leistungen können auch in Form eines Unterauftrags an entsprechend kompetente Institutionen vergeben werden." (Dieser Ausschreibungstext und der Auftraggeber sind dem TÜV Rheinland bekannt.)

Dieses Kapitel enthält zunächst eine Darstellung der vorbereitenden Schritte für eine Normkonformitätsprüfung (Abschnitt 3.1). Hierbei wird auf die Methoden verwiesen, die im Anhang A dieses Berichts beschrieben sind. Einige dieser Methoden können auch auf die Durchführung der Normkonformitätsprüfung angewendet werden (Abschnitt 3.2). Der für die Praxis der Produktprüfung bedeutsamste Teil, die Bewertung von Normabweichungen, ist in Abschnitt 3.3 dargestellt. Für die Dokumentation von Prüfergebnissen ist wichtig, dass der Prüfer nachvollziehbar darstellt, wie die in Abschnitt 3.4 beschriebenen Gütekriterien eingehalten wurden.

3.1 Vorbereitung der Normkonformitätsprüfung

Das Prüfverfahren ist in drei Abschnitte gegliedert:

- a) Prüfungsvorbereitung (Abschnitt 3.1, Bild 3.1.2-1),
- b) Durchführung einer Normkonformitätsprüfung (Abschnitt 3.2, Bild 3.2-1) und
- c) Bewertung von Abweichungen: Erhärtungstest (Abschnitt 3.3, Bilder 3.3-1, 3.3-2).

Folgende Methoden (siehe Anhang A) dienen sowohl der Vorbereitung als auch der Durchführung einer Prüfung:

- Aufgabenanalyse
- Inspektion
- Teilnehmende Beobachtung
- Benutzerbefragung
- Dokumentenanalyse

Je nach Prüfsituation werden diese Methoden mehr zur Vorbereitung oder mehr zur Durchführung einer Prüfung angewendet (siehe Anwendungsempfehlungen zu den einzelnen Methoden im Anhang A).

3.1.1 Vorbereitung mittels Benutzerfragebogen

Nicht immer muss gleich beim Auftauchen von Nutzungsproblemen am Benutzerarbeitsplatz ein professioneller Prüfer gerufen werden. Zur Vorbereitung einer Prüfung kann erst einmal ein Benutzerfragebogen eingesetzt werden (siehe Anhang D). Der Fragebogen gehört zu der oben genannten Methode Benutzerbefragung. Vorbereitende Fragebogenerhebungen sind jedoch nicht unter den oben genannten Abschnitten des Prüfverfahrens für Experten erwähnt, weil der Fragebogen nicht obligatorischer Bestandteil des Prüfverfahrens ist. Es soll nicht der Eindruck entstehen, als könne man mittels Benutzerfragebogen die Normkonformität von Softwareprodukten prüfen. Viele Fragebögen erwecken schon aufgrund ihres Namens den Eindruck, dass sie für diesen Zweck geeignet sind (z.B. ISONORM, Prümper 1997). Fragebögen sollten als Initialverfahren eingesetzt werden, d.h., man teilt Fragebögen an eine Stichprobe von Benutzern aus, um einen ersten Eindruck davon zu bekommen, welche Nutzungsprobleme von den Benutzern als störend empfunden werden (vgl. Abschnitt 5.1). Auf diese Weise kann man erste Hinweise auf mögliche Normabweichungen bekommen. Nicht alle der in Abschnitt 5.2 beschriebenen Fragebögen sind jedoch für den Einsatz als Initialverfahren geeignet.

Empfehlenswert ist die Verwendung des Benutzerfragebogens durch die Anwender eines Softwareprodukts. Beispielsweise können betriebliche Beauftragte (Sicherheitsbeauftragte, Arbeitsmediziner, Betriebsräte) den Fragebogen im Rahmen ihres Auftrags nach dem Arbeitsschutzgesetz (1996) oder der Bildschirmarbeitsverordnung (1996) an Benutzer verteilen und nach dem Einsammeln der ausgefüllten Fragebögen feststellen, zu welchen Fragen viele Kommentare abgegeben wurden und wie hoch der Prozentsatz der Benutzer ist, die die beanstandeten Mängel als sehr störend empfinden. Anhand der Kommentare zu den subjektiv sehr störenden Mängeln kann ein hinzugerufener Prüfer mit der Analyse der Nutzungsprobleme beginnen, um einzuschätzen, wie aufwendig eine Produktprüfung werden könnte. Wenn die Nutzungsprobleme breit gestreut sind und der Prüfaufwand beträchtlich zu werden scheint, so ist es empfehlenswert, das in diesem Bericht beschriebene Prüfverfahren vollständig anzuwenden. Wenn hingegen die Nutzungsprobleme vereinzelt auftreten, so kann sich der Prüfer auf die Beschreibung einzelner "critical incidents" in Form von Use-Szenarien (vgl. Abschnitt 4.2) beschränken, die passende Anforderung an die Software bestimmen und einer vermuteten Normabweichung mit Hilfe der Entscheidungstabellen nachgehen (vgl. Abschnitt 3.3).

Der periodische Einsatz des Fragebogens ist empfehlenswert, z.B. nach einem Releasewechsel. Auch für Wiederholungsuntersuchungen eignet sich der Fragebogen, etwa nach Beseitigung von ergonomischen Mängeln im Zuge eines Pflegeprozesses.

Zur Entwicklung des Fragebogens, zu den Maßnahmen vor der Verteilung an die Benutzer und zur Auswertung siehe Abschnitt 5.1.

3.1.2 Prüfungsvorbereitung

Die Prüfung einer Software auf Non-Konformität (oder Normkonformität) setzt voraus, dass ein geschulter Prüfer eingesetzt wird. Das hier vorgelegte Prüfverfahren kann man sich zwar autodidaktisch aneignen. Empfehlenswert ist jedoch eine Ausbildung durch erfahrene Prüfer aus akkreditierten Prüflaboratorien. Die Objektivität des Prüfverfahrens ist nur gewährleistet, wenn die Instruktionen zur Erhebung von

Szenarien beachtet werden und die Anforderungen unter Anwendung der Norm in der hier beschriebenen Weise abgeleitet werden. Bei der Schulung von Prüfern ist besonders darauf Wert zu legen, dass die oben genannten Methoden der Aufgabenanalyse, Inspektion usw. unter Beachtung der Gütekriterien angewendet werden (siehe Anhang A). Im Prüfbericht muss nachvollziehbar sein, dass die Gütekriterien beachtet wurden.

Jede Prüfung auf Normkonformität setzt voraus, dass Prüfkriterien definiert sind. Die Schritte 1 bis 4 beschreiben den Weg von der Analyse des Nutzungskontexts (DIN EN ISO 9241-11) bis zur Festlegung von Prüfkriterien (siehe auch Bild 3.1.2-1).

1. *Kontext-Szenario erheben:*

Leitfragen zur Erhebung von Kontext-Szenarien anwenden (siehe Abschnitt 4.1 und Anhang B).

2. *Anforderungen aus Kontext-Szenario ableiten:*

Aus dem Ergebnis von Schritt 1 werden Anforderungen extrahiert. Leitfragen für die Ableitung von Anforderungen (Abschnitt 4.1.3) und Auswertungsrahmen sind anzuwenden (siehe Anhang B). Während der Ableitung von Anforderungen werden die noch verbleibenden Unklarheiten mit dem befragten Benutzer geklärt. Kontext-Szenario vom befragten Benutzer validieren lassen.

3. *Anforderungen aus Use-Szenario ableiten:*

Jede im Kontext-Szenario festgestellte Kern-Aufgabe wird als Use-Szenario in ihrem vollständigen Tätigkeitsablauf am Bildschirmarbeitsplatz des Benutzers beschrieben (siehe Instruktion im Abschnitt 4.2.1 und Beispiel im Anhang C). Aus dem Use-Szenario werden weitere Anforderungen abgeleitet, sofern der Prüfer ein Nutzungsproblem identifiziert, das auf eine vermutete Normabweichung hindeutet.

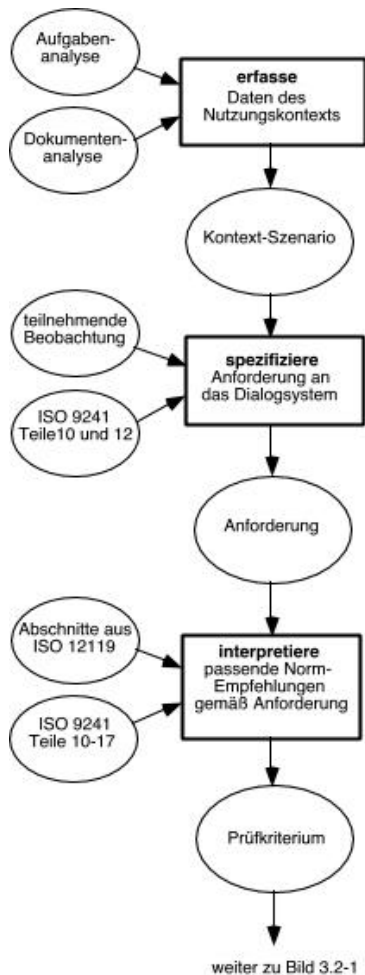
4. *Prüfkriterien definieren:*

Jede Empfehlung der Norm (z.B. DIN EN ISO 9241-10) wird auf Anwendbarkeit geprüft und im Kontext des Ergebnisses der Schritte 1, 2 und 3 konkretisiert. Daraus ergibt sich die Liste der Prüfkriterien.

Zur Vorbereitung einer Normkonformitätsprüfung werden in den Schritten 1 bis 3 bevorzugt die Methoden der Aufgabenanalyse, der teilnehmenden Beobachtung und der Dokumentenanalyse eingesetzt,

- um aus den Erfordernissen des Nutzungskontexts Anforderungen an das Dialogsystem abzuleiten,
- um mit Blick auf den Nutzungskontext (DIN EN ISO 9241-11) des Produkts die Anwendbarkeit von in der Norm formulierten Forderungen zu untersuchen und
- um Anforderungen als Prüfkriterien zu präzisieren, weil die abstrakt formulierten Normen teilweise keine Prüfkriterien enthalten.

Wenn die Norm eine Tätigkeitsanforderung enthält, so ist eine Interpretation dieser Anforderung mit Blick auf die Erfordernisse des Nutzungskontexts notwendig. Interpretieren heißt, die genormte Anforderung unter Berücksichtigung der Kontexterfordernisse zu transformieren, so dass ein Prüfkriterium definiert werden kann.



Erfassung: Da die Gebrauchstauglichkeit eines Produkts stets von den Erfordernissen des Nutzungskontexts abhängt, sind diese zunächst zu erfassen. Man beginnt deshalb mit einer Aufgabenanalyse und der Auswertung von Dokumenten über den Nutzungskontext und beschreibt den Nutzungskontext in Form eines Kontext-Szenarios. Leitfragen zur Datenerhebung (Anhang B) helfen die Objektivität der Daten zu sichern (Erhebungsobjektivität). Die Szenario-Form ist gut geeignet, die Daten von den befragten Benutzern validieren zu lassen.

Spezifikation: Aus dem Kontext-Szenario werden mit Blick auf die Aufgabenerfordernisse und Benutzerbelange Anforderungen an das Dialogsystem abgeleitet. Ein Auswertungsrahmen (Anhang B) hilft die Objektivität der Anforderungen zu sichern (Auswertungsobjektivität). Mittels teilnehmender Beobachtung am Dialogsystem und Dokumentation der Beobachtungsdaten in einem Use-Szenario werden bei identifizierten Nutzungsproblemen weitere Anforderungen ermittelt und unter Berücksichtigung der Dialogprinzipien (Teil 10 der Norm) und der Prinzipien der Informationsdarstellung (Teil 12) spezifiziert.

Interpretation: Die Anforderungen an das Dialogsystem sind kontextspezifisch formuliert, während die in den Normen stehenden Anforderungen kontextneutral sind. Durch Interpretation der neutralen Formulierungen im Lichte der kontextspezifischen Anforderungen gelangt man zur Festlegung von Prüfkriterien. Je nach Art der Anforderung kann man zum Zwecke der Interpretation alle Normen heranziehen, um das Kriterium möglichst präzise zu formulieren. Das Kriterium wird vorzugsweise als ausführbare Tätigkeit formuliert.

Bild 3.1.2-1 beschreibt die Aktionen und Ergebnisse der Prüfungsvorbereitung. Zu allen Aktionen sind die wesentlichen Voraussetzungen genannt. Es fällt auf, dass neben den ergonomischen Normen auch ISO/IEC 12119 (1994) als Voraussetzung für die Definition von Prüfkriterien genannt wird. Dies ist durch die Definition von Gebrauchstauglichkeit begründet (siehe Abschnitt 2.6). Die Leitfragen zur Erhebung von Kontext-Szenarios und der Auswertungsrahmen (siehe Anhang B) leiten den Prüfer an, Effektivitätskriterien gleich zu Anfang der Ableitung von Anforderungen zu definieren, und zwar für jede der im Kontext-Szenario beschriebenen Kernaufgaben. Effektiv im Sinne von DIN EN ISO 9241-11 ist eine Software, die dem Benutzer bei der Erledigung der Kernaufgaben die gewünschten Ergebnisse vollständig und korrekt liefert.

3.2 Durchführung der Normkonformitätsprüfung

Die Normkonformitätsprüfung bildet den Abschluss des Prüfverfahrens, es sei denn, eine Normkonformität kann nicht festgestellt werden, weil eine vermutete Normabweichung vorliegt. In diesem Fall wird die Prüfung durch eine Erhärtungsprüfung (Abschnitt 3.3) erweitert.

Die Normkonformitätsprüfung setzt voraus, dass mittels vorangegangener methodischer Schritte Prüfkriterien definiert worden sind. Prüfkriterien können als Produktmerkmale oder als Tätigkeiten definiert werden (vgl. Abbildung 3.1.2-1). Die Formulierung eines Kriteriums als ausführbare Tätigkeit ist einer Formulierung als Merkmal vorzuziehen, wenn mit dem Merkmal eine bestimmte Umsetzung in der Software nahe gelegt wird.

Bei der Durchführung der Normkonformitätsprüfung (Abbildung 3.2-1) wird das Prüfkriterium mit dem gegebenen Produktmerkmal oder der am Bildschirm ausführbaren Tätigkeit auf Übereinstimmung geprüft. Ist das Prüfkriterium als Merkmal definiert, so reicht i.d.R. eine Inspektion des Produkts aus, um Merkmalsübereinstimmung festzustellen. Ist das Prüfkriterium als Tätigkeit definiert, so ist teilnehmende Beobachtung zur Laufzeit des Systems die bevorzugte Methode.

Bild. 3.1.2-1: Definition eines Prüfkriteriums

Für die Durchführung der Normkonformitätsprüfung werden bevorzugt die Methoden der Inspektion, der teilnehmenden Beobachtung, der Benutzerbefragung und der Dokumentenanalyse eingesetzt,

- um die Erfüllung von in der Norm formulierten Forderungen festzustellen oder
- um einem Verdacht auf Normabweichung und seinen Gründen nachzugehen.

Anhand der kritischen Aufgaben wird Schritt für Schritt untersucht, ob und inwieweit das Produkt die Prüfkriterien einhält. Bei dieser Prüfung sind die aus der Erhebung des Kontext-Szenarios erworbenen Kenntnisse mit zu verwerthen. Die Beobachtung am Benutzerarbeitsplatz liefert eine (möglicherweise leere) Liste von Abweichungen. Werden jedoch vermutete Normabweichungen festgestellt, so werden sie als Nutzungsprobleme ("critical incidents") im Use-Szenario vermerkt; diese sind ggf. einer Erhärtungsprüfung zu unterziehen.

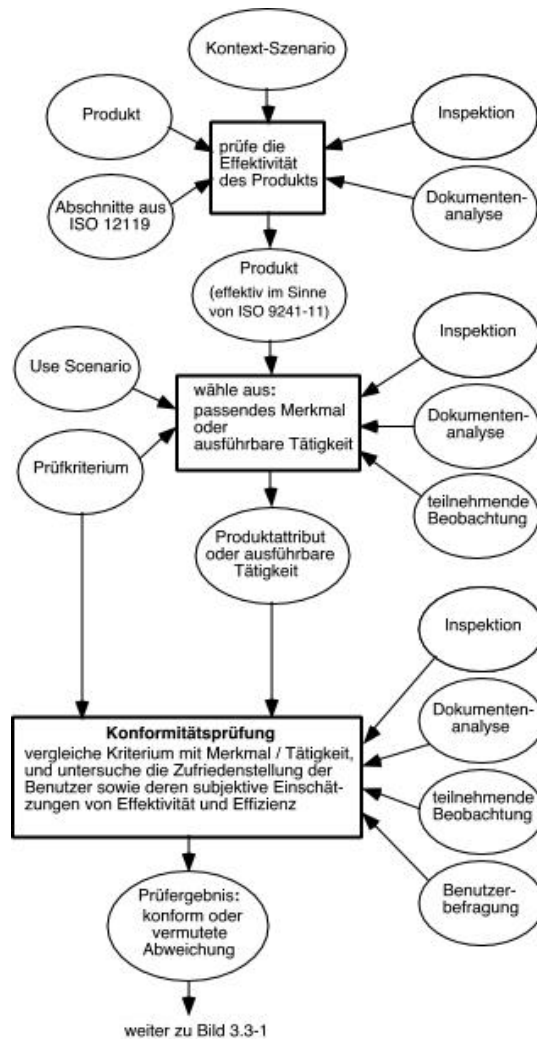
Von einer "vermuteten" Normabweichung zu sprechen, ergibt sich aus dem Falsifikationsansatz, der dem Prüfverfahren zugrunde liegt: jedem Produktmerkmal oder jeder ausführbaren Tätigkeit wird so lange Normkonformität unterstellt, bis diese Annahme durch den empirischen Nachweis negativer Wirkungen widerlegt wurde. Die Wirkungen eines Mangels sind gesondert zu beurteilen (siehe Erhärtungstest, Abschnitt 3.3).

Einige Methoden, z.B. Benutzertests, sind besser geeignet für Konformitätsprüfungen während der Entwicklung des Produkts (formative Evaluierung); die meisten anderen Methoden (Inspektion, teilnehmende Beobachtung, Dokumentenanalyse) sind besser geeignet für die abschließende Produktprüfung (summative Evaluierung).

Einige Methoden eignen sich besser für die Untersuchung der Effektivität des Produkts (z.B. Dokumentenanalyse), andere besser für die Untersuchung der Effizienz (z.B. teilnehmende Beobachtung) und andere wieder besser für die Untersuchung der Zufriedenstellung der Benutzer (z.B. Benutzerbefragung).

Die Normen enthalten sowohl Tätigkeits- als auch Merkmalsanforderungen. Die Methoden sind nicht alle in gleicher Weise für diese unterschiedlichen Anforderungen geeignet. Die Methoden sollten auch nicht unabhängig voneinander eingesetzt werden, sondern einander ergänzend. Beispiele: Eine Inspektion des Dialogs kann auf Grund einer vorangehenden Fragebogenerhebung oder einer Aufgabenanalyse zu eindeutigen Ergebnissen führen; eine Benutzerbefragung oder eine teilnehmende Beobachtung, die zur Aufklärung eines Nutzungsproblems beiträgt, kann ergänzt werden durch eine Merkmalsinspektion.

Bild 3.2-1 beschreibt die Aktionen und Ergebnisse der Normkonformitätsprüfung. Man beginnt mit der Effektivität des Produkts. Mit Hilfe der Norm ISO/IEC 12119 kann man die Effektivität der Software in bezug auf den Einsatzzweck prüfen. Für Prüfer existiert ein Prüfbaustein zur Anwendung dieser Norm im Rahmen der Gebrauchstauglichkeitsprüfung. Der Prüfbaustein ist veröffentlicht im DEKITZ-Prüfhandbuch Gebrauchstauglichkeit, Version 2.0 (2000). Unter Anwendung dieses Prüfbausteins wird geprüft, ob die Software effektiv im Sinne von DIN EN ISO 9241-11 ist, d.h., ob die Software dem Benutzer die gewünschten Arbeitsergebnisse vollständig und korrekt liefert. Es wird unterstellt, dass dies der Fall ist. Wenn aber nicht - und dies können die Benutzer klar beurteilen - so soll mit Hilfe der Erhärtungsprüfung (Abschnitt 3.3) untersucht werden, wie groß die Bedeutung der nicht erreichbaren Ergebnisse ist und wie oft die nicht erreichbaren Ergebnisse benötigt werden. Wird nach dieser Untersuchung festgestellt, dass die Effektivitätsminderung als hoch bewertet wird, dann ist die Normabweichung bedeutsam. Eine weitere Prüfung der Gebrauchstauglichkeit erübrigt sich dann, d.h., es brauchen auch keine weiteren Prüfkriterien definiert zu werden. Wenn beispielsweise ein Programm immer wieder unvorhergesehen zusammenbricht und dadurch Arbeitsergebnisse verloren gehen, so ist jede weitere Prüfung der Effizienz der Nutzung und der Zufriedenstellung der Benutzer überflüssig (siehe auch Bild 3.3-2).



Effektivitätsprüfung: Die Prüfung der Konformität eines Produkts mit ISO 9241-10 setzt voraus, dass das Produkt effektiv im Sinne von ISO 9241-11 ist. Hierzu werden die aus dem Kontext-Szenario abgeleiteten Ergebnisse der Arbeitsaufgaben mit der Leistung des Produkts verglichen (d.h. Inspektion der Funktionalität auf Effektivität und Zuverlässigkeit).

Auswahl passender Merkmale oder Tätigkeiten: Die Kern-Aufgaben der Benutzer wurden in Form von Use-Szenarien dargestellt; aus diesen werden per Inspektion, Dokumentenanalyse oder teilnehmender Beobachtung Merkmale des Dialogsystems oder ausführbare Tätigkeiten des Benutzers abgeleitet, die mit den Prüfkriterien korrespondieren.

Konformitätsprüfung ist ein Vergleich jedes Prüfkriteriums mit einem korrespondierenden Merkmal des Dialogsystems (bzw. der tatsächlich ausführbaren Tätigkeit). Bei mangelnder Übereinstimmung wird die unterstellte Normkonformität in Frage gestellt (vermutete Abweichung). Der Vergleich erfordert eine Inspektion des Dialogsystems oder eine Dokumentenanalyse (bezüglich der Merkmale) oder eine teilnehmende Beobachtung (insbesondere bezüglich der Tätigkeiten) und bei vermuteten Abweichungen eine Befragung der Benutzer.

Bild 3.2-1: Durchführung einer Normkonformitätsprüfung

Entscheidend für die abschließende Bewertung der Gebrauchstauglichkeit der Software ist die Zufriedenstellung der Benutzerzielgruppe. Aussagen über die Normkonformität im Sinne der Gebrauchstauglichkeit ohne Nachweis der Zufriedenstellung mit angemessenen Methoden (z.B. Benutzerbefragung) sind sehr fragwürdig. So wie es nicht möglich ist, ohne Beteiligung der Benutzer die Anforderungen an die Software zu definieren, so ist es nicht möglich, die Gebrauchstauglichkeit der Software zu bestätigen, ohne dass die Benutzer nachweislich zufriedengestellt sind. Zu beachten ist hierbei, dass Anwender und Prüfer zu anderen Bewertungen kommen können als die Benutzer. Wenn es jedoch um die Bewertung der Wirkungen einer festgestellten Non-Konformität geht, dann haben die Benutzer stets das letzte Wort.

3.3 Bewertung von Abweichungen: Erhärtungsprüfung

Die Konformitätsprüfung besteht in einem Vergleich des Prüfkriteriums mit dem relevanten Produktmerkmal oder der relevanten ausführbaren Tätigkeit. Nach dem Falsifikationsansatz wird für jede Tätigkeit (bzw. für jedes Merkmal) Normkonformität unterstellt. Eine festgestellte vermutete Abweichung vom Prüfkriterium ist einer Erhärtungsprüfung zu unterziehen.

Erhärtungsprüfung:

Für jede vermutete Normabweichung wird verifiziert, ob sie

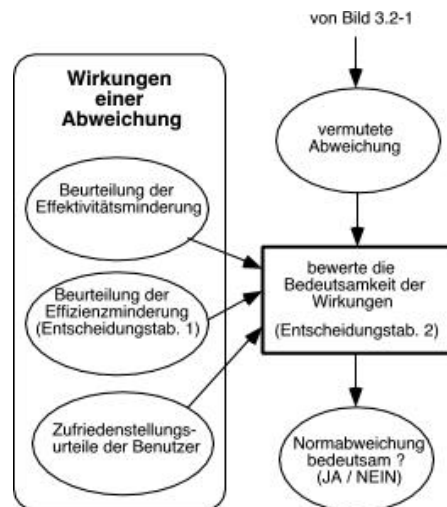
- (a) eine tatsächliche Normabweichung und
- (b) bedeutsam ist.

Bild 3.3-1 beschreibt die Voraussetzungen für eine Erhärtungsprüfung. Für diese Prüfung sind die Entscheidungsregeln der Erhärtungsprüfung anzuwenden. Bild 3.3-2 gibt einen Überblick zu deren Anwendung.

Um festgestellte ergonomische Mängel auf ihre Ursachen hin zu untersuchen, muss berücksichtigt werden, dass die Ursachen nicht nur in der geprüften Software, sondern auch in Merkmalen der Hardware oder in anderen Bedingungen des Nutzungskontexts

gegeben sein können. Eine Beurteilung der Arbeitsbedingungen (im Sinne des §3 der Bildschirmarbeitsverordnung oder aufgrund der DIN EN ISO 9241-2) kann nützlich sein, wenn hierdurch Mängelursachen aufgeklärt werden. Dies ist besonders wichtig, wenn Prüfer, Anwender und Benutzer zu unterschiedlichen Bewertungen eines Mangels neigen.

Mit Hilfe einer Entscheidungstabelle (in Anlehnung an DIN 66 271, 1995) wird festgestellt, ob sich die Vermutung einer Abweichung von der Norm verifizieren lässt. Diese Bewertung beruht hauptsächlich auf der Einschätzung der Wirkungen einer Abweichung.



Ergibt die Normkonformitätsprüfung eine Abweichung zwischen dem Prüfkriterium und dem Produktmerkmal (bzw. der tatsächlich ausführbaren Tätigkeit), so wird eine vermutete Normabweichung konstatiert. Es ist dann ein Anlass gegeben, die unterstellte Normkonformität des Produkts bezüglich dieses Merkmals zu falsifizieren. Um die Bedeutsamkeit der Wirkung der vermuteten Abweichung zu prüfen, wird eine Erhärtungsprüfung durchgeführt. Hierbei werden die Wirkungen der Abweichung auf die Aufgabenbearbeitung (Effektivitäts- und Effizienzminderung) sowie auf den Benutzer (geminderte Zufriedenstellung) untersucht.

Bild 3.3-1: Erhärtungsprüfung: Bewertung von Abweichungen

Folgende Schritte sind zur Durchführung der Erhärtungsprüfung notwendig:

Schritt 1: Vermutete Normabweichung wurde beobachtet.

Schritt 2: Verifikation: "Ist die Abweichung wirklich ein Verstoß
(a) gegen die Normempfehlung, aus der das Kriterium konkretisiert wurde, und
(b) gegen das Dialogprinzip, zu dem die Normempfehlung gehört?"

Hinweis: Wird eine zu den genormten Merkmalen alternative Gestaltung festgestellt, so ist sie nach ISO 9241 erlaubt und nicht als Normabweichung einzustufen, wenn sie weder die Leistung (Effektivität und Effizienz) noch die Zufriedenstellung mindert.

Wenn nein: Abweichung ist nicht bedeutsam. Ende.

Wenn ja: weiter mit Schritt 3.

Schritt 3: Wirkung der Abweichung beurteilen; man orientiert sich an ISO 9241-11, d.h. die Auswirkung der Abweichung wird auf Effektivität, Effizienz und Zufriedenstellung bewertet.

a) Wie groß ist die durch die Abweichung bewirkte Effektivitätsminderung (d.h. fehlende Funktionalität)?

Zunächst wird festgestellt:

"Welche Ziele der Produktnutzung sind aufgrund dieser Abweichung nicht anforderungsgemäß (vollständig und korrekt) erreichbar?"

Schlüsselfragen zu a):

- "Wie groß ist die Bedeutung der nicht erreichbaren Ergebnisse?" bzw.
- "Wie oft werden die nicht erreichbaren Ergebnisse benötigt?"

Anhand dieser drei Aspekte beurteilt der Prüfer die Effektivitätsminderung auf einer 2stufigen Skala: hoch, niedrig.

Ist die Effektivitätsminderung hoch, dann ist die (in Schritt 2 verifizierte) Normabweichung bedeutsam und die Erhärtungsprüfung für diese Abweichung zu Ende (vgl. Entscheidungstabelle 2 unten). Eine hohe Effektivitätsminderung (= relevante Ziele der Produktnutzung sind auch auf Umwegen nicht erreichbar) schlägt also im Gesamturteil stets durch.

b) Wie groß ist die durch die Abweichung bewirkte Effizienz-minderung?

Hier werden drei Unter-Beurteilungen vorgenommen:

ba) Schwere der Auswirkung (unnötiger Aufwand o.ä.):

Schlüsselfragen zu ba):

- "Sind Arbeitsschritte nötig, die nicht durch die eigentliche Arbeitsaufgabe bedingt sind (sondern durch Eigenschaften des Produkts)?"

bzw. falls die Abweichung zu Benutzungsfehlern führt(e):

- "Wie groß ist der Aufwand zur Beseitigung der Fehlerfolgen?"

Einstufung auf 2stufiger Skala: hoch, niedrig

bb) Anteil der betroffenen Benutzer:

Schlüsselfragen zu bb):

- "Welcher Prozentsatz der Benutzerzielgruppe ist betroffen?"
- "Wie oft tritt die Effizienz-minderung bei typischer Produktnutzung auf?"

Einstufung auf 2stufiger Skala: hoch, niedrig

bc) Umgehbarkeit, Kompensierbarkeit durch Nutzungskontext oder alternative Gestaltung:

Schlüsselfragen zu bc):

- "Ist durch Nutzung anderer als der hierfür vorgesehenen Funktionen eine Umgehung des Problems möglich?"
- "Kann die Abweichung durch entsprechende Anpassung des Nutzungskontexts kompensiert werden?" (Bsp.: zusätzliche Benutzerschulung kann fehlende oder schlechte Dokumentation kompensieren; Zusatzprogramm kann fehlende Funktionen ergänzen usw.)

Einstufung auf 2stufiger Skala: leicht, schwer

In der ersten Entscheidungstabelle werden die Bewertungskriterien (ba, bb, bc) übersichtlich angeordnet, um die "Höhe" oder "Schwere" der durch Schlüsselfragen ermittelten Effizienz-minderungen an den in der Tabelle vorgegebenen Ausprägungen zu messen.

Entscheidungstabelle 1:

Gesamturteil für Effizienz-minderung⁶

(2stufiges Ergebnis: hoch, niedrig)

ba)	Schwere der Auswirkung	hoch	-	-	niedrig
bb)	Betroffener Benutzeranteil	-	hoch	-	niedrig
bc)	Umgehbarkeit	schwer	schwer	leicht	-

⁶ Lesen der Entscheidungstabelle: Eine vermutete Abweichung wird nach drei Bewertungskriterien (siehe Zeilen der Tabelle) beurteilt (siehe Einträge "hoch/niedrig" oder "schwer/leicht" in den Spalten). In der letzten Zeile werden die Einzelbewertungen zu einer Gesamtbewertung der Effizienz-minderung zusammengeführt.

Effizienzminderung insgesamt	hoch	hoch	niedrig	niedrig
------------------------------	------	------	---------	---------

Die Effizienzminderung gilt somit insgesamt als hoch, wenn entweder die Schwere der Auswirkung oder der Anteil der betroffenen Benutzer hoch ist und keine leichte Umgehungsmöglichkeit besteht.

Anmerkung zur ökonomischen Vorgehensweise beim Bewerten:

Der Aufwand für die Entscheidungsvorbereitung kann verringert werden, wenn man mit der Bewertung der Umgehbarkeit beginnt; denn bei "leichter" Umgehbarkeit erübrigen sich die weiteren Bewertungen in der Entscheidungstabelle 1. Man kann sofort in die Entscheidungstabelle 2 gehen, um zu bewerten, ob die leichte Umgehbarkeit von den Benutzern als zufriedenstellend angesehen wird. Wird eine hohe Minderung der Zufriedenstellung festgestellt, dann ist trotz leichter Umgehbarkeit eine bedeutsame Normabweichung gegeben (siehe auch Bild 3.3-2).

Wird die Umgehbarkeit als "schwer" beurteilt, so kommt es zusätzlich auf die Bewertung der beiden weiteren Kriterien in der Entscheidungstabelle 1 an. Da die Bestimmung von "Schwere" der Auswirkung oder "Anteil" der betroffenen Benutzer je nach Bewertungsfall unterschiedlich kompliziert sein kann, ist es empfehlenswert, zunächst die leichter operationalisierbare Bewertung durchzuführen, weil sich dann u.U. die schwierigere erübrigt.

Ist die Effizienzminderung hoch, dann ist die Normabweichung bedeutsam (siehe Entscheidungstabelle 2), und die Erhärtungsprüfung für diese Abweichung zu Ende.

c) Wie groß ist die durch die Abweichung bewirkte Minderung der Zufriedenstellung?

Hinweis: Beurteilt wird subjektive Beeinträchtigung bei Benutzern, die das Produkt ausreichend lange benutzt haben, d.h. über die

Einarbeitungsphase hinweg sind, damit eine zeitliche Stabilität des Urteils gewährleistet ist.

Schlüsselfragen:

- "Wie zufrieden/unzufrieden sind die Benutzer mit der Nutzung des Produkts zur Durchführung der Arbeitsaufgabe?" (z.B. Rating-Skalen)
- "Wie hoch ist der Anteil der Benutzer, die eine subjektive Beeinträchtigung empfinden/äußern?"

Einstufung auf 2stufiger Skala: hoch, niedrig

(als gewichtetes Mittel über eine repräsentative Stichprobe der Zielgruppe)

In der zweiten Entscheidungstabelle werden die Bewertungskriterien (a, b, c) übersichtlich angeordnet, um aufgrund der Ergebnisse der Entscheidungstabelle 1 und der Schlüsselfragen zu Effektivität und Zufriedenstellung die Bedeutsamkeit der Normabweichung zu beurteilen.

Entscheidungstabelle 2:

Vorgehen bei der Signifikanzbeurteilung: Übersicht

(2stufiges Ergebnis: Normabweichung ist bedeutsam? Ja/Nein)

a)	Effektivitätsminderung	hoch	-	-	niedrig
b)	Effizienzminderung	-	hoch	-	niedrig
c)	Minderung d. Zufriedenstellung	-	-	hoch	niedrig
	Normabweichung bedeutsam	Ja	Ja	Ja	Nein

Insgesamt ist eine vermutete Abweichung also eine bedeutsame Normabweichung, wenn sie (a) als Verletzung der Normempfehlung und des Dialogprinzips verifiziert

werden kann und (b) in mindestens einer der drei Kategorien Effektivität, Effizienz und Zufriedenstellung zu einer hohen Minderung führt.

Aufgrund der Entscheidungstabelle 1 (Effizienzmindering) bedeutet das, dass mangelnde Effektivität oder mangelnde Zufriedenstellung stets "durchschlagen"; Effizienzmängel hingegen durch eine leichte Umgehungsmöglichkeit kompensiert werden können.

Aus Bild 3.3-2 geht hervor, dass eine Effizienzprüfung nur durchzuführen ist, wenn keine vermutete Normabweichung im Programm gefunden wird, die die Effektivität bedeutsam mindert. Die Zufriedenstellung muss nur in den Fällen erhoben werden, in denen die Effizienzmindering (nach Entscheidungstabelle 1) als niedrig bewertet wurde.

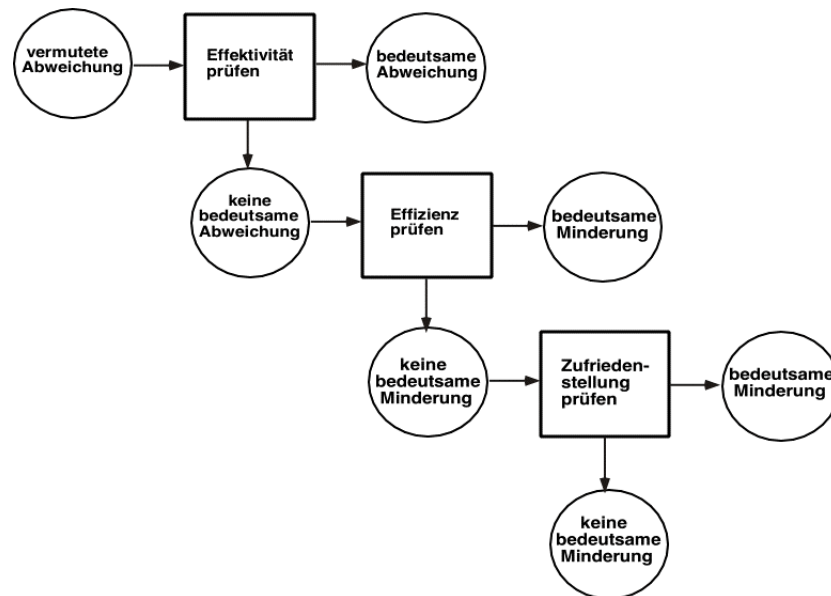


Bild 3.3-2 Die Entscheidungsregeln der Erhärtungsprüfung im Überblick

Um die Erhärtungsprüfungen zu einem konstruktiven Ausgang zu führen, sind folgende Maßnahmen denkbar, falls ausreichende Ressourcen verfügbar sind (siehe Bild 3.3-2):

- die fehlende Effektivität wird behoben, z.B. durch Wartung,
- die mangelnde Effizienz wird durch Merkmalsänderung beseitigt, z.B. durch Anpassung an die in den Teilen 12 bis 17 von ISO 9241 geforderten Merkmale,
- die Zufriedenstellung wird durch Merkmalsänderung oder durch kompensatorische Gestaltung des Nutzungskontexts erreicht.

Entscheidungsregeln kurz gefasst:

1. Zunächst die Effektivität der Nutzung bewerten.
Bei bedeutsamer Abweichung: Software anpassen.
2. Bei keiner oder nur geringfügiger Einschränkung der Effektivität: Effizienz bewerten.
Bei bedeutsamer Minderung: Benutzungsschnittstelle anpassen.
3. Bei nicht bedeutsamer Minderung der Effizienz: Zufriedenstellung prüfen.
Bei bedeutsamer Minderung der Zufriedenstellung: Merkmalsänderung notwendig oder kompensatorische Maßnahme.
4. Erfolg der Änderung / Anpassung oder der Maßnahme bewerten.

3.4 Gütekriterien der Konformitätsprüfung

Aus der Dokumentation einer Konformitätsprüfung muss hervorgehen, ob die nachstehenden Gütekriterien eingehalten wurden:

- Ist das Prüfkriterium so weit wie möglich als Tätigkeitsanforderung formuliert?
- Ist die Mindestanforderung einer Norm auf Anwendbarkeit geprüft?
- Sind genormte Tätigkeitsanforderungen mit Blick auf die Erfordernisse des Nutzungskontexts interpretiert worden?

- Ist die Erfüllung von Tätigkeitsanforderungen zur Laufzeit des Systems und im Nutzungskontext geprüft worden?
- Sind Normabweichungen bezüglich ihrer Relevanz auf Beeinträchtigungen der Benutzer und störende Umstände im Nutzungskontext analysiert worden?
- Differenziert die Normkonformitätsaussage ggf. zwischen der festgestellten ergonomischen Qualität im vom Hersteller beabsichtigen Nutzungskontext (ISO /IEC 12119) und dem realen Nutzungskontext⁷ beim Anwender (DIN EN ISO 9241-11)?
- Ist die Zufriedenstellung der Benutzerzielgruppe untersucht worden?
- Ist die Wirkung einer vermuteten Abweichung vom Prüfkriterium auf Bedeutsamkeit untersucht worden (Erhärtungsprüfung)?

⁷ Unter „Benutzbarkeit“ wird ein mit ISO/IEC 12119 normkonformes Produkt verstanden, das der Hersteller für einen beabsichtigten Nutzungskontext entwickelt hat. Ob sich dieses Produkt in einem realen Nutzungskontext, also am Arbeitsplatz des Benutzers, tatsächlich als „gebrauchstauglich“ erweist, wird durch einen Konformitätstest mit DIN EN ISO 9241 Teile 10 und 11 festgestellt.

4 Analyse des Nutzungskontexts und der Nutzung

Die Gebrauchstauglichkeit von Software-Produkten ist ein Qualitätsmerkmal, das die Nutzungsqualität beschreibt. Software-technische Qualitätsmerkmale wie Zuverlässigkeit, Änderbarkeit, Wartbarkeit usw. sind notwendige, aber nicht hinreichende Voraussetzungen für Nutzungsqualität. Diese hängt von den Einflüssen und Anforderungen des Nutzungskontexts eines Produkts ab. DIN EN ISO 9241-11 (1998) verlangt eine Analyse des Nutzungskontexts für die Ermittlung von Anforderungen. Merkmale des Nutzungskontexts sind z.B. die Aufgaben der Benutzer, Merkmale der Benutzer, das organisatorische und soziale Umfeld, Ziele der Organisation, die technische Ausstattung. Softwarehersteller können durch eine gründliche Analyse des Nutzungskontexts vor Beginn eines Projekts Entwicklungskosten sparen (Landauer, 1995). Anwender und Benutzer können wirksam an der Vorbereitung eines Entwicklungsprojekts sowie bei der Bewertung von Prototypen beteiligt werden, wenn Szenarien erhoben werden.

Eine Analyse des Nutzungskontexts nach den Empfehlungen der Norm (DIN EN ISO 9241-11) wurde hier in Form von Leitfragen für Szenarien umgesetzt. "Scenario-based analysis and design" ist eine seit Jahren im Requirements Engineering bewährte Methodik (Carroll, 1995). Wir unterscheiden zwei sich ergänzende Arten von Szenarien: Kontext- und Use-Szenarien (Dzida und Freitag, 1998). In diesem Kapitel werden zunächst die Erhebung und Auswertung von Kontext-Szenarien beschrieben (Abschnitt 4.1). Ferner wird beschrieben, wie Use-Szenarien erhoben werden können, um die Liste der Anforderungen zu ergänzen, die aus den Kontext-Szenarien abgeleitet wurde (Abschnitt 4.2). Schließlich wird dargelegt, wie die Objektivität der Erhebung und Auswertung von Szenarien gesichert werden kann (Abschnitt 4.3).

4.1 Erhebung von Kontext-Szenarien

Ein Szenario ist eine episodische Beschreibung von Tätigkeiten in deren Kontext. Bei der Erhebung eines *Kontext*-Szenarios kommt es im Unterschied zu einem *Use*-Szenario nicht darauf an, die genauen Arbeitsschritte am Dialogsystem zu

beschreiben. Die konkrete Nutzung von Arbeitsmitteln wird im Kontext-Szenario nicht erfasst. Allerdings werden alle Darstellungen des Benutzers über die Schwierigkeiten der derzeitigen Nutzung von Arbeitsmitteln erfasst sowie die Vorstellungen (Erwartungen) des Benutzers von Verbesserungsmöglichkeiten, die sich aus der Einführung von neuen Arbeitsmitteln (z.B. Anwendungsprogrammen) ergeben könnten. Voraussetzungen und Ziele der Ausführung von Tätigkeiten werden beschrieben, die normale und gestörte oder fehlerhafte Durchführung von Aufgaben, die hierbei relevanten organisatorischen Umstände usw. Damit bei der Erhebung von Kontext-Szenarien objektive Daten erhoben werden, wird der Analytiker durch Leitfragen angeleitet, die ähnlich einem strukturierten Interview geordnet sind (vgl. Tabelle 4.1-1). Diese Leitfragen müssen jedoch nicht streng sequentiell abgefragt werden, sondern dienen der nicht-direktiven Steuerung der Darstellungen des befragten Benutzers (siehe Instruktion zur Anwendung der Leitfragen, Abschnitt 4.1.1). Der "Charme" der Szenarien-Erhebung liegt in der freien Erzählform, die der Benutzer zur Darstellung seiner Tätigkeiten im gegenwärtigen Kontext wählt und die der Analytiker möglichst in der Sprache des Benutzers dokumentiert.

Bei der Erhebung von Kontext-Szenarien und der Ableitung von Anforderungen wird bewusst die Nutzung eines konkreten Systems außer acht gelassen. Eine oft gestellte Frage ist, warum Anforderungen an ein Dialogsystem ohne Berücksichtigung des tatsächlichen Systems ermittelt werden. Genau genommen geschieht dies erst bei der Erhebung von Use-Szenarien (Kapitel 4.2). Vorweg wird jedoch versucht, Anforderungen an das System nicht durch gegebene Systemmerkmale "einzufärben", sondern allein im Zusammenhang mit den Erfordernissen des Nutzungskontexts zu verstehen. Würde man Anforderungen mit Blick auf gegebene Systemmerkmale formulieren, so würde man die Prüfung dieser Merkmale gegen jede Kritik immunisieren. Dieser methodische Fehler wird häufig mit Merkmals-Checklisten gemacht, die unkritisch von Prüfern übernommen werden, ohne die gegebenen Checklisten-Items durch nachgewiesene Erfordernisse des Nutzungskontexts zu legitimieren. Um dieser Gefahr zu begegnen, wird den Analytikern bei der Ableitung von Anforderungen aus Szenarien empfohlen, diese nicht in der Form von Systemmerkmalen (Funktionen, implementierte Oberflächen- und Dialogmerkmale) auszudrücken, sondern als Arbeitstätigkeiten

(siehe Beispiel-Szenario). Jede abzuleitende Anforderung an das Dialogsystem ist mit Blick auf das zugrunde liegende Arbeitserfordernis zu formulieren. Die Praxis zeigt, dass sich Anforderungen, die als Tätigkeiten formuliert sind, in der Regel besser im Lichte der genormten Dialoganforderungen (DIN EN ISO 9241-10) interpretieren lassen.

In der Praxis hat sich herausgestellt, dass die Erhebung von Sachverhalten des Nutzungskontexts mit Hilfe der Leitfragen eingeübt werden muss. Damit die Erhebung objektive Ergebnisse liefert, sollte die im nachstehenden Abschnitt formulierte Instruktion beachtet werden. Man kann die Erhebung von Kontext-Szenarien auch autodidaktisch lernen. Es hat sich aber herausgestellt, dass eine Schulung unter Anleitung eines erfahrenen Analytikers zu besseren Ergebnissen führt. Abschnitt 4.1.2 enthält eine Trainingsanleitung.

Tabelle 4.1-1: Leitfragen zur Erhebung von Kontext-Szenarien

Abschnitt im Szenario	Leitfragen (Erläuterungen hierzu im Auswertungsrahmen, Beispiele hierzu im Beispiel-Szenario)
Einleitung	<ul style="list-style-type: none"> • Formulieren Sie die Tätigkeit in einem oder in zwei Sätzen. • Aus welchen Aufgaben ist die Tätigkeit zusammengesetzt (typische Kern-Aufgaben aufführen, d.h., wenn großer Zeitanteil oder häufig wiederkehrend oder sehr wichtig)? Welche dieser Kern-Aufgaben sollen durch die Software unterstützt werden? • Wie ist die Tätigkeit organisiert (z.B. als Mischarbeit, als Folge von Aufgaben, als monotone Einzelaufgabe)?
Voraussetzungen	<ul style="list-style-type: none"> • Welche Qualifikation ist zur Bewältigung der Aufgaben erforderlich (Aufgabenbewältigung / Softwarenutzung)? Welche Vorkenntnisse fehlen ggf.? • Wer (bzw. welches Ereignis) bestimmt, was zu tun ist? (Wer trifft die Auswahl? Selbständigkeit der Bearbeitung, Arbeitsteilung, externe Datenquellen) • Welche Hilfsmittel sind erforderlich (für die Aufgabenbewältigung / zur Softwarenutzung)? Welche davon fehlen ggf., welche sind zusätzlich gewünscht?
Normale Durchführung	<ul style="list-style-type: none"> • Welche Arbeitsschritte sind durchzuführen? • Welche Arbeitsschritte kehren häufig wieder? (Automatisierung gewünscht / erforderlich?) • Welche Arbeitsschritte werden automatisch durchgeführt? Sind bei automatisierten Arbeitsschritten Einflussmöglichkeiten des Benutzers vorhanden / erlaubt / gewünscht / erforderlich? • Kommt es vor, dass mehrere Benutzer gleichzeitig an dem gleichen Objekt (z.B. Vorgang, Akte, Dokument, Datensatz) arbeiten müssen? • Gibt es eine festgelegte Abfolge der Arbeitsschritte und wenn ja, wie sieht diese aus? (Ist eine Flexibilität sinnvoll / nötig?)

	<ul style="list-style-type: none"> • Welche Ergebnisse / Teilergebnisse entstehen und wie werden diese ggf. verwertet / weitergeführt? • Welches Feedback bekommt die befragte Person in Bezug auf die Arbeitsergebnisse und die Wirkung ihrer Arbeit?
Besonderheiten bei der Durchführung	<ul style="list-style-type: none"> • Welche Unterbrechungen gibt es und warum? Welche Störungen treten auf (organisatorisch / sozial / technisch)? • Wie werden Fehler zurückgemeldet und behoben (organisatorisch / sozial / technisch)? • Welche wichtigen Sonderfälle müssen berücksichtigt werden (bzw. fallen dem Benutzer spontan ein; z.B. zur Arbeitsteilung / Zusammenarbeit)?
Organisatorische Rahmenbedingungen	<ul style="list-style-type: none"> • Welche Organisationsziele gibt es im Hinblick auf die Tätigkeit? • Gibt es Mechanismen zur Leistungssteuerung / Leistungskontrolle? (Wenn ja, welche? Sind diese erforderlich?) • Welchen Überblick hat der Benutzer im Hinblick auf die Gesamttätigkeit? • Welche Änderungen, die die Aufgabenbearbeitung beeinflussen, sind zu erwarten oder werden gewünscht? Welche Vorschläge hat der/die Befragte dazu? • Von welchen Arbeitsergebnissen / Arbeitsschritten sind Dritte (z.B. Kunden) direkt betroffen? Und was folgt daraus? • Welche Stressfaktoren gibt es und wie wird damit umgegangen?
Sonstige Bemerkungen der befragten Person zu bereits aufgetretenen Nutzungsproblemen.	Hier Beispiele sammeln, falls die befragte Person bereits während des Interviews etwas über aufgetretene Nutzungsprobleme berichtet. Im Allgemeinen werden solche Probleme besser an Hand von Use-Szenarien analysiert.

4.1.1 Instruktion zur Erhebung von Kontext-Szenarien

Die zu befragende Person soll von einer Führungskraft ausgesucht werden, die beurteilen kann, ob die Person mitteilungsfreudig ist. Die Führungsperson selbst ist nicht gut geeignet. Es soll eine Person sein, die die Arbeit tatsächlich ausführt. Arbeitspersonen, die die Tätigkeit erst wenige Monate ausführen, sind für die Befragung nicht geeignet. Der Führungskraft soll diese Instruktion übergeben werden, damit keine Missverständnisse über Zweck und Verwendung der Datenerhebung entstehen.

Zu Beginn der Befragung wird der befragten Person erläutert, dass es sich um ein Gespräch handelt, in dem der Interviewer versuchen will, zu verstehen, welche Arbeit die Person erledigt. Es wird angekündigt, dass alle aufgezeichneten Sachverhalte von der Person auf Richtigkeit geprüft werden sollen, damit eventuelle Missverständnisse beseitigt werden. Es wird versichert, dass es nicht um die Bewertung der Arbeit geht, erst recht nicht um die Bewertung der Person, vielmehr um die alleinige Bewertung der Software. Es wird erklärt, dass man die Tauglichkeit der Software aber nur dann beurteilen kann, wenn vorher die zu erledigenden Aufgaben klar verstanden worden sind. Deshalb soll zunächst nicht über die aktuelle Software-Nutzung gesprochen werden, sondern nur über die Aufgaben selbst, ihre Bearbeitung, die Umstände, unter denen dies geschieht. Dabei sollen Voraussetzungen und Schwierigkeiten der Bearbeitung zur Sprache kommen, sowie Schwachstellen im Kontext der Aufgabe, persönliche Vorstellungen über Verbesserungsmöglichkeiten u.ä.

Nur geübte Interviewer sollen die Erhebung alleine durchführen. Besser ist es, wenn zwei Interviewer sich die Arbeit teilen: einer stellt die Fragen, macht sich Notizen, während der zweite das Gespräch beobachtet, sich ebenfalls Notizen macht, aber besonders darauf achtet, dass alle Leitfragen in den einzelnen Abschnitten tatsächlich gestellt werden und alle Ausführungen der befragten Person verstanden worden sind. Im Zweifel stellt der zweite Interviewer ergänzende Fragen.

Bei der Erhebung von Kontext-Szenarien hat es sich als zweckmäßig herausgestellt, einen Schreibblock DIN A4 zu benutzen. Auf jede neue Seite wird vorbereitend je eine der Leitfragen aufgeschrieben. Während des Interview-Gesprächs geht man dann Seite für Seite vor, so dass die Struktur des Gesprächs gewahrt wird und keine der Leitfragen übergangen (vergessen) werden kann.

Es ist nicht notwendig, sich beim Formulieren der Fragen an den genauen Wortlaut der Leitfragen zu halten. Die Zielsetzung der gestellten Frage sollte zum Ausdruck kommen (siehe Erhebungs- und Auswertungsrahmen, Anhang B). Die befragte Person sollte immer ausreden dürfen; sie darf nicht durch Zwischenfragen unterbrochen werden. Redet die Person über unwichtige Dinge, ist es zweckmäßig, sie weitersprechen zu lassen und vorsichtig zu versuchen, sie zum Thema zurückzuführen; dabei müssen die überflüssigen Informationen nicht mitprotokolliert werden.

Der Interviewer sollte nicht auf die Vollständigkeit der Erhebung fixiert sein. Es ergeben sich später (bei der Validierung) genügend Anlässe, um Daten zu ergänzen. Vielmehr sollte eine möglichst entspannte und flüssige Gesprächssituation aufrechterhalten werden. Es kommt nicht auf Perfektion an, sondern auf ein motivierendes Gesprächsklima. Hierin liegt der Charme der Erhebung von Szenarien im Unterschied zum strukturierten Interview, das dem Interviewer eine strengere Disziplin beim Abfragen auferlegt.

4.1.2 Trainingsanleitung zur Erhebung von Kontext-Szenarien

Ein Analytiker in der Rolle des Softwareprüfers oder Anforderungs-Entwicklers kann die Methode der Erhebung und Auswertung von Kontext-Szenarien entweder autodidaktisch oder unter Anleitung eines Trainers lernen. Dabei sind stets die Liste der Leitfragen, das Beispielszenario sowie der Auswertungsrahmen zur Hand zu nehmen.

Bei autodidaktischer Einarbeitung in die Erhebung von Szenarien sind folgende Lernschritte zu beachten:

1. Instruktion in Abschnitt 4.1.1 lesen.
2. Lernen am Beispiel: Im Beispielszenario jeden Szenario-Abschnitt durcharbeiten; dabei ist der Inhalt des Beispiels mit der Liste der Leitfragen, Abschnitt 4.1, zu vergleichen.
3. Zielsetzungen der Leitfragen besser verstehen: Im Erhebungs- und Auswertungsrahmen (siehe Anhang B) sind die Zielsetzungen der Leitfragen nachzulesen und mit den Abschnitten im Beispielszenario zu vergleichen, um sich klar zu machen, worauf es bei der Erhebung eines Szenarios und bei der anschließenden Reinschrift ankommt.
4. Mit Blick auf die Auswertungssituation für die Erhebungssituation lernen: Die im Beispielszenario abgeleiteten Erfordernisse und Anforderungen werden Schritt für Schritt nachvollzogen, um zu lernen, wie präzise ein Sachverhalt beschrieben werden muss, damit anschließend die vorgegebenen Erfordernisse und Anforderungen abgeleitet werden können.
5. Warnung vor schematischer Auswertung: Am Beispielszenario soll auch erkannt werden, dass neben den vorgegebenen Erfordernissen und Anforderungen weitere Ableitungen nötig sind, wenn dies vom Inhalt des Kontext-Szenarios her geboten erscheint.
6. Ableitung der Prüfkriterien nachvollziehen: Im Beispiel sind die Anforderungen an die Software als Prüfkriterien spezifiziert, d.h., die im Auswertungsrahmen vorgegebenen Anforderungen sind mit Blick auf den konkreten Szenario-Inhalt präziser formuliert worden.

Bei autodidaktischer Einarbeitung ist es sehr empfehlenswert, während der ersten Erhebung einen erfahrenen Analytiker dabei zu haben, der in Form einer Supervision die Gesprächssituation beobachtet, bewertet und anschließend mit dem beobachteten Analytiker bespricht. Da die Sachverhalte des Nutzungskontexts möglichst objektiv dargestellt sein sollten, ist eine kritische Durchsicht der Reinschrift des Szenarios durch den Supervisor sehr nützlich. Der Supervisor hat die mündliche Darstellung der Kontext-Sachverhalte selbst erlebt und kann die Objektivität der Reinschrift inhaltlich beurteilen.

Eine Einarbeitung in die Erhebungs-Methode durch einen Trainer kann weniger aufwendig sein als das autodidaktische Lernen.

Der zu Trainierende sollte zuvor folgende Unterlagen gelesen haben:

- Instruktion, Abschnitt 4.1.1,
- Leitfragen, Abschnitt 4.1
- Beispielszenario, Anhang B
- Erhebungs- und Auswertungsrahmen, Anhang B

Empfehlenswert ist es, diese Unterlagen auf dem Tisch zu haben, so dass jede Frage der Leitfragen mit Blick auf die Zielsetzung und die Gesprächssituation durchgearbeitet werden kann. Zweckmäßig ist es, die Situationen der Datenerhebung, der Reinschrift und der Auswertung nacheinander zu besprechen. Bei der Auswertungssituation angekommen, lohnt es sich, die im Auswertungsrahmen vorgegebenen Anforderungen an das Dialogsystem der jeweils passenden Leitfrage gegenüber zu stellen. Wenn dieser Bezug aus der Reinschrift nicht erkennbar ist, so ist die Objektivität der Ableitung von Dialoganforderungen gefährdet. Es lohnt sich also, die Erhebung und Reinschrift von Kontext-Szenarien aus der Sicht der Auswertung zu betrachten und somit die Sensibilität für die Zielsetzungen der Leitfragen zu schärfen. In einer trainer-geleiteten Einarbeitungssituation ist dieser Lerneffekt besser zu erreichen als beim Selbststudium.

Für die Reinschrift und Auswertung lohnt es sich, einen Auswertungsrahmen bereit zu legen, in dem nur die Zellen der zweiten Spalte leer sind, während in den übrigen Zellen die vorgegebenen Erfordernisse und Anforderungen enthalten sind. Nachdem in der zweiten Spalte für jede Leitfrage ein Sachverhalt eingetragen wurde, sollte die dritte Spalte darauf geprüft werden, ob sich das vorgegebene Erfordernis aus dem Sachverhalt ableiten lässt. Ist dies nicht der Fall, so ist der Inhalt des Szenario-Abschnitts nicht hinreichend klar beschrieben. Es kommt auch vor, dass ein von der Vorgabe abweichendes Erfordernis entdeckt wird; dann ist dieses einzutragen. Der Begriff des Erfordernisses ist ein Schlüsselbegriff der Qualitätssicherung, weil er den Analytiker / Prüfer dazu bewegt, sich in "die Natur der Sache" einzudenken. Nicht

immer wird ein dem Kontext-Sachverhalt innewohnendes Erfordernis sofort klar sein. Es ist besser, kein Erfordernis in die zweite Spalte einzutragen, als eines, das bezweifelt werden kann. Ein zutreffendes Erfordernis erkennt man daran, dass es einleuchtend und selbstverständlich erscheint, also zweifelsfrei ist.

An Bildschirmarbeitsplätzen in Büros hat sich empirisch herausgestellt, dass es bei der Erledigung von Aufgaben grundlegende Erfordernisse gibt, die fast immer vorkommen, wenn man ein Kontext-Szenario auswertet. Die im Auswertungsrahmen enthaltenen typischen Erfordernisse können nützlich sein, um beim Versuch angeleitet zu werden, eine für den gerade gegebenen Sachverhalt zutreffende Anforderung an die Software zu erkennen. Bei der Nutzung dieser Liste besteht jedoch die Gefahr, dass man nur noch schematisch vorgeht und die vorgegebenen Erfordernisse wie Bausteine eines Puzzles benutzt, die man den Sachverhalten des Kontext-Szenarios zuordnet.

Sobald ein vorgegebenes Erfordernis als zutreffend erkannt oder ein zusätzliches eingetragen wurde, kann das mit dem Erfordernis assoziierte Dialogprinzip festgestellt werden (manchmal treffen mehrere Dialogprinzipien zu). Dieser Schritt vom Erfordernis zum Dialogprinzip leitet den Auswerter an, das aus dem Sachverhalt "Erforderliche" in eine "Anforderung" an das Dialogsystem zu transformieren. Hierbei sollte auch die Norm DIN EN ISO 9241-10 benutzt werden, um unter dem festgestellten Dialogprinzip die passende Standard-Anforderung zu finden. Im Auswertungsrahmen ist die als passend erkannte Anforderung aus der Norm bereits eingetragen, diese kann mit Blick auf den Kontext-Sachverhalt inhaltlich noch weiter konkretisiert werden. Durch diese Präzisierung entsteht ein Prüfkriterium, das sowohl eine Referenz zur Norm hat als auch einen inhaltlichen Bezug zur beschriebenen Arbeitssituation.

4.1.3 Auswertung

Um die Auswertungsobjektivität von Kontext-Szenarien zu sichern, werden dem Analytiker Leitfragen zur Ableitung von Anforderungen und Prüfkriterien an die Hand gegeben.

Folgende Klassen von Anforderungen werden unterschieden:

- Anforderungen an die Gestaltung der Arbeitsaufgabe (Aufgabenerfordernisse).
BEISPIEL: Die Arbeitsperson soll Rückmeldung über das Ergebnis der ausgeführten Tätigkeit (Wahrnehmungsaufgabe, Arbeitstätigkeit) bekommen.
- Anforderungen an den Durchführungsrahmen der Arbeitsaufgabe (Organisationserfordernisse).
BEISPIEL: Es soll möglich sein, Hilfe durch Kollegen zu bekommen.
- Anforderungen an die vom Benutzer ausgeübte Tätigkeit (Kognitionserfordernisse).
BEISPIEL: Der Benutzer soll den Inhalt der Fehlermeldung verstehen können.
- Anforderungen an das Dialogsystem (Dialoganforderungen / Produktanforderungen).
BEISPIEL: Fehlermeldungen sollen in der Fachsprache des Benutzers, nicht jedoch in der des Systementwicklers formuliert sein.

ANMERKUNG: In der Regel lassen sich Kognitionserfordernisse in Form von Dialoganforderungen / Produktanforderungen formulieren, wenn Tätigkeitsmerkmal (z.B. verstehen) und Produktmerkmal (z.B. Text der Fehlermeldung) korrespondieren.

Neben den Klassen von Anforderungen soll der Prüfer berücksichtigen, dass jede Prüfung bestimmte Anforderungen an den Nutzungskontext (DIN EN ISO 9241-11) als erfüllt voraussetzen muss, da sonst kaum mit einer zufriedenstellenden Nutzung des Dialogsystems zu rechnen ist. Die Gebrauchstauglichkeit eines Softwareprodukts hängt von den Merkmalen des Nutzungskontexts ab. Werden wichtige Merkmalsanforderungen verletzt, so kann die hierauf beruhende Unzufriedenheit der Benutzer nicht dem Produkt angelastet werden. Gleichwohl sind festgestellte Verletzungen der nachstehenden Anforderungen stets dem Anwender eines Softwareprodukts mitzuteilen, da diese eine gebrauchstaugliche Nutzung des Produkts mit großer Wahrscheinlichkeit verhindern oder anzeigen, dass die Nutzung des Produkts nicht in Übereinstimmung mit gesetzlichen Anforderungen ist (BildscharbV, 1996). Nachstehende Anforderungen an den Nutzungskontext eines Produkts werden vorausgesetzt:

- Ist die Arbeit so organisiert, dass der Benutzer eine vollständige Tätigkeit ausführen kann (im Sinne von DIN EN ISO 9241-2, 1993) ?
- Sind aus dem Nutzungskontext erforderliche Möglichkeiten der Zusammenarbeit oder gegenseitigen Hilfe gegeben?
- Sind die Benutzer über eventuell vorhandene Mechanismen zur Leistungssteuerung oder Leistungskontrolle informiert?

Bevor der Prüfer eine Konformitätsprüfung mit DIN EN ISO 9241-10 unternimmt, sollte anhand des Kontext-Szenarios untersucht werden, ob und inwieweit bestimmte grundlegende Anforderungen an die Nutzung des Dialogsystems erfüllt sind. Nachstehende Anforderungen gelten in allen Nutzungskontexten eines Softwareprodukts. Je nach den spezifischen Merkmalen des Nutzungskontexts sind diese Anforderungen als Prüfkriterien zu formulieren. Werden diese Kriterien verletzt, so erhält der Prüfer hiermit sichere Hinweise auf Ursachen für mangelnde Effektivität oder Effizienz der Nutzung.

- Ist die Software mit den vorhandenen / vorgesehenen Qualifikationserfordernissen / Hilfsmitteln der Benutzer nutzbar?
- Sind die Informationen, die zur Erledigung des jeweiligen Arbeitsschrittes notwendig sind, auf dem Bildschirm übersichtlich verfügbar, soweit sie nicht sinnvollerweise außerhalb des Bildschirms zur Verfügung stehen?
- Können eventuelle Fehler (auch zu einem späteren Zeitpunkt) durch den verursachenden Benutzer korrigiert werden?

Um in der Auswertungstabelle (Tabelle 4.1.3-1) Anforderungen in angemessener Weise zu formulieren, sollte der Prüfer beachten, dass Arbeitserfordernisse und Prüfkriterien in Form von Arbeitstätigkeiten, angestrebten Arbeitsergebnissen oder Zielen zum Ausdruck kommen. In der Formulierung sind spezifische Merkmale zur technischen Realisierung (Implementierung) zu vermeiden.

Tabelle 4.1.3-1: Anforderungen an das Dialogsystem (aus Sicht der Arbeitsaufgabe!)

Gutes Beispiel: (weil aus Sicht der Arbeitsaufgabe)	Der Benutzer muss im Dokument Anmerkungen machen können.
Schlechtes Beispiel: (weil spezifische Lösungsmöglichkeit)	Das Dokument muss Verknüpfungen zu Anmerkungsdokumenten anbieten.

Tabelle 4.1.3-2: Leitfragen zur Ableitung von Anforderungen aus Kontext-Szenarien

- Welche offensichtlichen Anforderungen ergeben sich aus den besonderen Gegebenheiten des Nutzungskontextes?
BEISPIELE:
- Wenn mehrere Benutzer gleichzeitig an demselben Dokument arbeiten müssen, so muss die Software dies unterstützen
- Eine Software, die auch zur telefonischen Auskunftserteilung genutzt wird, muss die erforderliche Information schnell genug bereitstellen.
- Welche Arbeitsschritte müssen aufgrund des im Kontextszenario gegebenen Sachverhalts durch das Produkt unterstützt werden?
- Welche Hilfsmittel müssen aufgrund von Aufgabenerfordernissen und/oder Benutzerbelangen außerhalb der Softwarefunktionalität bleiben oder vorhanden sein?
- Welche Arbeitsschritte müssen in ihrer zeitlichen Erledigung frei wählbar sein?
- Bei welchen Aufgaben muss die Software zeitliche Abfolgen oder inhaltliche Abhängigkeiten der Arbeitsschritte abbilden / erzwingen, um den Benutzer zu führen oder Fehlern vorzubeugen?
- Welche Ähnlichkeiten zwischen gegebenen Arbeitsaufgaben / Arbeitsschritten müssen aus Konsistenzgründen ähnlich in der Software gestaltet sein?
- Für welche Defizite des Benutzers muss die Software Kompensationsmittel bereitstellen? (z.B. systeminitiierte / aktive Hilfe, kontextsensitive Hilfe)

- Welche wiederkehrenden Arbeitsschritte müssen aufgrund des gegebenen Sachverhalts im Kontextszenario von der Software automatisch erledigt werden?
- Welche Einfluss-/Steuerungs-/Unterbrechungsmöglichkeiten des Benutzers in Abläufen sind erforderlich?
- Welche Ergebnisse / Teilergebnisse müssen für den Benutzer ausgegeben werden?
- Welche besonderen Feedbackfordernisse ergeben sich aufgrund des im Kontextszenario gegebenen Sachverhalts (auch in Hinblick auf Benutzerfehler oder um die Übersicht zu behalten) ?
- ...

Die vorstehenden Leitfragen sind in jeder Auswertungssituation anzuwenden. In den Auswertungsrahmen (Anhang B) sind diese Leitfragen eingearbeitet, und zwar in die Fragen nach den Erfordernissen und/oder den Anforderungen an die Software. Im Beispielszenario (Anhang B) kann nachvollzogen werden, wie mit Blick auf den konkreten Inhalt eines Kontext-Szenarios die Prüfkriterien zustande kommen.

4.1.4 Validität von Kontext-Szenarien

Daten des Nutzungskontexts sind valide, wenn sie die Sachverhalte korrekt beschreiben. Die befragte Person sollten das vom Analytiker entworfene Kontext-Szenario vorgelegt bekommen, damit Korrekturen oder Ergänzungen eingetragen werden können. Sind die zu erfassenden Sachverhalte für den Analytiker schwierig zu verstehen, so kann die Validität der Aufzeichnungen leiden. Es empfiehlt sich, mehrere Analytiker einzusetzen und mehrere Personen zu befragen, damit ein aus mehreren Quellen entstehendes valides Kontext-Szenario entstehen kann.

4.2 Erhebung von Use-Szenarien

Zunächst ist eine Abgrenzung zwischen Kontext-Szenario und Use-Szenario notwendig. Während im Kontext-Szenario die reale Benutzung des Dialogsystems keine Beachtung findet (lediglich die gewünschte, erwartete Nutzung wird thematisiert),

wird im Use-Szenario die tatsächliche Tätigkeit am Dialogsystem beschrieben. Jede der im Kontext-Szenario aufgelisteten "key tasks" wird als Use-Szenario beschrieben. Beide Typen von Szenarien ergänzen sich somit.

Wenn für jede der im Kontext-Szenario genannten "key tasks" ein Use-Szenario aufgezeichnet werden soll, so kann dies sehr aufwendig werden. Es stellt sich die Frage, wozu alle unkritischen Dialogschritte zu beschreiben sind, wenn letztlich nur die kritischen interessieren; denn nur die kritischen Schritte weisen auf Nutzungsprobleme und auf mögliche Normabweichungen hin. Besonders aufwendig wird die Aufzeichnung, wenn man auch noch alle möglichen Abzweigungen, Umwege oder Fehlersituationen für jede "key task" beschreibt. In der Praxis hat sich herausgestellt, dass man für die unkritischen Dialogschritte mit einer Kurzaufzeichnung in der Form von Use-Cases (Jacobson et al., 1992) auskommt, d.h., es werden nur kurz der Fluss des Dialogablaufs in Interaktionssequenzen (Benutzeraktion und Systemaktion) festgehalten. Kritische Dialogschritte (als "critical incidents" bezeichnet) werden ausführlich in der Form eines Szenarios beschrieben. Ferner hat es sich in der Praxis bewährt, vollständige Use-Szenarien nur in Entwicklungsprojekten aufzunehmen. Geschieht dies im Prototyping-Prozess, so kann das Use Szenario als eine Dokumentation des validierten Prototypen aus der Benutzersicht dienen. Am Arbeitsplatz des Benutzers, wo es nur auf die Dokumentation der Nutzungsprobleme ankommt, reicht eine Aufzeichnung in der Form von Use-Cases, angereichert mit Szenarien für kritische Dialogschritte.

4.2.1 Use-Case angereichert mit Use-Szenarien

Die Praxis zeigt, dass der Prüfer ohnehin nur auf die Untersuchung vermuteter Normverletzungen ausgerichtet ist, so dass es ausreicht, Use-Szenarien nur für solche kritischen Fälle zu schreiben, in denen man bei der Erhebung auf vermutete Normverletzungen stößt. Würde man jedoch die Beschreibung von Use-Szenarien auf isolierte Mängel- oder Fehlerszenarien beschränken, so könnte man u.U. ihre Einbettung in eine vollständige Tätigkeit nicht immer nachvollziehen. Als Kompromiss hat sich bewährt, die vollständige Tätigkeit in Form eines "Use-Case" zu dokumentieren

und diesen an Stellen mit vermuteten Mängeln (oder Fehlern) in Form eines Use-Szenarios anzureichern (siehe Tabelle 4.2.1-1). Bei der Validierung des vom Analytiker beschriebenen Use-Case geht der Benutzer die zwar etwas spröde, aber immer noch nachvollziehbar beschriebene vollständige Tätigkeit durch und wird angeregt, die ausführlicher dargestellte Mangelsituation genauer zu lesen, da sie in Form eines Szenario-Fragments beschrieben ist. Die Beschreibung als Use-Case hat auch den Vorteil, dass die zur Erledigung von "key tasks" genutzten Funktionen differenzierter beschrieben werden, als dies in einem Use-Szenario nötig wäre; auf diese Weise werden Daten für die Prüfung der Effektivität miterfasst. Das Dokument, das aus einer Mischung von Use-Case und Fragmenten von Use-Szenarios besteht, soll "Use-Szenario" heißen, weil ein Use-Case als eine Kurzform des Szenarios angesehen wird. Die kritischen Use-Szenarios sollten mit "screen shots" angereichert werden, um die problematischen Merkmale des Dialogsystems besser zu veranschaulichen.

Aufgabe in Teilaufgaben zerlegt	Beschreibung der Interaktion in Form eines Use-Case		Anforderung oder zu klärende Frage
	Benutzeraktion	Aktion des Dialogsystems oder Reaktion	
Aufgabe			
- Teilaufgabe	Benutzertätigkeit	Systemreaktion	
- Teilaufgabe	Benutzertätigkeit	Systemreaktion	
- Teilaufgabe	Critical Incident: Fragmentarische Beschreibung der Benutzertätigkeit	Systemreaktion	evtl. noch zu klärende Frage oder Hinweis auf vermutete Normabweichung
usw.			

Tabelle 4.2.1-1 Am Produkt festgestellte Eigenschaften (ausführbare Tätigkeit oder Merkmal), werden als critical incidents im Use-Case erfasst und in

einem Use-Szenario-Fragment ausführlich beschrieben einschließlich der Feststellungen über vermutete Abweichungen angesichts von Anforderungen.

Zur Dokumentation eines typischen Use-Szenarios siehe Anhang C.

4.2.2 Erhebung und Prüftätigkeit in einem Arbeitsgang

Während der Erhebung von Use-Cases und kritischen Use-Szenarios verfolgt der Analytiker den tatsächlichen Tätigkeitsablauf am Bildschirm des Benutzers. Dabei hat der Analytiker das Wissen aus dem Kontext-Szenario "im Hinterkopf". Um den Erhebungs- und Prüfaufwand zu reduzieren, kann der erfahrene Analytiker die beobachtete Interaktion als Use-Case notieren, bis eine kritische Situation eintritt, für die ein Use-Szenario zu erheben ist; diese Erhebung geht einher mit einer (vorläufigen) Prüfung der auffälligen Merkmale oder Tätigkeiten. Dabei werden diese Merkmale oder Tätigkeiten als vermutete Normabweichungen gekennzeichnet. Es reicht zunächst aus, die vermutlich verletzten Dialogprinzipien zu notieren. Tabelle 4.2.1-1 illustriert diesen Arbeitsgang: Mit Blick auf das Produkt und unter Berücksichtigung der aus dem Tätigkeitsablauf und dem Kontext-Szenario abgeleiteten Anforderungen sieht der erfahrene Prüfer sofort, wo "etwas faul ist". Diese Fälle nimmt er zum Anlass, sich den Tätigkeitsablauf und die Systemmerkmale genauer anzuschauen, und als Use-Szenario zu beschreiben.

Wie bereits erwähnt, ist es weniger wichtig, ein Kontext-Szenario schon im ersten Ansatz vollständig zu erfassen, als den Benutzer flüssig reden zu lassen und ihn non-direktiv durch weitere Fragen zum Weiterreden zu veranlassen. Die Erfahrung hat gezeigt, dass man spätestens beim Erheben der Use-Cases und der fragmentarischen Use-Szenarios Lücken schließen kann, die noch im Kontext-Szenario bestanden haben mögen.

4.2.3 Validität von Use-Szenarios

Die dokumentierten fragmentarischen Mangelszenarios ("critical incidents") könnten subjektiv sein. Man sollte deshalb die Dokumentation dieser Szenarios nicht von einem

einzelnen Benutzer abhängig machen. Mehrere Benutzer sollten die sachliche Richtigkeit der "critical incidents" bestätigen. Zudem sollte jedes "critical incident" durch teilnehmende Beobachtung (u.U. sogar durch Video-Konfrontation) erfasst werden. Video-Konfrontation hätte den Vorteil, mehrere Benutzer mit derselben Benutzungssituation zu konfrontieren, um diese kommentieren zu lassen, damit Zweifel an der Objektivität ausgeräumt werden. Der Beobachter muss auch berücksichtigen, dass sich der beobachtete Benutzer in der Rolle eines "ausgewählten" Benutzers befindet und deshalb vielleicht Anstrengungen macht, über "critical incidents" elegant hinwegzukommen, um einen guten Eindruck von seinem Können zu hinterlassen.

4.2.4 Instruktion zur Erhebung von Use-Szenarien

Eine gesonderte Liste mit Leitfragen zur Erhebung von Use-Szenarien wird als überflüssig angesehen. Da die Arbeitsschritte während der Beobachtung des Dialogablaufs erfasst werden, wird mit hoher Objektivität der Datenerfassung gerechnet. Da die "critical incidents" ausführlich beschrieben werden und von mehreren Benutzern zu validieren sind, wird keine Objektivitätsminderung der Daten befürchtet. Im Unterschied zum Kontext-Szenario, das von zwei Interviewern erhoben werden sollte, kann ein Use-Szenario in der Regel von einem Beobachter erhoben werden. Da das Use-Szenario als eine Ergänzung des Kontext-Szenarios anzusehen ist, wird empfohlen, bei der Erhebung des Use-Szenarios unklare Stellen im Kontext-Szenario nachzubessern. Der Beobachter soll den Benutzer (die Benutzer) zu Beginn der Datenerhebung fragen, welches Modell sie von der Nutzung des Systems haben, etwa nach ihren Vorstellungen vom Aufbau des Systems fragen. Neben der Erfassung des normalen Ablaufs einer vollständigen Tätigkeit soll gefragt werden, wie häufig diese Tätigkeit vorkommt, wie wichtig sie im Unterschied zu anderen ist (falls solche Fragen nicht schon im Kontext-Szenario beantwortet sind). Kritische Dialogschritte sind gründlich zu untersuchen, auch solche, die der Benutzer gar nicht als kritisch bemerkt, etwa Fehlerfälle, Umwege, spezielle Nutzungsprobleme. Auch ist zu vermerken, wenn ein Tätigkeitsablauf nicht den Erwartungen entspricht, die sich der Prüfer aus dem Kontext-Szenario gebildet hat. Es ist ratsam, bei der Datenerhebung lieber ein vermutetes "critical incident" mehr zu erfassen als zu wenig; denn in unklaren Nutzungssituationen lohnt es

sich, diese erst einmal zu erfassen, mit besonderem Fragezeichen zu versehen und in Ruhe auszuwerten. Verzweigungen im Ablauf der Bearbeitung sind gesondert zu erfassen. Am Ende soll ein Use-Case zusammen mit dem Benutzer noch einmal im Schnellgang durchgesehen werden.

4.3 Erhebungs- und Auswertungsobjektivität

Weil DIN EN ISO 9241-10 keine Konformitäts-Klausel enthält, ist die Anwendung einer Prüfmethodik grundsätzlich Vereinbarungssache zwischen den Partnern (Hersteller und Anwender). Eine fatale Folge der mangelnden Konformitätsklausel ist, dass unterschiedliche Konformitätsmethodiken zu unterschiedlichen Konformitätsergebnissen führen werden und somit die Norm ad absurdum geführt wird. Bis zur Veröffentlichung des vorliegenden Prüfverfahrens war kein Verfahren bekannt, das für Normkonformitätsprüfungen geeignet ist, obwohl viele Verfahren Bezug auf Teil 10 der Norm nehmen und damit den Anschein einer Normkonformitätsprüfung erwecken. Das Anliegen der Objektivitäts-Untersuchungen ist es, den Nachweis zu führen, dass der Teil 10 der Norm mittels systematisch erarbeiteter Prüfgrundlagen für objektive Normkonformitätsprüfungen zugänglich ist.

Wenn mit der vorgelegten Prüfmethodik dieses Ziel nunmehr erreicht werden soll, so müssen strenge Maßstäbe an die Gütekriterien und die Handhabbarkeit gesetzt werden, damit Hersteller und Anwender davon überzeugt werden können, sich auf dieser Grundlage darauf zu einigen, das Prüfverfahren anzuwenden. Zur Handhabbarkeit des Verfahrens sei auf Abschnitt 4.4 verwiesen.

Zunächst steht die Objektivität der Prüfmethodik im Blickpunkt, und zwar unter folgenden Aspekten: 1. Erhebung der Prüfgrundlage (Erhebungsobjektivität) und 2. Ableitung der Prüfkriterien (Auswertungsobjektivität). Prüfkriterien, sofern sie nicht als kontext-neutrale Produktmerkmale definiert worden sind, müssen stets eine Rechtfertigung in Aufgaben- und Organisationserfordernissen besitzen. Die Objektivität der Prüfmethodik wird in mehreren Schritten nachgewiesen, und zwar die Objektivität der Erhebung von Daten des Nutzungskontexts (vgl. Abschnitte 4.3.1 bis 4.3.3) und die

Objektivität der Ableitung von Aufgaben-Erfordernissen sowie Anforderungen an das Dialogsystem (vgl. Abschnitt 4.3.4).

Die Objektivität der Erhebung von Sachverhalten des Nutzungskontexts, aus denen Anforderungen abzuleiten sind, um ein Softwareprodukt auf Normkonformität zu prüfen, ist ausschlaggebend für die Güte einer Prüfung. Man könnte sagen, die Qualität einer Prüfung steht oder fällt mit der Objektivität ihrer Prüfgrundlagen.

Objektivität ist ein Gütekriterium, das durch sorgfältige Erarbeitung der Prüfgrundlagen zu erfüllen ist. Dies ist erreichbar durch eine schrittweise Validierung und Verfeinerung der Prüfgrundlagen sowie durch Nutzung von Korrekturmechanismen, die in den Prüfprozess eingebaut sind. Grobe Objektivitätsmängel werden bereits bei der Validierung der Kontext-Szenarien ausgeräumt. Wird das Prüfverfahren in einem Entwicklungsprojekt eingesetzt, z.B. während des Usability-Prototyping, so ist absehbar, dass sich subjektive Verzerrungen in der Darstellung des Nutzungskontexts spätestens bei der Validierung der Anforderungen oder der Evaluierung von Prototypen herausstellen werden. Der Prozess der Anforderungsentwicklung und des Prototyping ist ja gerade darauf angelegt, solche Einschränkungen im Verstehen der Entwurfgrundlagen aufzudecken (siehe dialektisches Problemlösen im Entwurfsprozess, Dzida und Freitag, 1998). Auch in einer Prüfsituation am Arbeitsplatz des Benutzers können Korrekturmechanismen genutzt werden, die es gestatten, die Objektivität schrittweise zu verbessern. Die Korrekturmöglichkeiten liegen - wie schon beim Prototyping - in der Hand des Benutzers. Denn eine mögliche Normabweichung wird stets anhand der Prüf- und Beurteilungsgrundlagen zu verifizieren versucht, nämlich an der Effektivität und Effizienz der tatsächlichen Benutzertätigkeit (vgl. Abschnitt 3.3). Im Erhärtungstest (Entscheidungstabelle 2) hat der Benutzer das letzte Wort. Eine vermeintliche Normabweichung, die auf einer missverständlichen Beurteilungsgrundlage beruht, wird spätestens bei der subjektiven Beurteilung der Zufriedenstellung durch Benutzer "enttarnt". Die Vorsicht, die der Prüfer walten lässt, indem er in einem Mangelhinweis zunächst eine "vermutete" Normabweichung sieht (siehe Erhärtungstest), sowie die Zufriedenstellungsbeurteilungen der Benutzer sind Korrekturmechanismen, um Objektivitätsmängel aufzudecken.

Die Objektivität der Prüf- und Beurteilungsgrundlage kann beeinträchtigt werden, wenn die im Nutzungskontext befragte Person eine verzerrte Darstellung des Sachverhalts gibt. Deshalb ist zu empfehlen, mehrere Personen zu interviewen, wenn es im Nutzungskontext vergleichbare Arbeitsplätze gibt. Die Daten verschiedener Personen ergeben zusammen genommen ein vollständigeres und objektiveres Gesamtbild. Zusammenfassend lässt sich sagen, dass man um eine hohe Erhebungsobjektivität nicht besorgt zu sein braucht, wenn die Korrekturmechanismen des Prüfverfahrens greifen, die letztlich eine hohe Objektivität der Prüfgrundlage sichern helfen.

Ohne dass die Daten des Nutzungskontexts zutreffend erhoben wurden, kann es keine objektiv abgeleiteten Anforderungen geben. Für die Sicherung der Objektivität von Anforderungen hat die empirische Erprobung des Prüfverfahrens im Feld beigetragen. Dabei hat sich herausgestellt, dass es für Büroarbeitsplätze einen Satz von immer wieder vorkommenden, typischen Erfordernissen gibt, die sich aus dem Nutzungskontext zuverlässig ableiten lassen. Die im Feld arbeitenden Prüfer stellten fest, dass sich die Auswertung von Kontext-Szenarien in gewissem Umfang schematisieren lässt. In der Schematisierung liegt eine Gefahr und eine Chance. Die Gefahr ist, dass inhaltliche Besonderheiten eines Nutzungskontexts übersehen werden, wenn man das Schema unkritisch anwendet, worunter die Vollständigkeit der abgeleiteten Anforderungen leidet. Die Chance ist, dass die Objektivität zumindest für den vorgegebenen Auswertungsrahmen auf hohem Niveau gesichert werden kann. Darüber hinaus existieren für die Auswertungsobjektivität dieselben Korrekturmechanismen, die zur Sicherung der Erhebungsobjektivität in das Prüfverfahren eingebaut wurden. Es ist insgesamt mit einem hinreichend objektiven Verfahren zu rechnen, wenn die Instruktionen und Trainingsanleitungen sachgerecht umgesetzt werden. In den nachstehenden Abschnitten wird dargestellt, welche Anstrengungen gemacht wurden, das Prüfverfahren hinsichtlich Erhebungs- und Auswertungsobjektivität zu optimieren.

Zunächst wurde ein Versuchsplan erarbeitet (Abschnitt 4.3.1), in dem berücksichtigt wurde, dass auch das Training der Prüfer (hier in der Rolle als Analytiker des Nutzungskontexts) einen Einfluss auf die Objektivität haben könnte. Nachdem bereits bei der ersten Versuchsdurchführung klar gezeigt werden konnte, dass die Erhebungs-

methodik ohne gründliche Einarbeitung nicht einsetzbar ist, wurde die Vergleichsgruppe „ohne Training“ im Versuchsplan gestrichen und das Training der Prüfer zur Voraussetzung der Anwendung des Prüfverfahrens gemacht.

Die nachfolgenden Abschnitte zur empirischen Sicherung der Objektivität sind für Prüfer interessant, die sich in das Prüfverfahren einarbeiten möchten, da die Planung und Auswertung des Versuchs wichtige Erfahrungen widerspiegeln, die jeder Prüfer selber machen wird, wenn er oder sie sich um Objektivität bemühen.

4.3.1 Versuchsplan für Objektivitätsuntersuchungen

Erhebungsobjektivität

Zur Sicherung der Objektivität der Erhebung von Kontext-Szenarien wurden dem Analytiker Leitfragen an die Hand gegeben sowie eine Instruktion zur Einarbeitung und Anwendung.

Mit Hilfe der Leitfragen und mit Blick auf den Auswertungsrahmen sollten die Beurteilungsgrundlagen für die Ableitung von Anforderungen möglichst vollständig und sachlich zutreffend erfasst werden. Zwei unabhängig voneinander arbeitende Analytiker sollten in Form eines Kontext-Szenarios denselben Sachverhalt inhaltlich übereinstimmend beschreiben. Zunächst wurde die Objektivität der Erfassung hinsichtlich Vollständigkeit und sachlicher Korrektheit untersucht, dann hinsichtlich der Übereinstimmung zwischen den unabhängigen Analytikern.

Mitarbeiter des Projekts, in dem dieses Prüfverfahren entwickelt wurde, erhoben in einem Nutzungskontext ein Referenz-Szenario, das auf Vollständigkeit durch ein weiteres Projektmitglied und auf sachliche Richtigkeit (Validität) durch die befragte Arbeitsperson überprüft wurde. Drei Analytiker gingen in denselben Nutzungskontext und erhoben unter Anwendung der Leitfragen bei verschiedenen Arbeitspersonen und unabhängig voneinander die Sachverhalte eines Nutzungskontexts. Um die Vollständigkeit eines jeden der erhobenen Kontext-Szenarien zu beurteilen, wurden diese jeweils

mit dem Referenz-Szenario verglichen. Hierzu wurde das Referenz-Szenario pro Leitfrage in Abschnitte zerlegt. Diese Abschnitte wurden mit jedem Modul der erhobenen Kontext-Szenarien verglichen. Das Vorhandensein von Abschnitten sowie deren sachliche Übereinstimmung mit den Referenz-Abschnitten wurden beurteilt. Der Anteil der vorhandenen und sachlich richtigen Module wurde als Prozentsatz ausgedrückt. Für alle nicht oder nicht zutreffend erfassten Module wurde untersucht, ob dieser Mangel durch Verbesserung oder Ergänzung der Leitfragen behoben werden kann.

Auswertungsobjektivität

Zur Sicherung der Objektivität der Auswertung von Kontext- und Use-Szenarien wurden dem Analytiker Leitfragen an die Hand gegeben sowie eine Instruktion zur Einarbeitung und Anwendung. Ferner erhielten die Analytiker jeweils eine Liste mit typischen Aufgabenerfordernissen und Dialoganforderungen, die in Kontext-Szenarien oft vorkommen und wie Bausteine eingesetzt werden können. Es wurde jedoch betont, dass die Bausteine von Erfordernissen und Anforderungen nicht ausreichen, um eine vollständige Auswertung zu erreichen. Vielmehr sollten zusätzliche Erfordernisse aus dem Text des Kontext-Szenarios erschlossen werden, und es sollte die Norm DIN EN ISO 9241-10 zur Hand genommen werden, um weitere anwendbare Dialoganforderungen zu finden.

Mit Hilfe der Leitfragen und der vorgegebenen Anforderungs-Bausteine sollten die Anforderungen an das Dialogsystem aus den Szenarien abgeleitet werden. Wie schon für die Erhebungsobjektivität der Szenarien gefordert, sollten auch die Anforderungen vollständig und übereinstimmend abgeleitet werden können. Was bei den Szenarien an sachlicher Richtigkeit (Validität) zu beurteilen war, galt es auch bei den abgeleiteten Anforderungen zu erreichen, und zwar unter dem Gesichtspunkt der "Relevanz" der Anforderung für den betreffenden Sachverhalt des Nutzungskontexts. Die Relevanz wird im wesentlichen durch zwei Interpretationsprozesse beeinflusst, und zwar von der Interpretation der Sachverhalte durch den jeweiligen Analytiker und durch die

Anwendung des erarbeiteten Verständnisses auf die Auswahl einer als passend erkannten Anforderung aus DIN EN ISO 9241-10.

Um herauszufinden, bei welchem dieser beiden Interpretationsprozesse die Auswertungsobjektivität Mängel aufweist, wurde die Formulierung jeder Anforderung in zwei Schritten abverlangt: 1. Aufgabenerfordernis erkennen, 2. Anforderung an das Dialogsystem ableiten. Für jeden dieser Schritte wurden Listen mit vorgegebenen "Auswertungsbausteinen" von Erfordernissen bzw. Anforderungen zur Verfügung gestellt. Diese Bausteine ergaben sich aus den Erfahrungen der am Projekt beteiligten praktizierenden Prüfer. Die drei an der Erhebung beteiligten Analytiker wurden gebeten, die Szenarien auszuwerten, um deren Auswertungsergebnisse mit einer von Experten erarbeiteten Referenzliste vergleichen zu können.

Für jeden Abschnitt des Referenz-Szenarios wurden die abgeleiteten Anforderungen eines jeden Analytikers mit der Referenzliste der Anforderungen verglichen. Der Anteil der übereinstimmend abgeleiteten Anforderungen ergibt als Prozentsatz den Grad der Vollständigkeit, ein Indikator der Auswertungsobjektivität. Diese Beurteilung der "Übereinstimmung" wurde durch ein Mitglied der Expertengruppe vorgenommen, das an der Definition der Referenzliste beteiligt war. Wurden Abweichungen festgestellt, so wurde anschließend untersucht, ob diese Mängel durch Anpassung der Leitfragen oder der Instruktion abgestellt werden können.

Da die Versuchspersonen für die Auswertung der Szenarien dieselben Personen waren, die diese Szenarien erhoben hatten, sollten sie nicht ihre selbst erhobenen Szenarien auswerten, sondern das Referenz-Szenario zugrunde legen, damit für die Ermittlung der Auswertungsobjektivität vergleichbare Bedingungen vorlagen; außerdem konnten hierdurch die Schritte "Erhebung" und "Auswertung" unabhängig voneinander untersucht werden.

4.3.2 Durchführung der Untersuchung: Erhebungsobjektivität

Um die Erhebung von Daten des Nutzungskontexts mittels Kontext-Szenarien vorzubereiten, wurden Versuchspersonen (Vpn) auf ihre Tätigkeit als Analytiker

vorbereitet. Zwei der Vpn arbeiteten in einer Prüfstelle für Informationstechnik, wovon eine Person mit Fragen der Software-Ergonomie vertraut war, aber an einer Datenerhebung noch nicht teilgenommen hatte. Die andere Vp war vertraut mit Fragen der Softwarequalität aus technischer Sicht. Die dritte Vp stammte aus einer Bundesbehörde und war vertraut mit der Analyse und Behebung von Nutzungsproblemen nach Einführung von Software an Bildschirmarbeitsplätzen.

Die Datenerhebung fand an vier vergleichbaren Arbeitsplätzen des Liegenschaftsamts im Bonner Stadthaus statt. Die Zeiten für die Erhebung eines Kontext-Szenarios variierten zwischen 80 und 95 Minuten. Einer der Arbeitsplätze wurde für die Erhebung des Referenz-Szenarios reserviert. Zwei Experten aus dem Projekt-Team erhoben das Referenz-Szenario, ein dritter prüfte es auf inhaltliche Vollständigkeit relativ zu den Leitfragen. Eine Nacherhebung zum Referenz-Szenario führte zur Präzisierung der Kontext-Daten. (Das Referenz-Szenario wurde laut Versuchsplan für die Messung der Objektivität benötigt.) Die drei anderen Arbeitsplätze wurden von den ausgewählten Vpn aufgesucht. Das Referenz-Szenario ist im Anhang B als Beispielszenario nachlesbar.

Der Versuch wurde unter erschwerten Bedingungen durchgeführt. Jeder der trainierten Analytiker führte das Interview allein durch, obwohl in der Instruktion empfohlen wird, dass ein noch nicht genügend eingearbeiteter Analytiker diese Aufgabe unter Begleitung eines zweiten ausführen soll, wobei der erste Analytiker das Gespräch führt und Aufzeichnungen macht, während der zweite den Gesprächsinhalt nach Verständlichkeit und Plausibilität prüft, ggf. Zwischenfragen stellt und überwacht, ob alle Fragen der Fragenliste gemäß Zielsetzung gestellt wurden. Die erschwerten Bedingungen für die Vpn ergaben sich aus dem Umstand, dass es schwierig war, überhaupt motivierte Vpn dafür zu gewinnen, die notwendige Zeit für Training, Datenerhebung, Training der Auswertung, selbständige Auswertung und Rückfragen aufzubringen - alles in allem ein Aufwand von 2 1/2 bis 3 Tagen pro Vp.

Training der Analytiker

Das Training der Analytiker fand in zwei Abschnitten statt: 1. Datenerhebung, 2. Auswertung der Daten. Das Training der Datenerhebung fand jeweils vormittags (etwa 3 Stunden lang) am Tage der Datenerhebung in der GMD statt. Trainingsdefizite wurden nach der Datenerhebung anhand der Ergebnisse erkannt und nach Rücksprache mit den Vpn verifiziert. Schrittweise wurde das Training verbessert. Als Unterlagen für das Training dienten die Leitfragen zur Erhebung von Kontext-Szenarien, eine Instruktion zur Datenerhebung und ein Beispiel-Szenario. Infolge der aufgedeckten Trainingsdefizite wurden die Leitfragen schrittweise überarbeitet und eine kommentierte Liste der Leitfragen ausgearbeitet, in der die Zielsetzungen der Leitfragen erläutert wurden.

Das Training wurde nach folgender Anleitung durchgeführt: "Dem Analytiker (Interviewer) sind die für die Datenerhebung erforderlichen Unterlagen (Leitfragen, kommentierte Leitfragen, Instruktion und Beispiel-Szenario) auszuhändigen. Anhand dieser Unterlagen wird die Struktur der Leitfragen-Liste erläutert und am Beispiel-Szenario veranschaulicht. Abschnitt für Abschnitt des Beispiel-Szenarios wird durchgesprochen, und zwar sowohl am Textbeispiel des Beispiel-Szenarios sowie anhand der Leitfragen. Hierbei wird die Zielsetzung jeder Leitfrage erläutert. Negativ- und Positiv-Beispiele für erhobene Daten werden im Beispiel-Szenario veranschaulicht, soweit sich dies aus den Trainingserfahrungen als notwendig erwiesen hat. Anlässlich der Erläuterung der Leitfragen wird dem Analytiker immer wieder deutlich gemacht, dass die Leitfragen nicht wörtlich gestellt zu werden brauchen, da sich eine Gesprächssituation ergeben soll, die zwar einem strukturierten Interview ähnlich ist, jedoch weniger streng vom Analytiker (Interviewer) entlang der Fragenliste gesteuert werden soll. Der/die Befragte soll mehr Freiheit haben, das Gespräch selbst zu steuern, als in einer typischen Interviewsituation. Es soll z.B. geduldet werden, wenn der/die Befragte Themen wechselt, vertieft, oder anders fortführt, als es von den Leitfragen her notwendig wäre. Dem Analytiker wird gesagt, dass ein flüssiger Gesprächsverlauf kennzeichnend für die Form der Szenario-Erhebung ist. Gleichwohl sollen die

Analytiker darauf achten, dass aus jedem Abschnitt der Leitfragen sämtliche Fragen gestellt worden sind."

4.3.3 Erhebungsobjektivität: Objektivitätsmessungen und Ergebnisse

Gemäß Versuchsplan wurde die Objektivität der Erhebung von Kontext-Daten als inhaltliche Übereinstimmung mit dem Referenz-Szenario gemessen. Die Messung wurde in drei Schritten vorgenommen: 1. Vergleich der Kontext-Daten der ersten beiden Vpn mit dem Referenz-Szenario. 2. Überarbeitung der Leitfragen aufgrund der Auswertungsergebnisse des Schrittes 1, Änderung des Objektivitätsmaßes und 3. erneute Messung wie unter Schritt 1. Das im Versuchsplan festgelegte Objektivitätsmaß wurde geändert, weil erkannt wurde, dass nicht alle Kontext-Daten mit den Daten des Referenz-Szenarios übereinstimmen mussten, da es bei einigen der gestellten Leitfragen möglich war, von der befragten Person eine zwar subjektive, aber inhaltlich zutreffende Antwort zu bekommen. Wenn beispielsweise nach fehlenden Hilfsmitteln gefragt wurde, so kam es vor, dass die Befragten subjektiv unterschiedliche Hilfsmittel vermissten. Im dritten Schritt wurde mittels überarbeiteter Leitfragen die letzte der möglichen Datenerhebungen im Stadthaus durchgeführt und erneut gemessen, in der Annahme, dass nach Verfeinerung der Leitfragen und verbessertem Training ein besseres Objektivitätsergebnis erzielt werden konnte.

Die Objektivität im ersten Schritt der Messung betrug für die beiden ersten Vpn im Mittel 50%. Nach Verfeinerung der Leitfragen und Änderung des Objektivitätsmaßes betrug die Objektivität bei der ersten Vp 77%, bei der zweiten 59%. Die Objektivität betrug im dritten Schritt der Messung 85%. Die Übereinstimmungen zwischen den Vpn betrugen 30%, 48% und 59%. Die geringen Übereinstimmungen sind sowohl darauf zurückzuführen, dass unterschiedliche Vpn unterschiedliche Fehler bei der Erhebung der Kontext-Daten machten, als auch von den Befragten unterschiedliche Auffassungen von den Sachverhalten im Nutzungskontext mitgeteilt wurden. Hieraus ergibt sich die zwingende Folgerung, auf ein und denselben Nutzungskontext mehrere Analytiker anzusetzen, um ergänzende Kontext-Daten erheben zu können. Hiermit werden einerseits vollständigere Daten erhoben, andererseits Unterschiede oder gar Wider-

sprüche in den Daten aufgedeckt, weil die befragten Arbeitspersonen eine subjektive Auffassung von Sachverhalten im Nutzungskontext haben (und haben dürfen).

Aus den Objektivitätsmessungen wurden folgende Schlüsse gezogen:

- Die als Übereinstimmung der von den Vpn erhobenen Daten mit dem *Referenz*-Szenario gemessene Objektivität ist die inhaltlich aussagefähigere Messung in bezug auf die Güte der Leitfragen.
- Die Übereinstimmung zwischen den Vpn ist nur ein zusätzlicher Indikator für Objektivität, da ein Fehler-Anteil enthalten ist, der nicht auf die Leitfragen rückführbar ist.
- Mangelnde Übereinstimmung zwischen den Daten der Analytiker kann auch aus subjektiv variierenden Darstellungen der Befragten entstehen und ist dann kein Mangel der Erhebungsmethode.
- Mangelnde Übereinstimmungen zwischen Analytikern sowie subjektiv variierende Darstellungen der Befragten machen Mehrfacherhebungen im Nutzungskontext notwendig.
- Die Instruktion zur Erhebung von Kontext-Daten ist aufgrund dieser Erfahrungen anzupassen.

Aufgrund der Ergebnisse der dritten Messung wurden noch drei Schwachstellen in der Liste der Leitfragen festgestellt: eine Frage wurde umformuliert, zwei weitere wurden hinsichtlich der Zielsetzung kommentiert. Trotz der bisher erreichten Optimierung wurde erwartet, dass die Erhebungsmethode weiter optimiert werden kann. Dies liegt an der persönlichen Eignung eines Analytikers für diese Aufgabe. Die Vpn waren durchaus unterschiedlich für die Analysetätigkeit geeignet oder vorbereitet. Die erste Vp war vermutlich wegen des software-ergonomischen Hintergrundwissens besser vorbereitet als die zweite. Die dritte war wegen der kommentierten Leitfragen und des überarbeiteten Trainings besser vorbereitet, hatte aber vielleicht auch wegen der Berufserfahrungen in einer Behörde einen inhaltlich leichteren Zugang zu den Aufgaben im Stadthaus. Zudem ist aufgefallen, dass es vorteilhaft ist, wenn ein Analytiker darin geübt ist, Interviews oder Sitzungsprotokolle zu führen. Die Tätigkeit

des Zuhörens und des fast simultanen Mitschreibens der vernommenen Inhalte liegt nicht allen Menschen gleichermaßen.

4.3.4 Auswertungsobjektivität

Die Einarbeitung der Vpn in die Auswertung von Kontext-Szenarien geschah nach der Datenerhebung an einem anderen Ort, so dass zwischen Datenerhebung und Training eine Pause von ca. einer Stunde lag. Die beiden erstgenannten Vpn wurden noch einmal gesondert instruiert, konnten aber wegen hoher Arbeitsbelastung und Urlaub die Auswertungsarbeiten nicht sofort aufnehmen. Bei der letzten Vp betrug der Abstand zwischen Training und durchgeführter Auswertung nur 5 Tage. Hieran zeigt sich, wie schwierig die Umstände der Versuchsdurchführung waren.

Training der Analytiker

Um die Einarbeitung in die Datenauswertung zu erleichtern, bekam der Analytiker das ausgewertete Beispiel-Szenario ausgehändigt. Jeder der Auswertungsschritte wurde anhand des Beispiel-Szenarios erläutert. Hierbei wurden auch die Auswertungshilfen sowie der Teil 10 der Norm zu Hilfe genommen.

Die Anweisung zum Training lautete wie folgt: "Aus den Aufzeichnungen über den Nutzungskontext ist möglichst am Tag nach der Datenerhebung eine Reinschrift anzufertigen, bevor die frischen Gedächtnisspuren verblassen. Die Ziffern der Leitfragen sollen genutzt werden, um zunächst die Mitschrift zu ordnen. Dabei fällt im allgemeinen auf, ob eine der Fragen übersehen wurde oder welche Inhalte noch vervollständigt werden müssen. Zu diesem Zweck sind telefonische Nachfragen bei dem Befragten unerlässlich. Die Reinschrift soll in einem erzählenden Stil verfasst werden, und zwar in der Sprache der befragten Person, so dass es dieser Person Spaß macht, die Geschichte über die eigene Arbeit einmal in einem Zuge lesen zu können. Die Reinschrift soll dem Befragten zur Korrektur oder Vervollständigung gegeben werden (Inhaltsvalidierung).

Die Auswertung der Reinschrift wird in drei Schritten vollzogen: 1. Ableitung von Aufgaben-Erfordernissen, 2. Zuordnung von Dialogprinzipien zu den Erfordernissen und 3. Ableitung von Anforderungen an das Dialogsystem zu jedem der Erfordernisse. Für jeden der Auswertungsschritte werden Auswertungshilfen gegeben, die jedoch nicht wie Schablonen anzuwenden sind, obwohl die vorgegebenen Erfordernisse und Dialoganforderungen an typischen Büroarbeitsplätzen häufig vorkommen. Insbesondere bei der Präzisierung der Dialoganforderungen sollen die passenden Anforderungen in der Auswertungshilfe mit Blick auf den Nutzungskontext präzisiert werden (Definition des Prüfkriteriums). Bei der Präzisierung soll im Teil 10 der Norm unter dem gegebenen Dialogprinzip die passende, relevante Anforderung gesucht werden. Wenn die Präzisierung dieser Anforderungen mit Blick auf die Daten des Nutzungskontexts plausibel erscheint, hat man mit Sicherheit ein passendes Prüfkriterium gefunden. Ob es "das" passende ist, hängt davon ab, ob noch andere Dialoganforderungen aus demselben Sachverhalt des Nutzungskontexts plausibel ableitbar sind. Wenn es ungewohnt ist, die Norm zu lesen, so hilft es, sich je nach den Kontext-Daten die notwendige Interaktion des Benutzers mit dem Dialogsystem vorzustellen."

Objektivitätsmessungen und Ergebnisse

Die Objektivität der Auswertung wurde hinsichtlich der Ableitung der Erfordernisse und der Dialoganforderungen getrennt gemessen. Hiermit sollte kontrolliert werden, ob Auswertungsfehler bereits bei der Ableitung von Aufgabenerfordernissen entstehen.

Die erste Auswertung ergab eine Objektivität von 51% für die Ableitung der Erfordernisse und eine Objektivität von 49% für die Ableitung der Dialoganforderungen. Obwohl die Höhe der Objektivität bei weitem nicht ausreicht, kann doch festgestellt werden, dass für die meisten der richtig abgeleiteten Erfordernisse auch die zutreffenden Dialoganforderungen gefunden wurden. Insoweit hat sich die schrittweise Ableitung der Anforderungen über die Erfordernisse und das Dialogprinzip bewährt; nur die Vollständigkeit ließ noch zu wünschen übrig.

Auf Grund der Erfahrungen mit der ersten Objektivitätsmessung wurden die Auswertungsbausteine gründlich überarbeitet. Weitere Objektivitätsmessungen wurden auf der verbesserten Auswertungsgrundlage durchgeführt, in der Erwartung, dass die Objektivität steigt. Durch iterative Verbesserungen der Auswertungsbausteine konnte die Auswertungsobjektivität schrittweise auf etwa 70% gesteigert werden. Mit diesem Ergebnis konnten wir allerdings nicht zufrieden sein.

Sicherung der Auswertungsobjektivität durch einen Auswertungsrahmen

Nach den unbefriedigenden Objektivitätsuntersuchungen wurde ein Vorschlag praktizierender Prüfer umgesetzt, die festgestellt hatten, dass man das Auswertungsverfahren schematisieren kann, um wenigstens für einen Rahmen von vorgegebenen Anforderungen eine hundertprozentige Objektivität zu sichern. Dieser Auswertungsrahmen (siehe Anhang B) soll zusammen mit der Instruktion und der Trainingsanleitung die künftige Grundlage für die Auswertung von Kontext-Szenarien sein. Der Auswertungsrahmen ermöglicht es, bereits die Erhebung der Sachverhalte eines Nutzungskontexts vom Auswertungsschema her anzugehen. Darüber hinaus ist der Analytiker bei der Validierung der Daten des Nutzungskontexts gezwungen, die inhaltliche Darstellung des Szenarios zu den einzelnen Leitfragen so zu präzisieren, dass die vorgegebenen Erfordernisse und Anforderungen plausibel ableitbar sind. Auf diese Weise wird die Qualität der erhobenen Kontextdaten verbessert. Nur wenn der im Nutzungskontext festgestellte Sachverhalt inhaltlich nicht mit dem vorgegebenen Auswertungsschema in Verbindung gebracht werden kann, soll der Analytiker die besonderen Erfordernisse und Anforderungen festhalten. Wenn Zweifel an der Objektivität der Daten bestehen, soll an vergleichbaren Arbeitsplätzen die Erhebung solange wiederholt werden, bis es aufgrund der Ergänzungen im Kontext-Szenario zu einer inhaltlichen Sättigung kommt. Der Auswertungsrahmen legt ein Minimum an abzuleitenden Erfordernissen und Anforderungen vor, die erfahrungsgemäß in einem typischen Büro vorkommen. Somit kann man wenigstens für dieses Minimum eine hohe Erhebungs- und Auswertungsobjektivität erreichen.

4.4 Handhabbarkeitsprüfung

Mit der Handhabbarkeitsprüfung sollte neben der Ökonomie des Prüfverfahrens auch die Nützlichkeit in realen Projekten nachgewiesen werden. Anhand von Zeitaufwands-Einschätzungen sollte kalkuliert werden können, was der Einsatz des Prüfverfahrens kostet. Hierbei ist zu unterscheiden zwischen dem Einsatz des Verfahrens am fertigen Produkt und in einem Entwicklungsprojekt, in dem Prototypen zur Verfügung stehen. Mit der Integration des Prüfverfahrens in das DEKITZ-Prüfhandbuch (Abschnitt 4.4.1) wird der Einsatz am fertigen Produkt geprüft. Mit der Erprobung des Prüfverfahrens in einem großen Entwicklungsprojekt wird die Handhabbarkeit auch für die Anforderungs-Entwicklung und das Prototyping untersucht (Abschnitt 4.4.2). Diese Erprobung war im ursprünglichen Projektplan nicht vorgesehen. In der Praxis ist aber diese Art der Erprobung mindestens so wichtig wie die Erprobung des Prüfverfahrens an fertigen Produkten, denn die Tauglichkeit des Prüfverfahrens für Entwicklungsprojekte wurde im Rahmen der Arbeit der DEKITZ-Arbeitsgruppe zunehmend gefordert.

4.4.1 Integration in das DEKITZ-Prüfverfahren

Die Akzeptanz des Prüfverfahrens in den akkreditierten Usability-Prüflaboratorien kann als Indikator für Handhabbarkeit gewertet werden. Zu diesem Zweck wurde der Entwurf des bisher erarbeiteten Verfahrens in einer Arbeitsgruppe der Deutschen Akkreditierungsstelle für Informations- und Telekommunikationstechnik (DEKITZ) beraten und im Juni 1999 als Version 1.0 für die Prüfstellen freigegeben mit der Empfehlung zur Anwendung und mit der Bitte um kritische Rückmeldung. Bereits während der Beratungen in der DEKITZ-Arbeitsgruppe haben die beteiligten Prüfstellen Vorschläge zur Abfassung der Version 1.0 des DEKITZ-Prüfhandbuchs gegeben, aus denen Rückschlüsse auf eine mögliche Akzeptanz des Prüfverfahrens gezogen wurden. Mit der Einbindung von DEKITZ-akkreditierten Prüflaboratorien in die Handhabbarkeitsprüfung gelang eine rasche Umsetzung der Projektergebnisse in die Prüfpraxis.

4.4.2 Erprobung in einem Großprojekt

Die gemessene Objektivität der Prüfergebnisse war in einem Entwicklungsprojekt schlechter als bei einer Prüfung am fertigen Produkt. Dies wurde darauf zurückgeführt, dass die Use-Szenarien für noch unfertige Produkte mehr kritische Nutzungssituationen (critical incidents) aufweisen, als dies bei fertigen Produkten vorkommt. Kritische Nutzungssituationen sind bei fertigen Produkten augenscheinlich und besser reproduzierbar. Prototypen hingegen verhalten sich oft unberechenbar, weil das System noch nicht stabil genug ist. Deshalb fand die Prüfung der Objektivität der Erhebung von Use-Szenarien bei Prototypen unter erschwerten Bedingungen statt. Geringere Objektivität ist z.B. darauf zurückzuführen, dass die Erhebung durch noch unzulängliche Systembedingungen gestört wird, die bei dem einen Benutzer zum Problem führen, bei einem anderen nicht, weil die Störung nicht auftrat.

Bei der Entwicklung eines Prüfverfahrens für Software gilt es nicht nur zu gewährleisten, dass die zu überprüfenden Sachverhalte reliabel und valide beurteilt werden, sondern es muss zusätzlich darauf geachtet werden, dass das Verfahren handhabbar ist. Handhabbarkeit bedeutet in diesem Zusammenhang, dass die Anwendung des Verfahrens einerseits kosteneffizient ist, und andererseits, dass das Verfahren auch von den beabsichtigten Anwendern akzeptiert wird. Das Verfahren erfüllt diese beiden Kriterien aus folgenden Gründen: Das Verfahren ist kosteneffizient, da pro Benutzergruppe, die eine Tätigkeit mit mittlerer Komplexität ausführt, lediglich etwa fünf Personentage vom Erstkontakt bis zum abschließenden Prüfbericht notwendig sind. Der Prüfbericht liefert qualitative, d.h. inhaltlich beschriebene Mängel, so dass es möglich ist, gezielt und effizient die konkreten Mängel des Produktes zu beheben. Jeder Mangel wird aufgabenbezogen beschrieben, so dass es ohne weiteres möglich ist, eine direkte und gezielte Mängelanalyse zu wiederholen. Ein Beispiel für eine solche qualitative Aussage könnte sein: "Bei der Sachbearbeitung ist es nicht möglich, in der Übersicht auf dem Bildschirm die Priorität der Aktenbearbeitung zu erkennen, obwohl dies für die vorhandene Arbeitsorganisation erforderlich wäre."

In diesem Punkt hat das Prüfverfahren einen deutlichen Vorteil gegenüber anderen Verfahren, die vorgeben, eine Konformitätsprüfung mit der DIN EN ISO 9241-10 zu ermöglichen. Die bisherigen Verfahren liefern lediglich quantitative Aussagen über die Einhaltung der Dialogprinzipien. Eine quantitative Aussage wäre zum Beispiel "Das Produkt wird auf einer Skala von 1-7 bezüglich der Selbstbeschreibungsfähigkeit mit 4 benotet". Eine solche Aussage ist zur Mängelanalyse und anschließenden Verbesserung des Produktes nicht unmittelbar verwertbar und erfordert weiteren Analyseaufwand, der zwangsläufig weitere Kosten mit sich bringt.

Die Akzeptanz des Prüfverfahrens ergibt sich einerseits aus dem Nutzen qualitativ beschriebener Mängel für Korrekturmaßnahmen im Entwicklungsprozess, andererseits enthält es etablierte Analyse-Konzepte des Requirements-Engineering (Use-Cases und Szenarien). Das Verfahren dient somit in Softwareprojekten nicht nur der Konformitätsprüfung mit DIN EN ISO 9241-10. Darüber hinaus werden die Dokumente und Ergebnisse, die im Laufe des Prüfprozesses geliefert werden, für die Qualitätssicherung genutzt:

- die Aufgabenanalyse, die mit Hilfe der Szenarien durchgeführt wird, ist ein wesentlicher Beitrag zum Requirements-Engineering und ermöglicht die Erstellung eines von Benutzern validierten Fach- oder Nutzungskonzeptes;
- die aus dem Fach- oder Nutzungskonzept abgeleiteten Anforderungen können sehr früh im Entwurfsprozess daraufhin geprüft werden, ob sie normkonforme Anforderungen darstellen;
- somit kann das Verfahren als eine frühzeitig einsetzbare Maßnahme zur Qualitätssicherung aus Benutzersicht eingesetzt werden.

Bei Handhabbarkeitsuntersuchungen in Projekten des TÜV Rheinland konnte eine hohe Akzeptanz insbesondere bei Projektverantwortlichen festgestellt werden, da der Nutzen für die frühen Projektphasen erkannt worden ist. Projektmitarbeiter konnten schnell Konsens über die Einschätzung der aufgedeckten Mängel erlangen, da durch Verwendung der Szenarien die Mängel anschaulich und aufgabenbezogen nachvollziehbar waren und hinsichtlich ihrer Auswirkung im Dialog mit dem System bewertbar sind.

5 Bewertung durch Benutzer

In den voranstehenden Kapiteln ist wiederholt dargestellt worden, dass eine Normkonformitätsprüfung methodisch fragwürdig ist, wenn Benutzer an der Erarbeitung der Prüf- und Beurteilungsgrundlagen nicht beteiligt worden sind. Kurzum: Eine Prüfung allein durch Experten ist nicht ausreichend, um Non-Konformität festzustellen. Das mit diesem Bericht vorgelegte Prüfverfahren verfolgt deshalb das Ziel, mittels Befragung der Benutzer über ihre Aufgaben und Arbeitsbedingungen im Nutzungskontext eines Produkts valide Beurteilungsgrundlagen zu schaffen. Außerdem sind Benutzer über ihre subjektive Zufriedenstellung zu befragen; denn die Bewertung eines Mangels durch die Benutzer kann den Ausschlag dafür geben, eine Produkteigenschaft als bedeutsame Abweichung von der Norm ansehen zu müssen (siehe Erhärtungstest, Abschnitt 3.3). Bei allem Vertrauen in die Normkonformität eines Produkts sind es die Benutzer, die am besten beurteilen können, ob dieser Vertrauensvorschuss berechtigt ist, denn erst in der täglichen Nutzung am Arbeitsplatz bewährt sich die Gebrauchstauglichkeit eines Produkts.

Soll man es der Initiative der Benutzer überlassen, sich zu Worte zu melden, wenn sie mit Nutzungsproblemen kämpfen? Soll man warten, bis sich Benutzer über Mängel der Software beschweren? Benutzer neigen dazu, mit Nutzungsproblemen irgendwie fertig zu werden. Und wenn sie es geschafft haben, dann verdrängen sie diese Probleme im Tagesgeschäft; sie nehmen die technischen und ergonomischen Mängel nicht mehr wahr (Tendenz zum problemlosen Feld, Asher, 1963). Wenn man Benutzer einfach so fragt, wie sie denn mit dem System klar kommen, dann geben sie im allgemeinen stereotype, unkritische Antworten. Probleme werden verneint oder nicht mehr wahrgenommen. Diese Beurteilungstendenz kann als Stressbewältigung am Arbeitsplatz (coping) interpretiert werden, womit jedoch nur eine Scheinbewältigung der Probleme erreicht wird. Manche Benutzer tun sich sogar besonders damit hervor, dass alle Tricks kennen, um mit einem wenig gebrauchstauglichen Produkt umgehen zu können, auch wenn die Nutzung objektiv gesehen unproduktiv ist. Trotzdem sind auch diese Benutzer bereit, sich kritisch zu äußern, wenn man die angemessene Form findet, ihre kritische Kompetenz herauszufordern. Diese Form scheint ein Fragebogen zu sein, wenn er so

gestaltet ist, dass Benutzer angeregt werden, relativ zu einer wünschenswerten Arbeitstätigkeit die tatsächlich gegebene zu beschreiben. Aufgrund der Beobachtung dieses Verhaltens von Arbeitspersonen hat Ulich (1981) die subjektive Tätigkeitsanalyse entwickelt, ein Analysekonzept, das auch dem hier vorgelegten Fragebogen zu Grunde liegt. Die subjektive Tätigkeitsanalyse soll die Subjektposition des Benutzers zur Geltung bringen, im Unterschied zur Objektposition, in der sich Benutzer häufig als "Diener" einer zu bedienenden Maschine ausgeliefert sehen, wobei die gegebene Nutzungssituation als vorgegebene, ausreichende Lösung hingenommen werden muss. Mit der subjektiven Tätigkeitsanalyse sollen dem Benutzer die Möglichkeiten der Gestaltung vorstellbar gemacht werden, damit die gegenwärtige Nutzungssituation überhaupt als problemhaltig verstanden wird. Bevor Benutzer um eine Bewertung ihrer Arbeitssituation am Bildschirmarbeitsplatz gebeten werden, wird ihnen eine wünschenswerte Gestaltung des Mensch-Computer-Dialogs vorgestellt (Soll-Zustand), so wie sie im Teil 10 von DIN EN ISO 9241 beschrieben ist. Man sollte den Benutzern aber nicht einfach die Norm zum Lesen geben. Mit dem vorliegenden Fragebogen (Anhang D) wurden die meisten der genormten Anforderungen in eine Sprache übersetzt, die einfacher zu verstehen ist. Eine geeignete Provokation der Benutzer ist die Frageform, indem nach der Erfüllung einer genormten Forderung in Bezug zu einer bestimmten Arbeitsaufgabe gefragt wird. Fragen regen zum Antworten an. Anhand der Fragen wird vermutlich die selbst erlebte Nutzungssituation "in Frage gestellt", womit eine Neubewertung der Situation einhergehen kann, an die man sich vielleicht gewöhnt hat. Die Benutzer werden aufgefordert, die entdeckten oder wiedererkannten Soll-Ist-Differenzen zu beschreiben. Im Sinne von Ulich ist dies der erste Schritt in einem Prozess korrektiver Arbeitsgestaltung (a.a.O., S. 342).

Die subjektive Analyse und Bewertung ersetzt nicht die objektive, die mittels Prüfverfahren zu leisten ist. Aber sie kann den Anlass zu einer Prüfung geben. Die subjektive Befragung ist auch nicht Teil der Normkonformitätsprüfung, aber sie kann den Anstoß dazu geben, wenn mittels Fragebogen Nutzungsprobleme aufgedeckt werden, die von den Benutzern als sehr belastend eingeschätzt werden.

Die Befragung mittels Fragebogen soll Auskunft geben, wie die Software nach Effektivität oder Effizienz der beschriebenen Arbeitstätigkeiten subjektiv beurteilt wird. Diese Beurteilungen können als Indikatoren für Zufriedenstellung angesehen werden. Eine reine Zufriedenheitsmessung, ohne konkreten Bezug auf die Dialogprinzipien, würde den Anlass einer objektiven Prüfung nicht hinreichend begründen helfen, da Unzufriedenheit auch software-unabhängige Ursachen haben kann. Andererseits bedeutet Zufriedenstellung mit einem Softwareprodukt nicht zwangsläufig, dass es den Normen entspricht, denn Benutzer neigen zur Anpassung an Missstände, die mit der Zeit nicht mehr als solche wahrgenommen werden.

Als "Initialverfahren" soll der Fragebogen dazu dienen, den Anwendern und Benutzern ein Instrument an die Hand zu geben, mit dem sie die Initiative ergreifen können, die Aufmerksamkeit des Experten auf jene Situationen des Softwarenutzungs zu lenken, in denen die Benutzer Mängel bemerken, die sie sehr stören, damit einem begründeten Verdacht auf Normabweichung nachgegangen werden kann. Unter einem Initialverfahren verstehen wir einen anonym nutzbaren Fragebogen, der im Sinne einer subjektiven Tätigkeitsanalyse (sensu Ulich) den ersten Schritt machen hilft, die Problemhaltigkeit einer Nutzungssituation wahrzunehmen. Aus der beabsichtigten Anwendung des Fragebogens als Initialverfahren für eine sich möglicherweise anschließende Normkonformitätsprüfung sind die im nachfolgenden Abschnitt formulierten Anforderungen an einen Benutzerfragebogen entwickelt worden. In diesem Zusammenhang wird auch zu begründen sein, warum bereits existierende Fragebögen für diesen Anwendungszweck nicht geeignet sind (Abschnitt 6).

5.1 Anforderungen an den Benutzerfragebogen⁸

Fragebögen, in denen die Anforderungen des Teils 10 der Norm enthalten sind, gibt es genug (siehe nachstehende kommentierte Übersicht in Abschnitt 6). Warum also einen neuen entwickeln?

⁸ Der Fragebogen wurde im Rahmen einer Diplomarbeit entwickelt: Hofmann, (1999).

Eine Neuentwicklung ist nur dann berechtigt, wenn es noch keinen Fragebogen gibt, der als Initialverfahren dazu geeignet ist, der Bewertung einer kritischen Nutzungssituation am Dialogsystem voranzugehen und der sowohl an den expliziten Anforderungen der Norm als auch an den impliziten Erfordernissen einer bestimmten Arbeitsaufgabe verankert ist (Beurteilungsgrundlage). Für die Verwertung der Fragebogenergebnisse ist es unwichtig, ob die befragten Benutzer ihre Bewertungen in der Terminologie der Norm abgeben. Wichtig ist aber, dass sie angeregt werden, unter Vorlage einer korrektiven Gestaltungsaussage (Fragebogen-Item) und mit Blick auf eine vorweg beschriebene Aufgabe mögliche Nutzungsprobleme mit eigenen Worten kurz zu skizzieren.

Die Befragung dient nicht dazu, einen vom Benutzer (irgendwie) gebildeten subjektiven Gesamteindruck vom Dialogsystem mitgeteilt zu bekommen, denn wir wollen die Bewertungen des Benutzers nicht dazu verwerten, verschiedene Softwareprodukte anhand solcher subjektiver Pauschalbewertungen zu vergleichen. Wer einen Fragebogen für diesen Zweck einsetzen möchte (die meisten sind für diesen Zweck entwickelt worden), kann einen der in Kapitel 6 genannten Fragebogen benutzen, nur sollte er bei der Auswertung vermeiden festzustellen, dass das beste Produkt in der Rangreihe der untereinander verglichenen Produkte am besten mit der Norm DIN EN ISO 9241-10 konform geht. Solch eine Aussage ist unzulässig, da sie die Anwendung eines objektiven Prüfverfahrens voraussetzt.

Um das Ergebnis unserer Recherche von existierenden Fragebögen vorwegzunehmen: Wir haben nur einen Fragebogen gefunden (EU-CON II, 1999), der als Initialverfahren geeignet wäre. Allerdings enthält dieser Fragebogen keine Anforderungen aus dem Teil 10 der Norm, sondern aus der Council Directive (1990) zur Bildschirmarbeit (90/270/EEC). Zwar wird im EU-CON II-Fragebogen auch der Prozess der subjektiven Urteilsbildung berücksichtigt, aber in einer Weise, die wir nicht für den Zweck der Vorbereitung einer Bewertung akzeptieren (siehe Anmerkungen zum EU-CON II - Fragebogen in Abschnitt 6).

Als Ausgangspunkt für eine vergleichende Betrachtung der bereits vorliegenden Fragebögen (Abschnitt 6) werden zunächst einige Kriterien formuliert, die sowohl an eine Evaluierung der Gebrauchstauglichkeit von Software in dieser Form anzulegen sind als auch dem Zweck der Falsifikation dienen sollen, wonach einem Produkt auch aus Sicht des Benutzers Gebrauchstauglichkeit unterstellt wird, bis anhand eines Nutzungsproblems eine bedeutsame Normabweichung festgestellt worden ist.

5.2 Anforderungen an die Handhabbarkeit

Unterstützung der Phasen der Systementwicklung

Ausgehend von einem Modell der iterativen Entwicklung von Dialogsystemen ergibt sich an eine Evaluierung mittels Fragebogen die Anforderung, möglichst universell einsetzbar zu sein, sowohl für die Bewertung von Prototypen als auch des installierten Dialogsystems am Benutzerarbeitsplatz. Um in allen Phasen der Systementwicklung anwendbar zu sein, muss der Fragebogen auch für die Bewertung ausführbarer Teilaufgaben während des Prototypings anwendbar sein.

Anonymität der Befragung

Eine Benutzerbefragung, in der die Befragten kritische Bewertungen ihrer Arbeitssituation abgeben, sollte anonym sein. Unter dem Schutz der Anonymität sind Benutzer eher bereit, Nutzungsprobleme zu beschreiben. Niemand soll aus dieser Kritik Rückschlüsse auf die Person ziehen können, was je nach Betriebsklima von Einfluss auf die Befragungsergebnisse sein könnte.

Bearbeitung ohne Expertenbeteiligung

Der Fragebogen muss an Benutzerarbeitsplätzen ohne Beteiligung von Experten nutzbar und auswertbar sein. Bereits die Anwesenheit eines Experten bei der Benutzerbefragung schafft eine soziale Situation, in der manche Benutzer dazu neigen, im Sinne einer sozialen Erwünschtheit zu antworten oder im rechten Licht dastehen zu wollen.

Ein Experte soll nach der Auswertung der Fragebögen nur dann hinzugezogen werden, wenn die Ergebnisse auf Normabweichungen hindeuten. Dies erfordert eine weitgehende Formalisierung von Anwendung und Auswertung. Gleichzeitig muss darauf geachtet werden, dass der dafür notwendige Aufwand gering bleibt und sich zeitlich in der Größenordnung von etwa einer Stunde für Ausfüllen des Fragebogens und Auswertung bewegt.

Akzeptanz

Der Aufwand, den der Einsatz eines Fragebogens kostet, muss in einem vernünftigen Verhältnis zu dem erwarteten Nutzen stehen. Nur so kann eine Akzeptanz sowohl bei den befragten Benutzern als auch bei den Organisationen erwartet werden, in denen der Fragebogen eingesetzt werden wird.

5.3 Anforderungen an die Ergebnisqualität

Damit die Fragebogenergebnisse als valide Indikatoren für vermutete Abweichungen von der Norm gelten können, müssen verschiedene Anforderungen erfüllt sein:

DIN EN ISO 9241 – 10 als Gegenstand korrektiver Dialoggestaltung

Die Fragen müssen sich auf die Anforderungen des Teils 10 der Norm beziehen. Nur auf Basis dieses im internationalen Konsens entwickelten Standards können vergleichbare und nachvollziehbare Aussagen über die Gebrauchstauglichkeit der betrachteten Software gemacht werden. Relativ zu dieser Norm sollen Nutzungsprobleme reflektiert werden.

Verständlichkeit der Dialoggrundsätze

Der Benutzer muss die Dialoggrundsätze inhaltlich verstanden haben, damit er einen Soll-Zustand kennen lernt, gegen den er den Ist-Zustand der Nutzung des Dialogsystems beurteilen kann. Das Konzept des Erlernens eines Soll-Zustandes, *bevor* die

eigene Tätigkeit beurteilt wird, ist aus Erkenntnissen der Arbeitspsychologie begründet. Ulich (1981, 1998) hat diese Vorgehensweise im Rahmen seiner "Subjektiven Tätigkeitsanalyse" realisiert. Wie oben dargestellt, neigen Menschen dazu, schlechte Arbeitsbedingungen zu verdrängen oder zu verharmlosen, da sie meinen, doch nichts ändern zu können. Schafft man allerdings durch die Vermittlung eines Soll-Zustandes ein neues Problembewusstsein, wird es wahrscheinlicher, dass Missstände wiedererkannt und benannt werden. Der Soll-Zustand muss dabei in der Form von Dialoggrundsätzen als allgemein verbindlich dargestellt werden, so dass dem Benutzer die grundsätzlichen Realisierungsdetails als „gesetzlich vorgeschriebene“ Lösung suggeriert werden.

Eine gemeinsame Beurteilungsgrundlage bei Experten und Benutzern

Es muss sichergestellt werden, dass nicht nur die Dialoggrundsätze verstanden worden sind, sondern auch, dass Benutzer und Experte ihre Beurteilung auf denselben Sachverhalt beziehen. Dies bedeutet, dass der Benutzer nicht nur das Nutzungsproblem beschreiben soll, sondern gleichzeitig für den Experten nachvollziehbar machen muss, unter welcher der genormten Tätigkeitsanforderungen er das Problem einordnet und an welchen konkreten Dialogschritt (an welche Nutzungssituation) er gedacht hat, als er das Problem beschrieb.

Zur Sicherstellung einer übereinstimmenden Beurteilungsgrundlage von Benutzer und Experten ist es daher unerlässlich, dass die Software aufgabenspezifisch beurteilt wird. Für jede Arbeitsaufgabe muss also eine gesonderte Beurteilung vorgenommen werden, d.h. es sind gesonderte Fragebögen zu verteilen. Dies macht es dem Benutzer leicht, den Fragebogen zügig auszufüllen, weil die für eine Aufgabenbeurteilung nicht relevanten Fragen übergangen werden können. Nutzungsprobleme können anhand der vorgestellten Soll-Situation besser eingeordnet werden, und dem Experten wird dadurch eine gezielte und ökonomische Softwareprüfung ermöglicht, weil er in der Stichprobe der Fragebogenergebnisse schnell erkennen kann, wo sich Problembeschreibungen häufen.

Verständlichkeit der Dialoganforderungen für Benutzer

Bei der Konstruktion des Fragebogens muss darauf geachtet werden, dass die Items des Fragebogens zu den Dialoggrundsätzen inhaltlich passen und für Benutzer verständlich sind. Da die Fragen auf potentielle Nutzungssituationen abzielen, gilt es zusätzlich zu gewährleisten, dass sie inhaltlich zu den Nutzungssituationen passen. In allen drei Gesichtspunkten muss inhaltliche Validität gesichert sein.

Psychometrische Qualität

Zur Bestimmung der psychometrischen Qualität eines Benutzerfragebogens für Software ist grundsätzlich zu sagen, dass die klassische Testtheorie eigentlich gar nicht anwendbar ist. Die Theorie geht von der zeitlichen Stabilität des zu messenden Merkmals aus. Ergonomische Merkmale der Softwarenutzung wären wie Persönlichkeitsmerkmale anzusehen, die als Verhaltensdispositionen gegenüber kontextuellen Einflüssen relativ invariant bleiben. Diese Voraussetzung ist bei der Messung von Nutzungsmerkmalen kaum gegeben. Die Gebrauchstauglichkeit dieser Merkmale ist stark kontextabhängig (DIN EN ISO 9241-11). Ein psychometrisch zu bestimmendes Gütekriterium wie z.B. die Retest-Reliabilität einer Messung wäre demnach nicht anwendbar. Dennoch wird versucht, psychometrische Gütekriterien anzuwenden, um den Vergleich mit anderen Benutzerfragebögen zu ermöglichen, soweit diese durch gemessene Gütekriterien ausgewiesen sind. Außerdem wird dem Umstand Rechnung getragen, dass im Rahmen formaler Softwareprüfungen der Auftraggeber von der methodischen Seriosität überzeugt sein will, was sich durch Erfüllung von Gütekriterien erreichen lässt.

In bezug auf psychometrische Anforderungen ergibt sich, dass ein Fragebogen bei der Aufdeckung von Nutzungsproblemen reliabel sein soll, d.h., dass er bei wiederholter Anwendung auf eine Kombination Benutzer x – Aufgabe y – Dialogsystem z stets dieselben Probleme identifiziert. Da die zeitliche Stabilität des Erlebens und Bewertens

von Nutzungssituationen aus verschiedenen Gründen fraglich sein kann, eignet sich am besten die Paralleltestreliabilität als Indikator für die Reliabilitätsmessung. Es wird die Bewertung einer Nutzungssituation wie in einem verdoppelten Schnappschuss erhoben. Stimmen die Bewertungen der beiden Schnappschüsse gut überein, so ist Reliabilität gegeben.

In bezug auf die Validität wäre zu fordern, dass er zum einen nur norm-relevante Nutzungsprobleme identifiziert, zum anderen solche Probleme nicht übersieht, wenn sie im Expertenurteil vorliegen. Dies erfordert, den Fragebogen gegen das auf die Norm gestützte Urteil von Experten zu validieren, damit bei allen unvermeidbaren Unterschieden zwischen Laien- und Expertenurteil eine möglichst hohe Übereinstimmung in der Beurteilung von Nutzungsproblemen erreicht wird.

Da die von Benutzern geäußerten Beschwerden prinzipiell nicht gegeneinander aufgerechnet werden können und auch für das Fragebogenverfahren das Falsifikationsprinzip gilt (d.h. die Normkonformität wird zunächst unterstellt, kann aber durch eine einzige begründete Abweichung widerlegt werden), darf der Ausgang der Bewertung nicht durch einen Summenwert über die Items ermittelt werden, sondern muss sich an der Bewertung der Einzelitems orientieren. Bei der Berechnung von Summenscores könnte eine aus Expertensicht wichtige Beschwerde durch die besonders positive Einstufung einer anderen Eigenschaft kompensiert werden. Da das Vorgehen ggf. nach der Initialerhebung abbricht, muss für das sichere Erkennen etwaiger Abweichungen gesorgt werden. Dies erfordert, dass für die Einschätzung dessen, was ein Nutzungsproblem ist, eine konservative Auswahlregel angewendet werden muss, d.h., dass durch die befragten Benutzer tendentiell mehr Probleme gemeldet werden als sich später im Expertenurteil erhärten lassen. Auch die beschriebene Interpretationsweise des Fragebogens auf der Ebene einzelner Items erfordert wegen der geringeren statistischen Qualität von Einzelitems diese konservative Selektionsregel.

5.4 Entwicklungskonzept

Grundlage der subjektiven Bewertung der Nutzungsqualität eines Dialogsystems ist die zu erledigende Aufgabe im Nutzungskontext des Systems. Hierbei interessiert in erster Linie die Bewertung der Nutzungsprobleme. Es kommt somit beim Vergleich der Expertenbewertung mit der subjektiven Bewertung darauf an, dass für beide Bewertungen die gleiche Beurteilungsgrundlage gegeben ist (vgl. Abschnitt 2). Da die Expertenbewertung nach dem Falsifikationsprinzip abgegeben wird, müssen auch die Benutzer eines Systems nach demselben Prinzip ihre subjektiven Bewertungen machen. Nach dem Falsifikationsansatz wird einem Produkt gebrauchstaugliche Qualität unterstellt, und zwar so lange, bis diese Vermutung durch ein aufgetauchtes Nutzungsproblem in Frage gestellt wird. Nach dieser Richtschnur sollte auch das subjektive Urteil der Benutzer zustande kommen: Erst Aufgabe fokussieren, dann in der Durchführung einer Arbeitstätigkeit ein Problem erkennen, dann Problem einordnen, dann bewerten. Die Problemhaltigkeit muss klar den Nutzungsumständen zugerechnet werden. Bei der Formulierung von Fragen ist darauf zu achten, dass ein Benutzer ein Problem keinesfalls auf vermeintliche eigene Unzulänglichkeiten beziehen kann.

Um mittels Fragebogenkonstruktion den Urteilsprozess und die dargestellten Antwort-Tendenzen methodisch besser zu steuern, wurden folgende Überlegungen angestellt:

- Die Benutzer sollen in einer Phase der Problemfindung dazu angeregt werden, sich ihre Nutzungsprobleme bewusst zu machen und diese zu benennen.
- Eine leicht verständliche Einführung in die Bedeutung der Dialoggrundsätze für die Arbeit mit der Software soll den Benutzern helfen, einen wünschenswerten Zustand der Softwarenutzung mit ihrem gegenwärtigen Ist-Zustand zu vergleichen. Anhand der gestellten Fragen zur Bewertung dessen, was ein Dialogsystem eigentlich können sollte, haben die Benutzer Gelegenheit, den IST-Zustand ihrer eigenen Arbeitstätigkeit zu spiegeln. Dann sollen die Grundsätze und die Fragen auf die

tatsächliche Nutzung des Dialogsystems angewendet werden, um Probleme der Nutzung kurz durch Ankreuzen festzustellen.

- Es wird angenommen, dass die Benutzer allein aufgrund des Lesens des Fragebogens und der Dialoggrundsätze ihre Nutzungsprobleme besser erkennen können. Die Benutzer sollen aber durch die Instruktion angeregt werden, den Fragebogen in Ruhe zu lesen und nicht gleich mit dem Ausfüllen zu beginnen, damit ein Nachdenken über Nutzungsprobleme vertieft werden kann.
- Die Vorbereitung des durch den Fragebogen ausgelösten Urteilsprozesses ist abgeschlossen, wenn ein Benutzer den Fragebogen vor dem Ausfüllen vollständig gelesen hat und erkannt hat, dass im Inhalt der Fragen ein Potential für die korrektive Gestaltung der eigenen Arbeitstätigkeit enthalten ist. Der Fragebogen muss deshalb Gelegenheit bieten, neben dem Ankreuzen von Kästchen, die subjektive Bewertung eines erkannten Mangels aufzuschreiben.

5.5 Itemgenerierung und Itemselektion

Um zu einem ersten Itempool zu gelangen, wurden acht Software-Ergonomie-Experten gebeten, Items zu den Dialogprinzipien der DIN EN ISO 9241-10 zu generieren. Die Gruppe der Experten setzte sich zusammen aus:

- fünf Mitarbeitern der TÜViT-Essen und des TÜV-Rheinland, die seit mehreren Jahren Software auf ihre ergonomische Qualität prüfen,
- zwei Mitarbeitern der GMD - Forschungszentrum Informationstechnik, die sich seit vielen Jahren theoretisch und praktisch mit dem Thema Software-Ergonomie auseinandersetzen,
- einem Softwarehersteller aus Süddeutschland, der unter anderem an der Entwicklung ergonomischer Normen mitgearbeitet hat.

Die Experten sollten die Items passend zu den gegebenen Dialogprinzipien formulieren und dabei auf eine für Benutzer verständliche Ausdrucksweise achten. Der erste Itempool bestand aus insgesamt 192 Items: 50 Items für Aufgabenangemessenheit, 33 Items für Selbstbeschreibungsfähigkeit, 32 Items für Steuerbarkeit, 20 Items für

Erwartungskonformität, 24 Items für Fehlertoleranz, 10 Items für Individualisierbarkeit und 23 Items für Lernförderlichkeit.

Sicherung der Inhaltsvalidität der Items

Die unübersehbare Menge von Items galt es im folgenden zu validieren und auf eine für einen Fragebogen ökonomisch handhabbare Zahl zu reduzieren. Zunächst wurden die Items des Gesamt-Itempools durch Präferenzurteile der Experten in eine Rangreihe gebracht. Alle Items sollten auf einer fünfstufigen Skala bewertet werden, und zwar hinsichtlich Verständlichkeit (für Experten) und Repräsentativität für das jeweilige Dialogprinzip. Danach wurden Gesamtscores für die einzelnen Items errechnet. Jedes Item, das einen Gesamtscore von mindestens 20 erreichte, wurde zunächst in den Fragebogen aufgenommen. Der anfängliche Itempool von 192 Items reduzierte sich auf insgesamt 49 Items. Die Auswahl charakteristischer Items durch Expertenurteile diente gleichzeitig der inhaltlichen Validierung der Items des Fragebogens.

Die ranghöchsten 49 Items bildeten die Rohversion des Fragebogens. Es wurde eine Instruktion verfasst und zu jedem Dialogprinzip eine Einleitung geschrieben, um dem Benutzer vor dem Studium des Fragebogens und der Beantwortung der Fragen den Hintergrund des jeweiligen Prinzips und somit den "Soll-Zustand" zu vermitteln. Ferner wurden die Antwortkategorien "Ja" und "Nein" durch die Option "Frage trifft nicht zu" vorgesehen.

Verständlichkeits-Prüfung mit Benutzern

Zur Sicherung der Verständlichkeit der Fragen wurde die Rohversion des Fragebogens von einer Zufallsstichprobe von Benutzern im Beisein von Experten beantwortet. Die Experten ließen die Benutzer die Fragen des Fragebogens nach folgendem Schema beantworten: Nachdem eine Frage gestellt/gelesen war, wurde der Benutzer dazu aufgefordert zu erklären, was er unter dieser Frage versteht. Dann erst wurde die Frage konkret beantwortet. 7 Versuchspersonen (Vpn) mit Computererfahrung im Bürobereich wurde je ein Fragebogen durch einen Experten vorgelegt. Jede Vp sollte Frage für Frage durchlesen und das inhaltlich Gemeinte spontan interpretieren. Der Experte

verglich den durch die jeweilige Frage beabsichtigten Inhalt mit dem von der Vp interpretierten Inhalt. Nur im Falle völliger Übereinstimmung sollte die Frage als inhaltlich klar verständlich angesehen werden. Weniger klar verständliche Fragen sollten umformuliert oder gestrichen werden. Wenn eine Vp beim Interpretieren des Inhalts zögerte, so bat der Experte die Vp, sich eine Nutzungssituation vorzustellen, auf die der Inhalt der Frage passen könnte. Auf diese Weise konnte das Zögern manchmal überwunden und Inhaltsinterpretation eingeleitet werden. Wenn dies misslang, so galt die Frage als nicht unmittelbar verständlich und wurde gestrichen.

Die Untersuchung mit den Vpn ergab, dass die Instruktion als zu lang empfunden wurde. Wie sich in einer Nachbefragung mit 5 weiteren Vpn herausstellte, war die gekürzte Fassung der Instruktion besser verständlich. Die erläuternden Einleitungen zu jedem Dialogprinzip waren für alle Vpn verständlich. Die Terminologie der Fragen wurde vereinheitlicht. Statt "System" wurde in den meisten Fällen "Programm" verwendet. Einige Vpn erkannten orthographische Fehler. Das Wort "Software" wurde ebenfalls durch "Programm" ersetzt. Eine Frage wurde gestrichen, weil bei der Interpretation auffiel, dass die Vpn an "den dummen Benutzer" dachten, der evtl. gemeint sein könnte. Fünf weitere Fragen wurden gestrichen, weil sie entweder nicht spontan verstanden wurden, als überflüssig oder gar unsinnig beurteilt wurden. Nach der Verständlichkeitsprüfung durch die Benutzer wurde die Rohversion des Fragebogens gründlich überarbeitet. Es entstand die "Alpha-Version". Sie enthielt 43 Items.

Itemselektion anhand statistischer Kennwerte und Testrevision

Zum Zwecke der statistischen Itemanalyse und der Testrevision wurden jeweils 60 Fassungen der Alpha-Version des Fragebogens an drei Benutzergruppen verteilt, die mit unterschiedlicher Individualsoftware unterschiedliche Arbeitsaufgaben bewältigen. Es wurde darauf geachtet, dass die Benutzer innerhalb einer Gruppe ungefähr gleich lang mit dem Produkt arbeiten, damit Einschätzungen des Produkts nicht durch einen unterschiedlichen Grad an Expertise verzerrt werden. Weitere signifikante Einflüsse auf die Homogenität der Benutzergruppen wurden nicht angenommen und somit nicht

berücksichtigt. In der Instruktion für jede Benutzergruppe wurde jeweils die Aufgabe charakterisiert, an die beim Ausfüllen des Fragebogens gedacht werden sollte. Die Tätigkeiten zur Ausführung der Aufgabe wurden zusätzlich auf der ersten Fragebogen-seite beschrieben.

Bei den drei Benutzergruppen und zu beurteilenden Tätigkeiten handelte es sich um:

- Mitarbeiter eines Dienstleistungsunternehmens in Bonn. Die zu beurteilende Tätigkeit war die Erfassung von Störungsmeldungen.
- Mitarbeiter einer Prüfeinrichtung in Essen. Die zu beurteilende Tätigkeit war die Erfassung und Verwaltung der eigenen Arbeitszeit.
- Mitarbeiter einer Hamburger Bildagentur, die unter anderem für eine Fernsehzeitung Bildarchive verwaltet. Die zu beurteilende Tätigkeit war die Verwaltung von Mappen zu verschiedenen Medien.

Leider kamen nicht alle ausgeteilten Fragebogenexemplare ausgefüllt zurück. Trotzdem erfüllte sich die Hoffnung, pro Benutzergruppe wenigstens 25 Fragebögen zur Auswertung zur Verfügung zu haben. Der Rücklauf aus dem Bonner Dienstleistungsbetrieb betrug 32 Exemplare, aus Essen kamen immerhin 27 Exemplare und aus Hamburg 26 Fragebögen wieder zurück. Die Akzeptanz des Fragebogens war gering, da er von fast allen Benutzern als zu lang empfunden wurde. Die Dauer des Ausfüllens betrug im allgemeinen fast eine Stunde, was während der regulären Arbeitszeit als Zumutung empfunden wurde. Eine Reduktion der Anzahl der Fragen war also unabhängig von den Ergebnissen der statistischen Itemanalyse dringend notwendig.

Itemselektion durch Vergleich der Mittelwerte der einzelnen Items

In diesem Selektionsschritt wurden Items eliminiert, welche sich nicht eigneten, zwischen den drei verschiedenen Programmen zu diskriminieren und für die es keine anderen guten Gründe gab, trotzdem beibehalten zu werden. Inhaltlich "gute Gründe" für die Beibehaltung eines Items, trotz schlechter Kennwerte, liegen nach Fisseni (1997) zum Beispiel vor, wenn

- Items als Eisbrecher dienen,

- Items einen bestimmten Itemtyp repräsentieren,
- ansonsten keine Items vorliegen, die den entsprechenden Inhalt erfassen und
- Items besonders prägnant für das zu erfassende Merkmal sind.

Um einen Index für die "Diskriminationsfähigkeit" der Items zu erhalten, wurden die Mittelwerte der einzelnen Items pro Benutzergruppe verglichen. Es bestand die Annahme, dass sich die drei Programme hinsichtlich der Beantwortung der einzelnen Items unterscheiden. Mit anderen Worten, es gab keinen Grund zu der Vermutung, dass die drei Programme hinsichtlich der zu bewertenden Fragen gleich gut oder schlecht sind. Items, die bei allen drei Programmen nicht signifikant unterschiedlich beantwortet wurden, wurden neu überdacht und ggf. eliminiert.

Zum Vergleich der Mittelwerte pro Item zwischen den drei Benutzergruppen wurde eine einfaktorielle Varianzanalyse (ANOVA) gerechnet. Jedes Item, das bei einer Irrtumswahrscheinlichkeit von $\alpha = 10\%$ signifikante Unterschiede zwischen den drei Programmen aufwies, wurde in die Endfassung des Fragebogens übernommen. Tabelle 5.5-1 gibt einen Überblick über die Ergebnisse des Mittelwertvergleichs:

Alpha-Version des Fragebogens	df	F	Signifikanz
Aufgabenangemessenheit, Frage 1	2	2,537	,085
Aufgabenangemessenheit, Frage 2	2	1,130	,328
Aufgabenangemessenheit, Frage 3	2	4,011	,022
Aufgabenangemessenheit, Frage 4	2	5,466	,006
Aufgabenangemessenheit, Frage 5	2	7,583	,001
Aufgabenangemessenheit, Frage 6	2	1,742	,182
Aufgabenangemessenheit, Frage 7	2	,616	,543
Aufgabenangemessenheit, Frage 8	2	1,378	,258
Aufgabenangemessenheit, Frage 9	2	10,859	,000
Aufgabenangemessenheit, Frage 10	2	8,403	,001
Aufgabenangemessenheit, Frage 11	2	7,820	,002

Selbstbeschreibungsfähigkeit, Frage 1	2	,324	,724
Selbstbeschreibungsfähigkeit, Frage 2	2	2,994	,056
Selbstbeschreibungsfähigkeit, Frage 3	2	1,955	,149
Selbstbeschreibungsfähigkeit, Frage 4	2	5,318	,007
Selbstbeschreibungsfähigkeit, Frage 5	2	1,354	,268
Selbstbeschreibungsfähigkeit, Frage 6	2	,351	,705
Selbstbeschreibungsfähigkeit, Frage 7	2	6,670	,002
Steuerbarkeit, Frage 1	2	4,890	,010
Steuerbarkeit, Frage 2	2	1,004	,371
Steuerbarkeit, Frage 3	2	2,787	,068
Steuerbarkeit, Frage 4	2	,427	,654
Steuerbarkeit, Frage 5	2	1,926	,152
Steuerbarkeit, Frage 6	2	5,613	,005
Erwartungskonformität, Frage 1	2	5,165	,009
Erwartungskonformität, Frage 2	2	2,231	,114
Erwartungskonformität, Frage 3	2	,505	,605
Erwartungskonformität, Frage 4	2	,838	,436
Erwartungskonformität, Frage 5	2	7,637	,001
Erwartungskonformität, Frage 6	2	1,944	,150
Fehlertoleranz, Frage 1	2	2,243	,113
Fehlertoleranz, Frage 2	2	7,523	,001
Fehlertoleranz, Frage 3	2	,579	,563
Fehlertoleranz, Frage 4	2	,839	,436
Fehlertoleranz, Frage 5	2	4,388	,016
Fehlertoleranz, Frage 6	2	13,350	,000
Fehlertoleranz, Frage 7	2	2,711	,073
Individualisierbarkeit, Frage 1	2	2,161	,122
Individualisierbarkeit, Frage 2	2	2,875	,063

Lernförderlichkeit, Frage 1	2	3,629	,031
Lernförderlichkeit, Frage 2	2	,165	,848
Lernförderlichkeit, Frage 3	2	13,160	,000
Lernförderlichkeit, Frage 4	2	,403	,669

Tabelle 5.5-1: Ergebnisse der einfaktoriellen Varianzanalyse. Die Werte beziehen sich auf den Vergleich der Itemmittelwerte zwischen den drei Gruppen. Items, die bei einem $\alpha > 10\%$ keine signifikanten Unterschiede aufweisen, sind grau unterlegt.

Im folgenden galt es zu entscheiden, welche der "kritischen" Items ($\alpha > 10\%$) eliminiert oder beibehalten werden sollten. Hierzu wurde das inhaltliche Gewicht eines jeden kritischen Items für das jeweilige Dialogprinzip mit einem Experten diskutiert. Falls es keinen "guten Grund" für die Beibehaltung eines Items gab, wurde es gestrichen. Beispiele für gute Gründe waren: Die Fragen 7 und 8 unter dem Prinzip Aufgabenangemessenheit sind als inhaltlich beispielhafte Verankerungen anzusehen und somit repräsentativ für das Dialogprinzip. Gleiches gilt auch für die Fragen 3 und 6 unter dem Prinzip Selbstbeschreibungsfähigkeit, sowie für die Fragen 4 und 5 unter Steuerbarkeit, Fragen 2 und 4 Erwartungskonformität und Frage 1 Individualisierbarkeit.

Skalierung

Im nächsten Schritt galt es herauszufinden, anhand welcher *Skalen* der Fragebogen geeignet ist, Software zu beschreiben. Es bestand die Annahme, dass jedes Dialogprinzip für sich eine Skala mit inhaltlich konsistenten Items repräsentiert. Ein Maß für die Konsistenz ist Cronbachs Alpha. Gemessen wird die Konsistenz der Items, indem für alle möglichen Paare von Items ein Test bezüglich eines hypothetischen Merkmals (hier das Dialogprinzip) gemacht wird, das diese Items in konsistenter Weise kennzeichnet. Der Alpha-Koeffizient ist ein Maß für die Paralleltest-Reliabilität eines Fragebogens.

Eine Reliabilitätsanalyse auf der Basis von Cronbachs Alpha zerstörte allerdings die Hoffnung, dass sich die Dialogprinzipien auf der Grundlage der hier erhobenen Daten

als Skalen bestätigen lassen könnten. Die meisten der als Skalen vermuteten Dialogprinzipien wiesen nur eine sehr geringe innere Konsistenz auf, wie sich aus der nachfolgenden Tabelle 5.5-2 ersehen lässt. Ein für eine Skala akzeptables Cronbachs Alpha sollte mindestens einen Wert von annähernd 0,7⁹ aufweisen.

33 Items		
Dialogprinzip	Alpha-Wert	Anzahl der Items
Aufgabenangemessenheit (AA)	,64	9
Selbstbeschreibungsfähigkeit (SB)	,37	5
Steuerbarkeit (ST)	,35	5
Erwartungskonformität (EK)	,19	5
Fehlertoleranz (FT)	,40	5
Individualisierbarkeit (IN)	,69	2
Lernförderlichkeit (LF)	,35	2

Tabelle 5.5-2: Alpha-Werte und Anzahl der Items für jedes Dialogprinzip

Allein das Dialogprinzip Individualisierbarkeit könnte gemäß des Alpha-Wertes als Skala angesehen werden; allerdings setzt sich dieser Wert aus der Korrelation zweier Items zusammen, und in diesem Fall kann es gut möglich sein, dass diese zufällig hoch korrelieren. Es war also notwendig herauszufinden, welche Itemgruppierung eine hohe Konsistenz aufweist und somit zur Bildung neuer Skalen geeignet ist.

Die Dimensionalität des Fragebogens allein sagt noch nichts über die Nützlichkeit des Fragebogens aus. Je nach Art der befragten Stichprobe von Benutzern können sich die Fragebogenitems teilweise neu gruppieren, so dass andere abstrakte Bewertungskonzepte eingeführt werden müssten, um die Faktoren inhaltlich zutreffend zu interpretieren. Die nachfolgend dargestellten Ergebnisse dienen dazu, die Stichprobenabhängigkeit von Faktoren zu demonstrieren. Ein Vergleich mit den Faktoren der

⁹ Nach Weise (1975) handelt es sich hierbei um einen sehr geringen Reliabilitätswert. In Anbetracht der geringen Itemanzahl pro Skala (Dimension), soll ein Reliabilitätswert von annähernd ,7 jedoch als akzeptabel angesehen werden.

DIN EN ISO 9241-10 ist nicht möglich, da der Bestimmung der in dieser Norm enthaltenen Kategorien eine Stichprobe von Experten zugrunde lag (Dzida et al, 1978). Mit Hilfe der Faktorenanalyse und anschließender Interpretation konnten vier verschiedene Skalen aufgedeckt werden. Die statistisch zusammengehörigen Items ließen sich auch inhaltlich gut verknüpfen. Zunächst werden die vier neuen Skalen mit ihren jeweiligen Alpha-Werten vorgestellt. Danach wird gezeigt, dass die drei verschiedenen Computerprogramme, für die der Fragebogen ausgefüllt wurde, anhand dieser Skalen unterschieden werden können.

In den folgenden Tabellen (Tabellen 5.5-3 bis 5.5-6) werden die Items der vier Skalen vorgestellt. Es wird kenntlich gemacht, aus welchem Dialogprinzip die Items ursprünglich stammen (Abkürzungen, siehe Tabelle 5.5-2).

Skala 1: Interpretiert als **Verständlichkeit**
(Interpretiert als sicherer und selbständiger Umgang mit dem Programm.)

$\alpha = ,69$	
Ursprüngliche Itemzugehörigkeit	
AA9	1. Finden Sie im Programm Hilfetexte, die Ihnen auch tatsächlich weiterhelfen?
SB2	2. Können Sie bei der Arbeit mit dem Programm erkennen, welche Eingabe als nächstes von Ihnen erwartet wird?
SB3	3. Sind die Meldungen des Systems für Sie immer verständlich?
SB4	4. Werden Sie vor Aktionen, die nicht mehr rückgängig gemacht werden können, von der Software gewarnt?
SB6	5. Hilft Ihnen das Programm bei der Auswahl von Eingaben, z.B. durch Listen mit möglichen Eingaben?
SB7	6. Müssen Sie oft Kollegen oder ein Handbuch konsultieren ehe Sie weiterarbeiten können?
EK3	7. Verhält sich das Programm so, wie Sie es erwarten würden?
EK4	8. Sind Sie sich bei Wartezeiten immer noch sicher, ob das Programm weiter arbeitet?
FT2	9. Bekommen Sie bei fehlerhaften Eingaben Korrekturhinweise?
FT7	10. Hilft Ihnen das Programm Eingabefehler zu erkennen, bevor die Eingabe verarbeitet wird?

Tabelle 5.5-3: Items der Skala 1 mit Zuordnung zu den ursprünglichen Dialogprinzipien

Skala 2: Interpretiert als **Steuerbarkeit**

(Interpretiert als aktive Gestaltung des Dialogablaufs oder Wahl der Dialogform.)

$\alpha = ,75$	
Ursprüngliche Itemzugehörigkeit	
AA10	1. Können Sie die Hilfeinformationen jederzeit einfach erreichen?
ST4	2. Können Sie einen Arbeitsschritt wieder zurück nehmen, wenn es für Ihre Arbeitserledigung zweckmäßig ist?
ST6	3. Können Sie sich aussuchen, ob Sie mit der Maus oder Tastatur arbeiten wollen?
IN2	4. Können Sie am Computer alles so einstellen, dass Ihnen das Lesen von Information leicht fällt?
LF1	5. Ermöglicht Ihnen das Programm auch einmal gefahrlos etwas auszuprobieren?

Tabelle 5.5-4: Items der Skala 2 mit Zuordnung zu den ursprünglichen Dialogprinzipien

Skala 3: Interpretiert als **Aufwandsangemessenheit**

(Interpretiert als vom Benutzer zu leistender Aufwand bei der Aufgabebearbeitung.)

$\alpha = ,69$	
Ursprüngliche Itemzugehörigkeit	
AA3	1. Müssen Sie Dialogschritte machen, die eigentlich überflüssig wären?
AA5	2. Finden Sie, dass der erforderliche Aufwand für Ihr Arbeitsergebnis jeweils angemessen ist?
AA7	3. Müssen Sie Texte eingeben, die der Computer eigentlich wissen könnte?
ST1	4. Können Sie ihre Arbeitsschritte in der Reihenfolge erledigen, die Ihnen am sinnvollsten erscheint?
EK2	5. Finden Sie Menüpunkte oder Funktionen dort, wo sie Ihrer Meinung nach auch sein sollten?

Tabelle 5.5-5: Items der Skala 3 mit Zuordnung zu den ursprünglichen Dialogprinzipien

Skala 4: Interpretiert als **Aufgabenangemessenheit**

(Interpretiert als Angemessenheit des Dialogs für kontextspezifische Umstände, insbesondere die Aufgabe.)

$\alpha = ,75$	
Ursprüngliche Itemzugehörigkeit	
AA4	1. Ist es Ihnen möglich das wiederholte Eingeben von Texten zu vereinfachen?
AA8	2. Müssen Sie sich mit Tricks behelfen, um Ihre Arbeitsergebnisse so zu erzielen, wie Sie diese haben möchten?
AA11	3. Passt das Programm zu Ihren bisherigen Formularen und Formaten?
EK1	4. Können ähnliche Aufgaben ähnlich bearbeitet werden?
ST3	5. Können Sie bei Bedarf eine Aufgabe unterbrechen und später wieder fortsetzen, ohne alles neu eingeben zu müssen?

Tabelle 5.5-6: Items der Skala 4 mit Zuordnung zu den ursprünglichen Dialogprinzipien

Eliminierte Items im Rahmen der Skalierung

Im Rahmen der Skalierung konnten nicht alle Items einer Skala zugeordnet werden. Sieben Items wurden eliminiert. Im Ergebnis lässt sich feststellen, dass einige der vier neu gebildeten Skalen die bisher im Teil 10 der Norm enthaltenen Bewertungsaspekte beibehalten oder bündeln. Beibehalten werden Aufgabenangemessenheit und Steuerbarkeit. Unter der neuen Skala "Verständlichkeit" werden Teilaspekte der alten Skalen zusammengefasst. Diese Begriffsbildung ist auch schon in software-technischen Normen zu finden, in denen "usability characteristics" spezifiziert werden (siehe ISO/IEC 9126, Anhang A.2.3, 1991); dort korrespondiert "understandability" mit der hier festgestellten Skala "Verständlichkeit".

5.6 Vergleich der Softwareprodukte anhand von Skalenwerten

Nachdem statistisch akzeptable und inhaltlich plausible Skalen konstruiert waren und die statistisch weniger brauchbaren Items eliminiert waren, galt es herauszufinden, inwieweit die Hypothese zutraf, dass die drei bewerteten Softwareprodukte anhand der Skalen unterschieden werden können. Der Mittelwert aus den positiven Antworten auf die ausgewählten Items einer Skala wurde als Skalenwert interpretiert. Auf drei der vier oben genannten neuen Skalen konnten mit Hilfe der einfaktoriellen Varianzanalyse signifikante Unterschiede zwischen den Skalenwerten aufgedeckt werden.

Im folgenden werden die Skalenwerte der Produkte auf jeder einzelnen Skala vorgestellt und gezeigt, inwieweit die Skalen statistisch bedeutsame Unterschiede in der Bewertung der Programme aufdecken.

Skala 1 Skalenwerte der Softwareprodukte aus:	Ergebnisse der Mittelwertsvergleiche		
	df	F	Signifikanz
Essen 6,61	2	1,57	,219
Bonn 7,75			
Hamburg 6,57			

Tabelle 5.6-1: Messergebnisse auf Skala 1 (Verständlichkeit) mit den Ergebnissen der einfaktoriellen Varianzanalyse

Skala 1 eignet sich nicht, statistisch signifikante Unterschiede auf einem α - Niveau von .05 aufzudecken. Hinsichtlich der Verständlichkeit sind die drei bewerteten Programme etwa gleich gut. Hierin zeigt sich die Abhängigkeit der statistischen Itemselektion von einem Merkmal der Stichprobe der bewerteten Softwareprodukte.

Skala 2 Skalenwerte der Softwareprodukte aus:	Ergebnisse der Mittelwertsvergleiche		
	df	F	Signifikanz
Essen 3,00	2	6,06	,004
Bonn 4,31			
Hamburg 2,75			

Tabelle 5.6-2: Messergebnisse auf Skala 2 (Steuerbarkeit) mit dem Ergebnissen der einfaktoriellen Varianzanalyse

Die Unterschiede der Skalenwerte sind auf Skala 2 hoch signifikant. Mit Hilfe des Scheffé-Tests konnten die für den signifikanten Unterschied der Mittelwerte verantwortlichen Gruppen identifiziert werden. Die mittleren Differenzen der Bewertungen lassen die Softwareprodukte Essen-Bonn und Hamburg-Bonn auf einem α -Niveau von .05 signifikant unterscheiden.

Skala 3 Skalenwerte der Softwareprodukte aus:	Ergebnisse der Mittelwertsvergleiche		
	df	F	Signifikanz
Essen 2,96	2	4,41	,015
Bonn 4,03			
Hamburg 3,09			

Tabelle 5.6-3: Messergebnisse von Skala 3 (Aufwandsangemessenheit) mit dem Ergebnissen der einfaktoriellen Varianzanalyse

Die Unterschiede der Werte sind auf Skala 3 bereits auf einem α -Niveau von .015 signifikant. Für diese signifikanten Unterschiede sind laut Scheffé-Test wiederum die mittleren Differenzen der Bewertungen aus Essen-Bonn und Hamburg-Bonn verant-

wortlich, da sie auf einer Stufe von .05 statistisch bedeutsame Unterschiede aufweisen.

Skala 4 Skalenwerte der Softwareprodukte aus:	Ergebnisse der Mittelwertsvergleiche		
	df	F	Signifikanz
Essen 4,10	2	8,01	,003
Bonn 3,40			
Hamburg 1,88			

Tabelle 5.6-4: Messergebnisse von Skala 4 (Aufgabenangemessenheit) mit dem Ergebnissen der einfaktoriellen Varianzanalyse

Skala 4 misst signifikante Unterschiede zwischen den Bewertungen der Programme auf einem Signifikanzniveau von $\alpha = .003$. Verantwortlich für diesen Unterschied sind auf einem .05-Niveau allein die mittleren Differenzen der Programmbeurteilungen aus Essen und Hamburg.

5.7 Abschließende Itemselektion

Dominierendes Kriterium für die Entwicklung des Fragebogens war bis zu diesem Zeitpunkt die psychometrische Qualität. Diese ist bei der Itemselektion allerdings nicht das einzige und in bezug auf die praktische Nutzung des Fragebogens auch bei weitem nicht das wichtigste Qualitätskriterium. Um die Qualität des Fragebogens in der Anwendungspraxis zu sichern, müssen noch andere Aspekte berücksichtigt werden. Bei der psychometrischen Itemselektion stellte sich heraus, dass zugunsten der psychometrischen Qualität des Fragebogens einerseits Items extrahiert werden mussten, auf die aus der Sicht der Nützlichkeit nicht verzichtet werden konnte; andererseits wurden Items beibehalten, die im Hinblick auf den praktischen Nutzen des Fragebogens eigentlich zu vernachlässigen sind. Deshalb wurde der Fragebogen gemäß folgender praxisrelevanter Kriterien einer letzten Modifikation unterzogen, und dabei in Kauf genommen, dass die psychometrische Qualität des Fragebogens dadurch verwässert wurde:

Antwortresonanz: Die Fragebogen-Items sollen die Benutzer zu Antworten provozieren. Wenn ein Item eine zur Nennung von Nutzungsproblemen lebhaft anregte, so wurde es im Fragebogen gelassen, auch wenn die statistischen Daten weniger günstig waren. Dafür wurden statistisch wertvolle Items herausgenommen, wenn sie kaum Antwortresonanz hatten. Insoweit wurde zwar die Paralleltest-Reliabilität des Fragebogens vermindert, die Nützlichkeit des Fragebogens wurde jedoch optimiert, und zwar hinsichtlich der Tauglichkeit der Items, auf Nutzungsprobleme aufmerksam machen zu können, die auf mögliche Normabweichungen hindeuten.

Durchführungszeit: Es wurde bei der abschließenden Itemselektion noch darauf geachtet, dass der Fragebogen maximal 30 Items hat, damit die Durchführungszeit für die Ausfüllung des Fragebogens am Arbeitsplatz zumutbar ist. Zu diesem Zweck wurden Items eliminiert, deren Beantwortung durch Experten objektiv möglich ist. Ob z.B. ein Benutzerhandbuch verfügbar ist, kann objektiv festgestellt werden. Ob aber die in der Hilfefunktion angebotene Information verständlich ist, kann am besten der Benutzer beurteilen. Fragebogenitems, die sich nur aus der genauen Kenntnis des Nutzungskontexts eines Produkts beurteilen lassen, wurden solchen gegenüber bevorzugt, die auch ohne diese Kenntnisse beantwortbar sind. Hiermit wurde berücksichtigt, dass subjektive Bewertungen der Benutzer besonders wegen ihrer Kenntnis der Kontextbedingungen wertvoll sind. Zusätzlich wurden Items eliminiert, wenn sie signifikante Interkorrelationen mit anderen Items aufwiesen.

Tabelle 5.7-1 zeigt die Alpha-Werte der Endversion des Fragebogens für die Dialogprinzipien bei einer Item-Menge von $N = 28$. Bei einem Gesamt-Alpha von 0,85 weisen nur die Kategorien AA und SB Alpha-Werte auf, die unter Berücksichtigung der sehr kleinen Itemzahl pro Kategorie akzeptabel sind. Die wegen einfacher Itembesetzung nicht definierten Alpha-Werte für IN und LF sowie die niedrigen Alpha-Werte für ST, EK und FT gaben den Anlass, die Items neu zu gruppieren, um darzustellen, dass die empirisch konsistenten Dimensionen Aufgabenangemessenheit, Verständlichkeit und Steuerbarkeit auch die Endversion des Fragebogens zuverlässig prägen (siehe Tabelle 5.7-2).

28 Items		
Dialogprinzip	Alpha-Wert	Anzahl der Items
Aufgabenangemessenheit (AA)	,76	9
Selbstbeschreibungsfähigkeit (SB)	,62	6
Steuerbarkeit (ST)	,36	5
Erwartungskonformität (EK)	,16	3
Fehlertoleranz (FT)	,18	3
Individualisierbarkeit (IN)	nicht definiert	1
Lernförderlichkeit (LF)	nicht definiert	1

Tabelle 5.7-1: Alpha-Werte und Anzahl der Items für jedes Dialogprinzip

5.8 Beanspruchungsmessung

In die Endfassung des Fragebogens wurde unter die leeren Kommentarzeilen einer jeden Frage die Anmerkung hinzugefügt: "Ich empfinde dies als sehr störend." In die Instruktion wurde der Hinweis aufgenommen, dass der befragte Benutzer nach der Beschreibung eines Nutzungsproblems die Möglichkeit hat, hinter seine Anmerkung ein Kreuzchen zu machen, wenn der beschriebene Mangel als sehr belastend angesehen wird. Der Blick des Auswerters wird somit direkt auf solche Nutzungsprobleme gelenkt, die möglicherweise auch objektiv kritisch sind. Aus einem Stapel von zum Beispiel 20 oder 30 eingesammelten Fragebögen sollte es nicht mehr aufwendig sein, die von den Benutzern als beanspruchend erlebten Probleme schnell herauszufinden. Insoweit kann der Fragebogen auch gute Dienste bei der gesetzlich geforderten Arbeitsplatzanalyse leisten (vgl. ArbSchG, 1996 und BildscharbV, 1996).

Da die Fragebogendaten anonym erhoben wurden, ist es nicht möglich, jeder subjektiven Problembeschreibung am einzelnen Benutzerarbeitsplatz nachzugehen. Die Gesamtauswertung aller beschriebenen Probleme ergibt aber einen guten Überblick, so dass die am häufigsten genannten oder die als „sehr störend“ bewerteten Probleme rasch erfasst werden können. Wenn unter den extrahierten Problemen solche

sind, die ein Experte als vermutete Normabweichungen einschätzt, so können die von den Benutzern als sehr störend bewerteten Probleme als Indikatoren für mangelnde Zufriedenstellung angesehen werden. Diese Information ist sehr nützlich bei der Anwendung der Entscheidungstabellen des Erhärtungstests (Abschnitt 3.3), die vorsehen, den betroffenen Benutzeranteil sowie die Schwere der Auswirkung mit ins Kalkül zu ziehen.

5.9 Verwertung der Fragebogenergebnisse

Die Besonderheit des ErgoNorm-Fragebogens kann am besten im Zusammenhang mit der möglichen Verwertung der Fragebogenergebnisse dargestellt werden. Gegenüber anderen Fragebögen, die in Abschnitt 6 beschrieben sind, besteht das Alleinstellungsmerkmal des ErgoNorm-Fragebogens in den auf Nützlichkeit hin ausgewählten Items. Nützlich meint hier die Hilfe zur Bewusstmachung problemhaltiger Nutzungssituationen. Mit geringem Aufwand können kleine und mittlere Anwenderorganisationen den Fragebogen zu diesem Zwecke einsetzen. Gerade für diese Anwender ist es wegen der Nutzungskosten wichtig herauszufinden, welche ergonomischen Mängel einer Software unnötige Folgekosten verursachen. Gelingt es diese Mängel zu beseitigen, so können Anwender die Nutzungskosten der Software senken. Jede ergonomische Verbesserung eines Softwareprodukts trägt neben der Zufriedenstellung der Benutzer immer auch zur Steigerung der Produktivität bei (Zijlstra, 1993; Landauer, 1995; Arnold, 1998). Diese empirisch belegte Erkenntnis sollte bei der betrieblichen Umsetzung von Beschaffungs- oder Entwicklungsprojekten berücksichtigt werden. Mit dieser Zielsetzung wird Anwendern und Herstellern empfohlen, einen Pflegevertrag abzuschließen, in dessen Rahmen auch der Fragebogen eingesetzt werden kann. Es sollte besonders vereinbart werden, wie die während der Softwarenutzung aufgedeckten Nutzungsprobleme zu bewältigen sind. Kleine und mittlere Anwender sind gut beraten, wenn sie sich nicht nur mit einem Wartungsvertrag zufrieden geben. Nach dem heutigen Stand der Technik reicht es nicht mehr aus, dass eine Software korrekt funktioniert und für den gegen-teiligen Fall eine Wartung fällig wird. Unter Gebrauchstauglichkeit der Software wird heute die effiziente und zufriedenstellende Nutzung verstanden. Wird diese Qualitäts-

stufe nicht erreicht, so können mit dem hier vorliegenden Fragebogen und dem Prüfverfahren die Nutzungsprobleme aufgearbeitet werden.

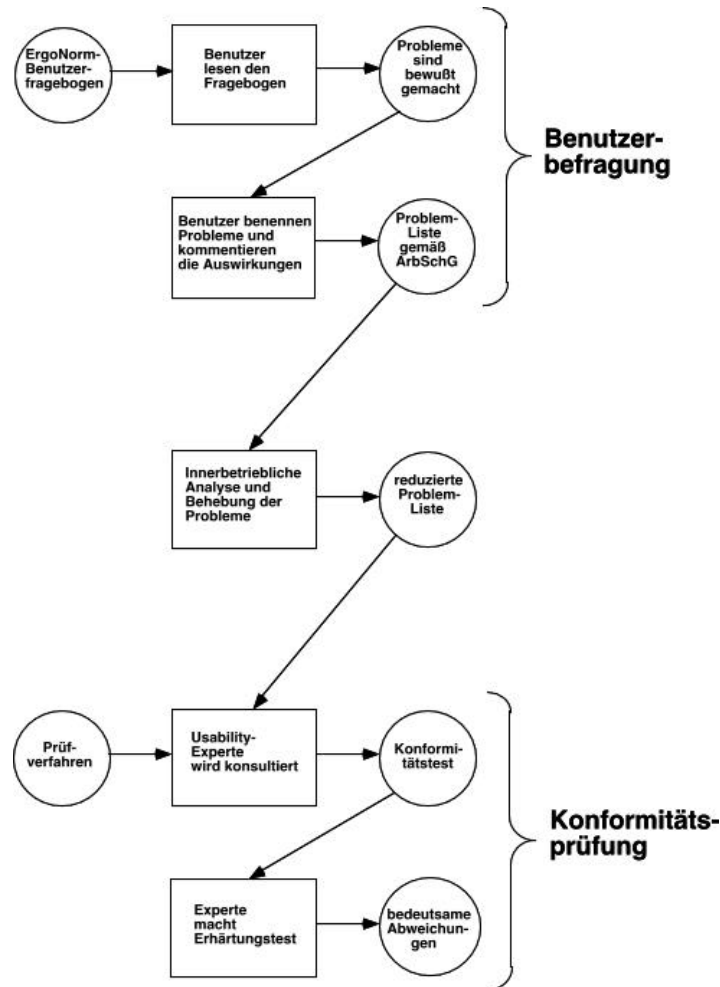


Bild 5.9-1: Vorgehensmodell für den betrieblichen Einsatz von Fragebogen und Prüfverfahren

In Bild 5.9-1 wird eine Vorgehensweise empfohlen, bei der die Benutzerbefragung mittels ErgoNorm-Fragebogen am Anfang steht. Betriebliche Beauftragte können somit im Rahmen der gesetzlichen Vorschriften (ArbSchG und Bildschirmarbeitsverordnung) die Benutzer von Softwareprodukten an der Aufdeckung und Bewältigung von Nutzungsproblemen beteiligen. Der Benutzerfragebogen ist ein effizientes Instrument der Beteiligung von Benutzern an vorbereitenden Maßnahmen des Arbeitsschutzes im Sinne des Arbeitsschutzgesetzes und der Bildschirmarbeitsverordnung. Betrieblich Verantwortliche können die Befragungsergebnisse zur Grundlage nehmen, mit dem Hersteller eines Produkts zu beraten, welche der Nutzungsprobleme bei der Vorbereitung des nächsten Release behoben werden sollen.

Wer viel fragt, bekommt unter Umständen viele Antworten. Man kann leicht in einer Datenflut ertrinken, insbesondere, wenn ein Softwareprodukt viele Nutzungsprobleme mit sich bringt. Deshalb ist der Fragebogen nur für eng formulierte Arbeitsaufgaben einzusetzen. Hierfür gilt es, die Probleme herauszufiltern, die am meisten störend wirken. Zu diesem Zweck liefert der Fragebogen Ergebnisse, die nützlicher sind als ein Skalenwert, z.B. auf der Skala Steuerbarkeit. Nützlich im Sinne der Bewusstmachung problemhaltiger Situationen ist nicht der Skalenwert, sondern das "dahinter liegende" Problem, das der Benutzer mit der Steuerung des Programms hat.

Im nächsten Abschnitt wird eine Schablone angeboten, die den Blick des Auswerter auf die Kreuzchen lenkt, mit denen die Befragten die als besonders störend empfundenen Probleme gekennzeichnet haben. Aber auch ohne Schablone wird der Auswerter diese Probleme schnell identifizieren. Es empfiehlt sich jedoch, die besonders störenden Probleme zu extrahieren und dann auszuzählen, wie viele Benutzer in der Stichprobe hinsichtlich der Problembewertung übereinstimmen (siehe Auswertungsleitfaden, Abschnitt 5.10).

In den meisten Fällen lohnt es sich, einen externen Experten als Prüfer hinzuzuziehen. Denn die Prüfung eines Produkts und die Bewertung der Nutzungsprobleme kann von

einem Experten konstruktiv angegangen werden. Schließlich soll es ja nicht bei der Mängelfeststellung bleiben. Es soll einen konstruktiven Anstoß zur Produktverbesserung geben. Ein Prüfer, der die Normen der Reihe DIN EN ISO 9241, Teile 10 bis 17 kennt, wird einen Vorschlag für eine normkonforme Lösung finden. Als Externer ist der Prüfer unbefangen und kann u.U. Ursachen für Nutzungsprobleme im Nutzungskontext erkennen, so dass Anwender oder Benutzer angeregt werden, die Nutzungssituation selber zu verbessern.

5.10 Auswertungsleitfaden

Zwei Arten der Auswertung des Fragebogens werden betrachtet:

1. Auswertung durch den Anwender "vor Ort", um die Entscheidung vorzubereiten, ob weitere Untersuchungen (durch einen Experten) notwendig sind.
2. Berücksichtigung subjektiver Bewertungen bei der konstruktiven Evaluierung durch einen Experten im Rahmen der Konformitätsprüfung oder eines anderen Qualitätssicherungsprozesses.

Der zweite Fall wird in diesem Handbuch bereits ausführlich beschrieben, so dass nur noch ein Leitfaden für den ergonomisch ungeschulten Anwender vor Ort notwendig ist; denn eines der Ziele bei der Entwicklung des Fragebogens war, diesen auch ohne Hinzuziehung eines ausgewiesenen Experten auszuwerten zu können. Dabei sollen die in der Tabelle 5.10.-1 dargestellten Abhängigkeiten sichergestellt werden:

	Fragebogenergebnisse legen Prüfung durch Experten nahe	Fragebogenergebnisse ergeben keine Hinweise auf Probleme
Gravierende Nutzungsprobleme vorhanden ¹⁰	A Typischer Fall	B Fehlerfall
Keine gravierenden Nutzungsprobleme vorhanden	C Der "übersichtliche" Fall	D OK

¹⁰ Diese Aussagen beziehen sich selbstverständlich nur auf den untersuchten Aufgabenbereich und treffen keinesfalls zwingend für ein ganzes Softwareprodukt zu.

Tabelle 5.10-1: Ziele bei der Auswertung des Fragebogens durch den Anwender

- Fall A und Fall B: Ziel der Auswertung der Fragebogenergebnisse ist es, tatsächliche Mängel anhand der in den beantworteten Fragen angezeigten potentiellen Mängel aufzudecken; bei der Auswertung wird davon ausgegangen, dass der Fall B aufgrund der Eigenschaften des Fragebogens vermieden wird. Zusätzlich kann bei der Auswertung der Antworten darauf geachtet werden, dass angezeigte Nutzungsprobleme nicht "unter den Tisch" fallen (Fall B).
- Fall C: Da bei der Auswertung noch keine Erhärtungsprüfung (Abschnitt 3.3) stattfindet, kann dieser Fall nicht zuverlässig verhindert werden, ohne dass die Wahrscheinlichkeit für das Auftreten des Fehlerfalles B steigt. Besser ist es, zunächst jede durch den Benutzer geäußerte Problemdarstellung ernst zu nehmen, auch wenn sich anschließend in der Erhärtungsprüfung zeigt, dass man „übersichtliche“ vorgegangen ist.
- Fall D: Analog zu Fall A.

Der Fragebogen unterscheidet implizit drei Antworten¹¹:

- Positive Antwort: "Es gibt kein Problem"
Dabei ist zu beachten, dass die "positive" Antwort abhängig von den Fragen die "Ja" oder die "Nein"-Antwort sein kann. Eine Aufstellung der jeweils negativen Antworten findet sich am Ende dieses Abschnittes.
- Negative Antwort: "Es gibt ein Problem, nämlich ..."
- Negative Antwort mit besonderer Gewichtung: "Es gibt ein Problem, nämlich... und ich empfinde dies als sehr störend"

Antworten der befragten Benutzer werden als Symptome für *gravierende* Mängel angesehen. Idealerweise sollte es bei der Auswertung möglich sein, die geäußerten Beschreibungen der Nutzungsprobleme für jede der Fragen zu sammeln und für jedes identifizierte Nutzungsproblem die Häufigkeit festzustellen. Dieses Vorgehen birgt aber für ungeübte Auswerter die Gefahr, dass verschiedene Beschreibungen, die in

¹¹ Der vierte Antwortfall, "Frage trifft nicht zu", braucht bei der Auswertung nicht berücksichtigt zu werden.

Wirklichkeit Symptome für ein- und dasselbe Problem sind, getrennt klassifiziert, als Einzelfälle angesehen und vernachlässigt werden. Nachfolgendes Auswertungsschema soll dieser Gefahr vorbeugen.

1. Fragen, die durchweg „positiv“ (im Sinne von nicht problemhaltig) beantwortet werden, können als Zeichen der Unbedenklichkeit angesehen werden; allerdings dürfen diese Antworten wegen der Natur einer subjektiven Befragung nicht als Konformitätsaussage genutzt werden dürfen.
2. Fragen, die einmal oder mehrfach „negativ“ (im Sinne von problemhaltig) beantwortet wurden *und* bei denen mindestens ein Befragter das Feld „Ich empfinde dies als sehr störend“ angekreuzt hat, sind als Indikatoren eines potentiellen Mangels zu deuten und legen eine Prüfung, ggf. unter Hinzuziehung eines Experten, nahe.
3. Fragen, die zwar einmal oder mehrfach „negativ“ beantwortet wurden, die aber nie als besonders störend angekreuzt wurden, deuten auf möglicherweise zu tolerierende Mängel hin. Um auch hier auf der sicheren Seite zu bleiben, empfiehlt es sich, eine weitere Untersuchung einzuleiten, wenn solche Antworten mehr als vereinzelt (d.h. mehr als 10%¹²) auftreten.

¹² Die Festlegung dieses Schwellenwertes erfolgt willkürlich auf Basis der Überlegung, dass Symptome, die von weniger als 10% der Befragten angezeigt werden *und* von diesen nicht als „sehr störend“ bewertet werden, nicht ein gravierendes Problem beschreiben.

Die Auswertungsergebnisse können in einer Tabelle zusammengefasst werden, die ein schematisches Abarbeiten der angezeigten Nutzungsprobleme erleichtert (siehe nachstehendes Beispiel):

Beispiel		Negativantwort ist	Anzahl der Negativantworten	Anzahl „Ich empfinde dies als sehr störend“	Anzahl „sehr störend“ > 0 ODER Anzahl Negativantworten > 10% der abgegebenen Fragebögen?
Aufgabenangemessenheit					
1. Enthält das Programm alle für Ihre Aufgabe benötigten Funktionen?	Nein	0	0		Nein
2. Müssen Sie Eingaben oder Dialogschritte machen, die eigentlich überflüssig wären?	Ja	3	1		Ja (da mindestens einmal als sehr störend eingestuft)
3. Ist es Ihnen möglich, das wiederholte Eingeben von Daten oder Texten zu vereinfachen?	Nein	4	0		Ja (angenommen es sind 20 Fragebögen eingegangen; Anzahl der Negativantworten > 10%)
4. ...					
Auswertung	Sind potenzielle gravierende Mängel festgestellt worden, d.h. sind in der rechten Spalte ein oder mehrere „Ja“ eingetragen?				Ja

Tabelle 5.10-2: Beispiel für schematische Auswertung von Fragebogenergebnissen

Die rechtste Spalte von Tabelle 5.10-2 liefert Indikatoren für potentiell gravierende Nutzungsprobleme, die nahe legen, einen Experten mit einer Erhärtungsprüfung zu beauftragen.

Die vollständige Tabelle zur schematischen Auswertung der Fragebogenergebnisse sollte dem Auswerter zusammen mit einer kurzen Anweisung zu Einsatz und Auswertung des Fragebogens ausgehändigt werden (siehe Tabelle 5.10-3).

Frage	Negativantwort ist	Anzahl der Negativantworten	Anzahl "Ich empfinde dies als sehr störend"	Anzahl "sehr störend" > 0 ODER Anzahl Negativantworten > 10% der abgegebenen Fragebögen?
Aufgabenangemessenheit				
1. Enthält das Programm alle für Ihre Aufgabe benötigten Funktionen?	Nein			
2. Müssen Sie Eingaben oder Dialogschritte machen, die eigentlich überflüssig wären?	Ja			
3. Ist es Ihnen möglich, das wiederholte Eingeben von Daten oder Texten zu vereinfachen?	Nein			
4. Finden Sie, dass der erforderliche Aufwand für Ihr Arbeitsergebnis jeweils angemessen ist?	Nein			
5. Haben Sie das Gefühl, dass Sie Arbeiten machen müssen, die besser das Programm erledigen sollte?	Ja			
6. Müssen Sie Werte und Texte eingeben, die der Computer eigentlich wissen könnte?	Ja			
7. Müssen Sie sich mit Umwegen oder Tricks behelfen, um Ihre Arbeitsergebnisse so zu erzielen, wie Sie diese haben möchten?	Ja			
8. Finden Sie in dem Programm Hilfetexte, die Ihnen auch tatsächlich weiterhelfen?	Nein			
9. Passt das Programm zu Ihren Formularen und bisherigen Formaten?	Nein			
Selbstbeschreibungsfähigkeit				
1. Sind die Informationen, die zur Erledigung der Aufgabe notwendig sind, auf dem Bildschirm übersichtlich verfügbar?	Nein			

Frage	Negativantwort ist	Anzahl der Negativantworten	Anzahl "Ich empfinde dies als sehr störend"	Anzahl "sehr störend" > 0 ODER Anzahl Negativantworten > 10% der abgegebenen Fragebögen?
2. Können Sie bei der Arbeit mit dem Programm erkennen, welche Eingabe als nächstes von Ihnen erwartet wird?	Nein			
3. Sind die Meldungen des Systems für Sie immer verständlich?	Nein			
4. Werden Sie vor Aktionen, die nicht rückgängig gemacht werden können, von der Software gewarnt?	Nein			
5. Hilft Ihnen die Hilfefunktion wirklich weiter, wenn einmal ein Dialogschritt oder Menüpunkt nicht ganz klar ist?	Nein			
6. Müssen Sie oft Kollegen oder ein Handbuch konsultieren, um weiterarbeiten zu können?	Ja			
Steuerbarkeit				
1. Können Sie Ihre Arbeitsschritte in der Reihenfolge erledigen, die Ihnen am sinnvollsten erscheint?	Nein			
2. Macht das Programm manchmal etwas, ohne dass Sie es zu dem Zeitpunkt wollen?	Ja			
3. Können Sie bei Bedarf eine Aufgabe unterbrechen und später wieder fortsetzen, ohne alles neu eingeben zu müssen?	Nein			
4. Können Sie einen Arbeitsschritt wieder zurücknehmen, wenn es für Ihre Aufgabenerledigung zweckmäßig ist?	Nein			
5. Fühlen Sie sich in Ihrem Arbeitstempo durch das Programm manchmal gebremst, z.B. durch zu lange Wartezeiten?	Ja			
Erwartungskonformität				
1. Finden Sie Menüpunkte oder Funktionen dort, wo sie Ihrer Meinung nach auch sein sollten?	Nein			
2. Sind Sie sich bei Wartezeiten immer noch sicher, ob das Programm weiterarbeitet?	Nein			

Frage	Negativantwort ist	Anzahl der Negativantworten	Anzahl "Ich empfinde dies als sehr störend"	Anzahl "sehr störend" > 0 ODER Anzahl Negativantworten > 10% der abgegebenen Fragebögen?
3. Sind Sie manchmal überrascht, wie das Programm auf Ihre Eingabe reagiert?	Ja			
Fehlertoleranz				
1. Bekommen Sie bei fehlerhaften Eingaben Korrekturhinweise?	Nein			
2. Können Sie die Folgen einer fehlerhaften Eingabe mit geringem Aufwand beheben?	Nein			
3. Arbeitet das Programm während der Ausführung Ihrer Aufgabe immer stabil und zuverlässig?	Nein			
Individualisierbarkeit				
1. Können Sie am Computer alles so einstellen, dass Ihnen das Lesen und Arbeiten leichter fällt?	Nein			
Lernförderlichkeit				
1. Ermöglicht Ihnen das Programm, auch einmal etwas gefahrlos auszuprobieren?	Nein			
Auswertung: Besteht weiterer Untersuchungsbedarf?	Sind potentiell gravierende Mängel festgestellt worden, d.h. sind in der rechten Spalte ein oder mehrere „Ja“ eingetragen?			

6 Existierende Prüf- und Bewertungsverfahren

In dem Vorläufer der Norm ISO 9241-10, der deutschen Norm DIN 66 234 Teil 8, wurde von den Autoren ehrlich zugegeben, dass noch kein Verfahren existiert, um diesen Teil der Norm für die Konformitätsprüfung von Softwareprodukten zu verwenden. Vielleicht hat dieser Satz dazu beigetragen, dass die Norm bei den deutschen Software-Herstellern weitgehend unbekannt geblieben ist; denn ein fehlendes Konformitätsverfahren kann leicht als Alibi für die Nichtbeachtung einer Norm dienen (Beimel, Schindler und Wandke, 1992).

Vielleicht hat der Hinweis auf fehlende Prüfverfahren viele Autoren dazu angeregt, eigene Verfahren zu entwickeln, um die Lücke zu schließen. Die Auswertung der Literatur zu diesen Verfahren soll aufklären helfen, inwieweit die vielfältigen Bemühungen zum Erfolg geführt haben und was man aus der Entwicklung dieser Verfahren für das hier vorgestellte Prüfverfahren lernen kann.

Um das Ergebnis der Recherche vorwegzunehmen: Ein Verfahren zur Konformitätsprüfung wird mit dem hier vorgelegten Bericht zum ersten Mal vorgestellt. Benutzerfragebögen allein hingegen gibt es genug. Benutzerfragebögen sind aber zur Konformitätsprüfung nicht geeignet, auch nicht der in diesem Bericht vorgelegte.

6.1 Literaturlauswertung

Verfahren zur Software-Evaluierung lassen sich unterteilen in solche, die subjektive Bewertungen der Benutzer über ein Softwareprodukt erfassen (Benutzerfragebögen) und in solche, die Experten bei der Evaluierung von Software anleiten und unterstützen. Benutzerfragebögen können lediglich zur Messung der Zufriedenstellung der Benutzer angewendet werden, auch wenn viele Autoren der Meinung sind, sie könnten damit den Grad der Gebrauchstauglichkeit der Software messen oder gar eine Konformitätsprüfung durchführen. Eine Konformitätsaussage kann nur durch ein objektives Expertenurteil zustande kommen, wobei der Experte die subjektive Zufriedenstellung der Benutzer mit ihrer Software auf keinen Fall vernachlässigen darf. Verminderte

Zufriedenstellung der Benutzer kann ein Indikator für Non-Konformität von Produktmerkmalen sein, jedoch muss der Experte bedenken, dass die Zufriedenstellung eines Benutzers nicht nur durch Softwareeigenschaften beeinflusst wird, sondern auch die Einstellung des Benutzers zur Aufgabe und zum Nutzungskontext spiegelt.

Die wichtigste methodische Grundlage für jede Art der Software-Evaluierung ist die eindeutige Bestimmung des Nutzungskontextes und der darin enthaltenen Erfordernisse (implied needs, vgl. ISO 8402, 1994). Eine Konformitätsprüfung zu DIN EN ISO 9241-10 (1995) kann ohne Berücksichtigung des Nutzungskontextes (DIN EN ISO 9241-11, 1998) nicht durchgeführt werden, da die genormten Forderungen mit Blick auf die Erfordernisse des Nutzungskontextes interpretiert werden müssen, bevor sie sich in Prüfkriterien transformieren lassen.

Bei der Analyse der schon existierenden Verfahren zur Software-Evaluierung ist von besonderem Interesse, inwieweit sich diese Verfahren zur Konformitätsprüfung eignen. Unter dieser Fragestellung ist jedes Verfahren nach folgenden Gesichtspunkten erfasst und bewertet worden:

1. Kurzbeschreibung des Verfahrens:

In der Kurzbeschreibung findet eine Einordnung in den Kontext der Konformitätsprüfung statt. Es wird beschrieben, welchen Stellenwert das Verfahren hat und für welchen Einsatz es von den Autoren vorgesehen ist. Von Bedeutung ist, ob das Verfahren den Anspruch erhebt, den Dialogprinzipien aus DIN EN ISO 9241-10 gerecht zu werden, oder ob andere Aspekte bei der Evaluierung im Vordergrund stehen.

2. Zielgruppe

Es wird festgestellt, wer das Verfahren anwendet (Benutzer oder Experte) und wer die Bewertung vornimmt.

3. Art der Methode

An dieser Stelle wird zum einen dokumentiert, ob es sich um ein subjektives Verfahren, um eine Anleitung zu einem Expertenurteil oder um eine Mischform handelt, bei der sowohl der Experte, als auch der Benutzer die Software beurteilen. Zum anderen wird die Art der zu erhebenden Daten festgestellt: handelt es sich um eine quantitative Evaluierung (z.B. "ratings") oder um eine qualitative (offene Fragen). Darüber hinaus wird festgestellt, ob der Nutzungskontext explizit bei der Evaluierung berücksichtigt wird und wenn ja, in welcher Form. Auch die psychometrische Qualität wird beschrieben.

4. Bewertung

Ein Verfahren zur Normkonformitätsprüfung von Software, welches den Nutzungskontext nicht berücksichtigt, ist grundsätzlich abzulehnen. Außerdem ist kein Verfahren valide, welches eine Softwarebeurteilung aufgrund von Benutzerfragebögen durchführt, deren Items keinen reproduzierbaren Bezug zu den tatsächlichen Aufgaben der Benutzer haben. Bei den Verfahren, die sich explizit am Teil 10 der Norm orientieren, ist noch zusätzlich zu bewerten, inwieweit der Bezug zur Norm methodisch korrekt umgesetzt ist.

Die im folgenden in alphabetischer Reihenfolge beschriebenen Ansätze stellen eine Auswahl dar, die wesentlich durch Bekanntheit und Zugänglichkeit, Repräsentativität und methodischen Entwicklungsstand bestimmt wurde.

6.2 Beschreibung der Bewertungsverfahren

BTQ Kassel Mainz (Beurteilung von Arbeitsbedingungen am Bildschirm/ Arbeitsplatzanalyse der BTQ nach Bildschirmarbeitsverordnung) 4. Teil: Software (Selbsterhebung)

Referenz: BTQ Kassel Mainz: Beratungsstelle der DAG für Technologiefolgen und Qualifizierung im Bildungswerk der DAG im Lande Hessen e.V., Stand Mai 1998.

Kurzbeschreibung

"Die Prüfliste zur Software" erfasst die subjektive Zufriedenheit der Benutzer im Umgang mit der Software, die sie zur Unterstützung ihrer Haupttätigkeit, d.h. für den

überwiegenden Teil ihrer Arbeit, einsetzen. Folgende 12 Dimensionen werden untersucht:

1. Einarbeitung/Erlernbarkeit
2. Informationsdarstellung
3. Selbstbeschreibungsfähigkeit
4. Erwartungskonformität
5. Aufgabenangemessenheit
6. Dialogführung
7. Steuerbarkeit
8. Individualisierbarkeit
9. Fehlertoleranz
10. Datenschutz / Datensicherheit / Leistungs- und Verhaltenskontrolle
11. Wirkung der Software auf mich
12. Masken

Zu jeder Dimension existiert eine Reihe von "Statements" wie z.B. "Ich empfinde die Farbwahl als angenehm." (Dimension: "Informationsdarstellung"). Die Benutzer sollen auf einer Rating-Skala von "3" (trifft genau zu) bis "0" (trifft nicht zu) beurteilen, inwieweit sie diesen Statements zustimmen. Ferner liegen, alternativ zu den Ratings, für jedes Statement die Antwortkategorien "unklar" und "Fragestellung entfällt" vor. Bei fast jeder Dimension wird der Benutzer, anschließend an die Ratings, zu freier Kritik oder eigenen Anmerkungen aufgefordert. Ausnahmen hierbei sind die Dimensionen "Einarbeitung/Erlernbarkeit", "Selbstbeschreibungsfähigkeit", "Erwartungskonformität".

Zielgruppe

Das Verfahren dient Benutzern zur selbständigen Beurteilung von Software.

Art der Methode

Die Prüfliste ist Teil eines Gesamtanalyseinstruments zur Beurteilung von Bildschirm-Arbeitsplätzen. Es fällt in die Kategorie "2.) Datenerhebung" und dort in die Unterkategorie "c) Selbsterhebung". Somit handelt es sich um ein subjektives Verfahren. Unklar ist, ob es noch eine "Fremderhebung der Software - Anleitung für Experten" oder

ähnliches gibt. Solange nicht geklärt ist, ob die "Software-Prüfliste (Selbsterhebung)" das einzige Software-Analyseinstrument im BTQ-Verfahren ist, kann kein abschließendes Urteil über die Evaluierungsmethode gefällt werden.

Bewertung

Als Prüfverfahren für Konformität mit DIN EN ISO 9241-10 ist BTQ nicht vorgesehen. Zur Bewertung vermuteter Normabweichungen durch Benutzer kann es eingesetzt werden. Fraglich ist, ob der Anspruch des Verfahrens, Übereinstimmung mit den Mindestanforderungen an die Software gemäß EU Bildschirmrichtlinie zu messen, aufrechterhalten werden kann, weil es prinzipiell nicht möglich ist, die "Übereinstimmung" eines subjektiv beurteilten Sachverhalts mit einer Rechtsnorm festzustellen.

Ergonomieprüfer: Prüfliste "Software-Ergonomie" ISONORM und ABETO

Referenz: Döbele-Martin et al., 1993 und 1997; Prümper & Anft, 1993; Prümper, 1997 und 1999; TBS, 1997; Richenhagen, 1996.

Kurzbeschreibung

Die Technologieberatungsstelle beim DGB Landesbezirk NRW (TBS) stellt den "Ergonomieprüfer" als Handlungshilfe zur ergonomischen Gestaltung und Bewertung der Bildschirmarbeit zur Verfügung, um insbesondere den betrieblichen und gewerkschaftlichen Interessenvertretungen bei der Umsetzung der Bildschirmarbeitsverordnung zu helfen. Die Prüfliste "Software-Ergonomie" setzt sich aus einer Kurzfassung und einer Langfassung zusammen. Die Kurzfassung dient zur Suche nach Problemen bei der Softwarenutzung, die mit Hilfe der Langfassung genauer lokalisiert werden können. In der ursprünglichen Form sind die Fragen gegliedert in Dialoggestaltung und Maskengestaltung. In der neuen Fassung des "Ergonomieprüfer" wird der Fragebogen ISONORM 9241/10 (Prümper & Anft, 1993) als Langfassung verwendet. Dieser Fragebogen ist als Benutzerfragebogen zur Unterstützung von iterativen Softwareentwicklungen vorgesehen. Der Fragebogen enthält 35 Fragen zu den Gestaltungsprinzipien der ISO 9241 Teil 10, die mit jeweils fünf Items in Form von semantischen Differentialen operationalisiert werden. ISONORM 9241/10 ist vollständig formalisiert. Es werden in Prümper (1997 und 1999) Gütekriterien zur Reliabilität und

Validität des Verfahrens mitgeteilt. Der Aufwand für Durchführung und Auswertung des Verfahrens ist sehr gering. Diese beiden Erhebungsverfahren sind eingebettet in ein umfassenderes Verfahren, die Arbeitsplatzbeurteilung nach Bildschirmarbeitsverordnung und EU-Richtlinie der TBS Oberhausen (ABETO).

Zielgruppe

Das Verfahren dient Betriebs- und Personalräten in der Rolle als Prüfer, um in Zusammenarbeit mit Benutzern Software zu beurteilen. Der in die Neufassung aufgenommene Fragebogen ISONORM ist als Benutzerfragebogen ohne Beisein von Personen der Personalvertretung einsetzbar.

Art der Methode

Beide Teile der ursprünglichen Fassung bestehen aus einer Reihe von Fragen zu den beiden Themenbereichen "Gestaltung der Dialoge" und "Gestaltung der Masken". Die Kurzfassung ist unterteilt in 5 Fragen zur Dialoggestaltung und 10 Fragen zur Maskengestaltung, dessen Bearbeitungsdauer etwa 10 Minuten beträgt (Burmester et al, 1997: vgl. Tabelle im SANUS-Handbuch: Grobanalyse-Software, Kap. 5.2.3). Die Fragen können von Betriebsräten *zusammen* mit Beschäftigten ausgefüllt werden. Die Auswertung erfolgt durch Auszählen der "Nein"-Antworten, die einen Hinweis auf einen Mangel geben. Mit Hilfe der Fragen in der Langfassung sollen die aufgedeckten Mängel weiter differenziert werden. Hierbei müssen Sachverständige hinzugezogen werden. Der Prüfer hat zum Schluss noch die Möglichkeit, Verbesserungsideen oder Anforderungen, gegliedert nach den Überschriften, in der Langfassung zu notieren. In der ursprünglichen Form können die Fragebogen nicht von den Benutzern selbständig und direkt ausgefüllt werden, da das Vokabular sehr technik-orientiert ist. Das Verfahren stellt also eher eine Mischform von Expertenverfahren und Benutzerbeurteilung dar. Der Nutzungskontext wird bei der Software-Evaluierung nicht ausdrücklich berücksichtigt. Die Antworten in beiden Teilen sind in "ja/nein/weiß nicht" vorgegeben. Offene Antworten sind nicht vorgesehen. Kommentare, Beispiele oder Verbesserungsvorschläge können vom Prüfer im dritten Teil dokumentiert werden.

Bewertung

In der Erfassung des "Ergonomieprüfer" orientieren sich die Fragen zum Dialog an den Dialogprinzipien aus ISO 9241 Teil 10; die Grundsätze der Individualisierbarkeit und der Lernförderlichkeit werden jedoch nicht berücksichtigt. In der Kurzfassung wird in allgemeiner Form nach der Einhaltung der Grundsätze gefragt; die in der Langfassung zu beantwortenden Fragen orientieren sich an den in der Norm formulierten Beispielen. Die Fragen (basierend auf den Beispielen) sind jedoch zum Teil anderen Dialoggrundsätzen zugeordnet als in der Norm. Die Fragen in beiden Fassungen sind mit hoher Wahrscheinlichkeit für Benutzer schwer verständlich, weil viele Fachbegriffe aus der Softwaregestaltung oder Softwareergonomie verwendet werden. Der Benutzer bekommt auch keine Unterstützung bei der Anbindung des Inhalts der Fragen in seinen Arbeits- und Aufgabenbereich. Die Übersetzung der Fragen in ein dem Arbeitsbereich angemessenes Vokabular und den Bezug zur Aufgabenstellung müsste somit der Prüfer herstellen, der mit dem Benutzer zusammen den Arbeitsplatz untersucht. Um die Abbildung der Fragen in den Arbeitskontext nachvollziehbar zu machen, sollte im Fragebogen zumindest Platz für Kommentare und Erläuterungen vorgesehen werden. Die Einteilung in "ja/nein/weiß nicht" ist daher nicht ausreichend. Das Verfahren ist für die Prüfung der Konformität mit ISO 9241-10 nicht geeignet, obwohl der Name "Ergonomieprüfer" dies suggeriert. Auch die Verwendung des Namens ISONORM für den integrierten Benutzerfragebogen ist irreführend, da er einen Konformitätstest nahe legt. - ABETO ist ein "Verfahren, mit dem sich eine komplette Beurteilung der Arbeitsbedingungen nach der BildscharbV und dem ArbSchG durchführen lässt." Dabei werden folgende Aspekte berücksichtigt: Hardware, Arbeitsplatz, Arbeitsaufgabe, Software, Arbeitsgestaltung, Beanspruchung. Im Rahmen von ABETO wird eine Grobanalyse durchgeführt, um offensichtliche Schwachstellen aufzudecken und die Arbeitsbedingungen soweit beurteilen zu können, dass eine Gruppierung der Arbeitsplätze für die Feinanalyse durchgeführt werden kann. Zur Grobanalyse können der 100-Fragen-Test, der Fragebogen zur Beanspruchung, die TBS-Kurzprüfliste oder der Fragebogen KFZA einzeln oder als "Methodenmix" eingesetzt werden. Die Feinanalyse muss von "qualifizierten Personen" durchgeführt werden. Einsetzbare Verfahren zur Feinanalyse sind die TBS-Langchecklisten, der Fragebogen ISONORM 9241/10 (siehe SANUS speziell in diesem Abschnitt) und Beobachtungsinterviews - ein Teil des KABA-

Verfahrens (Dunckel et al., 1993) zur kontrastiven Aufgabenanalyse. Letzteres Verfahren ist geeignet, um Ursachen für die Minderung der subjektiven Zufriedenstellung aufzuklären, die nicht in der Software begründet sind.

Eine Prüfung der Normkonformität ist mit ISONORM 9241/10 nicht möglich. Der Aufgabenbezug bei der Datenerhebung ist unklar. Er wird erst nachträglich im Gespräch mit dem Benutzer hergestellt. Dies wirft die Frage auf, ob die vorgeschlagenen Skalen tatsächlich *valide* Operationalisierungen der Dialoggrundsätze darstellen. Qualitative Nutzungsprobleme werden nicht während der Datenerhebung sondern erst nach einer quantitativen Auswertung in nachträglichen Gesprächen mit den Benutzern analysiert. Dies setzt voraus, dass die Fragebogenerhebung nicht anonymisiert wird. Prümper & Anft (1993) sehen vor, dass Nutzungsprobleme in Gesprächen mit den Benutzern weiter zu Gestaltungshinweisen konkretisiert werden. ISONORM 9241/10 beansprucht, das Kriterium der Einfachheit software-ergonomischer Bewertung durch die Benutzer zu erfüllen, da Anwendung und Auswertung der vorgesehenen Skalen nur geringen Aufwand erfordern. Da im Fragebogen nicht vorgesehen ist, dass die Benutzer zu den Items die Nutzungsprobleme mitteilen, die zur vorgenommenen Bewertung geführt haben, ist eine aufwendige Nacherhebung nötig.

EU-CON II

Software-ergonomische Bewertung von Bildschirmarbeitsplätzen gemäß der EU-Richtlinie

Referenz: Stary et al., 1997, 1999

Kurzbeschreibung

EU-CON II ist ein Verfahren zur Bewertung von Bildschirmarbeitsplätzen gemäß den Anforderungen der EU-Richtlinie 90/270/EWG (1990). EU-CON II setzt sich aus drei Teilen zusammen: einem Vorgehensmodell, welches das phasengerechte Vorgehen bei der Bewertung und Gestaltung der Software beschreibt, einem Leitfaden und einem Handbuch für Software-Prüfer und -Gestalter. Das Vorgehensmodell von EU-CON II legt der Bewertung und Gestaltung von Software vier Aktivitäten zu Grunde:

1. Vorbereitung

In dieser ersten Phase wird angeraten,

- ein innerbetriebliches Projektteam zu bilden,
- die Zielsetzung klar zu definieren und
- eine Schulung der Verantwortlichen durchzuführen.

Zur Qualifikation der Software-Prüfer liefert EU-CON II eine Anleitung in Form eines Handbuchs.

2. Durchführung

In der Durchführungsphase erhalten die Benutzer den Fragebogen und ein Informationspaket zum Fragebogen. Der Fragebogen ist für jede auszuführende Tätigkeit separat zu beantworten. Somit ist die Anforderung erfüllt, dass Software aufgabenspezifisch beurteilt werden sollte. Das Informationspaket liefert Erklärungen und Beispielantworten zu jeder einzelnen Frage.

3. Auswertung

Die Antworten auf die Fragen im EU-CON-Fragebogen können "per Kreuzchen" in Übersichtsbögen eingetragen werden. "Kritische Antwortkästchen" sind im Formular gekennzeichnet. Dies bedeutet, dass der Auswerter auf einen Blick erkennen kann, dass zum Beispiel ein "Nein" bei Frage X auf einen ergonomisch kritischen Zustand hinweist, der näher beleuchtet werden sollte.

4. Verbesserung

Zur Behebung der festgestellten Mängel existiert ein "Handbuch für Evaluatoren und Gestalter". Es gibt Auskunft über die Zusammenhänge zwischen den einzelnen Fragen des Fragebogens und zeigt Handlungsmöglichkeiten auf, welche bei der Umsetzung software-ergonomischer Mindestanforderungen gemäß der EU-Bildschirmrichtlinie relevant sind. Das Handbuch stellt somit eine Anleitung für betriebliche Verantwortliche, Ergonomen oder Fachkräfte für Arbeitssicherheit dar und sollte schon bei der Qualifizierung der Softwareprüfer zum Einsatz kommen.

Zielgruppe

Das Verfahren dient Benutzern zur selbständigen Beurteilung von Software.

Art der Methode

EU-CON ist ein subjektives Verfahren (*Benutzerfragebogen*). Zusätzlich existiert eine detaillierte Anleitung für den Experten, wie er mit den Ergebnissen des Fragebogens umgehen soll (*Handbuch für Evaluatoren und Gestalter*). Der erste Teil des Fragebogens bezieht sich auf die "Aufgabenangemessenheit" der Software. Dazu werden alle Aufgaben, die von dem Beschäftigten mit Hilfe von Software erledigt werden, einzeln analysiert. Es existiert für jede Frage eine "ja"-, "nein"- "weiß nicht-" und ggf. eine "trifft nicht zu"-Antwort zum Ankreuzen. Im Falle einer "nein"-Antwort gibt es bei jeder Frage die Möglichkeit, den Mangel mit eigenen Worten zu beschreiben und ggf. eigene Verbesserungsvorschläge für die Software zu machen. Somit wird für jeden potentiellen Mangel direkt der individuelle Bezug zur Arbeitsaufgabe hergestellt, und der Benutzer wird, durch die Aufforderung, selbst Verbesserungsvorschläge zu machen, in den Verbesserungsprozess der Software mit einbezogen. Im zweiten Teil werden aufgabenunabhängige Fragen gestellt. Hierbei werden "kognitive Faktoren und technische Aspekte der Adaptierbarkeit angesprochen". Die Antwortmöglichkeiten sind analog zum ersten Teil gehalten. Zum Schluss wird der Benutzer aufgefordert, Angaben zu seiner Person zu machen. Von Interesse ist hierbei, ob es sich um einen geübten Benutzer oder Anfänger handelt, ob eine ausreichende Schulung stattgefunden hat und ob der Benutzer an der Gestaltung seiner Software selbst mitwirken konnte. Der Fragebogen endet mit der Möglichkeit, eigene Kommentare und Anmerkungen zur Software zu machen.

Bewertung

Ein direkter Bezug zu DIN EN ISO 9241-10 liegt nicht vor, so dass EU-CON II nicht für Konformitätsprüfungen mit dieser Norm verwendet werden kann. Bezüglich der Übereinstimmungsprüfung mit der EU-Bildschirmrichtlinie gilt dasselbe, was für die Bewertung des BTQ gesagt wurde.

Der Benutzer soll, geführt durch das Informationsblatt, wie ein "Software-Experte" an die Fragestellungen herangehen. Er wird sensibilisiert für die ergonomischen Mängel von Software. Hierdurch ermöglicht der Fragebogen, mehr zu erfassen als bloße Zufriedenheits-Ratings und liefert Hinweise auf mögliche Normabweichungen. In EU-CON II ist allerdings die Trennung von Fragebogen und Informationspaket problematisch. Es ist nicht überprüfbar, inwieweit sich der Befragte vor dem Ausfüllen des Fragebogens wirklich eingehend mit dem Informationspaket auseinandersetzt, denn es handelt sich um ein separates Dokument. Es ist denkbar, dass viele Benutzer einfach nur nach bestem Wissen und Gewissen den Fragebogen beantworten, ohne sich wirklich die Mühe zu machen, sich mit dem jeweiligen Soll-Zustand und den durch die Beispiele beschriebenen Sachverhalten im Informationspaket auseinander zu setzen. Somit ist nicht gewährleistet, dass die Benutzer wirklich das zu beurteilende Gestaltungs-Konzept verstanden haben, bevor sie die Fragen beantworten. Ferner legt das Vorgehensmodell von EU-CON II nahe (vgl. Abbildung 3 des Verfahrens), den Fragebogen vor dem Informationspaket zu lesen. Dies bedeutet, dass zuerst die Fragen und dann der Soll-Zustand sowie der zu beurteilende Sachverhalt den Urteilsprozess des Benutzers beeinflussen. Dies sollte unserer Meinung nach umgekehrt sein: zuerst Lernen, während dieses Prozesses die Problemhaltigkeit der Nutzungssituation erkennen, dann Beurteilen. Zusammenfassend lässt sich jedoch sagen, dass EU-CON in bezug auf das Konzept, Benutzer erst über Soll-Zustände zu informieren, Beurteilungen aufgabenspezifisch vorzunehmen und die Ergebnisse in einem leicht zu interpretierenden Überblicksformular darzustellen, für die Entwicklung eines Benutzerfragebogens Vorbildcharakter hat.

EVADIS II - Evaluation von Dialogsystemen

Referenz: Oppermann et al., 1992

Kurzbeschreibung

EVADIS II ist ein Verfahren zur Bewertung der software-ergonomischen Qualität von Benutzungsschnittstellen im Bürobereich. Das Kernstück von EVADIS II ist ein Leitfaden mit detaillierten "Prüffragen", die als methodengeleitetes Expertenurteil gedacht sind. Die Prüffragen orientieren sich an "Kriterien" zu allen 4 Schnittstellen des IFIP-

Modells (Dzida, 1988). Darin enthalten sind die Dialogprinzipien aus Teil 10. Als begleitende Werkzeuge gibt es einen Fragebogen, um Benutzereigenschaften in die Analyse einbeziehen zu können. Außerdem existiert ein Leitfaden zur Erstellung der Prüfaufgaben (vereinfachte Arbeitsanalyse), der die Auswahl *standardisierter* Prüfaufgaben unterstützen soll. Das Verfahren wird in SANUS (Burmester et al., 1997, Kapitel 6) zum Einsatz bei der Feinanalyse von Software angeführt.

Zielgruppe

Die Bewertung einer Software mit EVADIS II muss ein Experte mit software-ergonomischem und grundlegendem software-technischen Wissen durchführen; eine Einarbeitung in das zu evaluierende Softwareprodukt ist nötig.

Art der Methode

Der Leitfaden besteht aus 150 Prüffragen. Etwa die Hälfte der Prüffragen sind durch Konsultation des Handbuchs oder gezielte Inspektion der Software zu beantworten, während die andere Hälfte zur Beantwortung einen Bearbeitungskontext erfordert. Die Fragen beinhalten Antwortvorgaben, die eine Orientierung zur Beantwortung geben sollen, aber durchaus ergänzt oder erweitert werden können. Es gibt ein Notizfeld für Bemerkungen des Evaluators. Die Auswertung der Fragen soll durch Kommentare erleichtert werden, in denen "Hilfen für die Interpretation und Bewertung der Prüffragen" gegeben werden. Der Prüfer soll die erfragte Eigenschaft mit einer Note zwischen 1 und 5 bewerten und eine Gewichtung (hoch-mittel-gering) zuordnen. EVADIS enthält schließlich eine Anleitung zur Erstellung des Prüfberichts, um die Prüfergebnisse nachvollziehbar und vergleichbar zu dokumentieren. EVADIS II versteht sich als "methodengeleitetes Expertenurteil". Die Evaluierung wird ohne Benutzer durchgeführt. Es wird also nicht in erster Linie nach Nutzungsproblemen oder Mängeln aus Benutzersicht gesucht, sondern eine Merkmals-Analyse vorgenommen, um "objektivierbare Eigenschaften der zu prüfenden Software" zu untersuchen. Ziel ist dabei eine "Beschreibung und Bewertung der software-ergonomischen Qualitäten einer Software" (Oppermann et al. 1992, S. 71). Der Leitfaden wird ergänzt durch einen Fragebogen zur Erfassung von Benutzereigenschaften und Fragen zur Aufgabengestaltung. Die Benutzerbefragung dient dazu, die (zukünftigen) Benutzer einer Kategorie zwischen "geübt" und

"regelmäßige Nutzung" bis "ungeübt" und "sporadische Nutzung" einzuordnen. Aufgrund dieser Einordnung können die software-ergonomischen Kriterien in ihrer Bedeutung für die aktuelle Bewertung gewichtet werden. Die Fragen zum Kontext der Softwarenutzung (Aufgabe, Organisation) sollen eine Bewertung der ergonomischen Qualität der Aufgabengestaltung ermöglichen und liefern die Basis zur Auswahl relevanter Aufgaben und deren detaillierter Beschreibung in Form einer standardisierten Prüfaufgabe. Der Kontextbezug wird anhand der standardisierten Prüfaufgabe hergestellt. Der Kontext wird also indirekt, d.h. unabhängig sowohl von der realen Arbeitssituation als auch von den individuellen Benutzereigenschaften und -belangen berücksichtigt.

Bewertung

Der Leitfaden enthält die Prinzipien aus Teil 10, umfasst aber noch weitere, die entweder als Verfeinerung einzelner Gestaltungsgrundsätze (z.B. Aufgabenangemessenheit) oder als Ergänzung in Bezug auf die Ein-Ausgabe-Schnittstelle und die Organisationsschnittstelle zu verstehen sind. Wenn es einen konkreten Bewertungsanlass gibt oder ein bestimmter Schwerpunkt gesetzt werden soll, lässt sich die Bewertung auf die in diesem Zusammenhang interessierenden Bereiche reduzieren, also z.B. auch auf die Dialogprinzipien aus Teil 10. (Der Gestaltungsgrundsatz der "Aufgabenangemessenheit" wird durch die beiden Kriterien "Nützlichkeit" und "Komfort" erfasst.) Es fällt jedoch auf, dass überwiegend nach dem Vorhandensein von Systemmerkmalen gefragt wird, denen Aufgabenangemessenheit oder andere Gestaltungsgrundsätze unterstellt werden. Die Antworten ergeben zwar eine detaillierte Beschreibung eines für ergonomisch gehaltenen Systems; es wird jedoch nicht untersucht, ob ein Systemmerkmal in einem bestimmten Nutzungskontext überhaupt relevant ist oder dazu beiträgt, ein Aufgabenerfordernis oder eine Tätigkeitsanforderung zu erfüllen. Deshalb ist EVADIS II nicht zur Konformitätsprüfung geeignet; denn konform kann ein Systemmerkmal immer nur mit einer aus dem Nutzungskontext abgeleiteten Tätigkeitsanforderung sein. Die Validität von EVADIS II ist unklar, da die Voraussetzung hierfür, nämlich eine konforme Beurteilungsgrundlage zwischen Experte und Benutzer, fehlt. Einarbeitung und Nutzung von EVADIS II sind sehr aufwendig. Es existiert eine Software-Unterstützung.

IfADo-Fragebogen zur Bewertung von Software

Referenz: Institut für Arbeitsphysiologie (IfADo), (1995)

Kurzbeschreibung

Der Fragebogen zur Bewertung von Software richtet sich an Benutzer. Es werden Daten aus drei Bereichen abgefragt: persönliche Daten über die Arbeit am Bildschirm, Einschätzung der Dialogprinzipien nach ISO 9241 Teil 10 - 12 bei der verwendeten Software und persönliche Bewertungen der verwendeten Software.

Zielgruppe

Das Verfahren dient Benutzern zur selbständigen Bewertung von Software.

Art der Methode

Der Fragebogen besteht aus 45 Fragen, die sich auf die Dialogprinzipien beziehen. Weitere Fragen beziehen sich auf DIN EN ISO 9241 Teile 11 und 12. Es gibt Fragen zur Effektivität, Effizienz und Zufriedenheit und weitere nach der Präsentation der Information (vgl. DIN EN ISO 9241-12). Die Fragen sind auf den Benutzer bezogen formuliert. Sie sollen auf einer Skala zwischen 1 und 7 bewertet werden. Die Wichtigkeit jeder Frage wird ebenfalls auf einer Skala zwischen 1 und 7 eingeschätzt. Über die Auswertung liegt uns keine Information vor. Der Fragebogen zur Bewertung von Software ist ein subjektives Verfahren. Neben Eigenschaften des Programms wird nach subjektiven Kriterien wie Zufriedenheit gefragt. Der Benutzer soll die Prüfaussagen und ihre Wichtigkeit einschätzen. Er hat nur bei der Gesamtbeurteilung die Möglichkeit, ergänzende Kommentare oder Beispiele anzugeben. Der Nutzungskontext wird nicht systematisch in die Untersuchung einbezogen oder erfasst. Es bleibt dem Benutzer überlassen, seine Aufgaben mit den Fragen in Verbindung zu bringen. Er kann dies aber nicht dokumentieren. Der Benutzer soll die Dialoggrundsätze aus Teil 10 auf Ratingskalen beurteilen. Er kann auf einem kontinuierlichen Spektrum angeben, ob die Prüfaussagen "sehr selten" oder "sehr häufig" zutreffen und ob der angesprochene Sachverhalt bzgl. seiner Wichtigkeit als "sehr gering" oder "sehr groß" eingeschätzt wird.

Bewertung

Der Fragebogen basiert auf den Dialogprinzipien aus Teil 10. Negative Urteile der Benutzer können als Hinweise auf vermutete Normabweichungen verwertet werden; dies erfordert jedoch eine Nachuntersuchung, etwa als Inspektion des Systems in Verbindung mit Beobachtungsinterview. Es finden sich nicht alle Empfehlungen des Teils 10 in den Prüfaussagen des Fragebogens wieder. Der Fragebogen ist deshalb für Konformitätstests nicht geeignet. Die Prüffragen sind in der "Ich-Form" formuliert und in der Terminologie dem Benutzer angepasst. Es gibt allerdings keine Erläuterungen zu den Prinzipien und ihrer Intention, die dem Benutzer helfen könnten, den Zusammenhang der Fragenkomplexe mit seiner Arbeit zu erkennen. Der Fragebogen wird vom Benutzer allein ausgefüllt. Der Benutzer hat nur bei der Gesamtbewertung die Möglichkeit, individuelle Kommentare abzugeben oder Beispiele anzuführen. Er kann jedoch keine differenzierten Antworten abgeben, sondern muss sich bezüglich der gefragten Aspekte ein Gesamturteil bilden. Dadurch besteht die Gefahr, dass negative Einzelaspekte der Software durch einen positiven Gesamteindruck verdeckt bleiben. Hinzu kommt, dass der Kontext nicht systematisch erfasst wird, so dass Probleme bei der Softwarenutzung, die durch den Nutzungskontext (Aufgabe, Organisation) verursacht werden, nicht erfasst werden. - Es gibt aus dem IfADo noch einen weiteren Fragebogen ("Uni Dortmund 2"), der zur Befragung der Benutzer durch einen Experten entwickelt wurde. Inwieweit er den ersten ergänzt, ist aus den Unterlagen nicht ersichtlich. Die 22 Prüffragen des zweiten Fragebogens sind noch allgemeiner formuliert; Fragen zur Lernförderlichkeit fehlen ganz. Über die Durchführung der Bewertung (Expertenevaluation, offene Fragen, Fragen nach der Aufgabe, etc.) und der Auswertung liegen uns keine Informationen vor.

IsoMetrics^S und IsoMetrics^L

Referenz: Willumeit et al., 1995, 1996

Kurzbeschreibung

Dieses von Willumeit, Gediga & Hamborg (1995, 1996) entwickelte Verfahren sieht zwei Vorgehensweisen zur Evaluierung der software-ergonomischen Qualität eines

Dialogsystems vor. Zunächst lässt sich mit Hilfe des IsoMetrics^S eine globale Bewertung des Systems vornehmen, die an den Dimensionen der DIN EN ISO 9241 Teil 10 orientiert ist. In Gediga & Hamborg (1999) werden dazu Vergleichswerte für bereits getesteten Dialogsysteme sowie Gütekriterien mitgeteilt. Außerdem werden umfangreiche Hinweise zur Berücksichtigung des Kontextes der Evaluation gegeben. Als kleinste Analyseeinheit für die Untersuchung der Qualitäten eines Dialogsystems wird die Ebene einer Aufgabe empfohlen. Das Verfahren ist vollständig formalisiert, der Aufwand zur Datenerhebung beschränkt sich im Rahmen von Benutzertests auf das Bearbeiten eines Fragebogens und seine Auswertung.

Mit dem Verfahren IsoMetrics^L können anschließend an die im IsoMetrics^S gemachten Einstufungen weitergehende Eindrücke von den Teilnehmern einer Evaluation gesammelt werden. So muss hier angegeben werden, wie wichtig ein Item für den Gesamteindruck ist, der vom Dialogsystem besteht. Außerdem werden die Teilnehmer gebeten, freitextlich konkrete Beispiele zu beschreiben, bei denen dieses Item ihrer Ansicht nach nicht zutrifft. Diese Anmerkungen werden gefiltert und später in eine Diskussion mit Benutzern und Entwicklern zurückgemeldet. Gegenüber dem IsoMetrics^S ist damit natürlich ein höherer Aufwand erforderlich. Dafür werden gezielte Gestaltungshinweise generiert.

IsoMetrics wurde in engem Bezug zur DIN EN ISO 9241 Teil 10 entwickelt, indem Experten gebeten wurden, einen Pool von 151 Items den sieben Dialoggrundsätzen zuzuordnen. Anschließend wurden alle Items eliminiert, bei denen aus Expertensicht kein Bezug zur DIN EN ISO 9241 Teil 10 feststellbar war. Es resultierten sieben Skalen, für die im Verfahren Einzelwerte berechnet werden. Die Variante IsoMetrics^L sieht neben der Einstufung eines Items durch den Benutzer auch Raum für die individuelle Beschreibung von Nutzungsproblemen vor, die später inhaltsanalytisch ausgewertet werden können.

Zielgruppe

Das Verfahren dient Benutzern zur selbständigen Beurteilung von Software.

Art der Methode

Bei IsoMetrics handelt es sich um einen Benutzerfragebogen. Es ist somit als subjektives Verfahren anzusehen. Jedes Item wird auf einer fünf-stufigen Ratingskala beurteilt. Eine weitere Beurteilungs-Kategorie, "keine Meinung", soll dazu dienen, einer Einstufung auf der Ratingskala auszuweichen, um willkürliches Ankreuzen zu vermeiden. Zusätzlich soll der befragte Benutzer bei jedem Item einschätzen, inwieweit das in dem Item angesprochene Merkmal oder der in dem Item angesprochene Sachverhalt seinen "Gesamteindruck" von der Software prägt. Um herauszufinden, wo genau die Schwachstellen des Systems liegen, wird im Anschluss an die "Gewichtung" des Items noch nach einem konkreten Beispiel gefragt. Somit kann ein zusätzliches Interview mit den Benutzern entfallen und die Befragung anonymisiert durchgeführt werden. Die Benutzer werden vor Anwendung des Fragebogens hinsichtlich ihres Alters, ihrer Berufserfahrung, ihrer EDV-Erfahrung usw. analysiert.

Bewertung

Eine explizite Aufgaben- und Kontextanalyse findet beim Anwenden von IsoMetrics nicht statt. Der Fragebogen liefert kein zusammenhängendes Bild von der Gesamtaufgabe oder den Teilaufgaben, die mit der Software verrichtet werden, obwohl ein gewisser Aufgabenbezug, durch die Möglichkeit, konkrete Beispiele anzugeben, hergestellt wird. IsoMetrics orientiert sich zwar an den Dialoggrundsätzen der ISO 9241-10, ist aber als Konformitätstest ungeeignet, da es einerseits lediglich die Zufriedenheit der Benutzer mit ihrer Software erfasst und andererseits keinen direkten Bezug zu den einzelnen Anforderungen der ISO 9241-10 herstellt. Zur Benutzerbewertung vermuteter Normabweichungen ist das Verfahren geeignet. Aufgrund der großen Itemanzahl ist IsoMetrics nicht ökonomisch einsetzbar. Leider wird das Falsifikationsprinzip nicht unmittelbar unterstützt, da itemweise keine begründeten Grenzwerte mitgeteilt werden. Trotzdem hat die Entwicklung von IsoMetrics Vorbildcharakter für die Entwicklung eines Benutzerfragebogens.

MUSiC and SUMI

Measuring Usability of Systems in Context Software Usability Measurement Inventory

Referenz: Bevan, 1994; Bevan et al., 1994; Maclead, 1994; Porteous and Kirakowski, 1992; Kirakowski and Corbett, 1993.

Kurzbeschreibung

MUSiC ist im Rahmen eines ESPRIT Projektes entstanden und ist ein Verfahren zur Erfassung der "Usability" bzw. "Quality in Use" von Software in Anlehnung an die EN ISO 9241-11. Zu diesem Zweck wird zunächst der Nutzungskontext systematisch analysiert, um anschließend anhand von Tätigkeits-Messungen in einem repräsentativen Kontext die Effektivität und Effizienz, mit der die Benutzer ihre Aufgaben erledigen, zu untersuchen. Die Indikatoren für Effektivität und Effizienz sind der Grad der Vollständigkeit, mit dem bestimmte Aufgaben bearbeitet werden und die Zeit, die zur Aufgabebearbeitung benötigt wird. Zur Bewertung der Tätigkeiten (performance) wird eine Benutzungssituation mit einem Videorecorder aufgenommen und mit Hilfe der DRUM (Diagnostic Recorder for Usability Measurement)-Software analysiert. Da im Englischen „performance“ sowohl Leistung als auch Tätigkeit bedeuten und da die Zeit zur Erledigung von Aufgaben gemessen wird, liegt eine einseitige Interpretation der Messergebnisse im Sinne von Benutzerleistung (Benutzertest) nahe. DRUM ermöglicht unter anderem eine Differenzierung zwischen produktiv verbrachter Zeit und der Zeit, die dafür aufgewendet wird, um z. B. Benutzungsfehler zu korrigieren oder Hilfe zu suchen. Ferner wird die mentale Beanspruchung bei der Bearbeitung der mit der Software zu verrichtenden Aufgaben erfasst. In diesem Rahmen können objektive Messungen (zum Beispiel Herzrate) und subjektive Messungen (persönlich erlebte Beanspruchung) vorgenommen werden.

Zusätzlich wird die subjektive Bewertung der Software durch den SUMI-(Software Usability Measurement Inventory) Benutzerfragebogen erhoben. Der Fragebogen besteht aus 50 Items, die sich folgenden Dimensionen zuordnen lassen:

- Affect (generelle, emotional gefärbte Beurteilung des Dialogsystems)

- Control (Ausmaß der wahrgenommenen Kontrolle über das System)
- Efficiency (Effizienz der Aufgabebearbeitung mit dem System)
- Helpfulness (Selbstbeschreibungsfähigkeit von Dialogsystem und Dokumentation)
- Learnability (Lernförderlichkeit des Systems)

Kirakowski (1995) stellt fest, dass mit Ausnahme der Skala Affect eine deutliche Übereinstimmung mit den Prinzipien der ISO 9241 Teil 10 vorliegt, obwohl SUMI nicht mit dem Anspruch entwickelt wurde, eine solche zu erreichen. Außer der summarischen Bewertung eines Dialogsystems auf den fünf Skalen wird von Kirakowski (1995) eine Verteilungsanalyse der Scores einzelner Items vorgeschlagen. Mit dieser Methode können Items identifiziert werden, bezüglich derer ein Dialogsystem besonders deutlich von den Normen abweicht. Darauf aufbauend können die Benutzer gezielt nach den besonderen Schwächen des Systems befragt werden. SUMI bietet keine besonderen Methoden zur Beschreibung des Kontextes einer Evaluation an. Das Verfahren bezieht sich explizit auf die Bewertung des Gesamtsystems. Es ist voll standardisiert; der Prüfungs- und Auswertungsaufwand ist als gering zu beurteilen. SUMI kann bei Vorliegen eines Prototypen und für fertige Dialogsysteme eingesetzt werden. SUMI sieht im Rahmen der Verteilungsanalyse vor, dass die Benutzer die Nutzungsprobleme schildern, die zur negativen Einstufung eines Items geführt haben. Es wird angeraten, die Ergebnisse des SUMI-Fragebogens nach ihrer Erhebung mit Blick auf den Nutzungskontext zu interpretieren. Besondere Sorgfalt wurde auf die Übertragung des Fragebogens in mehrere Sprachen verwandt, auch eine deutsche Version liegt vor. Die Auswertung des Fragebogens kann mit Hilfe der speziellen Software SUMISCO erfolgen.

Zielgruppen

Das MUSiC-Verfahren dient Experten zur Erfassung von Daten des Nutzungskontexts und der Messung der Effektivität und Effizienz von Tätigkeiten, die am Dialogsystem von Benutzern ausgeführt werden.

Das SUMI-Verfahren dient Benutzern zur selbständigen Beurteilung von Software.

Art der Methode

Das MUSIC-Verfahren beinhaltet sowohl Methoden zur Expertenevaluierung als auch subjektive Messinstrumente, die gemeinsam Aufschluss über die "Usability" der Software geben sollen: Die Gesamtbeurteilung der Software hängt einerseits von dem Expertenurteil ab, welches sich auf die Tätigkeits-Messungen bezieht und andererseits von dem Ausmaß der subjektiven Zufriedenstellung der Benutzer, das durch den SUMI-Fragebogen gemessen wird. Für die Gesamtbewertung spielt das Ausmaß der mentalen Beanspruchung eine Rolle, das sowohl objektiv als auch subjektiv gemessen wird. Die Kontextanalyse wird durch einen "Usability Context Analyse Guide" ("UCA") gestützt, in dem eine Anleitung gegeben wird, anhand welcher Merkmale die verschiedenen Komponenten des Nutzungsumfeldes ("Users", "Tasks", Equipment", "Environment") näher beschrieben werden können. Es wird angeraten die Kontextanalyse von mehr als einem Prüfer durchführen zu lassen. Nachdem der Kontext beschrieben ist, wird analysiert, ob er Merkmale enthält, die eine Usability (-Messung) gefährden könnten (ob z. B. Benutzer mit sehr unterschiedlicher Erfahrung gegeben sind). Anhand der ermittelten Kontextdaten wird entschieden, inwieweit man diese Merkmale während der Evaluation kontrolliert, überwacht oder ignoriert.

Bewertung

MUSIC weist keinen direkten Bezug zu EN ISO 9241-10 auf. Ein durch MUSIC festgestelltes hohes Maß an "Usability" lässt nicht zwangsläufig darauf schließen, dass die Software konform mit den Dialogprinzipien und den zugehörigen Anforderungen dieser Norm ist. MUSIC ermöglicht durch die Tätigkeits-Messungen lediglich Vergleichsmessungen zwischen schon bestehenden Produkten oder zwischen Versionen von Prototypen in verschiedenen Designstadien. MUSIC ist nicht zur Produktbewertung geeignet, da es lediglich die mit dem Produkt, im Nutzungskontext des Produkts, ausgeführten Arbeitstätigkeiten (quality in use) bewerten hilft. Werden diese Tätigkeitsbewertungen als Indikatoren für Produktqualität verwendet, so ist es meist schwierig zu unterscheiden, ob die gemessene Qualität der Tätigkeit haupt-

sächlich vom Produkt oder von einem anderen Faktor des Nutzungskontexts beeinflusst ist.

Das SUMI-Verfahren, das auch unabhängig von MUSIC eingesetzt werden kann, ist geeignet, subjektive Bewertung der Software zu erheben, um Hinweise auf mögliche Normabweichungen zu bekommen. Hierzu ist aber wegen des fehlenden Aufgabenbezugs des Fragebogens im Anschluss an die Ausfüllung des Fragebogens ein gesondertes Interview nötig, was die Anonymisierung der Befragung erschwert. Der Anspruch von SUMI, ein Inventar zur Messung von Software-Usability zu sein, ist nicht aufrechtzuerhalten, da es lediglich ein Verfahren zur Erhebung subjektiver Bewertungen ist. Subjektive Bewertungen allein erlauben keineswegs eine Beurteilung der Gebrauchstauglichkeit von Software. Kirakowski (1995) beschreibt die umfangreichen Entwicklungsarbeiten, die zu einem psychometrisch verlässlichen Instrument geführt haben. Für SUMI liegen neben den üblichen Gütekriterien umfangreiche Normwerte vor. Darüber hinaus hat sich das Verfahren bereits in vielen Validierungsstudien bewährt. Für die Evaluation von Prototypen ist wichtig, dass dabei auch geringe Veränderungen an einem Dialogsystem zu bedeutsamen Veränderungen in den Scores geführt haben. SUMI dürfte z.Zt. der psychometrisch am solidesten konstruierte Benutzerfragebogen zur globalen Bewertung von Dialogsystemen sein.

SANUS (allgemein)

Referenz: Burmester et al., 1997

Kurzbeschreibung

Das Projekt SANUS (Sicherheit und Gesundheitsschutz bei der Arbeit an Bildschirmen auf der Basis internationaler Normen Und Standards) hatte zum Ziel, Instrumente zur Verfügung zu stellen, die eine praktische Umsetzung der Anforderungen der Bildschirmarbeitsverordnung und des Arbeitsschutzgesetzes im Unternehmen unterstützen. Daher sollen die entwickelten Verfahren auch von Nicht-Experten der Software-Ergonomie anwendbar sein, z.B. von den Personen, die im Unternehmen für die Umsetzung des Arbeitsschutzgesetzes und der Bildschirmarbeitsverordnung verantwortlich sind, oder von Praktikern vor Ort, Arbeitgebern oder betroffenen Beschäftigten.

Zielgruppen

Das Verfahren dient Benutzern zur selbständigen Beurteilung von Software. Einige in SANUS eingesetzte Methoden, die der Feinanalyse von Mängeln dienen können, erfordern jedoch Expertenkenntnisse.

Art der Methode

Im Rahmen von SANUS wurden verschiedene, bereits existierende Instrumente zur Grob- und Feinanalyse von Hardware, Software (Qualitatives Software Screening, SHIVA) und Arbeitsorganisation sowie psychischer Belastung zusammengetragen und angepasst. Um die in einem Betrieb eingesetzte Software zu bewerten, wird im Rahmen eines Vorgehensmodells eine Grob- und anschließend eine Feinanalyse der Software durchgeführt (sog. Screening). Ziel ist die Aufdeckung von Schwachstellen und deren vertiefte Analyse, um die gefundenen Mängel zu beseitigen. Für die Grobanalyse der Software werden existierende Verfahren (IsoMetrics, Ergonomieprüfer und ISONORM 9241-10) und das im Rahmen von SANUS weiterentwickelte Werkzeug "Qualitatives Software Screening" empfohlen, das auf dem Fragebogen ISONORM 9241-10 basiert. Für Mängel, die im Rahmen der Grobanalyse entdeckt werden, kann eines der Verfahren zur Feinanalyse von Software (SHIVA, EVADIS II, Usability Tests) angewendet werden. Diese Verfahren haben eine umfassende Bewertung allgemeiner und aufgabenspezifischer Merkmale zum Ziel und erfordern zu ihrer Durchführung software-ergonomische Kenntnisse. Die zur Grobanalyse geeigneten Verfahren sind subjektive Verfahren, während die Feinanalyse eine Expertenevaluation erfordert. Bei der Software-Evaluierung wird der Nutzungskontext nicht explizit berücksichtigt. Ausgangspunkt für die Softwarebewertung ist das Programm und nicht die Aufgabe oder die Tätigkeit des Benutzers, aus der sich die Anforderungen an die Nutzung des Programms ergeben.

Bewertung

Die Bereiche Hardware, Software und Arbeitsorganisation werden getrennt analysiert. Beziehungen zwischen den einzelnen Bereichen sind nicht Untersuchungsgegenstand. Es wird also nicht untersucht, ob ein Mangel in einem Bereich durch eine Maßnahme in einem anderen kompensiert werden könnte. Im Zusammenhang mit ISO 9241-10 interessiert hier nur der Teil des Verfahrens (SANUS-Vorgehensmodell), der sich mit der Grob- und Feinanalyse von Software befasst. Diese Verfahren werden im folgenden einzeln dargestellt und bewertet.

SANUS (speziell): Qualitatives Software Screening Verfahren

Referenz: Burmester et al., 1997

Kurzbeschreibung

Verfahren zur Grobanalyse einer Software-Anwendung im Rahmen von SANUS. Durch zwei aufeinanderfolgende Befragungen der Benutzer mit Hilfe von Fragenkatalogen sollen Hinweise auf Schwachstellen gesammelt werden. Die Schwachstellen werden dann mit Hilfe anderer Verfahren zur Feinanalyse weiter untersucht. Die Grobanalyse kann von "Umsetzungsbeauftragten" ohne software-ergonomische Kenntnisse durchgeführt werden.

Zielgruppen: siehe SANUS allgemein

Art der Methode

Das Verfahren basiert auf dem Fragenkatalog "ISONORM 9241/10" (Prümper & Anft). Es besteht aus zwei Teilen: zuerst bearbeitet der Benutzer selbst einen Fragebogen mit 35 Fragen. Die Auswertung erfolgt anhand von Handlungsanleitungen, die im wesentlichen eine Mittelwertbildung über die Bewertungen aller Benutzer bezüglich eines Dialogprinzips ist. Die Dialogprinzipien, bei deren Bewertung ein bestimmter Mittelwert nicht unterschritten wird, müssen nicht weiter untersucht werden. Negative Einzelbewertungen können jedoch mit Hilfe eines nachträglichen Interviews des Benutzers durch den Auswerter weiter untersucht werden. Ist für ein oder mehrere

Dialogprinzipien ein Mittelwert, der unter dem kritischen Wert liegt, ermittelt worden, folgt eine direkte Befragung der Benutzer zu den im Fragebogen angegebenen Mängeln, sofern die Benutzer sich damit einverstanden erklärt haben. Dabei werden nur die Prüfaussagen untersucht, die von mindestens 40 % der Benutzer negativ bewertet worden sind. Die Bearbeitungsdauer für den ersten Teil beträgt etwa 20 Minuten; der zweite Teil erfordert 40 Minuten pro Befragung. Wenn viele Bildschirmarbeitsplätze vorhanden sind, wird empfohlen, Stichproben aus gleichen Aufgabenbereichen (die mit der gleichen Software zu bearbeiten sind) zu bilden.

Das "Qualitative Software Screening" ist ein subjektives Verfahren. Der Benutzer soll die kritischen Prüfaussagen allgemein einschätzen; er kann Kommentare abgeben oder Beispiele angeben. Im ersten Teil der Befragung wird kein expliziter Kontextbezug hergestellt. Im zweiten Teil hat der Benutzer die Möglichkeit, Beispiele aus seiner täglichen Arbeit heranzuziehen. Der Nutzungskontext wird nicht systematisch in die Untersuchung einbezogen. Es bleibt dem Benutzer überlassen, welche seiner Tätigkeiten er mit den Prüffragen in Verbindung bringen möchte.

Der erste Teil enthält Ratingskalen, mit deren Hilfe der Benutzer das Programm bezüglich der 7 Dialoggrundsätze aus ISO 9241-10 einschätzen soll. Zu jedem Dialoggrundsatz bekommt er 5 Fragen gestellt. Jeweils abschließend soll er einschätzen, ob dieser Grundsatz für seine Tätigkeit wichtig ist oder nicht. Die Auswertung erfolgt durch Mittelwertbildung über den Aussagen zu einem Dialogprinzip aller Benutzer. Liegt der Mittelwert über einem definierten Kriteriumswert, besteht "kein dringender Verbesserungsbedarf" an dieser Stelle.

Der zweite Teil enthält vertiefende Fragen zu jeder Aussage des ersten Teils. Der Unternehmensbeauftragte befragt die Benutzer, die sich dazu bereit erklärt haben, bezüglich der von diesem (und weiteren) Benutzer negativ bewerteten Bereiche. Die Befragung findet am Arbeitsplatz mit der zu prüfenden Software statt. Der Unternehmensbeauftragte stellt anhand des Fragebogens offene Fragen, bittet um Beispiele und auch um Verbesserungsvorschläge. Die Ergebnisse des ersten und zweiten Teils werden in einem Prüfbericht zusammengefasst.

Bewertung

Beide Fragenkataloge basieren auf den Dialogprinzipien aus ISO 9241-10. Sie sind für Normkonformitätsprüfungen nicht geeignet. Die mit Hilfe der Prüffragen aufgedeckten Schwachstellen können jedoch Hinweise auf mögliche Normverletzungen geben. Es gibt jedoch keine direkte Abbildung zwischen den Prüfaussagen des Fragebogens und den Anforderungen des Teils 10. Die Prüfaussagen im ersten Teil sind sehr allgemein gehalten und beziehen sich hauptsächlich auf die Funktionalität des Programms; dieser Eindruck verstärkt sich, wenn man die spezifischen Fragen zu jeder Aussage im zweiten Teil betrachtet. Es kann nicht vorausgesetzt werden, dass diese Operationalisierung für alle Benutzer unmittelbar verständlich ist und ihnen die Intention der Dialogprinzipien genügend verdeutlicht, die ja nicht funktionsbezogen, sondern tätigkeitsbezogen ist. Der Fragebogen wird vom Benutzer allein ausgefüllt. Der Benutzer hat im ersten Teil der Befragung keine Möglichkeit, individuelle Kommentare abzugeben oder Beispiele anzuführen. Er kann keine differenzierten Antworten abgeben, sondern muss sich bezüglich der gefragten Aspekte ein Gesamturteil bilden. Dadurch besteht die Gefahr, dass negative Einzelaspekte der Software durch einen positiven Gesamteindruck verdeckt bleiben. Hinzu kommt, dass der Kontext nicht systematisch erfasst wird, so dass Probleme bei der Softwarenutzung, die durch den Kontext (Aufgabe, Organisation) verursacht werden, nicht erfasst werden. Die quantitative Auswertung durch Mittelwertbildung stellt nicht sicher, dass in der abschließenden Analyse alle bedeutsamen Mängel aufgedeckt werden. Da Aufgabe und Kontext nicht miteinbezogen werden, kann ein schwerwiegender Mangel, den ein Benutzer bei der Lösung seiner Aufgaben feststellt, durch die positive Bewertung anderer Benutzer (die an anderen Aufgaben arbeiten) aufgehoben werden. Dies gilt insbesondere, wenn bei der Stichprobenbildung verschiedene Aufgabenbereiche zusammengefasst wurden. In der anschließenden qualitativen Befragung wird nur nach zuvor festgestellten negativen Einschätzungen gefragt. In der persönlichen Befragung, die sich der schriftlichen anschließen kann, sollen die festgestellten Schwachstellen genauer (möglichst anhand von Beispielen) ermittelt werden und mögliche Verbesserungsvorschläge und Anmerkungen der Benutzer dokumentiert werden. Es ist fraglich, ob die Prüffragen bzw. -aufträge, die dabei zu bearbeiten sind, von einem

Unternehmensbeauftragten bearbeitet werden können, der keine software-ergonomischen Kenntnisse und keinen genauen Einblick in die Tätigkeit der Benutzer hat.

SHIVA: Structured Human Interface Validation Technique

Referenz: Ziegler und Burmester, 1995; Burmester et al., 1997.

Kurzbeschreibung

SHIVA ist ein rechnergestütztes Werkzeug (Windows 95), das den Evaluator dabei unterstützt, die zu untersuchende Software-Anwendung auf verschiedene Themenschwerpunkte hin zu analysieren. Das Werkzeug soll bei der Feinanalyse im Rahmen von SANUS eingesetzt werden, insbesondere zur Feinanalyse der durch das Qualitative Software Screening aufgedeckten Mängel. SHIVA kann nur im Rahmen einer Expertenevaluation eingesetzt werden.

Zielgruppe

Software-Ergonomie-Prüfer.

Art der Methode

Die Evaluation eines Softwareprodukts mit Hilfe von SHIVA erfolgt in mehreren Durchgängen. Der Prüfer beantwortet zu jedem Durchgang Prüffragen. Diese basieren auf ISO 9241 Teil 10, sowie Teile 12 - 17. Die Prüffragen sind unterteilt in Fragen zur Navigation im Dialog, zum Aufgabenablauf, zur Informationsgestaltung und Konsistenz der Merkmale des Systems. Im ersten Durchgang durchläuft der Evaluator alle "Views" des Systems und beantwortet allgemeine Fragen zur "Benutzbarkeit", die sich zum Teil an ISO 9241-10 orientieren. "Views" sind die verschiedenen Bereiche der Benutzungsschnittstelle, die der Benutzer durchläuft, um seine Aufgaben zu lösen. Das Navigationsmodell zeigt auf, wie der Benutzer die Views aufsuchen kann. Die Views selbst bieten eine spezielle Sicht auf die zu bearbeitenden Objekte. Die Fragen, die der Experte während der Prüfung bearbeitet, beziehen sich entweder auf einen speziellen

View (Informationsdarstellung, Aufgabenangemessenheit, etc.) oder auf die Navigationsmöglichkeiten, die sich von diesem View aus ergeben (Steuerbarkeit, Erwartungskonformität, etc.). Der Evaluator kann auch in Anlehnung an die "Heuristische Evaluation" (Nielsen, 1994) Kommentare und Gestaltungsvorschläge eingeben. Ein weiterer Durchgang orientiert sich an Aufgabenszenarien, die unter Beteiligung der Benutzer erstellt werden. Der Evaluator durchläuft die Szenarien und untersucht die (Sequenz von) Views, die dazu benötigt werden. Dabei stellt er sogenannte "Was wäre, wenn"-Fragen, die sich auf veränderte Aufgabenziele beziehen, um den Funktionsraum zu erforschen, der durch die Szenarien nur fragmentarisch beleuchtet wird. Während im ersten Durchgang untersucht wird, für welche Aufgaben das System geeignet ist, wird im zweiten Durchgang geprüft, ob es auch für die vorgegebenen Szenarien angemessen ist. Aus den Antworten auf die Prüffragen und den Kommentaren oder Gestaltungsvorschlägen des Prüfers wird abschließend automatisch ein Prüfbericht erstellt. SHIVA ist ein Werkzeug, das die Expertenevaluation unterstützt. Ziel ist die ergonomische Bewertung eines "Informationssystems" bzw. die Prüfung der ergonomischen Anforderungen der Bildschirmarbeitsverordnung: "Mit SHIVA können die ergonomischen Anforderungen der BildscharbV auf der Basis des internationalen Ergonomiestandards für Bildschirmarbeitsplätze im Bürobereich ISO 9241 Teil 10 und 12 - 17 überprüft werden." Bei der Prüfung wird ein expliziter Aufgabenbezug hergestellt. Anhand von Szenarien, die von oder mit den Benutzern erstellt werden, werden wichtige und typische Aufgaben der Benutzer erfasst und deren Durchführung im System geprüft. Da der Experte die Aufgabendurchführung übernimmt, bleiben jedoch andere Kontextfaktoren (Organisation, Motivation, technische Ausrüstung usw.) außerhalb der Betrachtung.

Bewertung

Die Bearbeitungsdauer für die Bewertung einer Anwendung wird mit 3 - 4 Tagen angegeben. Daher ist besonders die Frage interessant, wie die Ergebnisse, die durch die Grobanalyse (Qualitative Software Screening) ermittelt wurden, dazu beitragen können, die vermuteten Mängel im Rahmen der Feinanalyse weiter einzugrenzen und genauer zu untersuchen. Dabei sollten auch die Antworten und Kommentare der interviewten Benutzer miteinbezogen werden. Dazu gibt es im SANUS-Handbuch keine

Angaben. Für die Untersuchung der Non-Konformität von Software-Produkten kann SHIVA eingesetzt werden. Der diesem Verfahren zugrunde liegende Qualitätsbegriff ist jedoch unklar, weil nicht ausdrücklich verlangt wird, Qualitätsbewertungen auf der Grundlage von Anforderungen zu machen. Der Bezug zu den Aufgaben des Benutzers ist zwar gegeben, aber es wird nicht ausdrücklich verlangt, aus den Szenarien Anforderungen abzuleiten. Eine Aussage über die Konformität eines Produkts mit einer Rechtsnorm (EU-Bildschirmrichtlinie) ist unzulässig, da ein festgestellter Sachverhalt nicht mit einer Rechtsnorm *übereinstimmen* kann. Eine festgestellte Konformität mit ISO 9241-10 oder mit anderen Teilen der Norm kann als gutachtlicher Indikator verwendet werden, der dafür spricht, dass die Rechtsnorm bei der Softwaregestaltung *angewendet* wurde.

SynBA-DQ

Referenz: Wieland-Eckelmann et al., 1996

Kurzbeschreibung

Das Verfahren SynBA-DQ (Synthetische Beanspruchungsanalyse - Dialogqualität) wurde im Rahmen des SANUS-Projektes von Ademmer (1995) entwickelt. Es bedient sich des von Schwarz & Schmitz (1994) und Wieland-Eckelmann et al. (1994, 1996) entwickelten SynBA-Formalismus, der eine simultane Erhebung von Aufgabenmerkmalen, Belastungsfaktoren und der subjektiven Beanspruchung vorsieht. Dazu werden von den Benutzern des Dialogsystems drei Fragebögen bearbeitet. Fragebogen A erfasst die allgemeine Beanspruchungswirkung von 30 Belastungsfaktoren, die eine Operationalisierung der Prinzipien der ISO 9241 Teil 10 darstellen. Fragebogen B befragt anschließend das tatsächliche Auftreten dieser Belastungsfaktoren bei der Arbeit mit dem Dialogsystem. Fragebogen C sieht schließlich eine Bewertung der Belastungsfaktoren hinsichtlich ihrer globalen Beanspruchungswirkung vor. Die Auswertung von Fragebogen C ergibt zunächst eine allgemeine Rangreihe der Belastungsfaktoren hinsichtlich ihrer Beanspruchungswirksamkeit. Damit kann festgestellt werden, welche Belastungsfaktoren aus Sicht der Benutzer besonders wirksam sind und welche weniger. Anhand des Fragebogens B kann dann festgestellt werden, welche dieser Belastungsfaktoren bei der Arbeit mit dem Dialogsystem tatsächlich vorkommen. Für

eine Umgestaltung des Dialogsystems müssen nun zunächst nur solche Belastungsfaktoren näher betrachtet werden, die eine hohe Wirksamkeit haben und auch tatsächlich vorkommen. Fragebogen A kann eingesetzt werden, um eine differenzierte Bewertung der Wirkung zu erhalten, die von den Belastungsfaktoren ausgeht. So soll mit SynBA-DQ das Gestaltungsziel erreicht werden, funktionale Beanspruchungen zu fördern und dysfunktionale zu vermeiden.

Zielgruppe

Benutzer, Arbeitsmediziner.

Art der Methode

SynBA-DQ baut auf einer fundierten Theorie der psychischen Beanspruchung auf (Wieland-Eckelmann, 1992; Wieland-Eckelmann et al., 1994; 1996) und liefert eine vollständig formalisierte Benutzerbefragung.

Bewertung

Im Verfahren werden keine Hinweise zur Erhebung des Kontextes einer Analyse gegeben. Es ist davon auszugehen, dass die niedrigste Analyseebene, die von SynBA-DQ unterstützt wird, eine Arbeitsaufgabe ist, obwohl die Autoren bislang nur Ergebnisse für Gesamtsysteme mitteilen. Damit kann das Verfahren in Benutzungstests mit Prototypen eingesetzt werden, generiert jedoch keine expliziten Hinweise zur Umgestaltung. Der Prüfungsaufwand ist relativ gering, es wird etwa eine Stunde Bearbeitungsaufwand veranschlagt. Durch das Vorliegen von Auswertungsprozeduren kann eine schnelle Auswertung der Resultate erreicht werden. Auch beim SynBA-DQ ist ein Bezug zur DIN EN ISO 9241 Teil 10 intendiert, aber nicht mit methodischen Maßnahmen hineinkonstruiert oder anderweitig belegt worden. Die Auswertung der umfangreichen Fragebögen kann nur durch einen Experten vorgenommen werden; Platz für subjektive Schilderungen von Nutzungsproblemen ist nicht vorgesehen. Es werden zwar Grenzwerte mitgeteilt, aber als Grundkonzept ist die gegenseitige Kompensation von Beschwerden und nicht die Falsifikation anhand einzelner belegbarer Schwächen vorgesehen.

6.3 Zusammenfassende Beurteilung

Abgesehen von den spezifischen Problemen, die mit der Anwendung der subjektiven Bewertungsverfahren verbunden sind, ist ein Mangel verallgemeinerbar: die Fragebogenitems sind durch keine Anforderungsanalyse legitimiert und sind somit nicht als valide Prüfkriterien im Sinne des Konformitätsmodells (vgl. Abschnitt 2) definiert. Folglich sagen die mit diesen Fragebögen erhobenen Daten nichts über die Normkonformität eines Produkts aus, obwohl die Namen vieler Fragebögen diese Annahme nahe legen. Die meisten Fragebögen eignen sich gut zum Produktvergleich; eine hierdurch erzielte Rangfolge der Produkte sagt aber nichts über die Rangfolge der Normkonformität der Produkte aus. Nichtsdestoweniger sind auch die bereits eingeführten Bewertungsverfahren für eine bestimmte Prüfstrategie nützlich: sie dienen dem Aufspüren von Nutzungsproblemen und evtl. zugrundeliegenden ergonomischen Mängeln. Dieses Ergebnis liefern aber die meisten Fragebögen nicht direkt; vielmehr sind anschließende Benutzerbefragungen und teilnehmende Beobachtungen am Bildschirmarbeitsplatz erforderlich, d.h., die Beurteilungsgrundlagen für die subjektiven Bewertungen sind noch zusammen mit den Benutzern zu rekonstruieren.

Auch die Expertenverfahren sind für eine Konformitätsprüfung nicht geeignet. Diese Prüfung setzt eine Entwicklung von Anforderungen voraus, ein methodisches Muss, das keines der untersuchten Verfahren erfüllt. Der Hinweis, dass genormte Anforderungen vorausgesetzt werden, ist unbefriedigend, da die genormten Forderungen stets mit Blick auf die Erfordernisse des Nutzungskontexts eines Produkts interpretiert werden müssen, bevor Prüfkriterien festgelegt werden (Dzida, 1996,1997). Dieser Interpretationsakt ist vergleichbar mit der Auslegung von Grundsätzen unserer Rechtsnormen (Dzida, 1994), die stets mit Blick auf die Erfordernisse eines genau analysierten Sachverhalts anzuwenden sind.

7 Literatur

- Ademer, C. (1995). Beurteilung der Dialogqualität von Bildschirmarbeit – Entwicklung eines Verfahrens vor dem Hintergrund europäischer Normen und Standards. Unveröffentlichte Diplomarbeit. Bergische Universität – Gesamthochschule Wuppertal.
- Arbeitsschutzgesetz (ArbSchG) (1996). Gesetz über die Durchführung von Maßnahmen des Arbeitsschutzes zur Verbesserung der Sicherheit und des Gesundheitsschutzes der Beschäftigten bei der Arbeit.
- Arnold, A.G. (1998). Action Facilitation and Interface Evaluation. Thesis TU Delft. Delft University Press, Delft.
- Asher, J.J. (1963). Towards a neo-field theory of problem solving. *Journal of General Psychology*, 68, 3-8.
- Beimel, J., Schindler, R. and Wandke, H. (1992). Understanding and acceptance of the first draft international standard ISO 9241 Part 10, Dialogue Principles. Internal Report Humboldt University Berlin, Institute of Applied Psychology, Germany.
- Bevan, N. (1995). Measuring usability as quality of use. *Journal of Software Quality*, 4, 115-130.
- Bevan, N. & Macleod, M. (1994). Usability Measurement in Context. *Behaviour & Information Technology*, 13, 132-145.
- Bildschirmarbeitsverordnung (BildscharbV) (1996). Verordnung über Sicherheit und Gesundheitsschutz bei der Arbeit an Bildschirmgeräten.
- BTQ Kassel Mainz (1998). Beurteilung von Arbeitsbedingungen am Bildschirm / Arbeitsplatzanalyse der BTQ nach Bildschirmarbeitsverordnung. Beratungsstelle der DAG für Technologiefolgen und Qualifizierung im Bildungswerk der DAG im Lande Hessen e.V.
- Burmester, M. et al. (1997). Das SANUS-Handbuch - Bildschirmarbeit EU-konform, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, FB 760I, Dortmund/Berlin.
- Çakir, A. and Dzida, W. (1997). International ergonomic HCI standards. In: Helander et al. (eds.): *Handbook of Human-Computer Interaction*. Elsevier, Amsterdam, 407-420.

- Carroll, J.M. (ed.) (1995). *Scenario-Based Design. Envisioning Work and Technology in System Development*. Wiley, New York.
- Council Directive (1990). Council Directive on the minimum safety and health requirements for work with display screen equipment (fifth individual Directive within the meaning of Article 16(1) of Directive 87/391/EEC). Number: 90/270/EEC of 29 May 1990. Official Journal of the European Communities, No L 156, p.14-18.
- DEKITZ Prüfhandbuch Gebrauchstauglichkeit (2000). Leitfaden für die software-ergonomische Evaluierung von Software auf der Grundlage von DIN EN ISO 9241, Teile 10 und 11. Version 2.0. DEKITZ c/o TGA GmbH, Gartenstraße 6, 60584 Frankfurt.
- DIN EN ISO 9241-2 (1993). Ergonomic requirements for office work with display terminals (VDTs): Guidance on task requirements.
- DIN EN ISO 9241-10 (1995). Ergonomic requirements for office work with display terminals (VDTs): Dialogue principles.
- DIN EN ISO 9241-11 (1998). Ergonomic requirements for office work with display terminals (VDTs): Guidance on usability.
- DIN ISO DIS 9241-12 (1999). Ergonomic requirements for office work with display terminals (VDTs): Presentation of information.
- DIN EN 45014 (1990). Allgemeine Kriterien für Konformitätserklärungen von Anbietern.
- DIN 66 271 (1995). Software-Fehler und ihre Beurteilung durch Lieferanten und Kunden.
- Döbele-Martin, C., Martin, P. (Hrsg.). (1997) *Ergonomie-Prüfer. Technologieberatungsstelle beim DGB Landesbezirk NRW e.V.*
- Döbele-Martin, C., et al. (Hrsg.). (1997). *Ergonomie-Prüfer. Technologieberatungsstelle beim DGB Landesbezirk NRW e.V.*
- Dunckel, H., Volpert, W., Zölch, M., Kreutner, U., Pleiss, C. und Hennes, K. (1993). *Kontrastive Aufgabenanalyse im Büro - Der KABA-Leitfaden*. Zürich/Stuttgart, Verlag der Fachvereine, Teubner, Stuttgart, 1993.
- Dzida, W., Herda, S., and Itzfeldt, W-D. (1978). User-perceived quality of interactive systems. *IEEE Transactions on Software Engineering*, vol. SE-4, no. 4, 270-276.

- Dzida, W. (1994). Qualitätssicherung durch software-ergonomische Normen. In: E.Eberleh et al. (Hrsg.): *Einführung in die Software-Ergonomie*. de Gruyter, Berlin, 373-406.
- Dzida, W. (1996). International Usability Standards. *ACM Computing Surveys*, 28, No. 1, 173-175.
- Dzida, W. (1997). International User-Interface Standardization. In: A.B. Tucker (ed.): *The Computer Science and Engineering Handbook*. CRC Press, Boca Raton, 1997, 1474-1493.
- Dzida, W. and Freitag, R. (1998). Making use of scenarios for validating analysis and design. *IEEE Transactions on Software Engineering*, 24, No. 12, 1182-1196.
- Dzida, W., Geis, T. und Redtenbacher, W. (2000). DEKITZ-Prüfverfahren für Softwareprodukte – Konformität der Nutzungsqualität mit internationalen Normen. In: K.-P. Timpe, H.-P. Willumeit und H. Kolrep (Hrsg.) (2000): *Bewertung von Mensch-Maschine-Systemen*. Fortschritt-Berichte VDI, Reihe 22, Nr.1. Düsseldorf: VDI-Verlag, 104-114.
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik*. 2., überarbeitete und erweiterte Auflage. Hogrefe, Göttingen.
- Gediga, G. & Hamborg, K. (1999). IsoMetrics: An usability inventory supporting summative and formative evaluation of software systems. In H.-J. Bullinger & J. Ziegler (Eds.) *Human-Computer Interaction. Ergonomics and User Interfaces*. Vol. 1., pp 1018-1022. Erlbaum, Mahwah, NJ.
- Hofmann, B. (1999). *Ansätze der testtheoretischen Absicherung eines Instrumentes zur Erfassung wahrgenommener ergonomischer Qualität am Bildschirmarbeitsplatz nach ISO 9241-10*. Unveröffentlichte Diplomarbeit am Psychologischen Institut der Universität Bonn.
- Institut für Arbeitsphysiologie (IfADo) (1995). Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): Fragebogen zur Bewertung von Software.
- ISO 8402 (1994). *Quality - Vocabulary*.
- ISO/IEC 9126 (1991). *Information technology – Software product evaluation – Quality characteristics and guidelines for their use*.
- ISO/IEC 12119 (1994). *Information technology - Software product evaluation - Quality requirements and testing*.

- Jacobson, I., Christerson, M., Jonsson, P., and Övergaard, G. (1992). Object Oriented Software Engineering. A Use Case Driven Approach. Reading: ACM press.
- Kirakowski, J. and Corbett, M. (1993). SUMI - The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), 210-212.
- Kirakowski, J. (1995). The software usability measurement inventory: background and usage. In: P. Jordan, B. Thomas, and B. Weerdmeester (eds.): *Usability Evaluation in Industry*. Taylor & Francis, London.
- Landauer, T.K. (1995). *The Trouble with Computers. Usefulness, Usability and Productivity*. The MIT Press, Cambridge MA.
- Macleod, M. (1994). Usability in Context: Improving Quality of Use. In: G. Bradley, and H.W. Hendricks (Eds.): *Human Factors in Organizational Design and Management-IV. Proceedings of the International Ergonomics Association 4th International Symposium on Human Factors in Organizational Design and Management*, Amsterdam: Elsevier, pp. 375-380.
- Microsoft Windows (1995). *The Windows Interface Guidelines for Software Design: An Application Design Guide*. Microsoft Press.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In: J. Bennett and G. Lynch (eds.): *Proceedings of the CHI'92 Conference*, ACM Press, New York, pp. 373-380.
- Nielsen, J. and Mack, R.L. (1994). *Usability Inspection Methods*. Wiley, New York.
- Oppermann, R., Murchner, B., Reiterer, H., Koch, M. (1992). *Software-ergonomische Evaluation*. de Gruyter, Berlin/New York.
- Petrasch, R. (2000). Style Guides am Beispiel der "Java Look and Feel Design Guidelines" von Sun Microsystems – eine kritische Betrachtung. *Softwaretechnik Trends*, 20, 1, 17-21.
- Porteous, M.A. and Kirakowski, J. (1992). SUMI - Software Usability Measurement Inventory. Department of Applied Psychology, University College, Cork, Ireland, 1992.
- Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In: R. Liskowsky et al. (Hrsg.): *Software-Ergonomie '97*, Teubner, Stuttgart, 253 - 262.

- Prümper, J. (1999) TestIT: ISONORM 9241/10. In H.-J. Bullinger & J. Ziegler (Eds.) *Human-Computer Interaction. Ergonomics and User Interfaces*. Vol. 1, Erlbaum, Mahwah, NJ, pp. 1028-1027.
- Prümper, J., Anft, M. (1993). Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. In: K-H. Rödiger (Hrsg.): *Software-Ergonomie '93*, Teubner, Stuttgart, 145 - 156.
- Richenhagen, G. (1996). ABETO - Ein Verfahren zur Arbeitsanalyse bei Bildschirmarbeitsplätzen. *Der Personalrat*, 1, 10-12.
- Schwarz, R., Schmitz, U. (1994). Die Synthetische Beanspruchungsanalyse (SynBa); Methode und Anwendungsbeispiel. *Zeitschrift für Arbeitswissenschaft*, 48, 225-236.
- Stary, C., Riesenecker-Caba, T., Kalkhofer, M., Flecker, J. (1997). EU-CON. Ein Verfahren zur EU-konformen Software-ergonomischen Bewertung und Gestaltung von Bildschirmarbeit. vdf, Hochschul-Verlag an der ETH, Zürich.
- Stary, C., Riesenecker-Caba (1999). EU-CON II – Softwareergonomische Bewertung und Gestaltung von Bildschirmarbeit. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Wirtschaftsverlag NW, Verlag für neue Wissenschaft, Bremerhaven, Dortmund, Berlin.
- Sun Microsystems (1999). *Java Look and Feel Design Guidelines*. Addison Wesley.
- The Open Group CDE 2.1/Motif 2.1 (1997). *Style Guide and Glossary, Style Guide Reference and Style Guide Certification Check List*.
- Ulich, E. (1981). Subjektive Tätigkeitsanalyse als Voraussetzung autonomieorientierter Arbeitsgestaltung. In: F. Frei und E. Ulich (Hrsg.): *Beiträge zur psychologischen Arbeitsgestaltung*. Huber, Bern, 327-347.
- Ulich, E. (1998). *Arbeitspsychologie*. Schäffer-Poeschel, Stuttgart.
- Weise, G. (1997). *Psychologische Leistungstests*. Hogrefe, Göttingen.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung. Theoretische und empirische Studien zu informationsverarbeitenden Tätigkeiten*. Göttingen: Hogrefe.

- Wieland-Eckelmann, R., Baggen, R., Schwarz, R., Schmitz, U. (1994) Systemergonomische Beanspruchungs- und Arbeitsanalyse (SEBA). Schriftenreihe der Bundesanstalt für Arbeitsmedizin. Bremerhaven: Wirtschaftsverlag NW.
- Wieland-Eckelmann, R., Baggen, R., Saßmannshausen, A., Schwarz, R., Schmitz, U., Ademmer, C., Rose, M. (1996) (unter Mitarbeit von J. Mielke und K. Schuhmacher) Gestaltung beanspruchungsoptimaler Bildschirmarbeit. Schriftenreihe der Bundesanstalt für Arbeitsmedizin. Bremerhaven: Wirtschaftsverlag NW.
- Willumeit, H., Gediga, G. & Hamborg, K. (1995). Validation of the IsoMetrics Usability Inventory. Forschungsbericht Nr. 105 der Universität Osnabrück, Fachbereich Psychologie.
- Willumeit, H., Gediga, G., Hamborg, K.-C. (1996). IsoMetrics^L - Ein Verfahren zur formativen Evaluation von Software nach ISO 9241/10. Mitteilungen des Fachausschusses 2.3. Ergonomie & Informatik, März, 5-12.
- Ziegler, J., Burmester, M. (1995). Structured Human Interface Validation Technique - SHIVA. In: Azai, Y., Ogawa K., Mori, H. (eds.): Symbiosis of Human and Artifact. Proceedings of the 6th International Conference on Human-Computer-Interaction. Tokyo, Vol. 2, Elsevier, Amsterdam, 899-906.
- Zijlstra, F.R.H. (1993). Efficiency in Work Behaviour. A Design Approach for Modern Tools. Thesis TU Delft, The Netherlands, Delft University Press, Delft.

Anhang A: Methoden zum Prüfverfahren

A.1 Aufgabenanalyse

A.1.1 Ziele

Dient der Feststellung der Nutzungsziele (gewünschte Ergebnisse), die mit einem Produkt in einem Anwendungsbereich erreicht werden sollen, um die Leistung des Produkts aufgabengerecht zu projektieren, und dient der Feststellung, ob Arbeitsaufgaben nach den Empfehlungen von ISO 9241 Teil 2 gestaltet sind.

A..1.2 Gütekriterien

- Sind die Analysedaten hinsichtlich der Empfehlungen von ISO 9241 Teil 2 bewertet?
- Sind die Zusammenhänge zwischen Kernaufgaben (key tasks) und Nutzungszielen beschrieben?
- Sind die Analysedaten in Form von Szenarien, Use Cases oder in anderer Form dokumentiert, so dass sie für die Anforderungsdefinition verwertbar sind?
- Sind Anforderungen in Abhängigkeit von Merkmalen des Nutzungskontexts spezifiziert und gewichtet?
- Sind die Anforderungen durch Anwender oder Benutzer validiert?
- Ist die Anforderungsdefinition für die Leistungsbeschreibung des Produkts oder die Erstellung des Pflichtenheftes verwertbar?

A.1.3 Erläuterungen zur Aufgabenanalyse

Die Aufgabenanalyse ist entbehrlich, wenn in den Dokumenten der Anforderungsdefinition (Pflichtenheft) oder in der Leistungsbeschreibung des Produkts nachvollziehbar ist, welche Anforderungen an das Produkt sich auf welche Aufgaben oder Szenarien beziehen. Die Aufgabenanalyse muss nachgeholt werden, wenn die genannten Dokumente keinen Bezug zu Aufgaben und Nutzungszielen enthalten oder

wenn dieser ungenügend dargestellt ist. Zumindest muss ein Vergleich von Zielen oder Ergebnissen der Arbeitsaufgaben mit den Leistungen des Produkts möglich sein. Eine Konformitätsprüfung ist stets auf die Anforderungen an ein Produkt (Prüfkriterien) zu beschränken, die auf Aufgabenerfordernissen oder Benutzerbelangen beruhen. Somit besteht das Ergebnis einer Aufgabenanalyse in der Darstellung der Erfordernisse und Belange sowie korrespondierender Prüfkriterien. Eine Normkonformitätsprüfung¹³ bezieht darüber hinaus die genormten Anforderungen mit ein.

Andere Erfordernisse des Nutzungskontexts eines Produkts (z.B. arbeitsorganisatorische oder technische) sowie Anforderungen aus anderen Normen (z.B. Sicherheitsnormen) sind zunächst in Aufgabenerfordernisse oder Benutzerbelange zu transformieren, damit sie anschließend als Prüfkriterien definiert werden können. Eine Analyse der Arbeitsbedingungen (im Sinne des § 3 der Bildschirmarbeitsverordnung) ist nicht Ziel der Aufgabenanalyse, kann aber nützlich sein, wenn hierdurch Erfordernisse des Nutzungskontexts erfasst werden, die zur Präzisierung der Aufgabenerfordernisse beitragen.

Die Ergebnisse der Aufgabenanalyse sind unter Beteiligung repräsentativer Benutzer zu validieren und hinsichtlich möglicher Verletzungen gegen ISO 9241-2 zu prüfen.

Aufgabenanalyse und Aufgabendesign sind zu unterscheiden. Es ist nicht Gegenstand der Aufgabenanalyse, bessere Aufgaben nach arbeitspsychologischen Grundsätzen zu entwickeln. Aber es kann notwendig sein, Abweichungen von den Empfehlungen in ISO 9241-2 festzustellen. Werden Abweichungen festgestellt, so sind Anwender oder Benutzer darauf hinzuweisen, dass eine mangelnde Zufriedenstellung der Benutzer gemäß ISO 9241-11 nicht auf Produkt-, sondern auf Aufgabenmerkmale zurückführbar sein kann. Gegebenenfalls kann auf die Verletzung der Mindestanforderungen der Bildschirmarbeitsverordnung (1996) verwiesen werden.

¹³ Die Normkonformitätsprüfung dient zur Feststellung der ergonomischen Mindestqualität.

A.1.4 Anwendungsempfehlungen

Die Aufgabenanalyse dient der Vorbereitung der Konformitätsprüfung, wenn die Entwicklungsdokumente keine brauchbaren Analysedaten enthalten. In der Regel findet man zwar Datenmodelle und funktionale Spezifikationen, diese sagen aber nichts über Struktur (Ziele und Merkmale der intendierten Arbeitsergebnisse) und Ablauf (Aufwand und eingesetzte Hilfsmittel) der Aufgaben aus, die ein Benutzer zu erledigen hat. Auch Kontextmerkmale, die Tätigkeiten des Benutzers beeinflussen, sind in solchen Dokumenten meist nicht enthalten. Manchmal sind non-funktionale Anforderungen spezifiziert, aus denen man nützliche Hinweise auf Merkmale des untersuchten Nutzungskontexts des Produkts bekommt.

Arbeitspsychologische Aufgabenanalysen können geeignet sein, wenn Merkmale des Nutzungskontexts berücksichtigt wurden. In der Regel dienen diese Analysemethoden jedoch dem Design von Arbeit und Organisation und erweisen ihre Nützlichkeit bei Bewertungen im Sinne von ISO 9241-2. Arbeitserfordernisse und Benutzerbelange, die zu diesem Zweck festgestellt wurden, können jedoch auch für eine Transformation in Dialoganforderungen verwendet werden.

Gute Erfahrungen werden mit Kontext-Szenarien gemacht, da Benutzer leicht zu motivieren sind, die erhobenen Daten zu validieren, ohne dass sie irgendwelche Vorkenntnisse mitbringen müssen, außer ihrer Expertise im Anwendungsfeld. Kontext-Szenarien enthalten in episodischer Form die Beschreibung einer vollständigen Tätigkeit in der Sprache des Benutzers (siehe Leitfaden für die Erhebung von Kontext-Szenarien).

A.2 Inspektion¹⁴

A.2.1 Ziele

Dient der Feststellung, ob geforderte Merkmale vorhanden sind und / oder ob geforderte Tätigkeiten (user performance) ausführbar sind. Je nach erforderlichem Prüfungsumfang werden Merkmale oder Tätigkeiten vollständig oder stichprobenartig untersucht.

A.2.2 Gütekriterien

- Ist für jeden festgestellten Mangel geprüft, ob die Gestaltung des Merkmals oder der Tätigkeit von Aufgabenerfordernissen abhängt?
- Ist für jeden festgestellten Mangel der Bezug zu mindestens einem Gestaltungsgrundsatz (des Dialogs oder der Informationsdarstellung) berücksichtigt?
- Ist jeder festgestellte Mangel im Kontext einer Aufgabenbearbeitung ermittelt und nachvollziehbar dokumentiert?
- Sind für jeden festgestellten Mangel die verletzten Normen und Gestaltungsgrundsätze aufgeführt?
- Sind die festgestellten Mängel in bezug auf die Benutzerzielgruppe relevant?

A.2.3 Erläuterungen zur Inspektion

Eine Inspektion setzt voraus, dass Aufgabenerfordernisse und Benutzerbelange in korrespondierende Produktmerkmale transformiert worden sind (z.B. Aufmerksamkeitsanforderungen des Benutzers in Kodierungsmerkmale der dargestellten Information transformieren oder Tätigkeitserfordernisse in Dialogschritte).

Wird eine Abweichung von geforderten Merkmalen (Prüfkriterien) festgestellt, so ist die Abweichung mit einer Verletzung von Aufgabenerfordernissen oder Benutzerbelangen zu begründen (design rationale). Wird eine Normverletzung festgestellt, so ist

¹⁴ Inspektion umfasst die Methodik der analytischen Evaluierung und Beobachtung gemäß ISO 9241 (Teile 12-17). Siehe auch Fußnote zur Methode der teilnehmenden Beobachtung.

zusätzlich auf die relevanten Anforderungen der Norm zu verweisen sowie auf die verletzten Gestaltungsgrundsätze. Festgestellte Mängel sind umsichtig und fair zu bewerten. Hierbei sind alle relevant erscheinenden Merkmale des Produkts sowie Umstände des Nutzungskontexts zu berücksichtigen, die den Mangel zumutbar kompensieren können (z.B. andere Schriftgröße einstellen, wenn Lesbarkeit mangelhaft, oder einen möglichen Umweg im Dialog gehen).

Mängel sind zu gewichten. Kompensierbare Mängel sind geringer zu gewichten. Bei Inspektionen zur Feststellung der Benutzbarkeit gemäß ISO/IEC 12119 sind Mängel in bezug auf die Benutzerzielgruppen einzuschätzen und entsprechend zu gewichten.

Für festgestellte Mängel muss nachvollziehbar sein, ob sie in einem Nachbesserungsprozess behoben werden oder wurden (siehe Dokumentenanalyse).

Die Methodik der Inspektion zum Zwecke der Normkonformitätsprüfung ist zu unterscheiden von jedweder Inspektion mittels erstellter Merkmalslisten, da diese Checklisten oft keine legitimen Prüfkriterien enthalten (siehe Gütekriterien der Inspektionsmethode).¹⁵ Merkmals-Items einer Checkliste müssen als Prüfkriterien legitimiert sein, d.h. einen klaren Bezug zu Aufgabenerfordernissen oder Benutzerbelangen haben.

A.2.4 Anwendungsempfehlungen

Die Inspektion ist für die Durchführung von Konformitätsprüfungen am besten geeignet, da man die Übereinstimmung von Merkmalen oder Tätigkeiten mit den Prüfkriterien durch Beobachtung am Ort der Arbeitstätigkeit feststellen kann. Häufigster Fehler bei der Inspektion ist, dass der Prüfer versäumt, einen konkreten Aufgabenbezug zu definieren, denn Merkmale des Dialogs oder der Informationsdarstellung sowie das Prüfkriterium sind stets im Aufgabenzusammenhang zu interpretieren. Ist der Prüfer mit

¹⁵ Merkmalschecklisten können für Produktvergleiche nützlich sein, wenngleich festgestellte Rangunterschiede zwischen Produkten nichts über Normkonformität aussagen. Eine "heuristische Inspektion" nach Nielsen kann nützlich sein, sagt aber ebenfalls nichts über Normkonformität aus; auch ist zu beachten, daß die heuristischen Bewertungskategorien mit den Gestaltungsgrundsätzen der Normen nicht deckungsgleich sind.

der Aufgabe nicht vertraut, so sollte ein Benutzer die Aufgabenbearbeitung am System vorführen und kommentieren. Insoweit geht die Inspektion in die teilnehmende Beobachtung über.

ANMERKUNG: Die häufig verwendete "heuristische Evaluierung"¹⁶ wendet anstelle der Prüfkriterien sogenannte Heuristiken an, die teilweise den Dialogprinzipien entsprechen. Diese Methode ist für Konformitätsprüfungen nicht geeignet, da Prüfkriterien anhand von Aufgabenerfordernissen, Benutzerbelangen und Normen konkretisiert werden, während Nielsens Heuristiken kontextneutral angewendet werden. Beispielsweise wird "Aufgabenangemessenheit" gar nicht als Heuristik verwendet, was jedoch die wichtigste Voraussetzung einer Konformitätsprüfung ist (siehe Feststellung der Effektivität). Die Methode von Nielsen ist eher eine Inspektion zur Erfassung des "ersten Eindrucks", den man von einem System gewinnen möchte. Sie wird meist beim Prototyping angewendet. Wendet man die Vorgehensweise der heuristischen Evaluierung anhand der Dialogprinzipien (DIN EN ISO 9241-10) an, so kann man grobe Normabweichungen erkennen.

¹⁶ Siehe Nielsen and Mack (1994): Usability Inspection Methods. New York: Wiley.

A.3 Teilnehmende Beobachtung¹⁷ / ¹⁸

A.3.1 Ziele

Dient der Erfassung von Merkmalen einer Tätigkeit im Zeitverlauf (z.B. zögerlicher Ablauf, Nachschlagen im Benutzerhandbuch, Fehler machen und beheben, mündliche Äußerungen, umständlicher Ablauf) sowie der begleitenden Arbeitsumstände am Bildschirmarbeitsplatz, um mögliche Benutzungsprobleme im Nutzungskontext zu untersuchen.

A.3.2 Gütekriterien

- Sind die Repräsentativität des beobachteten Benutzers, der Nutzungsumstände sowie der durchzuführenden Aufgabe (bezogen auf den Verwendungszweck des Produkts) gegeben?
- Werden die aufgezeichneten Beobachtungsdaten hinsichtlich ihrer kritischen Merkmale durch die beobachteten Personen bestätigt?
- Wird die Bewertung der kritischen Beobachtungsdaten von ihrer sachlichen Beschreibung (Aufzeichnung) getrennt? Dies ist insbesondere bei summativer Evaluierung notwendig.
- Ist die erlaubte Beeinflussung der beobachteten Person durch den Beobachter definiert (z.B. Ziel der Beobachtung mitteilen, Aufforderung zur Wiederholung einer Tätigkeit, Bitte um mündliche Erläuterung der Tätigkeit)?
- Ist durch geeignete Vorbereitung der Aufzeichnung dafür gesorgt, dass die zu beobachtende Person von der Aufzeichnungstechnik (z.B. Mikrofon, Kamera) nicht beeinträchtigt wird?

¹⁷ Während bei der Inspektionsmethode die Beobachtung eher auf eine Zustandsanalyse der Tätigkeit und die Feststellung von Produktmerkmalen gerichtet ist, zielt die teilnehmende Beobachtung eher auf die Untersuchung des Tätigkeitsverlaufs und der Wechselwirkungen zwischen Verlauf und Nutzungskontext.

¹⁸ Die teilnehmende Beobachtung kann zum "Benutzertest" (im Sinne eines Quasi-Experiments) werden, siehe Erläuterungen im Abschnitt 3.3.

A.3.3 Erläuterungen zur teilnehmenden Beobachtung

Die teilnehmende Beobachtung ist meist zielgerichtet. Beispielsweise wird untersucht, welchen Aufwand ein Benutzer bei der Behebung eines Nutzungsfehlers benötigt. Dabei ist es notwendig, die vollständige Ausführung einer Aufgabe zu beobachten, da die Bewertung der Beobachtungsdaten von den Verlaufseigenschaften der gesamten Tätigkeit abhängen kann. Beispielsweise ist ein ergonomischer Mangel schwerwiegender, wenn seine Wirkungen nur mit großem Aufwand rücknehmbar sind.

Der Beobachtungsablauf wird weniger systematisch vorbereitet als beim Benutzertest. Es muss z.B. nicht streng darauf geachtet werden, dass die Verlaufsanalyse bei jeder beobachteten Person in gleicher Weise durchgeführt wird, da die teilnehmende Beobachtung eher der Erkundung dient. Die teilnehmende Beobachtung kann in einen Benutzertest übergehen, wenn der Beobachtungsablauf vom Beobachter nicht unterbrochen wird und die Beobachtung bei allen beobachteten Benutzern unter den gleichen Beobachtungs-Bedingungen stattfindet. Dadurch wird eine quasi-experimentelle Untersuchungssituation geschaffen: Jeder Benutzer wird instruiert, welche Aufgabe auszuführen ist; es wird außerdem mitgeteilt, dass nicht der Benutzer getestet wird (insofern ist der Begriff "Benutzertest" irreführend), sondern die Nutzung des Programms, und es wird darauf geachtet, dass die Beobachtungs-Bedingungen, etwa die Aufzeichnung mittels Video-Kamera, die Benutzer nicht beeinflussen.

Auch bei der teilnehmenden Beobachtung muss der Beobachter sorgfältig darauf achten, dass seine Gegenwart, seine Aufzeichnungstechnik, seine persönlichen Einstellungen usw. einen verfremdenden Einfluss auf die Verlaufsanalyse nehmen können. Der Beobachter muss sich bewusst sein, dass er während der Beobachtung selbst Teil des Nutzungskontexts der beobachteten Bildschirmtätigkeit ist.

Die aufgezeichneten Beobachtungsdaten werden meist hinsichtlich des Beobachtungsziels ausgewertet. Deshalb sind die zielkritischen Daten zu dokumentieren und von der beobachteten Person als sachlich zutreffend zu bestätigen. Zum Zwecke der Veranschaulichung können "screen shots" geeignet sein. Mit steigender Komplexität eines

Tätigkeitsverlaufs können die zu dokumentierenden Daten unübersichtlich werden; dann empfiehlt es sich, den Verlauf als Use-Szenario zu beschreiben. Hierdurch werden die untersuchten Merkmale im Tätigkeitszusammenhang leichter verständlich und reproduzierbar. Die beobachtete Person muss die sachliche Richtigkeit des dokumentierten Verlaufs bestätigen.

Bei der Merkmalsdokumentation hat sich der Beobachter jeder Bewertung zu enthalten. Seine anschließenden Bewertungen und die der beobachteten Benutzer sind getrennt zu dokumentieren.

A.3.4 Anwendungsempfehlungen

Die teilnehmende Beobachtung dient sowohl der Vorbereitung als auch der Durchführung einer Konformitätsprüfung, wobei Vorbereitung und Durchführung meist ineinander übergehen. Das Augenmerk liegt auf der Erfassung kritischer Merkmale oder Tätigkeiten im Ablauf. Vorbereitend werden mit dieser Methode die Umstände des Dialogablaufs sowie die dabei auftretenden kritischen Situationen exploriert. In der Regel sind es äußere Anlässe, die eine Untersuchung dieser Art nahelegen, z.B. aufgetretene Benutzungsprobleme oder mangelnde Zufriedenstellungen der Benutzer, denen nachgegangen werden muss, um vermutete Normabweichungen festzustellen und deren Wirkungen zu beurteilen.

Manchmal wird die Methode der teilnehmenden Beobachtung auch als "Beobachtungsinterview" bezeichnet, wenn der Schwerpunkt der Untersuchung auf der Befragung des Benutzers liegt. Die methodische Vorgehensweise wechselt dann oft zwischen Beobachtung und Benutzerbefragung.

Wird die Methode der teilnehmenden Beobachtung ohne die genannten Anlässe angewandt, also ohne die Zielsetzung, eine vermutete Abweichung zu untersuchen, so wechselt der Prüfer (Analytiker) von der Methode der teilnehmenden Beobachtung zur Inspektion über.

A.4 Benutzerbefragung

A.4.1 Ziele

Dient der Ermittlung der subjektiven Zufriedenstellung der Benutzer sowie der Feststellung von Ursachen für mangelnde Zufriedenstellung.

A.4.2 Gütekriterien

- Sind die befragten Benutzer repräsentativ?
- Haben die Fragebogen-Items einen definierten Aufgabenbezug?
- Sind relevante Merkmale des Nutzungskontexts beschrieben?
- Ist der Übungsgrad der Benutzer festgestellt?
- Ist die bisherige Nutzungsdauer festgestellt?
- Falls sich der Benutzer gerade einarbeitet: Ist eine Wiederholungsmessung nach der Einarbeitungsphase geplant?
- Bei mangelnder Zufriedenstellung: Sind die verursachenden Merkmale des Produkts oder die störenden Nutzungsumstände dokumentiert?
- Ist für jeden beobachteten Mangel festgestellt, ob dieser negativ erlebt wird?

4.1.4.3 Erläuterungen zur Benutzerbefragung

Dies ist die unter ökonomischen Gesichtspunkten günstigste Methode, um einem Mangelverdacht nachzugehen oder den Grad der Zufriedenstellung festzustellen. Fragen an Benutzer sind so zu stellen, dass keine Verzerrung (bias) durch die Fragenformulierung entstehen kann (etwa durch Suggestivfragen). Fragebogenitems (z.B. in Checklisten) müssen validiert sein (Aufgabenbezug). Der Aufgabenbezug darf nicht durch eine Standardaufgabe definiert sein. Vielmehr ist die Repräsentativität von Aufgaben und Benutzern durch den realen Nutzungskontext nachzuweisen, d.h., es muss für eine Frage eindeutig nachvollziehbar sein, unter welcher Aufgabenbedingung sie gestellt und beantwortet wurde.

Oft ist es darüber hinaus erforderlich, eine Frage nur im aktuellen Nutzungskontext zu stellen (z.B. zur Laufzeit des Systems), damit diagnostiziert werden kann, welche Merkmale des Systems im Zusammenhang mit der Antwort des Benutzers stehen. Wird die Zufriedenstellung erfragt, so ist zu untersuchen, welchen Einarbeitungsgrad die Befragten haben; unter Umständen sind Befragungen zu wiederholen, um die Zuverlässigkeit der Befragungsergebnisse nachzuweisen. Die Zuverlässigkeit spielt bei wenig geübten Benutzern oder in der Einführungsphase eines Produkts eine geringere Rolle als unter konsolidierten Nutzungsbedingungen; reliable Mängelhinweise unter konsolidierten Bedingungen sind meist bedeutsam. Werden diese negativ erlebt, so kann es erforderlich sein, die subjektiven Wirkungen zu untersuchen, wenn nicht ohnehin zweifelsfrei feststeht, dass der Mangel behoben wird.

A.4.4 Anwendungsempfehlungen

Die Methode der Benutzerbefragung ist zur Vervollständigung eines Konformitätsnachweises in jedem Fall anzuwenden. Benutzerbefragungen ersetzen nicht die Konformitätsprüfung, sondern sie ergänzen sie hinsichtlich des unverzichtbaren subjektiven Urteils der Benutzer. Weil sich Experten über die Effektivität und Effizienz eines Produkts irren können, sind stets die Benutzer nach ihrem subjektiven Eindruck über Effektivität und Effizienz zu befragen, ein Urteil, das als Zufriedenstellung interpretiert wird. Wenn Experten und Benutzer im Urteil abweichen, so haben bei mangelnder Zufriedenstellung immer die Benutzer recht - was nicht bedeuten muss, dass die Benutzer immer die Ursache mangelnder Zufriedenstellung richtig erkennen.

Typische Fragen, die man Benutzern stellen kann, sind solche nach

- Fehlersituationen,
- häufig vorkommenden Fehlern,
- subjektiv erlebten Benutzungsproblemen,
- subjektiven Einschätzungen über die Qualität ausgewählter Masken usw.

Ungeeignet sind pauschalisierte Fragen nach der Zufriedenheit. Überschätzt wird oft der Wert quantitativer subjektiver Daten, die keinen unmittelbaren Bezug zu einem Tätigkeits- oder Produktmerkmal haben.

Der häufigste Fehler, der bei Benutzerbefragungen gemacht wird, besteht darin, dass die Ergebnisse als undifferenzierte Aussagen über die Softwarequalität oder gar über die Normkonformität verwendet werden. Richtig ist, dass Befragungsergebnisse etwas über subjektive Einschätzungen der Effektivität und Effizienz aussagen. Um den subjektiven Akzent der Aussage hervorzuheben, werden diese als Zufriedenstellungsurteile behandelt.

Für eine Benutzerbefragung braucht man u.U. gar nicht viele Benutzer. Man hört auf, weitere Benutzer zu befragen, wenn man feststellt, dass die bisher befragten Benutzer soviel ergänzende Informationen beigesteuert haben, dass von weiteren Befragungen keine neuen Daten zu erwarten sind. Häufig reichen 3 bis 5 Benutzer aus. In jedem Fall muss die Stichprobe der Befragten repräsentativ sein.

Wird bei der Befragung mangelnde Zufriedenstellung festgestellt, so kann es nützlich sein, die Stichprobe der Befragten zu erweitern, um die Verallgemeinerbarkeit des Ergebnisses zu sichern; denn manchmal sind subjektive Benutzerbeschwerden auf lokale Ursachen zurückzuführen, die nicht für die gesamte Zielgruppe gelten.

Wichtig ist, die Initiative nicht den Benutzern zu überlassen, etwa zu warten, bis sie sich über Nutzungsprobleme beschweren. Die Initiative zur Benutzerbefragung sollte in der Regel vom Prüfer ausgehen.

A.5 Dokumentenanalyse

A.5.1 Ziele

Dient der Feststellung, ob Ergebnisse von Aktivitäten, z.B. die Anwendung von erforderlichen Analyse- und Prüfmethoden, ihren Niederschlag gefunden haben und ob die Gütekriterien dieser Methoden erfüllt wurden.

A.5.2 Gütekriterien

- Ist der dokumentierte Inhalt sachlich nachvollziehbar und verständlich und gibt es keine Hinweise auf Widersprüche (und zwar innerhalb jedes Dokuments und über verschiedene Dokumente hinweg)?
- Sind die erzielten Prüfergebnisse auf der Grundlage der Dokumente reproduzierbar?
- Ist die Verwertung der dokumentierten Prüfergebnisse in anderen Dokumenten nachvollziehbar?

A.5.3 Erläuterungen zur Dokumentenanalyse

Eine Dokumenteneinsicht ist meist notwendig, um Aussagen über den Leistungsumfang eines Softwareprodukts mit den Erfordernissen des Nutzungskontexts zu vergleichen. Auf dem Wege der Dokumenteneinsicht wird geprüft, ob und inwieweit die softwaretechnische Qualität eines Produkts den Anforderungen der Effektivität (ISO 9241-11) genügt. Es muss beispielsweise sachlich nachvollziehbar sein, für welchen vom Hersteller beabsichtigten Nutzungskontext das Produkt entwickelt ist und welche Aussagen zu den Qualitätsanforderungen der ISO /IEC 12119 gemacht sind. (Ein vom Hersteller beabsichtigter Nutzungskontext kann in die Prüfung einbezogen werden, und zwar durch Beteiligung eines Pilotanwenders.)

Dokumente, die für eine Analyse nützlich sein können, sind Produktbeschreibungen (im Sinne von ISO / IEC 12119), z.B. Funktionsbeschreibungen, Leistungsbeschreibungen, Anforderungsanalysen, Pflichtenhefte, Beschreibungen des Fachkonzepts, Ergebnisse

von Nutzungskontext-/Aufgabenanalysen, Benutzerdokumentation (z.B. Benutzerhandbuch).

Wenn die Dokumente nachvollziehbare Prüfergebnisse zu ergonomischen Qualitätsmerkmalen enthalten, so kann der Aufwand für die Normkonformitätsprüfung reduziert werden, soweit die dokumentierten Merkmale unabhängig vom Nutzungskontext des Produkts (z.B. in Übereinstimmung mit einem normverträglichen Style-Guide) sind. Um die Reproduzierbarkeit der dokumentierten Prüfergebnisse zu prüfen, können stichprobenartige Merkmalsinspektionen am Produkt zweckmäßig sein.

Eine Dokumentenanalyse bezieht i.d.R. die Übersichtlichkeit des Dokumentationsystems mit ein. Mangelnde Systematik ist meist ein Anzeichen für weitere Dokumentationsmängel. Auch ist zu prüfen, ob Querverweise auf Dokumente nutzbar sind u.ä. Insbesondere müssen Aussagen zur Produktqualität stets auf ihre Grundlagen (Tests, Prüfungen) rückführbar sein.

A.5.4 Vergleich mit einem Referenzprodukt

Ein normkonformes Referenzprodukt (auch ein Teil eines solchen Produkts) kann wie ein Dokument für den Vergleich mit einem zu prüfenden Produkt herangezogen werden, um den wiederholten Aufwand für die Konformitätsprüfung zu reduzieren. Voraussetzung ist, dass die bereits geprüften Merkmale des Referenzprodukts mit den Merkmalen des zu prüfenden Produkts in einem vergleichbaren Nutzungskontext weitgehend übereinstimmen. Normkonformität ist dann nur noch für die abweichenden Merkmale zu testen.

A.5.5 Anwendungsempfehlungen

Die Dokumentenanalyse dient eher der Vorbereitung, seltener der Durchführung einer Konformitätsprüfung. Vorbereitend werden alle geeigneten Dokumente (Entwicklungsdokumente, Benutzerdokumentation) hinsichtlich der Verwertbarkeit von Anforderungsdefinitionen analysiert. Im Idealfall findet der Prüfer nicht nur eine

Leistungsbeschreibung des Produkts vor, sondern eine Beschreibung des Nutzungskontexts (im Sinne von ISO 9241-11), aus dem Arbeitserfordernisse und Anforderungen an das Dialogsystem abgeleitet wurden.

Meist muss jedoch aufgrund der Mängel der Dokumentation die konkrete Beschreibung der Tätigkeit im Nutzungskontext nachgeholt werden (siehe Methode der Aufgabenanalyse), um Prüfkriterien herleiten zu können.

Wird ein Referenzprodukt als Dokument analysiert, so reicht das Produkt als Anschauungsobjekt nicht aus. Vielmehr muss die dem Referenzprodukt zugehörige Anforderungsspezifikation eingesehen werden, um festzustellen, ob diese für die Definition von Prüfkriterien verwertbar ist.

Für die Durchführung der Konformitätsprüfung sind Dokumente nur verwertbar, wenn diese Prüfkriterien enthalten oder sich Prüfkriterien ableiten lassen, z.B. aus erhobenen Kontext-Szenarien, Benutzerbefragungen oder Anforderungsanalysen.

Anhang B: Erhebung und Auswertung von Kontext-Szenarien

Dieser Anhang enthält einen Erhebungs- und Auswertungsrahmen für Kontext-Szenarien und ein Beispiel-Szenario. Sie dienen der Einarbeitung in die Erhebung von Kontext-Szenarien und in die Auswertung, d.h. in das Erkennen von Erfordernissen und die Ableitung von Anforderungen. Auswertungsrahmen und Beispiel-Szenario sind im Zusammenhang mit der Instruktion und der Trainingsanleitung anwendbar (siehe Abschnitt 4.1). Bei genauer Beachtung der Zielvorgaben für die mittels Leitfragen zu erhebenden Inhalte eines Nutzungskontexts können Kontext-Szenarien eine ausreichende Erhebungsobjektivität aufweisen. Wenn die erhobenen Sachverhalte des Szenarios darüber hinaus mit Blick auf die abzuleitenden Erfordernisse und Anforderungen des Auswertungsrahmen sorgfältig beschrieben sind, so dass sich aus den Abschnitten des Kontext-Szenarios die vorgegebenen Erfordernisse und Anforderungen plausibel ableiten lassen, so kann die Erhebungsobjektivität noch verbessert werden, und die Auswertungsobjektivität wird für den vorgegebenen minimalen Rahmen hinreichend gesichert.

Erhebungs- und Auswertungsrahmen

Leitfragen	Kontext-Szenario Hinweise zum Inhalt	Aufgaben- erfordernisse	Dialogprinzip	Anforderungen an die Software
Einleitung	Einleitung	Einleitung	Einleitung	Einleitung
1. Formulieren Sie die Tätigkeit in einem oder in zwei Sätzen.	Es gilt, die Grobstruktur der Aufgaben zu erfassen. Ziel ist es, einen Überblick über die Gesamttätigkeit zu erlangen. Details interessieren erst im Abschnitt "normale Durchführung". Die Beschreibung der Aufgaben ist abstrakt und knapp.	1. Werden alle Arbeitsergebnisse vollständig und korrekt erzielt? (Hier interessiert nur die Effektivität der Arbeitsergebnisse im Sinne von ISO 9241-11.)	Aufgabenangemessenheit	Werden die auszuführenden Aufgaben durch das Dialogsystem unterstützt?
2. Aus welchen Aufgaben ist die Tätigkeit zusammengesetzt? (typische Kern-Aufgaben auf-führen, d.h., wenn großer Zeitanteil, häufig wiederkehrend oder sehr wichtig) Welche dieser Kern-Aufgaben sollen durch Software unterstützt werden?	Die Gesamttätigkeit setzt sich aus Kern-Aufgaben (key tasks) zusammen. Dies sind typische und wichtige Aufgaben. Sie sollen spiegelstrichartig aufgelistet werden. Kurz anzumerken ist, welche der Aufgaben viel Zeit in Anspruch nehmen oder oft vorkommen. Inhaltlich festhalten, was der/die Befragte dazu sagen kann, für welche der Aufgaben eine Softwarenutzung wünschenswert wäre und warum.	2. Können die aufgeführten Key-Tasks effizient durchgeführt werden? (Hier interessiert, ob die Kern-Aufgaben mit angemessenem Aufwand erledigt werden können. Für welche Aufgaben wird erwartet, dass Software den Aufwand vermindern hilft?)	Aufgabenangemessenheit	Werden die auszuführenden "key tasks" durch das Dialogsystem effizient unterstützt?
3. Wie ist die Tätigkeit organisiert (z.B. als Mischarbeit, als Folge von	Hier ist zu erfragen, ob die Gesamttätigkeit als Mischarbeit oder anders organisiert ist, etwa als Abfolge von unzusammenhängenden Einzel-	3. Ist die Arbeit so organisiert (und zwar in Planung, Ausführung und Feedback), dass	Alle Dialogprinzipien können hier anwendbar sein.	Kann die Software ggf. schlecht gestaltete Arbeit kompensieren?

Aufgaben, als monotone Einzelaufgabe)?	aufgaben, als monotone Wiederhol-tätigkeit, als Springertätigkeit usw. Eventuell gegebene ergonomische Mängel der Arbeitsgestaltung festhalten, die nicht der Software anzulasten sind.	vollständige Tätigkeiten ausgeübt werden (ISO 9241-2)?		
Voraussetzungen	Voraussetzungen	Voraussetzungen	Voraussetzungen	Voraussetzungen
4. Welche Qualifikation ist zur Bewältigung der Aufgaben erforderlich (Aufgabenbewältigung / Softwarenutzung)? Welche Vorkenntnisse fehlen ggf.?	Die Qualifikation der befragten Person ist zu erfassen. Anschließend ist der Bezug zur Nutzung von Software herzustellen (jetzt und künftig). Dabei soll erfragt werden, welche Vorbedingungen für die Softwarenutzung erforderlich und gegeben sind, gewünscht sind oder (noch) fehlen.	4. Kann die Arbeitsperson die erworbenen Fachkenntnisse zur Erledigung der Aufgaben nutzen? 5. Sind die erforderlichen Vorkenntnisse vorhanden?	Selbstbeschreibungsfähigkeit Erwartungskonformität Lernförderlichkeit	Ist die Software mit den erworbenen fachlichen Vorkenntnissen nutzbar?
5. Wer (bzw. welches Ereignis) bestimmt, was zu tun ist? (Wer trifft die Auswahl)? Selbständigkeit der Bearbeitung, Arbeitsteilung, externe Datenquellen)	Anlässe feststellen, aus denen sich Tätigkeiten am Arbeitsplatz ergeben. Wichtig ist auch festzustellen, welche Datenquellen den Input für die auszuführenden Tätigkeiten liefern.	6. Kann die Arbeitsperson darüber entscheiden, welche der Aufgaben sie ausführen will (es sei denn, die Folge der Aufgaben ist durch Prioritäten vorgegeben)?	Steuerbarkeit	Kann der Benutzer die nächste zu bearbeitende Aufgabe am System frei wählen? Kann der Benutzer den Dialog so steuern, wie es die Aufgabenbearbeitung erfordert?
6. Welche Hilfsmittel sind erforderlich (für die Aufgabenbewältigung / zur Software-	Es gilt festzuhalten, welche Hilfsmittel für die Erledigung von Aufgaben genutzt werden, welche nützlich sind, welche weniger und warum. Auch feststellen, ob und welche	7. Sind alle Hilfsmittel gegeben, die für das Erledigen der Aufgabe erforderlich sind?	Aufgabenangemessenheit	Sind alle für die am Softwareprodukt auszuführenden Arbeitstätigkeiten

warennutzung)? Welche davon fehlen ggf., welche sind zusätzlich gewünscht?	Hilfsmittel fehlen. Der/die Befragte soll auch Wünsche äußern über zusätzliche Hilfsmittel.			erforderlichen Hilfsmittel gegeben?
Normale Durchführung	Normale Durchführung	Normale Durchführung	Normale Durchführung	Normale Durchführung
7. Welche Arbeitsschritte sind durchzuführen?	Dies ist die umfangreichste Frage der Erhebung. Hinsichtlich der Kernaufgaben sind die Tätigkeitsschritte zu erfragen, die zur Erreichung von Ergebnissen notwendig sind. Der "rote Faden des Tätigkeitsablaufs" ist zu erfassen. Es geht bei dieser Frage um eine Detaillierung der Frage 2. Die Antworten werden als Grundlage für die spätere Analyse der Arbeitsschritte am Softwareprodukt verwertet (siehe Use-Szenario).	8. Können die Ergebnisse der Arbeitsschritte vollständig und korrekt erzielt werden? 9. Ist der Stand der bisherigen Bearbeitung erkennbar?	Aufgabenangemessenheit Aufgabenangemessenheit Selbstbeschreibungsfähigkeit	Werden die anfallenden Arbeiten durch das Dialogsystem unterstützt? Ist der Bearbeitungsstand so angezeigt, dass dem Benutzer die Bearbeitung des nächsten Arbeitsschritts klar ist?
8. Welche Arbeitsschritte kehren häufig wieder? (Automatisierung gewünscht / erforderlich?)	Hier gilt es herauszufinden, welche Tätigkeiten eher dem Menschen überlassen bleiben sollen und welche besser der Computer erledigen soll. Wiederkehrende Arbeiten können oft automatisiert werden. Es gilt festzustellen, welche "Computerisierung" als nützlich empfunden wird.	10. Sind die für die Aufgabenbearbeitung erforderlichen Informationen unmittelbar verfügbar? 11. Wird die Arbeitsperson bei wiederkehrenden Aufgaben besonders unterstützt?	Aufgabenangemessenheit Selbstbeschreibungsfähigkeit	Sind die Informationen, die zur Erledigung der Aufgaben notwendig sind, auf dem Bildschirm übersichtlich verfügbar? Werden wiederkehrende Aufgaben vom Dialogsystem besonders unterstützt? Werden Aufgaben, die das System selbst

Feedback erhält die Arbeitsperson in bezug auf ihre Arbeitsergebnisse und die Wirkung ihrer Arbeit?	Kollegen oder Vorgesetzten oder von seiten der Software festzustellen. Wichtig ist auch, ob und inwieweit das Feedback ermöglicht, den Erfolg oder die Wirkungen der eigenen Arbeit zu erkennen oder zu überschauen. (Externes Feedback, siehe Frage 21.)	person eine Rückmeldung über erreichte Arbeitsergebnisse? 20. Ist der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson der zu bearbeitende Vorgang überschaubar ist?	bungsfähigkeit Aufgaben-angemessenheit	Benutzer an den Systemrückmeldungen das Ergebnis seiner Arbeit? Ist der Bearbeitungsstand so angezeigt, dass dem Benutzer die Beurteilung des Vorgangs möglich ist?
Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung
14. Welche Unterbrechungen gibt es und warum? Welche Störungen treten auf? (organisatorisch / sozial/technisch)	Zu erfragen ist, auf welche Weise Arbeitstätigkeiten unterbrochen werden. Es interessieren die Störungen, die sich aus der Arbeitsorganisation ergeben und die Art, wie damit umgegangen wird. Wichtig ist, dass während des Interviews der Blick auf die Software oder andere technische Möglichkeiten gerichtet wird, um herauszufinden, ob man dadurch mit Störungen besser umgehen kann.	21. Ist die für die Aufgabenbearbeitung erforderliche Information – auch bei nötigem Aufgabenwechsel – unmittelbar verfügbar? 22. Kann die Arbeitsperson die Ausführung der Aufgabe unterbrechen, um auf externe Anforderungen zu reagieren?	Aufgaben-angemessenheit Steuerbarkeit Steuerbarkeit	Sind die Informationen, die zur Erledigung der Aufgabe notwendig sind, auf dem Bildschirm übersichtlich verfügbar? Bietet das System die Möglichkeit, eine aktuelle Bearbeitung ausreichend schnell zu unterbrechen, um sich einer anderen Aufgabe zu widmen?

				Kann die unterbrochene Aufgabe später ohne unnötigen Aufwand wieder aufgenommen werden?
15. Wie werden Fehler zurückgemeldet und behoben? (organisatorisch / sozial / technisch)	Hier interessieren in erster Linie die fachlichen Fehler sowie deren Aufdeckung und Behandlung im Umfeld der Arbeitstätigkeit. Mit Blick auf eine möglicherweise bessere Beherrschung von Fehlersituationen durch Software-Unterstützung soll diese Frage erweitert werden. Fehler jedoch, die bei der Nutzung von Software auftreten, bleiben unberücksichtigt, da sie besser im Use-Szenario untersucht werden können.	23. Können Fehler von der Arbeitsperson selbst behoben werden? 24. Kann der Zeitpunkt der Behebung selbst bestimmt werden, soweit er nicht anderweitig festgelegt ist? 25. Bekommt die Arbeitsperson eine Rückmeldung über fehlerhafte Arbeitsergebnisse?	Fehlertoleranz, Steuerbarkeit Fehlertoleranz Selbstbeschreibungsfähigkeit, Fehlertoleranz	Können Fehler ohne unnötig großen Aufwand behoben werden? Erlaubt die Software den Aufschub der Fehlerbehandlung? Erkennt der Benutzer an den Systemrückmeldungen fehlerhafte Ergebnisse?
16. Welche wichtigen Sonderfälle müssen berücksichtigt werden? (bzw. fallen dem Benutzer spontan ein; z.B. zur Arbeitsteilung / Zusammenarbeit)	Sonderfälle werden in der üblichen Interviewsituation, in der Regelfälle des Arbeitsalltags erfragt werden, selten genannt. Trotzdem sollte die Frage gestellt werden und dem / der Befragten Zeit zur Antwort gelassen werden.	Ein generelles Erfordernis kann hier nicht vorgegeben werden. Es muss vielmehr aus den je speziellen Sonderfällen ein Erfordernis erschlossen werden. Viele Sonderfälle haben etwas mit flexibler Anpassung zu tun.	Aufgaben-angemessenheit	Werden diese Sonderfälle von der Software angemessen unterstützt?

Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen
17. Welche Organisationsziele gibt es im Hinblick auf die Tätigkeit?	Der befragten Person sind die abstrakten Ziele einer Organisation bekannt, etwa so, wie sie vom Management artikuliert werden. Hier kommt es darauf an, den Zusammenhang der Organisationsziele mit der Computerisierung festzustellen. Ein sehr oft genanntes Ziel ist die Produktivitätssteigerung der Mitarbeiter, d.h. Rationalisierung durch Computereinsatz.	26. Ist das Ziel der Organisation durch Einsatz der Arbeitsmittel effizient erreichbar?	Aufgabenangemessenheit	Werden die Ziele der Arbeitsorganisation (z.B. ähnliche Aufgaben ähnlich zu bearbeiten oder Aufgaben mit weniger Aufwand zu erledigen) durch die Software unterstützt?
18. Gibt es Mechanismen zur Leistungssteuerung / Leistungskontrolle (Wenn ja, welche? Sind diese erforderlich?)	Diese Frage steht im Zusammenhang mit einer Anforderung der BildscharbV. Für die Beurteilung der Software ist besonders wichtig, ob und inwieweit die fraglichen Mechanismen die Geschwindigkeit der Arbeitstätigkeit beeinflussen.	27. Steht die Geschwindigkeit der Bearbeitung allein unter der Kontrolle der Arbeitsperson (es sei denn, gesetzte Prioritäten sind zu beachten)?	Steuerbarkeit	Wird die Geschwindigkeit des Dialogs (hier: Arbeitsleistung) allein vom Benutzer gesteuert? Sind im Dialogsystem integrierte Leistungskontrollen abschaltbar?
19. Welchen Überblick hat der Benutzer im Hinblick auf die Gesamttätigkeit?	Wenn mangelnder Überblick festgestellt wird, z.B. hinsichtlich der Verwertung / Weiterleitung / Weiterbearbeitung der Arbeitsergebnisse, so interessiert, ob der mangelnde Überblick durch Softwarefunktionen behoben werden könnte.	28. Ist der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson der zu bearbeitende Vorgang überschaubar ist?	Aufgabenangemessenheit	Ist der Bearbeitungsstand so angezeigt, dass der Benutzer einen Überblick über die Gesamttätigkeit erhält?

20. Welche Änderungen, die die Aufgabenbearbeitung beeinflussen, sind zu erwarten oder werden gewünscht? Welche Vorschläge hat der / die Befragte dazu?	Hier sind nicht nur organisatorische Änderungen zu erfassen, sondern jegliche Änderungen im Umfeld, z.B. auch rechtliche Entwicklungen. Hier ist auch wichtig zu erfahren, welche Änderungen und welchen Nutzen man vom Einsatz der Software erwartet. Auch visionäre Änderungsvorstellungen erfragen.	29. Wenn es Änderungen gibt: Werden die ggf. anstehenden Änderungen bei der Erledigung von Aufgaben eingeplant?	Aufgabenangemessenheit	Werden bei der Gestaltung /Anpassung des Dialogs die ggf. anstehenden Änderungen berücksichtigt?
21. Von welchen Arbeitsergebnissen / Arbeitsschritten sind Dritte (z.B. Kunden) direkt betroffen? Und was folgt daraus?	Es gilt festzustellen, wie sich die Arbeit des / der Befragten auf die Welt außerhalb der eigenen Arbeitsorganisation auswirkt. Beispielsweise wirkt sich die Dienstleistung eines Büros auf die Klienten aus. Hier ist der Blick darauf zu richten, ob und inwieweit die Computerisierung einen Einfluss hat (positiv oder negativ). Gegebenenfalls sind die Folgen zu beschreiben oder was man anders oder besser machen könnte.	30. Bekommt die Arbeitsperson eine Rückmeldung über erreichte oder fehlerhafte Arbeitsergebnisse?	Selbstbeschreibungsfähigkeit Fehlertoleranz	Erkennt der Benutzer an den Systemrückmeldungen das Ergebnis seiner Arbeit? Erkennt der Benutzer an den Systemrückmeldungen fehlerhafte Ergebnisse?
22. Welche Stressfaktoren gibt es und wie wird damit umgegangen?	Stressfaktoren können durch die Aufgabe und Organisation sowie durch die Software gegeben sein.	31. Hat die Arbeitsperson die Möglichkeit, die Stressfaktoren zu beeinflussen, um die entstehenden Belastungen zu regulieren? (Zum Beispiel: Arbeitszeit einteilen, eingehende Telefonate umleiten.)	Steuerbarkeit	Kann der Benutzer den Dialog so steuern, wie es die Aufgabenbearbeitung / Belastung erfordert?
Sonstige Bemerkungen der befragten Person zu bereits aufgetretenen	Hier Beispiele sammeln, falls die befragte Person bereits während des Interviews etwas über aufgetretene Nutzungsprobleme berichtet. Im allgemeinen werden solche Probleme	32. Ist eine Rückkopplung über Nutzungsprobleme so organisiert, dass Systemverbesserungen ermöglicht werden?	Aufgabenangemessenheit	Werden gemeldete Nutzungsprobleme vom Hersteller bei der

Nutzungs- problemen.	besser an Hand von Use-Szenarien analysiert.			Weiterentwicklung angemessen berücksichtigt?
-------------------------	---	--	--	--

Beispiel-Szenario

Leitfragen	Kontext-Szenario Hinweise zum Inhalt	Aufgaben- erfordernisse	Dialogprinzip	Anforderungen an die Software - Prüfkriterien -
Einleitung	Einleitung	Einleitung	Einleitung	Einleitung
1. Formulieren Sie die Tätigkeit in einem oder in zwei Sätzen.	Herr B. bearbeitet, koordiniert und betreut in einer Verwaltung die städtischen Mietangelegenheiten für Wohnungen und gewerbliche Räume, und zwar vom ersten Kontakt bis zur Beendigung des Mietverhältnisses. Er ist für einen bestimmten Stadtbezirk zuständig und hat vier KollegInnen, die für die anderen Stadtbezirke zuständig sind.	1. Werden alle Arbeitsergebnisse vollständig und korrekt erzielt? Werden vergleichbare Arbeitsergebnisse erzielt? Werden die Ergebnisse auf vergleichbare Weise erzielt?	Aufgaben- angemessenheit Aufgaben- angemessenheit Erwartungs- konformität	Werden die auszuführenden Aufgaben (Koordination und Verwaltung von Mietangelegenheiten) durch das Dialogsystem unterstützt? Können die für die Stadtbezirke ähnlichen Aufgaben ähnlich bearbeitet werden?
2. Aus welchen Aufgaben ist die Tätigkeit zusammengesetzt? (typische Kern-Aufgaben auf-führen, d.h., wenn großer Zeitanteil, häufig wiederkehrend oder sehr wichtig) Welche dieser	Typische Aufgaben sind: <ul style="list-style-type: none">• Schriftverkehr mit Mietern, deren Nachbarn oder mit Anwälten,• Formulargesteuerte Bearbeitung von Mietverträgen und beim Verwalten laufender Mietverhältnisse,• Kündigungen bearbeiten,• Koordination mit anderen Verwaltungsstellen,• Außendienst (Kontakt mit Mietern)	2. Können die aufgeführten Key-Tasks effizient durchgeführt werden?	Aufgaben- angemessenheit	Werden die aufgeführten "key tasks" (mit Ausnahme des Außendienstes) durch das Dialogsystem effizient unterstützt?

Kern-Aufgaben sollen durch Software unterstützt werden?	Alle Arbeiten werden und sollen durch den Computer unterstützt werden, außer im Außendienst.			
3. Wie ist die Tätigkeit organisiert (z.B. als Mischarbeit, als Folge von Aufgaben, als monotone Einzelaufgabe)?	Diese Arbeiten werden für alle der etwa 700 zu verwaltenden Mietobjekte erledigt, sobald sie anfallen. Es handelt sich um einen typischen Mischarbeitsplatz. Die Arbeiten werden selbständig erledigt.	3. Ist die Arbeit so organisiert (und zwar in Planung, Ausführung und Feedback), dass vollständige Tätigkeiten ausgeübt werden (ISO 9241-2)?	Alle Dialogprinzipien können hier gemeint sein.	Kann die Software ggf. schlecht gestaltete Arbeit kompensieren?
Voraussetzungen	Voraussetzungen	Voraussetzungen	Voraussetzung	Voraussetzungen
4. Welche Qualifikation ist zur Bewältigung der Aufgaben erforderlich (Aufgabenbewältigung / Softwarenutzung)? Welche Vorkenntnisse fehlen ggf.?	Notwendig sind allgemeine Rechts- und Verwaltungskenntnisse, besonders aber Kenntnisse des Mietrechts. Mietrechtskommentare und Mietrechtsseminare helfen, auf dem Laufenden zu bleiben. Die Einarbeitung am PC wurde durch eine Benutzerschulung gefördert. Es existieren Benutzerhandbücher. Kenntnisse in Windows NT, Textverarbeitung und Excel sind vorhanden.	4. Kann die Arbeitsperson die erworbenen Fachkenntnisse zur Erledigung der Aufgaben nutzen? 5. Sind die erforderlichen Vorkenntnisse vorhanden?	Selbstbeschreibungsfähigkeit Erwartungskonformität Lernförderlichkeit	Ist die Software mit den erworbenen Vorkenntnissen nutzbar?
5. Wer (bzw. welches Ereignis bestimmt, was zu tun ist? (Wer trifft die Auswahl? Selbständigkeit der Bearbeitung, Arbeitsteilung, externe Datenquellen)	Der laufende Posteingang bestimmt, was zu bearbeiten ist. Entweder werden die Arbeiten selbständig durchgeführt oder deren teilweise Durchführung veranlasst, wenn Koordination mit anderen Stellen der Verwaltung oder Kollegen erforderlich ist.	6. Kann die Arbeitsperson darüber entscheiden, welche der Aufgaben sie ausführen will (es sei denn, die Folge der Aufgaben ist durch Prioritäten vorgegeben)?	Steuerbarkeit	Kann der Benutzer anhand der Posteingänge oder Telefonkontakte bestimmen, welches die nächste zu bearbeitende Aufgabe am System ist? Kann der Benutzer die

				Koordination mit anderen Verwaltungsstellen am System so steuern, wie es die Aufgabenbearbeitung erfordert?
6. Welche Hilfsmittel sind erforderlich (für die Aufgabenbewältigung / zur Softwarenutzung)? Welche davon fehlen ggf., welche sind zusätzlich gewünscht?	Außer einem Taschenrechner sind keine weiteren Hilfsmittel für die Arbeit am Schreibtisch oder am Bildschirmarbeitsplatz notwendig. Die Nutzung der früher üblichen Karteikarten geht immer mehr zurück und wird vielleicht eines Tages ganz überflüssig sein. Mietspiegel sind am Rechner verfügbar und geplant ist eine CD über Mietrecht. Häufiger Wechsel zwischen dem Rechner und schriftlichen Unterlagen (Hilfsmitteln) wird dadurch reduziert. Ein Arbeitsplatz-Drucker wäre nützlich.	7. Sind alle Hilfsmittel gegeben, die für das Erledigen der Aufgabe erforderlich sind? Ist die für die Aufgabenbearbeitung erforderliche Information unmittelbar verfügbar?	Aufgabenangemessenheit Aufgabenangemessenheit	Ist ein "Calculator" am Bildschirm vorhanden? Ist ein traditioneller Taschenrechner notwendig und ggf. nutzbar? Ist die Nutzung des Systems ohne die traditionellen Karteikarten möglich und wünschenswert? Ist die Nutzung des Systems ohne Mietspiegel als Hardcopy möglich und wünschenswert? Wird ein ggf. störender Medienwechsel zwischen Online-Infos und schriftlichem Mietspiegel vermieden?
Normale Durchführung	Normale Durchführung	Normale Durchführung	Normale Durchführung	Normale Durchführung

7. Welche Arbeitsschritte sind durchzuführen?	Regelmäßig oder periodisch anfallende Arbeiten sind z.B. bei der Neuvermietung: Wohnung "frei" melden, die Vorbereitung und Überwachung von Modernisierungen freigelegener Wohnungen, Wohnungsangebote schreiben, für Bewerber Besichtigungstermine vereinbaren, Bezugstermine festlegen, ortsübliche Vergleichsmiete festlegen, klären, wer die Miete zahlt (z.B. manchmal das Sozialamt), klären, wer die Kautions zahlt, Kautions berechnen, Wohnungen übergeben, Mietverträge ausstellen. Bei Kündigung fallen regelmäßig Arbeiten an wie Wohnungsbesichtigung, Abnahme der Wohnung, Vorbereitung der Neuvermietung.	8. Können die Ergebnisse der Arbeitsschritte vollständig und korrekt erzielt werden? 9. Ist der Stand der bisherigen Bearbeitung erkennbar? 10. Sind die für die Aufgabenbearbeitung erforderlichen Informationen unmittelbar verfügbar?	Aufgaben-angemessenheit Aufgaben-angemessenheit Selbstbeschreibungsfähigkeit Aufgaben-angemessenheit Selbstbeschreibungsfähigkeit	Werden die aufgeführten periodisch anfallenden Arbeiten durch das Dialogsystem unterstützt? Ist der Bearbeitungsstand der aufgeführten Aufgaben so angezeigt, dass dem Benutzer die Bearbeitung des nächsten Arbeitsschritts klar ist? Sind die Informationen, die zur Erledigung der aufgeführten Aufgaben notwendig sind, auf dem Bildschirm übersichtlich verfügbar?
8. Welche Arbeitsschritte kehren häufig wieder? (Automatisierung gewünscht / erforderlich?)	Häufig anfallende Arbeiten entstehen bei Wohnungsanfragen, beim Ausstellen von Mietbescheinigungen, beim Bearbeiten von Reparaturmeldungen oder Mieterbeschwerden (hauptsächlich bei Nachbarschaftsstreitigkeiten). Jährlich wiederkehrende Arbeiten sind Betriebskosten berechnen, Heizkostenabrechnungen veranlassen; alle drei Jahre werden Mieterhöhungen	11. Wird die Arbeitsperson bei wiederkehrenden Aufgaben besonders unterstützt?	Aufgaben-angemessenheit	Werden die aufgeführten wiederkehrenden Aufgaben vom Dialogsystem besonders unterstützt? Werden unter den wiederkehrenden Aufgaben, die vom System

	berechnet. Alles, was der Rechner können soll, das kann er. Mir fallen keine Arbeiten auf, die mir lästig wären und die der Computer übernehmen sollte.			selbst auszuführen, vom System automatisch ausgeführt?
9. Welche Arbeitsschritte werden automatisch durchgeführt? Sind bei automatisierten Arbeitsschritten Einflussmöglichkeiten des Benutzers vorhanden / erlaubt / gewünscht / erforderlich?	Regelmäßig durchzuführende Aufgaben sind zwar rechnerunterstützt, aber nicht automatisiert. Da eine mit allen beteiligten Verwaltungsstellen gemeinsame Vorgangsbearbeitung eingerichtet werden soll, kann der Rechner prüfen, ob an alles gedacht wurde, bevor der Vorgang weiterbearbeitet oder abgeschlossen wird. Hierbei wird der Stand der Vorgangsbearbeitung besser überschaubar. Alle mit einem Mietverhältnis zusammenhängenden Arbeiten gehen in die Vorgangsbearbeitung ein. Es gibt Vorgänge, die nur das Fachamt betreffen, und solche, in die andere Ämter einbezogen sind, etwa, wenn das Stadtbauamt an der gemeinsamen Bearbeitung von Schadensfällen mitarbeitet.	12. Ist der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson der zu bearbeitende Vorgang überschaubar ist? 13. Kann die Arbeitsperson in automatische Vorgänge steuernd eingreifen (z.B. Abschalten, Anhalten).	Aufgaben-angemessenheit Steuerbarkeit	Ist der Bearbeitungsstand der regelmäßig durchzuführenden Aufgaben so angezeigt, dass dem Benutzer die Beurteilung des Vorgangs möglich ist? Entfällt
10. Kommt es vor, dass mehrere Benutzer gleichzeitig an dem gleichen Objekt (z.B. Vorgang, Akte, Dokument, Datensatz) arbeiten müssen?	Ein lesender Zugriff auf die Objekte der Kollegen ist möglich. Verwalten der Mietobjekte heißt oft, mit anderen Verwaltungsstellen koordinieren, z.B. bei Instandsetzungsarbeiten mit der Bauunterhaltung, bei Mietrückständen mit dem Amt für Soziales und Wohnen oder mit der Stadtkasse, oder bei Betriebskostenabrechnungen. Zur Zeit	14. Sind die für die Koordinations-Aufgaben erforderlichen Informationen unmittelbar verfügbar?	Aufgaben-angemessenheit	Sind die Informationen, die von anderen Verwaltungsstellen geliefert werden und zur Erledigung der Aufgabe notwendig sind,

	<p>wird in einer Pilotanwendung die zunehmende Vernetzung der PCs erprobt. Vorteilhaft ist z.B., dass man Antworten an die Bürger anhand eines Antwort-Entwurfs mit mehreren Ämtern gemeinsam erarbeiten oder abstimmen kann. Oder man kann z.B. bei Schadensmeldungen der Mieter die Reparatur- oder Wartungsaufträge besser kontrollieren, soweit andere Stellen konsultiert werden müssen. Auch mit der Stadtkasse kann z.B. die Führung des Personenkontos besser abgestimmt werden.</p> <p>Bei dieser Koordinierung entstand früher viel Schreibaufwand, der durch Einführung einheitlicher Vordrucke stark reduziert wurde. Verschiedene Ämter sind mit demselben Vorgang befasst. Daher wird für alle beteiligten Stellen ein Zugriff auf die erforderliche Information notwendig sein. Im Zuge der Vernetzung sollen für alle beteiligten Verwaltungsstellen einheitliche Vordrucke am Bildschirm verwendet werden. Zugriffsberechtigungen / Leseberechtigung / Zuständigkeiten für Daten und Programmen werden erteilt. Man erwartet von der computergestützten Koordinierung der Arbeiten eine erhebliche Beschleunigung bei der Vorbereitung schriftlicher Arbeiten, bei den Berechnungen und bei der Kontrolle der Vorgangs-</p>	<p>Ist der Stand der Bearbeitung erkennbar, den eine kooperierende Stelle erreicht hat, so dass der Arbeitsperson die erwartete weitere Bearbeitung klar ist?</p>	<p>Aufgaben-angemessenheit Selbstbeschreibungsfähigkeit</p>	<p>auf dem Bildschirm übersichtlich verfügbar? Ist der Bearbeitungsstand eines Vorgangs am Bildschirm so angezeigt, dass dem Benutzer die Bearbeitung des nächsten Arbeitsschritts ohne weiteres möglich ist?</p>
		15. Sind die Berechtigungen angemessen erteilt?	<p>Aufgaben-angemessenheit</p>	<p>Unterstützt das Dialogsystem die für die Aufgabendurchführung erforderlichen Berechtigungen? Wird die Geschwindigkeit der Vorgangsbearbeitung allein</p>
		Steht die Geschwindigkeit der Bearbeitung allein unter der Kontrolle der Arbeitsperson (es sei denn, gesetzte Prioritäten sind zu beachten)?	<p>Steuerbarkeit</p>	

	bearbeitung.			vom Benutzer gesteuert?
11. Gibt es eine festgelegte Abfolge der Arbeitsschritte und wenn ja, wie sieht diese aus? (Ist eine Flexibilität sinnvoll / nötig?)	<p>Der Arbeitstag ist selbstorganisiert. Natürlich sind Terminsachen besonders zu beachten. Aber alle Vorbereitungen für Termine werden eigenständig getroffen. Es gibt zwar einen gewissen Entscheidungsspielraum bei der Bearbeitung bestimmter Mietprobleme, z.B. bei Ratenvereinbarungen für Mietrückstände, aber die gesetzlichen Bestimmungen setzen den Rahmen.</p>	16. Kann die Arbeitsperson darüber entscheiden, welche der Aufgaben sie ausführen will (es sei denn, die zweckmäßige Folge der Aufgaben ist vorgegeben)?	<p>Steuerbarkeit</p> <p>Steuerbarkeit</p> <p>Aufgaben-angemessenheit Fehlertoleranz</p>	<p>Kann der Benutzer die nächste zu bearbeitende Aufgabe am System frei wählen? Kann der Benutzer die Bearbeitung von Terminsachen so steuern, wie es die Aufgabebearbeitung erfordert? Erzwingt die Software dort, wo es festgelegte Ratenvereinbarungen gibt oder Mietrückstände entstanden sind die entsprechende Einhaltung der Reihenfolge der Arbeitsschritte? Kann der Benutzer den Dialog so steuern, wie es die Kooperation erfordert?</p>
12. Welche Ergebnisse / Teilergebnisse entstehen und wie	<p>Auch mit den Kollegen der gleichen Dienststelle muss Arbeit koordiniert werde, etwa, wenn eine Haussanierung die vorübergehende Umsetzung von Mietern erfordert.</p>	Ist trotz geregelter Zuständigkeiten eine Kooperation bei der Aufgabebearbeitung möglich?	<p>Steuerbarkeit</p>	<p>Unterstützt die Software die Weiterleitung von Ergebnissen an</p>
12. Welche Ergebnisse / Teilergebnisse entstehen und wie	<p>Es existiert ein zentraler Posteingang. Der Amtsleiter sieht sich den wichtigen Schriftverkehr an. Auf diese Weise behält er die Übersicht und kontrolliert,</p>	17. Können Ergebnisse / Teilergebnisse in einer vom Empfänger verwertbaren Form weitergegeben	<p>Aufgaben-angemessenheit</p>	

werden diese ggf. verwertet / weitergeführt?	was sich an den Arbeitsplätzen so tut. In gleicher Weise kontrollieren auch Sachgebietsleiterin und Abteilungsleiterin. Es gibt eine elektronische Wiedervorlage. Bei den Mietverhältnissen gibt es verschiedene Bearbeitungszustände, die im Wohnungsverwaltungssystem erkennbar sind.	werden? 18. Ist bei eingegangenen Teilergebnissen der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson die erwartete Tätigkeit klar ist?	Selbstbeschreibungsfähigkeit	Vorgesetzte, Kollegen und andere Verwaltungsstellen in verwertbarer Form? Ist der Bearbeitungsstand in der Wiedervorlage und im Vorgang so angezeigt, dass dem Benutzer die Bearbeitung des nächsten Arbeitsschritts möglich ist?
13. Welches Feedback erhält die Arbeitsperson in bezug auf ihre Arbeitsergebnisse und die Wirkung ihrer Arbeit?	Ferner gibt es eine Unterschriften-Regelung; in wichtigen Mietangelegenheiten zeichnen Vorgesetzte mit. Im System kann man z.B. feststellen, ob eine Buchung auf dem Personenkonto durchgeführt wurde oder ob eine Mahnung raus ist.	19. Bekommt die Arbeitsperson eine Rückmeldung über erreichte Arbeitsergebnisse? 20. Ist der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson der zu bearbeitende Vorgang überschaubar ist?	Selbstbeschreibungsfähigkeit Aufgabenangemessenheit	Erkennt der Benutzer an den Systemrückmeldungen über Buchungen oder Mahnungen das Ergebnis seiner Arbeit? Ist der Bearbeitungsstand so angezeigt, dass dem Benutzer die Beurteilung des Vorgangs möglich ist?

Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung	Besonderheiten bei der Durchführung
14. Welche Unterbrechungen gibt es und warum? Welche Störungen treten auf? (organisatorisch / sozial / technisch)	Um eine möglichst störungsfreie Bearbeitung von Vorgängen zu ermöglichen, gibt es Sprechstunden für das Publikum. Aber die Bearbeitung komplizierter Vorgänge wird durch Telefonanrufe immer noch gestört. Mittels Anrufumleitung wurde zwar versucht, alle Anfragen bei einem Sachbearbeiter zu bündeln, um die anderen Kollegen in Ruhe arbeiten zu lassen. Aber bewährt hat sich nur die gelegentliche Anrufumleitung an den jeweiligen dienstlichen Vertreter.	21. Ist die für die Aufgabenbearbeitung erforderliche Information – auch bei nötigem Aufgabenwechsel – unmittelbar verfügbar? 22. Kann die Arbeitsperson die Ausführung der Aufgabe unterbrechen, um auf externe Anforderungen zu reagieren?	Aufgabenangemessenheit Steuerbarkeit Steuerbarkeit	Sind die Informationen auch zu den am Telefon zu erledigenden Aufgaben auf dem Bildschirm übersichtlich verfügbar? Bietet das System die Möglichkeit, bei Telefonanrufen oder in der Sprechstunde für das Publikum eine aktuelle Bearbeitung ausreichend schnell zu unterbrechen, um sich einer anderen Aufgabe zu widmen? Kann die unterbrochene Aufgabe später ohne unnötigen Aufwand wieder aufgenommen werden?
15. Wie werden Fehler zurückgemeldet und behoben?	Fehler bei der täglichen Arbeit können durch den Sachbearbeiter selbst ausgebügelt werden. Fehler entstehen z.B. bei Betriebskostenberechnungen	23. Können Fehler von der Arbeitsperson selbst behoben werden?	Fehlertoleranz, Steuerbarkeit	Können Fehler ohne unnötig großen Aufwand behoben werden?

(organisatorisch / sozial / technisch)	aufgrund fehlerhafter Dateneingaben. Fehler fallen z.B. dem Mieter oder dem Vorgesetzten auf.	24. Kann der Zeitpunkt der Behebung selbst bestimmt werden, soweit er nicht anderweitig festgelegt ist?	Fehlertoleranz	Entfällt
		25. Bekommt die Arbeitsperson eine Rückmeldung über fehlerhafte Arbeitsergebnisse?	Selbstbeschreibungsfähigkeit, Fehlertoleranz	Erkennt der Benutzer an den Systemrückmeldungen fehlerhafte Dateneingaben oder falsche Berechnungen der Betriebskosten?
16. Welche wichtigen Sonderfälle müssen berücksichtigt werden? (bzw. fallen dem Benutzer spontan ein) (z.B. Arbeitsteilung / Zusammenarbeit)	Wichtige Sonderfälle müssen nicht berücksichtigt werden. - In der Zusammenarbeit mit den Kollegen werden rechtliche Probleme besprochen. Zusammenarbeit ist auch erforderlich, wenn z.B. ein Mieter in den Bereich eines anderen Sachbearbeiters umzieht oder wegen Modernisierungsarbeiten umgesetzt werden muss.	Ist infolge geänderter Zuständigkeit eine effektive Übergabe / Übernahme und Bearbeitung der Vorgangsdaten möglich?	Erwartungskonformität	Kann der Benutzer die übernommenen Daten in der erwarteten Weise weiterbearbeiten?
Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen	Organisatorische Rahmenbedingungen
17. Welche Organisationsziele gibt es im Hinblick auf die Tätigkeit?	Ziel der Verwaltung ist die Gleichbehandlung aller Mieter. Abstimmungen zwischen den Kollegen dienen dazu, "dieselbe Linie zu fahren". Mit der Computerisierung soll der Service für die Bürger verbessert werden, hauptsächlich die Schnelligkeit der Bearbeitung.	26. Ist das Ziel der Organisation durch Einsatz der Arbeitsmittel effizient erreichbar?	Aufgabenangemessenheit	Werden die Ziele der Arbeitsorganisation (z.B. ähnliche Aufgaben ähnlich zu bearbeiten oder Aufgaben mit weniger Aufwand zu erledigen) durch

				die Software unterstützt?
18. Gibt es Mechanismen zur Leistungssteuerung / Leistungskontrolle (Wenn ja, welche? Sind diese erforderlich?)	Eine direkte Leistungskontrolle oder Mechanismen der Leistungssteuerung gibt es nicht.			
19. Welchen Überblick hat der Benutzer im Hinblick auf die Gesamttätigkeit?	Vor Einführung der integrierten Vorgangsbearbeitung bleibt die Übersicht über die Gesamttätigkeit schlecht.	28. Ist der Stand der Bearbeitung erkennbar, so dass der Arbeitsperson der zu bearbeitende Vorgang überschaubar ist?	Aufgabenangemessenheit	Ist der Bearbeitungsstand eines Vorgangs so angezeigt, dass der Benutzer einen Überblick über die Gesamttätigkeit erhält?
20. Welche Änderungen, die die Aufgabenbearbeitung beeinflussen, sind zu erwarten oder werden gewünscht? Welche Vorschläge hat der / die Befragte dazu?	Organisatorische Änderungen sind nicht zu erwarten			Entfällt
21. Von welchen Arbeitsergebnissen / Arbeitsschritten sind Dritte (z.B. Kunden) direkt betroffen? Und was folgt daraus?	Am meisten wirkt sich die Betriebskostenabrechnung aus, da die Mieter darauf reagieren. Auch Mahnungen der Stadtkasse wirken sich aus, etwa, wenn sich die Betroffenen beschweren. Ich kann mir vorstellen, dass die heutigen Computerausdrucke auf die Altmietler unpersönlicher wirken als die früheren Schriftstücke	30. Bekommt die Arbeitsperson eine Rückmeldung über erreichte oder fehlerhafte Arbeitsergebnisse?	Selbstbeschreibungsfähigkeit	Erkennt der Benutzer an den Systemrückmeldungen über Mahnungen, Berechnungen, Schriftstücken und Formularen das Ergebnis seiner

	des Amtes. Aber Beschwerden sind mir nicht bekannt. Wenn es Beschwerden von außen gibt, die im Zusammenhang mit dem Computer stehen, so gebe ich diese an den DV-Beauftragten weiter. Und dieser gibt die Sache an den Hersteller weiter. Aber das liegt ganz allein an meiner Initiative.	Ist eine Rückkopplung über Nutzungsprobleme so organisiert, dass Systemverbesserungen ermöglicht werden?	Fehlertoleranz Lernförderlichkeit	Arbeit, so dass Beschwerden Dritter nachvollziehbar sind? Erkennt der Benutzer an den Systemrückmeldungen fehlerhafte Ergebnisse? Können Benutzer die Nutzungsprobleme am System zur Förderung des Lernens verwerten?
22. Welche Stressfaktoren gibt es und wie wird damit umgegangen?	Die Publikumstage, Montag und Donnerstag, können Stress mit sich bringen. Ein Großraumbüro ist für Publikumsverkehr ohnehin problematisch, aber für die Kooperation unter Kollegen wieder sehr praktisch.	31. Hat die Arbeitsperson die Möglichkeit, die Stressfaktoren zu beeinflussen, um die entstehenden Belastungen zu regulieren? (Zum Beispiel: Arbeitszeit einteilen, eingehende Telefonate umleiten.)	Steuerbarkeit	Kann der Benutzer den Dialog so steuern, wie es die Belastungen an Publikumstagen erfordert?
Sonstige Bemerkungen der befragten Person zu bereits aufgetretenen Nutzungsproblemen.		32. Ist eine Rückkopplung über Nutzungsprobleme so organisiert, dass Systemverbesserungen ermöglicht werden?	Aufgabenangemessenheit	Werden gemeldete Nutzungsprobleme vom Hersteller bei der Weiterentwicklung angemessen berücksichtigt?

Anhang C: Erhebung und Auswertung von Use-Szenarien

Dieser Anhang enthält ein typisches Use-Szenario. Use-Szenarien werden in Ergänzung zu Kontext-Szenarien erhoben. In Kenntnis der aus dem Nutzungskontext abgeleiteten Anforderungen wird die tatsächliche Erledigung von Aufgaben beobachtet. An kritischen Stellen der Benutzer-System-Interaktion (critical incidents) werden weitere Anforderungen abgeleitet oder die Vermutung einer Normabweichung kommentiert. Die zusätzlichen Anforderungen gehen in die Non-Konformitätsprüfungen ein. Zur Veranschaulichung der "critical incidents" können die kritischen Dialogschritte mittels "screen shots" illustriert werden.

Aufgabe in Teilaufgaben zerlegt	Beschreibung der Interaktion in Form eines Use Case		Anforderung oder zu klärende Frage
	Benutzeraktion	Aktion des Dialogsystems oder Reaktion	
Freistellungsantrag bearbeiten: - Freistellungsantrag lesen Die Überprüfung des Antrags ergibt, dass der Freistellung statt gegeben werden kann.	Arbeitsauftrag auswählen	BAföG-Sachbearbeitungsprogramm aufrufen	Ist die Form der Eingabe
	Posteingangsmaske öffnen	Maske angezeigt	

- Bewilligungsbescheid für die Freistellung anfertigen	Auf einem Notizzettel hält sie den Tag der Antragstellung und das Eingangsdatum fest. Diese Daten benötigt sie später im Bearbeitungsprozess.		der Aufgabe angemessen? Existiert z.B. ein elektronischer Notizzettel?
- "maßgebliches (anrechenbares) Einkommen" des Darlehnsnehmers errechnen und seinen Tilgungsplan prüfen	Zum Errechnen des anrechenbaren Einkommens benutzt sie einen Taschenrechner und notiert das Ergebnis zusätzlich auf ihrem Notizzettel.		Hat der Benutzer die Möglichkeit, zwischen alternativen Eingabemitteln zu wählen?
- Tilgungsplan des Darlehnsnehmers einsehen	Zur BAFöG-Sachbearbeitungsmaske wechseln (Doppelklick?) Frau Muster tippt die Kürzel "tp an" für "Tilgungsplan Ansicht" ein. Tilgungsansicht schließen	Maske angezeigt: das Aktenzeichen automatisch übernommen; eine Kommandozeile ist verfügbar Tilgungsplan angezeigt BAFöG-Dialogmaske angezeigt	Hat der Benutzer die Möglichkeit, zwischen alternativen Eingabemitteln zu wählen? Existiert eine alternative, selbsterklärende Eingabeform oder eine Hilfe?
- Ausstellung des Freistellungsantrags beginnen; dem Antrag stattgeben	Frau Muster beginnt mit der Ausstellung des Freistellungsbescheids. Um die Option "Antrag stattgeben" auszuführen, tippt Frau Muster in die Kommandozeile "fr 31" ein und drückt auf "Return". "fr"	Maske angezeigt: Eingabefelder für Beginn und Ende des nächsten Freistellungszeitraumes angezeigt	Existiert eine alternative, selbsterklärende Eingabeform?

- neuen Termin für das Ende der Freistellung ermitteln	bedeutet "Freistellung", "31" bedeutet "Antrag stattgeben". (Falls Frau Muster eine Zahl mal nicht im Kopf hat, muss sie einfach nur "fr" eintippen und die Return-Taste drücken, dann erscheint eine Übersicht über alle Kennziffern und ihre Bedeutung.) Frau Muster wechselt per Funktionstaste zur Maske "Darlehnsnehmer", um den bisherigen Tilgungsplan einzusehen. Danach wechselt sie wieder per Kommando "fr31" in die vorherige Maske. Dort bestätigt sie ihre Eingaben mit "Return". Keine weitere Eingabe; einfach nur "Enter" Frau Muster hat keine Ahnung was diese Formulierung bedeutet, bestätigt aber immer mit "ja".	Maske "Dispositionen anordnen" erscheint Als Reaktion wird ein Prompt angezeigt mit der Frage: "Möchten Sie die Integration zum Funktionskennzeichen fr31 durchführen?" Schreiben an Darlehnsnehmer erzeugen; Rückmeldung auf Sachbearbeitungsmaske: Status des Vorgang mit der Bemerkung "in Vorbereitung" angezeigt. Maske "Freistellungsbrief fr 31" anzeigen	Ist die erneute Kommandoingabe ein unnötiger Dialogschritt? Ist die Terminologie der Rückmeldung in der Sprache des Benutzers formuliert?
--	---	---	--

- Daten für den Bescheid eintragen	An dieser Stelle benötigt Frau Muster die Daten von ihrem Notizzettel.	Brief automatisch erstellen. Rückmeldung auf Sachbearbeitungsmaske: "fr 31" und Datum; die Bemerkung "in Vorbereitung" verschwindet. Prompt mit der Frage: "Möchten Sie das automatisch erstellte Schreiben nochmal überarbeiten?"	Hat der Benutzer Information über die zu erwartende Eingabe?
- Brief als "Eilbrief" kennzeichnen	Mit "Ja" bestätigen Es wäre jetzt theoretisch möglich, einfach zusätzlich Text in den Brief einzufügen, doch das ist sehr gefährlich. Der Text ist mit sogenannten "Feldfunktionen" formatiert, und wenn man ein Wort einfügen will, so muss man das im Text der Feldfunktionen tun, wo man aus Versehen einen der Feldfunktionsbefehle löschen kann und damit den Briefaufbau zerstören würde.	Text wird angezeigt	Ist der Benutzer über die schwerwiegenden Folgen des Dialogschritts informiert?
- Schreiben an den Darlehnsnehmer verschicken	Frau Muster geht darum lieber in die Menüzeile auf "Ansicht-Feldfunktionen". Das Wort "Eilbrief" eingeben und speichern Button "Postausgang erzeugen"	Feldfunktionsbefehle zusätzlich angezeigt "Postausgangsobjektmaske" angezeigt. Rückmeldung: das	

	auswählen	gerade erstellte Schreiben ist angezeigt	
	Pfeil in Richtung Postausgangsfenster aktivieren	Rückmeldung: Bescheid erscheint im Postausgangsfenster	
	Versandart "Normal" wählen und mit "ok" bestätigen	Prompt mit der Rückfrage, ob das Postausgangsobjekt gespeichert werden soll	
- Den Postausgang veranlassen (Schlussverfügung)	Mit "Ja" Absendung auslösen	Rückmeldung: Das Schreiben auf der Sachbearbeitungsmaske als "PAO" (Postausgangsobjekt) und mit der Bemerkung "vorbereitet" gekennzeichnet	
	Zurück auf die Sachbearbeitungsmaske	Rückmeldung: letzte Maske, auf der unter anderem nochmal zwischen den drei Verfügungsarten gewählt werden kann	Warum diese extra Maske? Wird hier das bereits gewählte "ZdA" bestätigt?
	Zu den Akten (ZdA) auswählen		Wird der Benutzer vom System unterstützt, Eingabefehler zu entdecken und ggf. zu vermeiden?
	Das ist Frau Muster ganz recht, denn sie hat auch schon mal den falschen Button gedrückt und ein Objekt zu den Akten gelegt, was sie eigentlich nochmal einsehen wollte.		
Arbeitsvorgang abgeschlossen	Angaben mit "ok" bestätigen		

Anhang D: ErgoNorm-Benutzerfragebogen zu "Arbeit & Software"

Anwendungshinweise

Dieser Fragebogen ist nicht Teil des Prüfverfahrens. Mit einem Fragebogen, den man an Benutzer austeilt, um subjektive Bewertungen der Software zu erheben, kann man keine Software auf Normkonformität prüfen. Dennoch ist der nachstehende Fragebogen ein unverzichtbares Instrument in der Hand des Prüfers. Mit Hilfe der Fragen an die Benutzer kann der Prüfer einen ersten Eindruck von den Nutzungsproblemen bekommen. Der Fragebogen ist somit als "Initialverfahren" gedacht. Der Prüfer sollte den Fragebogen einem betrieblichen Beauftragten in einer Anwenderorganisation übergeben und sich nach dem Einsammeln der ausgefüllten Fragebögen mitteilen lassen, zu welchen Fragen viele Kommentare abgegeben wurden und wie hoch der Prozentsatz der Benutzer ist, die die beanstandeten Mängel als sehr störend empfinden.

Anhand der Kommentare zu den subjektiv sehr störenden Mängeln kann der Prüfer mit der Analyse von Nutzungsproblemen beginnen, um einzuschätzen, wie aufwendig eine Produktprüfung werden könnte. Wenn die Nutzungsprobleme breit gestreut sind und der Prüfaufwand beträchtlich zu werden scheint, so ist empfehlenswert, das in diesem Bericht beschriebene Prüfverfahren vollständig anzuwenden. Wenn hingegen die Nutzungsprobleme vereinzelt auftreten, so kann sich der Prüfer auf die Beschreibung in Form eines Use-Szenarios beschränken, d.h. zu einem "critical incident" die passende Anforderung bestimmen und einer vermuteten Normabweichung mit Hilfe der Entscheidungstabellen nachgehen.

Es ist nicht empfehlenswert, sich einen Vorrat von ausgedruckten Fragebogenkopien anzulegen, um sie bei Bedarf schnell verteilen zu können. Vielmehr soll der Fragebogen als Datei gespeichert werden. Vor dem Ausdruck des Fragebogens soll sowohl in die Instruktion als auch in die darauf folgende Seite die genaue Aufgabenbeschreibung eingetragen werden. Ein Fragebogen ohne Aufgabenbeschreibung ist

wertlos, weil der Auswerter nicht feststellen kann, an welche Aufgabe die Benutzer gerade gedacht haben, als sie die Fragen beantwortet haben. Nur bei klar vorgegebenem Aufgabenbezug können zuverlässige Benutzerangaben und nachvollziehbare Nutzungsprobleme erfasst werden.

Der Fragebogen sollte nicht ausgeteilt werden, wenn sich die Benutzer gerade in der Einarbeitungsphase befinden. Erfahrungsgemäß werden die meisten einfachen Einarbeitungsprobleme schnell überwunden und die möglicherweise verursachenden Mängel nach der Lernphase nicht mehr als beanspruchend empfunden. Nach der Einarbeitung bleiben die echten Nutzungsprobleme übrig, die überdauernd oder wiederkehrend als Beeinträchtigung erlebt werden und die es sich von der Aufwandsökonomie her lohnt, mittels Fragebogen aufzuspüren.

Der Prüfer kann der Anwenderorganisation den periodischen Einsatz des Fragebogens empfehlen, z.B. nach einem Releasewechsel. Auch für eine Wiederholungsuntersuchung eignet sich der Fragebogen, etwa nach Beseitigung von ergonomischen Mängeln im Zuge eines Pflegeprozesses.

ErgoNorm- Benutzerfragebogen zu „Arbeit & Software“

Lieber Benutzer,

dieser Fragebogen dient dazu, Ihre persönliche Einschätzung des Computers, mit dem Sie täglich arbeiten, zu erfassen. Nur Sie können beurteilen, wie gut oder schlecht der Computer Sie in Ihrer Arbeitssituation unterstützt. Es geht darum herauszufinden, bei welchen Tätigkeiten Ihnen die Software Schwierigkeiten bereitet, bei welchen Arbeitsschritten Sie sich ärgern oder vielleicht nicht weiter wissen.

Es kann sein, dass Ihnen bei der Erledigung Ihrer Arbeit die Schwachstellen des Programms gar nicht mehr unangenehm auffallen, weil Sie sich daran gewöhnt haben oder vielleicht denken: "So ist das halt mit dem Computer." Der Fragebogen hilft Ihnen, auch solche Schwachstellen im Programm zu identifizieren und zu benennen. Ihre Antworten auf die Fragen werden dazu verwendet, Qualitätsmängel zu erfassen. Ziel ist es, den Computer besser an Ihre Bedürfnisse anzupassen und Ihnen somit die Arbeit am Bildschirmarbeitsplatz zu erleichtern.

Alle Daten werden selbstverständlich anonym erhoben, so dass keine Ihrer Aussagen auf Sie persönlich zurückgeführt wird.

Handhabung des Fragebogens

Vermutlich nutzen Sie den Computer, um verschiedene und in sich abgeschlossene Aufgaben zu erledigen. Bitte denken Sie beim Ausfüllen des Fragebogens an die Ausführung folgender Aufgabe:

Bevor Sie damit beginnen, den Fragebogen auszufüllen, sollten Sie erst einmal alle Fragen durchgelesen haben. Sie werden feststellen, dass alle Fragen auf sehr nützliche Eigenschaften aufmerksam machen, die der Computer haben sollte. Beim Ausfüllen des Fragebogens ist es wichtig, dass Sie an die eingangs beschriebene Aufgabe denken. Nur solche Fragen sind zu beantworten, die im Zusammenhang mit dieser Aufgabe bedeutsam sind. Wenn Sie feststellen, dass einige Fragen mit der genannten Aufgabe inhaltlich nichts zu tun haben, so kreuzen Sie rechts "Frage trifft nicht zu" an. Beim Ausfüllen können Sie auf Mängel hinweisen. Falls Sie diese als "sehr" belastend erleben, können Sie dies ankreuzen.

Beginnen Sie erst mit dem Ausfüllen des Fragebogens, nachdem Sie in Ruhe alle Fragen gelesen haben.

Beschreibung der Aufgabe: (Bitte benennen und charakterisieren Sie in Stichworten die Tätigkeit, an die Sie beim Ausfüllen des Fragebogens denken.)

Aufgabenangemessenheit

Ein Computerprogramm ist aufgabenangemessen, wenn es zur Erledigung Ihrer konkreten Tätigkeit brauchbar ist. "Brauchbar" bedeutet, dass alle Tätigkeiten, die Sie erledigen müssen, vom Programm unterstützt werden. Es ist Ihnen dabei wirklich eine Hilfe und kein nötiges Übel, das Ihre Arbeit in manchen Situationen eher erschwert oder umständlicher macht.

1. Enthält das Programm alle für Ihre Aufgabe benötigten Funktionen?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Bitte benennen Sie den Arbeitsschritt, bei dem Sie sich wünschen würden, dass das Programm "mehr kann" als gerade möglich ist.

☐ Ich empfinde dies als sehr störend

2. Müssen Sie Eingaben oder Dialogschritte machen, die eigentlich überflüssig wären?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Bitte benennen Sie die in ihren Augen überflüssigen Eingaben und Dialogschritte.

☐ Ich empfinde dies als sehr störend

3. Ist es Ihnen möglich, das wiederholte Eingeben von Daten oder Texten zu vereinfachen?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

In welcher Situation würden Sie sich wünschen, dass Sie nicht so oft dasselbe eingeben müssten?

☐ Ich empfinde dies als sehr störend

4. Finden Sie, dass der erforderliche Aufwand für Ihr Arbeitsergebnis jeweils angemessen ist?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

In welcher Situation haben Sie schon mal gedacht "Das könnte man auch mit weniger Aufwand bewerkstelligen."

☐ Ich empfinde dies als sehr störend

5. Haben Sie das Gefühl, dass Sie Arbeiten machen müssen, die besser das Programm erledigen sollte?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Bitte benennen Sie diese Arbeiten.

☐ Ich empfinde dies als sehr störend

6. Müssen Sie Werte und Texte eingeben, die der Computer eigentlich wissen könnte?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Bitte beschreiben Sie die Situationen, in denen Sie sich zum Beispiel denken: "Das müsste der Computer jetzt eigentlich wissen, wieso muss ich das noch einmal hinschreiben."

☐ Ich empfinde dies als sehr störend

7. Müssen Sie sich mit Umwegen oder Tricks behelfen, um Ihre Arbeitsergebnisse so zu erzielen, wie Sie diese haben möchten?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Beschreiben Sie bitte die Situationen, in denen Sie das Gefühl haben, umständlich "tricksen" zu müssen, um Ihr Arbeitsergebnis zu erreichen.

☐ Ich empfinde dies als sehr störend

8. Finden Sie in dem Programm Hilfetexte, die Ihnen auch tatsächlich weiterhelfen?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Benennen Sie die Situationen, in denen Sie die Hilfeinformation nicht weitergebracht hat.

☐ Ich empfinde dies als sehr störend

9. Passt das Programm zu Ihren Formularen und bisherigen Formaten?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Benennen Sie die Tätigkeit, bei der das Programm nicht zu Ihren Papierunterlagen oder Formaten passt.

☐ Ich empfinde dies als sehr störend

Selbstbeschreibungsfähigkeit

Eine Computerprogramm ist selbstbeschreibungsfähig, wenn Sie jederzeit informiert sind, was der Computer gerade macht und was er als nächstes von Ihnen als Eingabe oder Reaktion erwartet. Dies bedeutet unter anderem, dass Sie alle Rückmeldungen verstehen können, immer wissen, wo Sie als nächstes etwas eingeben müssen und sich jederzeit klar über die Folgen sind, die eine Eingabe von Ihnen haben wird.

1. Sind die Informationen, die zur Erledigung der Aufgabe notwendig sind, auf dem Bildschirm übersichtlich verfügbar?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte die Informationen, die Sie benötigen, aber nicht "auf einen Blick" zur Verfügung stehen.

☐ Ich empfinde dies als sehr störend

2. Können Sie bei der Arbeit mit dem Programm erkennen, welche Eingabe als nächstes von Ihnen erwartet wird?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Schildern Sie bitte kurz die Situation, in der Sie unsicher sind, was als nächstes bei der Arbeit mit dem Computer zu tun ist.

☐ Ich empfinde dies als sehr störend

3. Sind die Meldungen des Systems für Sie immer verständlich?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie die Situationen, in denen Ihnen unverständliche Meldungen aufgefallen sind.

☐ Ich empfinde dies als sehr störend

4. Werden Sie vor Aktionen, die nicht rückgängig gemacht werden können, von der Software gewarnt?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Bitte benennen Sie Situationen, in denen Sie keine Warnung erhalten haben.

☐ Ich empfinde dies als sehr störend

5. Hilft Ihnen die Hilfefunktion wirklich weiter, wenn einmal ein Dialogschritt oder Menüpunkt nicht ganz klar ist?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Beschreiben Sie die Situationen, in denen die Hilfe nicht verständlich ist.

☐ Ich empfinde dies als sehr störend

6. Müssen Sie oft Kollegen oder ein Handbuch konsultieren, um weiterarbeiten zu können?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Nennen Sie bitte Situationen, in denen Sie auf die Hilfe von Kollegen oder eines Handbuchs angewiesen waren.

☐ Ich empfinde dies als sehr störend

Steuerbarkeit

Eine Computerprogramm ist steuerbar, wenn Sie als Benutzer die Abfolge der Arbeitsschritte weitgehend selbst bestimmen können. Wenn es die Arbeitssituation erfordert, können Sie die Arbeit am Computer unterbrechen und diese dann ohne Verlust der bis dahin erreichten Arbeitsergebnisse wieder aufnehmen.

1. **Können Sie Ihre Arbeitsschritte in der Reihenfolge erledigen, die Ihnen am sinnvollsten erscheint?**

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte Arbeitsschritte, bei denen Ihnen eine andere Reihenfolge sinnvoller erscheinen würde.

☐ Ich empfinde dies als sehr störend

2. **Macht das Programm manchmal etwas, ohne dass Sie es zu dem Zeitpunkt wollen?**

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Nennen Sie bitte das Verhalten des Programms, das ungewollt auftritt.

☐ Ich empfinde dies als sehr störend

3. **Können Sie bei Bedarf eine Aufgabe unterbrechen und später wieder fortsetzen, ohne alles neu eingeben zu müssen?**

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Schildern Sie bitte, in welcher Situation Sie durch eine Unterbrechung bereits eingegebene Daten wieder verloren haben.

☐ Ich empfinde dies als sehr störend

4. **Können Sie einen Arbeitsschritt wieder zurücknehmen, wenn es für Ihre Aufgabenerledigung zweckmäßig ist?**

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte die Situationen, in denen es aus Ihrer Sicht zweckmäßig wäre, einen Arbeitsschritt wieder ungeschehen machen zu können.

☐ Ich empfinde dies als sehr störend

5. **Fühlen Sie sich in Ihrem Arbeitstempo durch das Programm manchmal gebremst, z.B. durch zu lange Wartezeiten?**

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Beschreiben Sie bitte die Situationen, in denen Sie gerne zügiger arbeiten würden.

☐ Ich empfinde dies als sehr störend

Erwartungskonformität

Ein Computerprogramm ist erwartungskonform, wenn Sie bei der Arbeit mit dem Computer keine "Überraschungsmomente" erleben. Solche Momente können zum Beispiel sein, dass sich eine Funktion an einer ganz anderen Stelle im Menü befindet, als Sie gedacht hätten oder dass Aufgaben nicht, wie Sie es gewohnt sind, ausgeführt werden können.

1. Finden Sie Menüpunkte oder Funktionen dort, wo sie Ihrer Meinung nach auch sein sollten?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte die konkrete Stelle im Menü oder in einer anderen Übersicht, in der die Anordnung der Information **nicht** Ihren Erwartungen entspricht.

☐ Ich empfinde dies als sehr störend

2. Sind Sie sich bei Wartezeiten immer noch sicher, ob das Programm weiterarbeitet?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte die Situationen, in denen Sie sich nicht sicher sind, ob das Programm noch arbeitet, z.B., wenn das Programm sehr lange benötigt, um Daten zu speichern.

☐ Ich empfinde dies als sehr störend

3. Sind Sie manchmal überrascht, wie das Programm auf Ihre Eingabe reagiert?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "ja":

Beschreiben Sie die Situationen, in denen sie über die Reaktionen des Systems erstaunt sind.

☐ Ich empfinde dies als sehr störend

Fehlertoleranz

Ein Computerprogramm ist fehlertolerant, wenn Sie ihr Arbeitsergebnis trotz fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand erreichen können. Dies bedeutet, dass es durchaus erlaubt sein muss, sich zu vertippen oder einen falschen Arbeitsschritt zu machen, ohne dass das Programm gleich abstürzt, oder Sie den Fehler nur mit Mühe wieder gut machen können. Außerdem sollte das Programm Sie darauf aufmerksam machen, wenn es einen Fehler bemerkt und Ihnen mögliche Korrekturhinweise liefern.

1. Bekommen Sie bei fehlerhaften Eingaben Korrekturhinweise?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie bitte Situationen, in denen Sie sich vielleicht wünschen würden, dass das Programm Ihnen einen Vorschlag für eine richtige Eingabe macht.

☐ Ich empfinde dies als sehr störend

2. Können Sie die Folgen einer fehlerhaften Eingabe mit geringem Aufwand beheben?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Schildern Sie bitte kurz die Situationen, in denen Ihnen der Aufwand für die Korrektur einer fehlerhaften Eingabe zu hoch erscheint.

☐ Ich empfinde dies als sehr störend

3. Arbeitet das Programm während der Ausführung Ihrer Aufgabe immer stabil und zuverlässig?

☐ ja ☐ nein ☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie die Situationen, in denen Sie der Software nicht trauen oder Sie einen "Absturz" befürchten.

☐ Ich empfinde dies als sehr störend

Individualisierbarkeit

Ein Computerprogramm ist individualisierbar, wenn Sie Einstellungen des Programms an Ihre individuellen Bedürfnisse anpassen können.

- 1. Können Sie am Computer alles so einstellen, dass Ihnen das Lesen und Arbeiten leichter fällt?**

□ ja

☐ nein

☐ Frage trifft nicht zu

wenn "nein":

Nennen Sie die Stellen, bei denen Ihnen das Arbeiten mit dem Programm schwer fällt.

☐ Ich empfinde dies als sehr störend

Lernförderlichkeit

Ein Computerprogramm ist lernförderlich, wenn es Ihnen unter anderem ermöglicht, selbständig einfach mal "rumzuprobieren", ohne dass Sie Angst haben müssen etwas "kaputt" zu machen. Zusätzlich sollten Sie durch das Programm die für Sie relevanten Informationen erhalten, die Sie Ihrer Meinung nach benötigen, um das Programm besser zu verstehen.

1. Ermöglicht Ihnen das Programm, auch einmal etwas gefahrlos auszuprobieren?

□ ja

☐ nein

☐ Frage trifft nicht zu

wenn "nein":

Beschreiben Sie bitte die "Strafen", die Sie von dem Programm durch "Rumprobieren" schon bekommen haben.

☐ Ich empfinde dies als sehr störend

Der letzte Teil des Fragebogens ist für Ihre individuellen Anmerkungen reserviert. Hier ist Platz für weitere Kritik an dem Computerprogramm oder für die Probleme, die Sie bei Beantwortung der Fragen nicht losgeworden sind.

[illegible]

Anhang E: Glossar

Abweichung (nonfulfilment / departure)

Unterschied zwischen einem gegebenen Produktmerkmal oder einer am Dialogsystem ausführbaren >Arbeitstätigkeit und einem >Prüfkriterium. Siehe auch Normabweichung, Fehler, Mangel.

Analytiker (analyst)

Rollenbezeichnung für eine Person, die in der Vorbereitung eines Software-Entwicklungsprojekts den >Nutzungskontext kennen lernt, um die >Erfordernisse zu erkennen, die als Grundlage für die Entwicklung von >Anforderungen dienen. Analytiker sind wegen ihrer Kenntnisse des Nutzungskontexts und der >Anforderungen gut vorbereitet, um in einem Entwicklungsprojekt in der Rolle des >Prüfers die Gebrauchstauglichkeit von Software festzustellen oder ggf. >Normabweichungen und >Mängel zu bewerten.

Anforderung (requirement)

Ein qualitativer oder quantitativer Wert einer Qualitätsforderung, die aus >Erfordernissen der Arbeit oder Organisation (>Nutzungskontext) des Benutzers sowie seiner Belange abgeleitet ist. Anforderungen sind i.d.R. als geforderte >Arbeitstätigkeit zu formulieren. Nur wenn gesicherte ergonomische Erkenntnisse für die Merkmale eines Produkts (z.B. Farbkodierung, Hintergrundfarbe des Bildschirms) gegeben sind, können diese Merkmale als Anforderung dienen. Eine Anforderung kann in ein >Prüfkriterium transformiert werden.

Anwender (customer / user organization)

Eine Organisation bzw. Institution, die Rechensysteme zur Erfüllung von Datenverarbeitungsaufgaben bzw. zur Unterstützung von Informationsverarbeitungsprozessen einsetzt. Hierbei ist es unerheblich, welchen Umfang die Rechneranwendung hat. Der Begriff Anwender wird bewusst nicht auf eine einzelne Person oder auf eine Personengruppe bezogen, die Rechner für ihre speziellen Aufgaben einsetzt. In letzterem Fall wird der Begriff >Benutzer verwendet. Die Begriffe Anwender und Benutzer werden nicht synonym verwendet; bei einem Anwender gibt es üblicherweise viele verschiedene Benutzer.

Arbeitstätigkeit (work activity / user performance)

Eine beobachtbare Tätigkeit des >Benutzers am Bildschirmarbeitsplatz, die zur Erledigung der Arbeitsaufgabe ausgeführt wird. Nicht beobachtbar sind kognitive Prozesse wie lernen, erinnern, eine Vorstellung entwickeln (mentales Modell), denken, suchen usw.

Benutzbarkeit (usability)

Eine Menge von Merkmalen zur Erleichterung der Benutzung mit Rücksicht auf die individuelle Bewertung durch eine festgelegte oder vorausgesetzte Gruppe von >Benutzern. (Vgl. ISO/IEC 12119)

Benutzer (user / end user)

Eine Person oder eine Personengruppe, die Rechner zur Unterstützung ihrer Aufgabenbearbeitung einsetzt.

Benutzerbelange (user needs)

Bedürfnisse des >Benutzers, die sich aus Eigenschaften des Benutzers ergeben, und im Zusammenhang mit der Nutzung einer Software im >Nutzungskontext zu berücksichtigen sind. Beispiele sind: Alter, Linkshändigkeit, sporadischer Nutzer, Sprachkenntnisse.

Benutzungsoberfläche (user interface)

Der Begriff "Benutzungsoberfläche" hat den früher benutzten missverständlichen Begriff "Benutzeroberfläche" abgelöst.

"Bezeichnung für die Gesamtheit der >Eingabemittel und Anzeigeeinrichtungen eines Systems (z.B. Tastatur, Cursor, Bildschirm, Anzeigen). Die Benutzungsoberfläche ist Teil der >Benutzungsschnittstelle.

Benutzungsschnittstelle (user interface)

Gesamtheit der Eingabe-Merkmale, der Ausgabe-Merkmale, der Dialog-Merkmale und der Werkzeug-Merkmale, mittels derer der >Benutzer auf das System einwirkt (z.B. Eingabefeld, Kommandozeile, Unterbrechungskommando, UNDO) sowie der Merkmale, in denen sich das Verhalten des Systems gegenüber dem Benutzer äußert (z.B. Meldungen, Nutzungshinweise, Menüs am Bildschirm, nutzungsrelevante Werkzeug-eigenschaften).

Design-Use-Cycle

Die Phasen, die ein Softwareprodukt durchläuft, beginnend mit der Produktidee über die Entwicklung des >Nutzungskonzepts, über den Entwurf, die Implementierung und die Installierung des Produkts beim >Anwender, bis hin zur >Pflege des Produkts im >Nutzungskontext, in dem aufgrund der Nutzungserfahrungen das nächste Release vorbereitet wird. Eine iterative Folge der Phasen ist üblich. Entwicklungsaktivitäten und Maßnahmen der >Qualitätssicherung wechseln sich ab.

Dialog (dialogue)

Ein Ablauf, bei dem der Benutzer zur Abwicklung einer Arbeitsaufgabe - in einem oder mehreren Dialogschritten - Daten eingibt und jeweils Rückmeldung über die Verarbeitung dieser Daten erhält (DIN 66 234, Teil 8).

Dialogschritt (dialogue step)

Jeder Dialogschritt umfasst einen Wechselschritt von Aktionen des Menschen und des Computers. Ein Dialogschritt besteht aus Eingabedaten, den zugehörigen Verarbeitungsprozessen und den zugehörigen Ausgabedaten.

Einarbeitungsproblem (initial usage problem)

Eine bei der anfänglichen Nutzung des Softwareprodukts festgestellte Barriere, die eine effiziente Erledigung der Arbeitsaufgabe unnötig erschwert, aber bei regelmäßiger Nutzung beeinträchtigungsfrei umgangen wird.

Effektivität (effectiveness)

Die Genauigkeit und Vollständigkeit, mit der >Benutzer ein bestimmtes Ziel (Arbeitsergebnis) erreichen (ISO 9241-11).

Effizienz (efficiency)

Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem >Benutzer ein bestimmtes Ziel (Arbeitsergebnis) erreichen (ISO 9241-11).

Eingabemittel (input device)

Als Eingabemittel werden alle peripheren Geräte eines Dialogsystems bezeichnet, die es dem >Benutzer erlauben, dem System Daten zu übermitteln, wie z.B. Tastatur, Tablett, Stift, Maus, Mikrophon, etc.

Erfordernis (implied need)

Eigenschaft einer Arbeit oder einer Organisation (>Nutzungskontext), die selbstverständlich vorausgesetzt wird und aus der sich spezifische >Anforderungen ableiten lassen.

Erhärtungsprüfung (evidence test)

Verifizierung der Vermutung, dass ein identifiziertes Nutzungsproblem durch eine >Normabweichung verursacht ist und die Wirkungen des festgestellten >Mangels nicht zumutbar sind.

Fehler (failure)

Ein fehlendes Qualitätsmerkmal, das eine >effektive Nutzung des Produkts verhindert oder ein > Mangel, der als Fehler angesehen wird, weil ihm eine definierte >Anforderung zugrunde liegt.

Funktion (function)

Eine Funktion ist eine in sich geschlossene Folge von >Operationen zur Ausführung einer Teilaufgabe des >Benutzers.

Funktionalität (functionality)

Gesamtheit aller >Funktionen, die mit dem Dialogsystem ausgeführt werden können, bezogen auf die Arbeitsaufgabe.

Gebrauchstauglichkeit (usability in context of use)

Das Ausmaß, in dem ein Produkt durch bestimmte >Benutzer in einem bestimmten >Nutzungskontext genutzt werden kann, um bestimmte Ziele >effektiv, >effizient und >zufriedenstellend zu erreichen (ISO 9241-11).

Konformität (conformity / conformance / compliance)

Übereinstimmung eines Produktmerkmals oder einer am Dialogsystem ausführbaren >Arbeitstätigkeit mit einem >Prüfkriterium. Ist das >Prüfkriterium aus einer Norm abgeleitet (z.B. ISO 9241-10), so spricht man im Falle von Übereinstimmung von Normkonformität. Siehe auch Non-Konformität.

Kontext-Szenario (context scenario)

Eine episodische Beschreibung der >Arbeitstätigkeiten einer Arbeitsperson (z.B. der künftige >Benutzer), um den Einfluss der Bedingungen des >Nutzungskontexts auf die Ausführung der Tätigkeiten zu analysieren. Diese Erhebung dient der Vorbereitung der Anforderungsentwicklung zu Beginn eines Projekts. Es werden in der Sprache des Benutzers die Nutzungsbedingungen eines noch zu entwickelnden Systems beschrieben, um ein klares Verständnis der >Erfordernisse der Arbeit und Organisation zu erarbeiten. Die sachliche Richtigkeit und Vollständigkeit der Beschreibung muss vom Benutzer bestätigt werden (Validierung).

Mangel (defect)

Ein Qualitätsmerkmal, das qualitativ oder quantitativ vom >Prüfkriterium abweicht und somit eine >effiziente Nutzung des Produkts behindert. Ein Mangel kann als nicht zumutbar bewertet werden, wenn die Wirkungen des Mangels nicht umgebar sind und die >Arbeitstätigkeit des Benutzers beeinträchtigen (siehe >Erhärtungsprüfung).

Non-Konformität (nonconformity)

Die Nichterfüllung eines >Prüfkriteriums (vgl. ISO 8402). Nichterfüllung kann Abwesenheit eines geforderten Qualitätsmerkmals (>Fehler) oder >Abweichung von einem geforderten Tätigkeits-Merkmal (>Mangel) bedeuten. Siehe auch Normabweichung.

Normabweichung (nonconformity)

Unterschied zwischen einem gegebenen Produktmerkmal oder einer am Dialogsystem ausführbaren >Arbeitstätigkeit und einem aus der Norm (z.B. ISO 9241-10) abgeleiteten >Prüfkriterium. Eine verifizierte Normabweichung wird gemäß dem Prinzip der Falsifikation so lange als vermutete Normabweichung angesehen, bis die >Erhärtungsprüfung zu der Feststellung führt, dass die Normabweichung nicht zumutbare Wirkungen hat. Siehe auch Abweichung, Non-Konformität, Mangel, Fehler.

Nutzungskontext (context of use)

Faktoren, die die >Gebrauchstauglichkeit eines Produktes beeinflussen können, z.B. die Qualifikation des >Benutzers, die Merkmale des Arbeitsinhalts, der Arbeitsplanung, der Arbeitsteilung sowie die physische und soziale Umgebung (vgl. ISO 9241-11).

Nutzungskonzept (useware)

Die Struktur oder Strategie eines Lösungswegs zur Erledigung einer Aufgabe und die dafür benötigten Hilfsmittel (nicht gemeint als DER eine, ideale Weg der Aufgabenerledigung). Hierbei sind die mentalen Modelle der Nutzer (Benutzer) von der Aufgabe und dem gewünschten Arbeitsergebnis maßgebend.

Nutzungsproblem (usage problem)

Eine bei der Nutzung des Softwareprodukts festgestellte Barriere, die eine effiziente Erledigung der Arbeitsaufgabe unnötig erschwert, und nicht auf ein >Einarbeitungsproblem rückführbar ist.

Marktanalyse (market analysis)

Eine Form der Marktforschung zur Feststellung des potentiellen Bedarfs für die Nutzung eines Produkts in einem >Nutzungskontext.

Pflege (usability maintenance)

Die Verbesserung eines Softwareprodukts nach Installation am Benutzer-Arbeitsplatz, um ergonomische >Mängel (Abweichungen von ISO 9241) zu beseitigen. Die Pflege kann sehr aufwendig sein, wenn versäumt wurde, die >Anforderungen im Rahmen der Anforderungsentwicklung und des >Prototyping zu validieren.

Produktprüfung (product evaluation)

Vergleich der Merkmale eines Produkts (z.B. der >Benutzungsschnittstelle) mit den >Anforderungen, um die Erfüllung der Forderungen festzustellen.

Projekt (project)

Ein Entwicklungsvorhaben in einer Herstellerorganisation zum Zwecke des Entwurfs, der Entwicklung, Dokumentation und >Pflege eines Softwareprodukts.

Prototyping (usability prototyping)

Ein iterativer Prozess zur Herstellung von Teilen einer >Benutzungsschnittstelle anhand eines >Use-Szenarios sowie der Herstellung oder Simulation einer dahinter liegenden Funktionalität zum Zwecke der Veranschaulichung der vom >Usability-Designer verstandenen >Anforderungen. Dient ein >Prototyp der Erkundung noch nicht verstandener Anforderungen, so nennt man diesen Prozess exploratives Prototyping.

Prüfer (usability tester)

Rollenbezeichnung für eine Person, die ein Softwareprodukt (ggf. auch Prototyp) auf >Gebrauchstauglichkeit prüft. Während der Prüfungsvorbereitung ist der Prüfer in der Regel als >Analytiker tätig, um die >Anforderungen und >Prüfkriterien festzustellen, gegen die ein Produkt geprüft werden soll.

Prüfkriterium (test criterion)

Vorzugsweise eine geforderte >Arbeitstätigkeit, die unter Anwendung von ISO 9241-10 aus einer >Anforderung abgeleitet wurde, oder ein Produktmerkmal, das zu einer geforderten >Arbeitstätigkeit in Bezug steht.

Qualität (quality)

Die Gesamtheit der Eigenschaften eines Produkts, die geeignet ist, >Erfordernisse oder festgelegte >Anforderungen zu erfüllen (ISO 8402).

Qualitätssicherung (quality assurance)

Alle geplanten, systematischen Maßnahmen, die erforderlich sind, um an einem Produkt die geforderten >Prüfkriterien zu erfüllen.

Styleguide

Eine Herstellernorm zur Gestaltung von Oberflächen-Merkmalen der >Benutzungsschnittstelle zum Zwecke der konsistenten Gestaltung dieser Merkmale über alle Anwendungsprogramme hinweg.

Szenario (scenario)

Eine episodische Beschreibung von >Arbeitstätigkeiten unter Berücksichtigung des >Nutzungskontexts (>Kontext-Szenario) oder unter Berücksichtigung der tatsächlichen Ausführung der >Arbeitstätigkeit am Dialogsystem (>Use-Szenario).

Usability-Designer

Rollenbezeichnung für eine Person, die an der Gestaltung der >Benutzungsschnittstelle mitwirkt, z.B. durch Entwicklung der >Anforderungen mittels >Szenarien und Prototypen, und die an der >Qualitätssicherung der >Gebrauchstauglichkeit der Software beteiligt ist.

Usability-Engineering (usability engineering)

Eine Entwicklung von >Anforderungen, >Prototypen und Softwareprodukten, deren Validierung und Verbesserung, unter systematischer Anwendung von >Usability-Engineering-Methoden im Rahmen eines >Design-Use-Cycle.

Usability-Engineering-Methoden (usability-engineering methods)

Methoden zur Entwicklung von innovativen >Nutzungskonzepten, >Anforderungen und >Prototypen, deren >Validierung, Bewertung und schrittweise Verbesserung während des Entwurfs- und der >Pflegeprozesses.

Use Case

Die Kurzform eines >Use-Szenarios. Jede Interaktion des >Benutzers mit dem Dialogsystem (>Dialogschritt) wird in zwei Spalten festgehalten, links die >Arbeitstätigkeit des Benutzers, rechts die darauf folgende Reaktion des Dialogsystems.

Use-Szenario (use scenario)

Eine episodische Beschreibung von >Arbeitstätigkeiten, die ein >Benutzer zur Erledigung einer Kern-Aufgabe am Dialogsystem erledigt. Diese Erhebung dient der Anforderungsentwicklung während des Prototyping oder der Ermittlung von Anforderungen am fertigen Produkt zum Zwecke der Softwareprüfung. Während der Softwareprüfung an einem existierenden System braucht die Interaktion mit dem Dialogsystem nicht ausführlich erfasst zu werden; es reicht eine Erfassung in Form eines >Use Case aus. >Dialogschritte, die jedoch auf Nutzungsprobleme oder auf vermutete >Normabweichungen hinweisen (critical incidents), sind ausführlich zu beschreiben und anschließend auszuwerten (Prüfung der >Konformität).

Validierung (validation)

Ein Prozess der >Qualitätssicherung, um festzustellen, ob und inwieweit >Anwender (>Benutzer) und Softwarehersteller in der Festlegung von >Anforderungen übereinstimmen und inwieweit die Eigenschaften eines Produkts mit den Anforderungen an die >Effektivität übereinstimmen.

Zufriedenstellung (user satisfaction)

Beeinträchtigungsfreiheit der Nutzung und positive Einstellungen der >Benutzer zur Produkt-Nutzung (ISO 9241-11).