

## Projektberichte zum Forschungsprojekt

# **B**enutzer- **O**rientierte **S**oftwareentwicklung und **S**chnittstellengestaltung

Herausgegeben von:

P. Spinas, M.Rauterberg, O.Strohm, D.Waeber & E.Ulich

Projektbericht Nr. 6

**Benutzungsorientierte Benchmark-Tests:  
eine Methode zur Benutzerbeteiligung bei  
der Entwicklung von Standardsoftware**

Matthias Rauterberg

**Zürich**

**1991**

**Vom Bundesministerium für Forschung und Technologie (BMFT)  
der Bundesrepublik Deutschland (BRD)  
gefördertes Verbundprojekt**

**Förderkennzeichen: 01HK706/0**

**Projekttitel:**

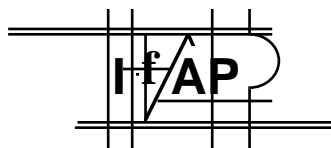
**"Entwicklung und empirische Überprüfung von Kriterien, Methoden und Modellen zur benutzerorientierten Software-Entwicklung und Dialog-Gestaltung"**

**Projektträger:**

**DLR  
Projektträgerschaft  
Arbeit und Technik (AuT)  
Südstrasse 125  
D-5300 BONN 2**

**Verbundpartner:**

**IfAP-ETH, Zürich (CH)**



**Prof. Dr. Eberhard Ulich (Projektleitung)**

**+41-1-254-7070**

**Dr. Philipp Spinas (Ansprechpartner)**

**+41-1-254-7077**

**Dipl.Inf.,Dipl.Psych. Matthias Rauterberg**

**+41-1-254-7082**

**Dipl.Psych. Oliver Strohm**

**+41-1-254-7084**

**lic.phil. Daniel Waeber**

**+41-1-254-7077**

**ADI Software GmbH, Karlsruhe (BRD)**



**Dr. Karl Schlagenhaut (Projektleitung)**

**Raimund Mollenhauer (Ansprechpartner)**

**+49-721-57000-66**

**Jürgen Dobrinski**

**© Eidgenössische Technische Hochschule Zürich -**

**Institut für Arbeitspsychologie**

**ISBN 3-906509-09-5**

# Projektberichte zum Forschungsprojekt

## **B**enutzer- **O**rientierte **S**oftwareentwicklung und **S**chnittstellengestaltung

Herausgegeben von:

P. Spinas, M.Rauterberg, O.Strohm, D.Waeber & E.Ulich

### Projektbericht Nr. 6

**Benutzungsorientierte Benchmark-Tests:  
eine Methode zur Benutzerbeteiligung bei  
der Entwicklung von Standardsoftware**

Matthias Rauterberg  
unter Mitarbeit von  
Christian Rolfsen

**Zürich**  
**1991**



## INHALTSVERZEICHNIS

TEIL-I: Benutzungs-orientierte Benchmark-Tests - eine Methode zur Benutzerbeteiligung bei Standardsoftware-Entwicklungen.....	3
1. Einleitung .....	3
2. Stand der Forschung zur Benutzerbeteiligung.....	3
3. Benutzerbeteiligung bei Standardsoftware-Entwicklungen .....	3
4. Benutzungs-orientierte Benchmark-Tests (bBTs).....	5
4.1. Induktive benutzungs-orientierte Benchmark-Tests (bBTs).....	5
4.1.1. Vorgehensweise bei induktiven bBTs .....	5
4.1.1.1. Anforderungen an die zu testende Software.....	5
4.1.1.2. Gestaltung der Beobachtungssituation .....	5
Konstruktion der Benchmark-Aufgaben: .....	5
Auswahl der TestpartnerInnen:.....	6
Umgebungsbedingungen während der Beobachtung:.....	6
4.1.1.3. Vorbereitung der TestpartnerInnen .....	7
Beschreibung von Ziel und Zweck:.....	7
Das Recht der TestpartnerIn auf Beendigung:.....	7
Vertrautmachen mit der Untersuchungssituation: .....	7
Die Technik des "lauten Denkens": .....	8
4.1.1.4. Durchführung der Beobachtung .....	8
Messung von intervenierenden Variablen: .....	8
Messung der abhängigen Variablen:.....	9
Aufgabenbearbeitung durch die TestpartnerInnen:.....	9
Verhalten des Testleiters:.....	9
"Sokratischer Dialog":.....	10
4.1.1.5. Nachbereitung mit den TestpartnerInnen .....	10
4.1.1.6. Auswertung der Beobachtungen.....	11
Aufbereitung der Daten:.....	11
Qualitative Auswertungen:.....	11
Quantitative Auswertungen: .....	11
4.2. Deduktive benutzungs-orientierte Benchmark-Tests (bBTs).....	12
4.2.1. Vorgehensweise bei deduktiven bBTs.....	12
4.2.1.1. Voraussetzungen an die zu testende Software.....	12
4.2.1.2. Gestaltung der Beobachtungssituation .....	12
Test-Design: .....	12
Konstruktion der Benchmark-Aufgaben: .....	12
Auswahl der TestpartnerInnen:.....	12
4.2.1.3. Auswertung der Beobachtungen.....	13
4.3. Aufwandabschätzung der Methode.....	13
4.4. Kosten-Nutzen-Analyse .....	13
5. Das Fallbeispiel ADIMENS.....	16
6. Zusammenfassung .....	18

TEIL-II: Benutzungs-orientierte Benchmark-Tests - Anwendung auf die Gestaltung der Benutzeroberfläche von ADIMENS-GT hin zu ADIMENS-GTplus .....	19
1. Einleitung .....	19
2. Stand der Forschung.....	19
3. Eigenschaften des Produktes, Gestaltungsmaßnahmen und Test-Fragen .....	20
4. Test-Design .....	22
4.1. Auswahl der TestpartnerInnen .....	22
4.2. Ablauf des Benchmark-Tests.....	23
4.3. Auswahl und Beschreibung der Benchmark-Aufgaben.....	23
5. Darstellung der Ergebnisse zur Bearbeitungszeit.....	25
6. Belastungs- und Beanspruchungs-Parameter.....	27
7. Subjektive Bewertungen der beiden Oberflächen .....	29
8. Berechnung des Oberflächen- und des Lern-Effektes .....	30
9. Kosten-Nutzen-Analyse.....	33
10. Diskussion.....	34
11. Literatur.....	35
12. Anhang.....	38

## **TEIL-I: Benutzungs-orientierte Benchmark-Tests - eine Methode zur Benutzerbeteiligung bei Standardsoftware-Entwicklungen\***

### **1. Einleitung**

Es lassen sich vier Arten von Software-Entwicklungsprojekten unterscheiden (STROHM 1990; WAEBER 1990): 1. spezifische Anwendungen für firmeninterne Fachabteilungen (*Typ-A*); 2. spezifische Anwendungen für externe Kunden (*Typ-B*); 3. Standardbranchenlösungen für externe Kunden (*Typ-C*); 4. Standardsoftware für einen anonymen Kundenkreis (*Typ-D*). Während bei Typ-A bis C die potentiellen Benutzer weitgehend bekannt sind, müssen bei Typ-D neue Wege der Benutzerbeteiligung beschritten werden. Es wird nun eine Möglichkeit vorgestellt, aus Ergebnissen empirischer Studien generalisierbare Aussagen in Form von benutzer-orientierten Gestaltungsvorschlägen auf der Grundlage des Kriterienkonzeptes von ULICH (1985, 1986, 1991) ableiten zu können (BEWLEY et al. 1983; RAUTERBERG 1988a, 1990b).

### **2. Stand der Forschung zur Benutzerbeteiligung**

Je nach Form, Grad, Inhalt, etc. der Benutzerbeteiligung (ACKERMANN 1988, SPINAS 1990) werden schon heute schon bei der Entwicklung von Standardsoftware ansatzweise Benutzer beteiligt (STROHM 1990). Die am häufigsten eingesetzten Methoden sind: 'hot-line', alpha-, beta-Tests, 'user group'-Workshops, Besprechungstermine (STROHM & SPINAS 1991) und sonstiger Kontakt zu Kunden. Diese Methoden dienen jedoch fast ausschließlich der Überprüfung auf funktionale Vollständigkeit und Korrektheit. Benutzungsprobleme im handlungs- und arbeitspsychologischen Sinne werden dabei kaum berücksichtigt. Hier setzt die Methode der *benutzungs-orientierten Benchmark-Tests*<sup>1</sup> an. Benutzerbeteiligung ist deshalb notwendig, weil bestimmte Eigenschaften interaktiver Systeme nur in der konkreten Interaktion meßbar sind (DZIDA 1990, S.10).

### **3. Benutzerbeteiligung bei Standardsoftware-Entwicklungen**

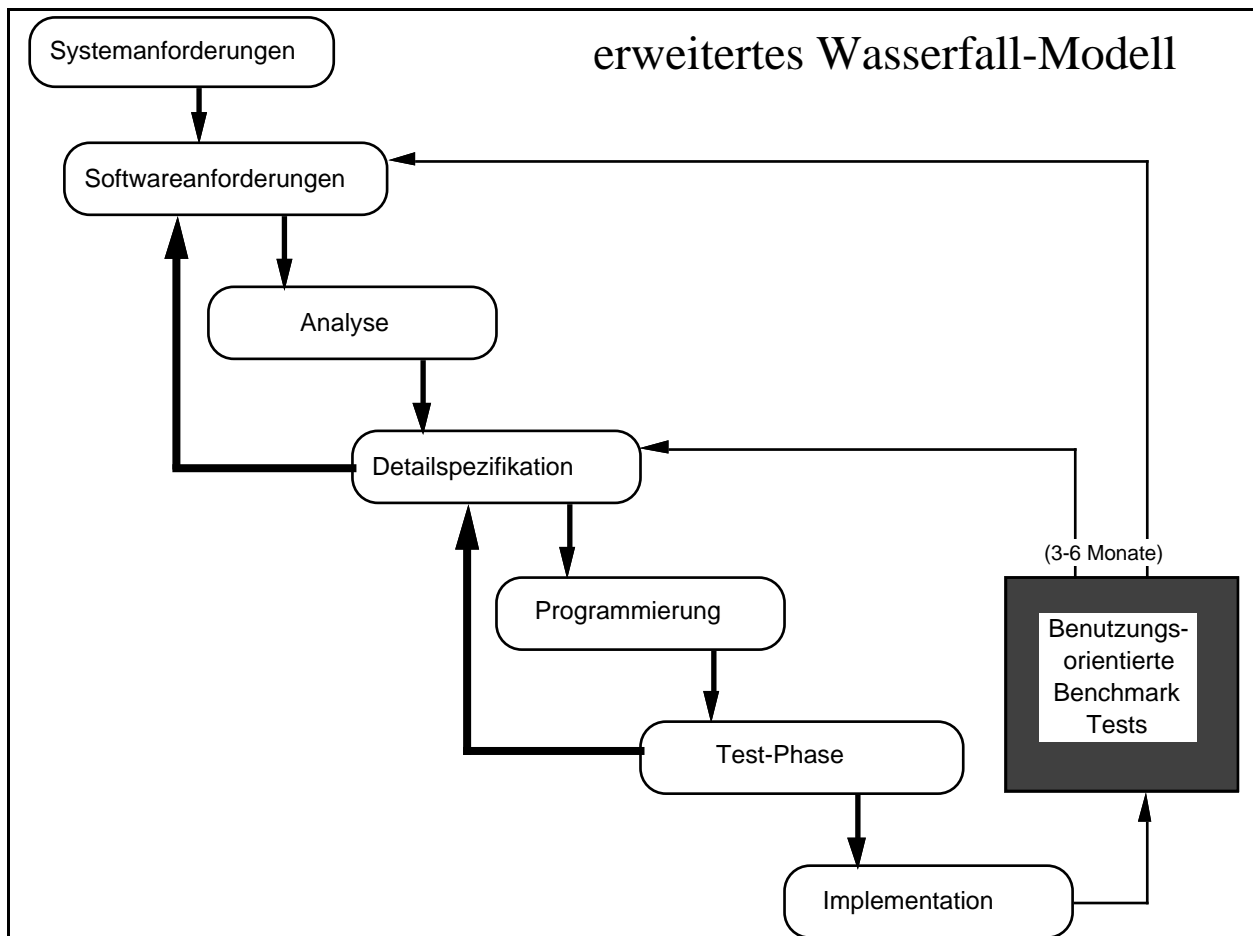
Als ein besonderes Problem im Zusammenhang mit der Beteiligung von Benutzern bei Standardsoftware-Entwicklungen muß die große *Heterogenität des Benutzerkreises* angesehen werden. Um nun benutzungs-orientierte Benchmark-Tests durchführen zu können, sollte das zu testende System bestimmte Voraussetzungen erfüllen (siehe weiter unten bei der Beschreibung der Metho-

---

\* Die Ergebnisse dieser Studie sind im Rahmen des vom BMFT geförderten Forschungsprojektes "Benutzerorientierte Software-Entwicklung und Schnittstellengestaltung (BOSS)" am Institut für Arbeitspsychologie (IfAP) der ETH-Zürich im Verbund mit der ADI-GmbH aus Karlsruhe gewonnen worden.

<sup>1</sup> zur Verwendung des Begriffes "Benchmark" im Kontext von Benutzerbeteiligung siehe: WILLIGES, WILLIGES & ELKERTON (1987, S. 1434); WHITESIDE, BENNETT & HOLTZBLATT (1988, S. 796); MACAULAY et al. (1990, S. 109).

de). Dies führt dazu, daß benutzungs-orientierte Benchmark-Tests vorwiegend im Entwicklungs- oder Lebenszyklus des Standardsoftware-Produktes eingesetzt werden (siehe Abb. 3.1).



**Abbildung 3.1:** Das traditionelle "Wasserfall"-Modell der Software-Entwicklung (nach BOEHM 1981) erweitert um Rückkopplungszyklen über benutzungs-orientierte Benchmark-Tests (siehe auch WILLIGES, WILLIGES & ELKERTON 1986, S. 1418).

Um ein möglichst repräsentatives Abbild der verschiedenen Benutzungskontexte von Standardsoftware-Produkten zu erhalten, lassen sich Befragungen mittels *Fragebögen* unter dem registrierten Benutzerkreis durchführen (KLAS, KUICH & LAES 1990; HUNZIGER & HÄSSIG 1990; STUTZ et al. 1990). In einer verkürzten Form können auch wichtige Informationen von TestpartnerInnen über eine entsprechend erweiterte Version der *Registrierkarten* fortlaufend erhalten werden (MOLLENHAUER 1991). Um End-Benutzer möglichst direkt und unmittelbar einzubeziehen, können *Benutzer-Entwickler-Treffen* durchgeführt werden (WAEBER 1990). Bei reinen Neu-Entwicklungen lassen sich im Rahmen von Marktanalysen auch stichprobenartig Befragungen (mittels Fragebogen oder Interview) unter dem potentiellen Benutzerkreis durchführen.



## 4. Benutzungs-orientierte Benchmark-Tests (bBTs)<sup>1</sup>

Benutzungs-orientierte Benchmark-Tests (bBTs) lassen sich in zwei Typen unterteilen: *induktive* und *deduktive bBTs* (RAUTERBERG 1990a). Die induktiven bBTs sind bei der Evaluation eines (z.B. vertikalen) Prototypen, oder einer (Vor)-Version zur Gewinnung von Gestaltungs- und Verbesserungsvorschlägen, bzw. zur Analyse von Schwachstellen in der Benutzbarkeit einsetzbar. Induktive bBTs können immer dann zum Einsatz kommen, wenn nur *ein* Prototyp, bzw. *eine* Version der zu testenden Software vorliegt. Demgegenüber verfolgen deduktive bBTs primär den Zweck, zwischen mehreren Alternativen (mindestens zwei Prototypen, bzw. Versionen) zu entscheiden (SPENCER 1985; KARAT 1988, S. 894). Zusätzlich lassen sich jedoch auch mit deduktiven bBTs Gestaltungs- und Verbesserungsvorschläge gewinnen (RAUTERBERG 1990b).

### 4.1. Induktive benutzungs-orientierte Benchmark-Tests (bBTs)

Bei der Durchführung eines induktiven bBTs sind die im folgenden aufgeführten Bedingungen zu beachten. Da die meisten Bedingungen zur Durchführung eines induktiven bBTs mit denen eines deduktiven bBTs übereinstimmen, werden dann später bei der Darstellung deduktiver bBTs nur noch die Änderungen und Ergänzungen erwähnt.

#### 4.1.1. Vorgehensweise bei induktiven bBTs

##### 4.1.1.1. Anforderungen an die zu testende Software

Um Artefakte bei der Benutzung - hervorgerufen durch unvollständige Systemfunktionalität - zu vermeiden, sollte im Bereich der zu testenden Systemfunktionen ein möglichst realitätsgerechtes Systemverhalten gegeben sein. Um die einzelnen Aktionen der TestpartnerInnen später auswerten zu können, empfiehlt es sich, eine automatische Aufzeichnung der Tastendrucke in das Test-System einzubauen ("logfile recording": MOLL 1987, MÜLLER-HOLZ et al. 1991).

##### 4.1.1.2. Gestaltung der Beobachtungssituation

Konstruktion der Benchmark-Aufgaben:

Als erstes sollte man mindestens eine, oder besser noch mehrere Benchmark-Aufgaben abgestimmt auf die zu testenden Systemteile festlegen. Diese Benchmark-Aufgaben sind dem typischen Aufgabenkontext des zukünftigen End-Benutzers zu entnehmen. Sofern dieser Aufgabenkontext nicht oder nur recht vage bekannt ist, sollten die Benchmark-Aufgaben zumindest jedoch typische, zu erwartende Teil-Aufgaben enthalten. Eine einzelne Benchmark-Aufgabe sollte nicht zu komplex, aber auch nicht zu einfach sein; die Bearbeitungszeiten sollten ungefähr zwischen fünf und fünfzehn Minuten liegen. Die Formulierung der Aufgaben sollte als schriftliche Fassung den TestpartnerInnen während der Aufgabenbearbeitung zugänglich sein. Je nach fach-spezifischem Vorwissen der TestpartnerInnen über den Aufgabenkontext ist es sinnvoll, die Aufgabenbeschrei-

---

<sup>1</sup> bBT = benutzungs-orientierter Benchmark-Test; mit dem Adjektiv "benutzungs-orientiert" soll eine Verwechslung mit den sonst üblichen, rein system-technischen Benchmark-Tests vermieden werden.

bung unterschiedlich abzufassen; bei hohem fach-spezifischem Vorwissen ist die Beschreibung möglichst *problem-orientiert*<sup>1</sup> abzufassen, bei geringem, bzw. keinem fach-spezifischen Vorwissen ist die Beschreibung *handlungs-orientiert*<sup>2</sup> zu halten. Die handlungs-orientierte Aufgabenbeschreibung soll verhindern helfen, daß die beobachteten Benutzungsprobleme überwiegend aufgrund fehlenden Fachwissens zustande gekommen sind. Es ist bei der Konstruktion von Benchmark-Aufgaben generell zu berücksichtigen, wieviel Vor-Information über die Handhabung des zu testenden Systems in die Aufgabenbeschreibung einfließen darf. Je weniger Hinweise auf die Lösung der jeweiligen Benchmark-Aufgabe in der Aufgabenstellung enthalten sind, desto stärker fällt der auf die Sachaufgabe<sup>3</sup> bezogene Problemlöse-Anteil, welcher von den TestpartnerInnen aufgebracht werden muß, ins Gewicht und verringert somit die Bedeutung des auf die Interaktionsaufgabe bezogenen Problemlöse-Anteil.

#### Auswahl der TestpartnerInnen:

Die TestpartnerInnen-Gruppe sollte möglichst *repräsentativ* für die Population der End-Benutzer sein (siehe CAMPBELL & STANLEY 1970). Die Auswahl der TestpartnerInnen muß daher *zufällig* aus dem potentiellen oder aktuellen Benutzerkreis erfolgen. Die ausgewählten TestpartnerInnen sollten das zu testende System nicht kennen. Da sich diese Bedingung oftmals bei größeren Entwicklungsphasen, bzw. Weiterentwicklungen nicht einhalten läßt, muß die Vorerfahrung der TestpartnerInnen mit dem System, bzw. ähnlichen Systemen kontrolliert werden (siehe dazu mehr im Abschnitt zur "Messung von intervenierenden Variablen"). Die Gruppengröße sollte aus statistischen Gründen mindestens sechs und aus ökonomischen Gründen maximal zwanzig bis dreißig Personen (pro Gruppe) betragen<sup>4</sup>. Wenn kein oder nur sehr wenig Wissen über die Population der End-Benutzer vorliegt, lohnt es sich, die TestpartnerInnen-Gruppe hinsichtlich der folgenden Parameter möglichst *heterogen* zusammensetzen: EDV-Vorerfahrung, Alter, Geschlecht, Ausbildung, Beruf (Kirakowski & Corbett 1990 29).

#### Umgebungsbedingungen während der Beobachtung:

Anzustreben ist eine möglichst den realen Einsatzbedingungen entsprechende Beobachtungs-Umgebung (siehe zum Thema "kontext-sensitive Beobachtung" WHITESIDE, BENNETT & HOLTZBLATT 1988). Da es für die spätere Auswertung jedoch oftmals sehr wichtig ist, das Verhalten der einzelnen TestpartnerInnen sowie das entsprechende Systemverhalten auf Video, Tonband,

---

<sup>1</sup> z.B. "Erstellen Sie bitte einen Brief mit folgendem Inhalt: ... und bereiten Sie ihn zum Eintüten und Versenden an folgende Adressen: ... vor. Benutzen Sie dabei bitte als Briefvorlage das Dokument mit dem Namen: ..."

<sup>2</sup> z.B. "Laden Sie bitte das Textdokument mit folgendem Namen: ... und ergänzen es um den folgenden Inhalt: ... Versehen Sie dieses Textdokument mit allen für die Serienbriefferstellung notwendigen Steueranweisungen. ... usw."

<sup>3</sup> Zur Unterscheidung von Sach- und Interaktions-Aufgabe siehe STREITZ (1986).

<sup>4</sup> Je größer die Stichprobe ist, desto sensibler ist der statistische Test; man kann somit kleinere Effekte messen. Zur Berechnung der genauen Stichprobengröße unter Annahme einer zu erwartenden Effektstärke siehe bei BORTZ (1989).

"Screen-recording"<sup>1</sup>, etc. aufzuzeichnen (siehe auch VOSSEN 1991), empfiehlt es sich, einen ruhigen, abgeschlossenen Raum mit entsprechendem Mobiliar zu benutzen. Stehen keine speziellen Aufzeichnungsgeräte zur Verfügung, sind für den Testleiter einfach auszufüllende Protokollbögen (z.B. mittels Strichlisten für vorgegebene Problem-, Fehler-Kategorien, etc.) schon sehr nützlich.

#### 4.1.1.3. Vorbereitung der TestpartnerInnen

Der Testleiter sollte sich stets darum bemühen, jeder TestpartnerIn das Gefühl der Wichtigkeit und Wertschätzung zu vermitteln (WEISER & SHNEIDERMAN 1987, S. 1401). Die folgenden drei Aspekte sind für die Schaffung eines vertrauensvollen und für Kritik offenen Klimas hilfreich.

##### Beschreibung von Ziel und Zweck:

Mit möglichst allgemein verständlichen Worten wird der TestpartnerIn Ziel und Zweck des bBTs erläutert (z.B. Test eines Prototypen in einem frühen Entwicklungsstadium, etc.)<sup>2</sup>. Es ist besonders wichtig, der TestpartnerIn verständlich zu machen, daß das System und *nicht* die TestpartnerIn getestet werden soll. Jede Art von Schwierigkeiten seitens der TestpartnerIn bei der Aufgabebearbeitung sind von besonderem Interesse!

##### Das Recht der TestpartnerIn auf Beendigung:

Jeder TestpartnerIn muß zugesichert werden, daß sie/er die Durchführung des bBTs jederzeit unter- und sogar abbrechen kann, ohne daß irgendwelche negativen Konsequenzen (z.B. Wegfall einer zugesagten Bezahlung, etc.) erfolgen. Die vollständige Freiwilligkeit der Teilnahme und Zusage der Anonymität ist unabdingbar für die Gewinnung von brauchbaren Beobachtungen.

##### Vertrautmachen mit der Untersuchungssituation:

Jeder TestpartnerIn werden alle für die Durchführung des Benchmark-Tests in dem Beobachtungsraum vorhandenen Geräte (Computer, Video-Kamera, Mikrophone, etc.) hinsichtlich Einsatzzweck und Funktionsweise erläutert. Als besonders wichtig erweist sich eine möglichst standardisierte Einführung in die Handhabungs- und Bedienungsweise der zu testenden Software. Je nach Vorerfahrung der TestpartnerIn können unterschiedliche Schwerpunkte gesetzt werden. Um die Motivation der TestpartnerIn bei einer längeren Einführungsphase (in die Benutzung der Software) aufrechtzuerhalten, hat es sich als sinnvoll gezeigt, die TestpartnerIn zur Generierung von Frage- und Problemstellungen über die "Welt der Anwendungsobjekte" zu stimulieren und zur selbstständigen Exploration anzuregen. Dennoch muß unbedingt darauf geachtet werden, daß jede TestpartnerIn das gleiche Vorwissen erhält. Da diese Bedingung nur schwer zu kontrollieren ist, begnügt man sich oftmals mit Personen *ohne* jegliches EDV-Vorwissen.

---

<sup>1</sup> diese Software für den MacIntosh ist zu beziehen bei: FARALLON (1988): Screenrecording vers. 1.0. FARALLON COMPUTING INC., 2150 Kittredge Street, Berkley, CA 94704, USA.

<sup>2</sup> für weitere Formulierungsvorschläge siehe GOMOLL (1990, S. 87).

#### Die Technik des "lauten Denkens":

Viele TestpartnerInnen haben Schwierigkeiten, ihre Gedanken während der Bearbeitung einer Aufgabe laut zu äußern. Es empfiehlt sich daher, der TestpartnerIn zu verdeutlichen, was mit "lautem Denken" gemeint ist, und mit ihr/ihm einige Übungsbeispiele gemeinsam durchzugehen. Trotzdem neigen TestpartnerInnen einerseits bei hoch routinisierten Handlungen, andererseits bei komplexen Problemlösungsprozessen dazu, das "laute Denken" einzustellen. Dies kann man durch folgende Maßnahmen umgehen: a) der Testleiter fordert den Testpartner in solchen Situationen zum "lauten Denken" auf; b) es werden *zwei* TestpartnerInnen als Paar mit der Aufgabenbearbeitung betraut, wodurch die verbale Kommunikation zwischen beiden Partnern das "laute Denken" ersetzt; c) man führt nach Beendigung des bBTs der TestpartnerIn problematische Ausschnitte per Videoaufzeichnungen vor und bespricht mit ihr/ihm seine jeweiligen Ziele und Handlungsabsichten (MOLL 1987).

#### 4.1.1.4. Durchführung der Beobachtung

Es empfiehlt sich, vor Beginn einer Beobachtungsserie, ein bis zwei Probe-Durchläufe zur Optimierung der gesamten Beobachtungsprozedur durchzuführen.

#### Messung von intervenierenden Variablen:

Als bedeutsame Einflußgrößen auf die Art und Weise der Aufgabenbearbeitung haben sich die EDV-Vorerfahrung der TestpartnerIn und die Hilfestellungen durch den Testleiter herausgestellt (RAUTERBERG 1988a, 1990b). Beide Variablen sollten gemessen werden, um sie später bei der statistischen Auswertung berücksichtigen zu können. Die Vorerfahrung läßt sich mittels Fragebogen erfassen (YAVERBAUM & CULPAN 1990). Dabei ist es hilfreich, getrennt nach Nutzungsdauer (ND; in Jahren, Monaten, Wochen, etc.) und Nutzungshäufigkeit (NH; in Stunden/Woche, etc.) von verschiedenen interaktiven Systemen zu fragen, und anschliessend einen Nutzungs-Intensitäts-Index ( $NII = ND * NH$ ; in Stunden) zu berechnen.

Die Art und die Anzahl der gegebenen Hilfestellungen des Testleiters wird während des bBTs auf einem Protokollbogen oder mittels eines entsprechenden Protokollierungssystems (VOSSEN 1991) registriert. Die Anzahl an Hilfestellungen muß *positiv* mit der Aufgaben-Bearbeitungszeit korrelieren, um zu garantieren, daß die Hilfestellungen durch den Testleiter *nicht* zu einer artifiziellen Verkürzung der Bearbeitungszeit geführt haben.

Die kognitiv-emotionale Beanspruchung der TestpartnerInnen sollte unbedingt gemessen werden, um zu verhindern, daß zwar auf der Leistungsebene eine Verbesserung stattgefunden hat, diese Verbesserung aber auf Kosten der kognitiv-emotionalen Beanspruchung zustande gekommen ist. Hier haben sich neben Zustandsfragebögen (zB. EZ-Skala, siehe APENBURG 1986), auch Video-ratingverfahren (siehe RAUTERBERG 1988b) bewährt. Die Beanspruchung kann auch als ein Bewertungskriterium für die erreichte Verbesserung eingesetzt werden; dann ist sie jedoch als "abhängige Variable" anzusehen.

Da bisher bei der verbreiteten interaktiven Software im wesentlichen das visuelle Wahrnehmungssystem der TestpartnerIn in Anspruch genommen wird, ist es ratsam, die Sehschärfe und gege-

benenfalls die Farbwahrnehmung mit entsprechenden Sehtesttafeln vor Beginn der Untersuchung zu testen. Zusätzlich scheint die Persönlichkeits-Dimension der "visual ability"<sup>1</sup> einen Einfluß auf das Erlernen interaktiver Software zu haben (SEIN, BOSTROM & OLFMAN 1987).

#### Messung der abhängigen Variablen:

Als "abhängige" Variablen werden alle erhobenen Meßwerte bezeichnet, welche bei der Auswertung Aufschluß über die Güte der Benutzbarkeit des zu testenden Systems Auskunft geben können (WHITESIDE et al. 1988; BOOTH 1990, S. 123). Die Menge der abhängigen Variablen teilt sich auf in die Menge der *qualitativen* Variablen (problematische Benutzungssituationen, handlungspsychologische Bedingungskomplexe, etc.) und die Menge der *quantitativen* Variablen auf (Bearbeitungsdauer, Einlernzeit, Überlegungszeit<sup>2</sup>, Anzahl Tastendrucke, Anzahl Fehler<sup>3</sup>; siehe MÜLLER-HOLZ et al. 1991).

#### Aufgabenbearbeitung durch die TestpartnerInnen:

Es gibt grundsätzlich zwei verschiedene Vorgehensweisen für die Gestaltung des zeitlichen Rahmens: 1. die TestpartnerIn wird aufgefordert, die gestellte Aufgabe solange zu bearbeiten, bis sie/er sie vollständig gelöst hat, bzw. die Bearbeitung auf eigenen Wunsch vorzeitig abbricht; 2. der TestpartnerIn wird eine so knapp bemessene Zeitspanne für die Aufgabenbearbeitung zur Verfügung gestellt, daß sie/er in der Regel die Benchmark-Aufgabe nicht vollständig lösen kann. Falls die Aufgabenbearbeitungszeit als ein relevanter Indikator für Benutzungseigenschaften des zu testenden Systems herangezogen werden soll, empfiehlt sich die erste Variante. Bei der zweiten Variante muß man für eine Auswertung den erreichten Lösungsgrad der Aufgabe bewerten. Diese Bewertung ist oft nicht einfach möglich, sodaß sich daher in vielen Fällen die erste Variante empfiehlt. Bei der ersten Variante kommt man jedoch auch nicht um die Festlegung einer maximalen, ausreichend bemessenen Testdurchführungszeit für den Test insgesamt umhin. Nach Ablauf dieser Gesamtzeit sollte die weitere Testdurchführung abgebrochen werden.

#### Verhalten des Testleiters:

Der Testleiter hält sich während der Aufgabenbearbeitung ruhig im Hintergrund. Für ihn ist es sehr wichtig, seinen natürlichen Impuls, der TestpartnerIn in problematischen Situationen sofort zu helfen, "im Zaume zu halten". Hier kann eine klare Absprache zwischen Testleiter und TestpartnerIn hilfreich sein: nur wenn sich die TestpartnerIn explizit an den Testleiter wendet und um Hilfe nachfragt, wird der Testleiter aktiv. Ein zu frühes Eingreifen des Testleiters hindert die TestpartnerIn, eine eigene Lösung zu finden. Als Testleiter sollten daher möglichst keine an der Entwicklung des zu testenden Systems unmittelbar beteiligten Personen eingesetzt werden. Es wird sogar empfohlen, um eine weitgehend realistische Beobachtungssituation herzustellen, daß der

---

<sup>1</sup> gemessen mit dem Test nach EKSTROM, FRENCH & HARMAN (1976).

<sup>2</sup> siehe hierzu ACKERMANN & GREUTMANN (1987), bzw. GREUTMANN & ACKERMANN (1987).

<sup>3</sup> zum Thema "Fehler"-Analyse siehe DUTKE (1988) und ZAPF & FRESE (1989).

Testleiter sich völlig passiv verhält und dies der TestpartnerIn vorher deutlich macht (GOMOLL 1990, S.89).

"Sokratischer Dialog":

Es ist insbesondere bei der Durchführung von Benchmark-Tests mit weitgehend EDV-unerfahrenen TestpartnerInnen ratsam, ihnen in scheinbar ausweglosen Situationen mit Hilfestellungen seitens des Testleiters zur Seite zu stehen. Dies ist insbesondere dann wichtig, wenn man möglichst vollständige Aufgabenbearbeitungen erreichen will. Hier hat es sich nun bewährt, daß der Testleiter gemäß dem folgenden, fünf-fach gestuften Schema der TestpartnerIn Hilfestellung zukommen läßt (RAUTERBERG 1988a). Wir nennen dieses Schema "sokratischer Dialog", weil der Testleiter auf ein Hilfesuch seitens der TestpartnerIn nach dem "*Frage-Prinzip mit minimaler Information*" vorgeht:

1. Stufe: Der Testleiter weist die TestpartnerIn auf das Handbuch und/oder das Hilfesystem hin. "Bitte schauen Sie im Handbuch (Hilfesystem) nach!".
2. Stufe: Der Testleiter versucht die Aufmerksamkeit der TestpartnerIn auf die Einweisungs-/ Instruktionsphase zu lenken. "Können Sie sich noch erinnern, was Sie während der Einweisung/ Instruktion an dieser Stelle getan haben?".
3. Stufe: Der Testleiter lenkt die Aufmerksamkeit auf den relevanten Suchbereich, indem er den Suchraum einschränkt. "Können Sie sich noch erinnern, welche Funktionstaste (bzw. Menü-Item, Ikon, etc.) in Frage kommen könnte?"
4. Stufe: Der Testleiter schränkt den Suchraum auf die konkrete Dialog-Operation weiter ein. "Können Sie sich noch erinnern, was passieren würde, wenn Sie die Funktionstaste F3 (bzw. das Menü-Item "DB-Info", etc.) betätigen würden?"
5. Stufe: Der Testleiter weist die TestpartnerIn direktiv an, die entsprechende Dialog-Operation durchzuführen. "Bitte drücken Sie jetzt die Funktionstaste F3!".

Meistens reicht es aus, wenn der Testleiter bis zur dritten Stufe Hilfestellung gibt, weil dann viele TestpartnerInnen schon ausreichend Informationen haben, um sozusagen "von selbst" auf die Lösung ihres interaktiven Problems zu kommen (zB. erkennbar an Ausrufen wie "Ach ja", etc.). Dieses Vorgehen empfiehlt sich deshalb, weil die TestpartnerInnen bis zur vierten Stufe das Empfinden der Kontrolle über die Situation behalten können.

#### 4.1.1.5. Nachbereitung mit den TestpartnerInnen

Jeder TestpartnerIn muß die Gelegenheit gegeben werden, sofern dies nicht schon im Rahmen der Video-Konfrontation<sup>1</sup> eingeplant ist, für ihr/ihn wichtige und noch offene Fragen zu Zwecken und Zielen des bBTs, problematischen Situationen während der Aufgabenbearbeitung, etc. in einem Gespräch nach Beendigung der Aufgabenbearbeitung mit dem Testleiter klären zu können. Meistens erhält man sehr zutreffende Problemsichten aus diesen Gesprächen. In dieser Nachbereitung

---

<sup>1</sup> siehe mehr zur Methode der Videokonfrontation bei MOLL 1987, S. 183f.

lassen sich auch Fragebögen mit standardisierten oder offenen Fragen zur Beantwortung spezieller Benutzungsaspekte einsetzen.

#### 4.1.1.6. Auswertung der Beobachtungen

Während der Durchführung eines bBTs wird man immer wieder überraschende und sehr informative Benutzungsweisen beobachten können (KARAT 1988, S. 895). Es lohnt sich daher, die einzelnen Aufgabenbearbeitungen der verschiedenen TestpartnerInnen auf Video, U-matic, etc. aufzuzeichnen, um sie dann mit den Entwicklern später detailliert in entsprechenden Workshops (WAEBER 1990) diskutieren zu können. Wichtig ist dabei, daß die beobachtbaren Schwierigkeiten niemals der TestpartnerIn, sondern ausschließlich der Benutzbarkeit des zu testenden Systems angelastet werden.

##### Aufbereitung der Daten:

Die Beobachtungsdaten (Video, Tonband, Logfile, 'screenrecording') müssen hinsichtlich der design-relevanten Informationen ausgewertet werden. Dazu empfiehlt es sich, auf der Grundlage der durchgeführten Beobachtungen ein Auswertungsraster aufzustellen und dieses Raster systematisch auf die Beobachtungsdaten aller TestpartnerInnen anzuwenden (siehe auch VOSSEN, 1991). Aufwendig und schwierig sind die qualitativen Auswertungen von Logfile-Daten, weil es sich hierbei oftmals um komplexe Mustererkennungsprozesse handelt, welche bisher nur begrenzt automatisch ausführbar sind (ACKERMANN & GREUTMANN 1987, MÜLLER-HOLZ et al. 1991).

##### Qualitative Auswertungen:

Als eine besonders informative Quelle für design-relevante Entscheidungen hat sich die Analyse von Videoaufzeichnungen ergeben. Hierbei werden Erklärungsmodelle für bestimmtes Benutzungsverhalten aufgestellt, um das beobachtbare Verhalten in problematischen Situationen handlungspsychologisch begründen zu können. Die eingehendere Beschäftigung mit diesen Situationen führt dann unter Berücksichtigung des Kriterienkonzeptes von ULICH (1985, 1986, 1991) zu konstruktiven, tragfähigen Gestaltungsvorschlägen (siehe hierzu mehr im entsprechenden Kapitel in Teil-II).

##### Quantitative Auswertungen:

Um die in der qualitativen Auswertung gewonnenen Erklärungsmodelle für einzelne Benutzungsprobleme auf eine möglichst breite Basis zu stellen (Problem der 'Generalisierbarkeit'), sind statistische Auswertungen unumgänglich. Diese können von einfachen Häufigkeitszählungen bestimmter Problemklassen, bis hin zu komplexen inferenz-statistischen Zusammenhangsanalysen gehen (siehe mehr bei LANDAUER 1988). Dazu ist es notwendig, daß die verschiedenen Benutzungsigenschaften (Dauer der Aufgabenbearbeitung, mittlere Überlegungszeit, Fehlerart, Ausmaß an Beanspruchung, etc.) pro TestpartnerIn in quantifizierter Form vorliegen.

## **4.2. Deduktive benutzungs-orientierte Benchmark-Tests (bBTs)**

Deduktive bBTs dienen primär der Entscheidungsfindung zwischen verschiedenen System-Alternativen (z.B. WANDKE 1990), bzw. zur Kontrolle der erreichten Verbesserung (siehe zum Stich-

wort "Oberflächen-Effekt" im entsprechenden Kapitel in Teil-II) und erst sekundär der Gewinnung von Gestaltungsvorschlägen (KARAT 1988). Es ergeben sich daher einige unterschiedliche Anforderungen an die Durchführung von deduktiven bBTs.

#### **4.2.1. Vorgehensweise bei deduktiven bBTs**

##### 4.2.1.1. Voraussetzungen an die zu testende Software

Im Gegensatz zu induktiven bBTs müssen die verschiedenen zu testenden, alternativen Systemversionen beim deduktiven bBT verstärkt der Forderung nach einem - an realen Einsatzbedingungen gemessenen - realistischen Systemverhalten (d.h. möglichst funktionale Vollständigkeit, adäquates Systemantwortzeitverhalten, etc.) genügen. Diese Anforderungen sind deshalb wichtig, weil die Entscheidung zwischen den System-Alternativen primär mittels quantitativer Meßgrößen gefällt wird (Kirakowski & Corbett 1990).

##### 4.2.1.2. Gestaltung der Beobachtungssituation

Test-Design:

Je nachdem, ob viele TestpartnerInnen wenig Zeit haben (Fall-I), oder, ob wenige TestpartnerInnen viel Zeit haben (Fall-II), an einem bBT teilzunehmen, wird die Gruppenbildung unterschiedlich vorgenommen. Im Fall-I wird pro System-Alternative eine eigene Gruppe gebildet. Im Fall-II hat lediglich eine Gruppe alle System-Alternativen zu testen. Um Lerneffekte messen und kontrollieren zu können, muß dann die Reihenfolge der zu testenden Systeme innerhalb dieser einen Gruppe ausbalanciert werden (BORTZ 1984, S. 422ff; siehe auch Teil-II weiter unten). Es gibt somit zwei unterschiedliche Test-Designs: I) pro Oberflächenvariante wird je eine eigene Testgruppe ( $N \geq 6$ ) gebildet; II) alle TestpartnerInnen testen alle  $n$  Oberflächenvarianten; um hierbei Lerneffekte durch die wiederholte Aufgabenbearbeitung kontrollieren zu können, müssen die TestpartnerInnen in  $(n!)$  verschiedene Unter-Gruppen (mit jeweils  $N \geq 1$ ) aufgeteilt werden (bei  $n=2$ : (AB) (BA); bei  $n=3$ : (ABC) (ACB) (BAC) (BCA) (CAB) (CBA); bei  $n=4$ : (ABCD) (...)).

Konstruktion der Benchmark-Aufgaben:

Im Fall-II müssen für jede System-Alternative *parallele* Benchmark-Aufgaben entwickelt werden (siehe im Kapitel 4.3 in Teil-II).

Auswahl der TestpartnerInnen:

Die Gruppengröße sollte aus statistischen Gründen mindestens sechs Personen pro Gruppe betragen. Wenn man plausible Annahmen über den zu erwartenden Oberflächeneffekt hinsichtlich einer quantitativ zu messenden Benutzungseigenschaft (z.B. Dauer der Aufgabenbearbeitung, Anzahl Fehler, etc.) für die Alternativen hat, kann man die genau benötigte Gruppengröße auch ausrechnen (BORTZ 1984, S. 504ff; LANDAUER 1988, S. 918ff). Die einzelnen Gruppen müssen hinsichtlich verschiedener Parameter möglichst homogen sein: EDV-Vorerfahrung, Geschlecht, Alter, Beruf. Die Homogenität der Test-Gruppen trägt dafür Sorge, daß die "Stör"-Varianz minimiert wird und damit die Test-Stärke steigt. Wenn man jedoch den Grad der EDV-Kenntnisse mit testen



möchte, empfiehlt es sich, entsprechende Test-Gruppen extra speziell zusammenzustellen (RAUTERBERG 1990b).

#### 4.2.1.3. Auswertung der Beobachtungen

Mittels inferenzstatistischer Auswertungsverfahren können die gewonnenen Beobachtungsergebnisse auf ihre *Generalisierbarkeit* hin getestet werden (LANDAUER 1988; BORTZ 1989).

### **4.3. Aufwandabschätzung der Methode**

Der Durchführungsaufwand von bBTs hängt im wesentlichen von der Anzahl der zu beteiligenden TestpartnerInnen ab. Je mehr TestpartnerInnen beteiligt sind, desto repräsentativer und umfassender sind die Ergebnisse. Dennoch ist es ratsam, den Aufwand auf ein notwendiges Minimum zu beschränken. Am einfachsten lassen sich induktive bBTs durchführen. Für die Planung, Durchführung und Auswertung sind in diesem Fall einige Personen-Tage bis -Wochen zu veranschlagen. Dagegen ist bei deduktiven bBTs in der Regel mit einigen Personen-Wochen bis -Monaten zu rechnen.

Ein sehr interessantes und auch sehr wichtiges Ergebnis besteht darin, daß allein durch den einfachen induktiven bBT (N = 8 TestpartnerInnen) (RAUTERBERG 1988a, RAUTERBERG 1990b) ca. 50% der insgesamt gefundenen Gestaltungsvorschläge entdeckt werden konnten. Es genügen in der Regel schon "eine handvoll" TestpartnerInnen, um die wichtigsten Benutzungsprobleme aufdecken zu können.

### **4.4. Kosten-Nutzen-Analyse**

Wenn man deduktive bBTs durchführt und Performanz-Masse als abhängige Meßgrößen erhebt, kann man den erreichten Grad der Verbesserung auch in finanzieller Hinsicht abschätzen. Hierzu ist eine erweiterte Kosten-Nutzen Betrachtung notwendig. Die finanziellen Aufwendungen, welche durch den Einsatz von bBTs entstehen, werden den zu erwartenden Einsparungen beim Gebrauch der verbesserten Version gegenübergestellt.

Es lassen sich folgende Berechnungsgrößen für eine erweiterte Kosten-Nutzen-Analyse zugrunde legen (MANTEI & TEOREY 1988; KARAT 1990), welche zunächst investiert werden müssen:

- Kosten für den Testleiter (als Jahresgehalt) KTL,
- Anzahl an eingesetzten Testleitern ATL,
- Zeitdauer der Durchführung (in Personen-Monaten) ZPM,
- Kosten für die Teilnahme pro TestpartnerIn KTP,
- Anzahl der teilnehmenden TestpartnerInnen ATP,
- Kosten für zusätzliche Geräte (Videoanlage, etc.) KZG,
- Kosten für sonstige Ausgaben (Sekretariat, etc.) KSA.

**Formel 1** zur Berechnung der gesamten Investitionskosten:

$$IGK = (ATL * KTL / 12) * ZPM + ATP * KTP + KZG + KSA$$

Zur Berechnung des Grades der erreichten Verbesserung (VG) wird der Quotient aus dem ausgewählten Performanz-Maß (PM; z.B. die Test-Bearbeitungs-Zeit "TBZ") für die neue und die alte Software-Version gebildet:

**Formel 2** zur Berechnung der Verbesserungs-Grades der neuen Version:

$$VG = 100\% - (PM_{\text{neu}} / PM_{\text{alt}}) * 100\%$$

Die folgenden Berechnungsgrößen können als Grundlage zur Abschätzung der Einsparungen durch den Einsatz der verbesserten Softwareversion herangezogen werden:

- Anzahl der End-Benutzer AEB,
- durchschnittlicher Stundenlohn des End-Benutzers DSL,
- Anzahl an Arbeitsstunden der End-Benutzer AAS,
- Anschaffungspreis der Software APS,
- "Upgrade"-Preis für die verbesserte Version UPS,
- Verbesserungs-Grad (Angabe in Prozent) VG.

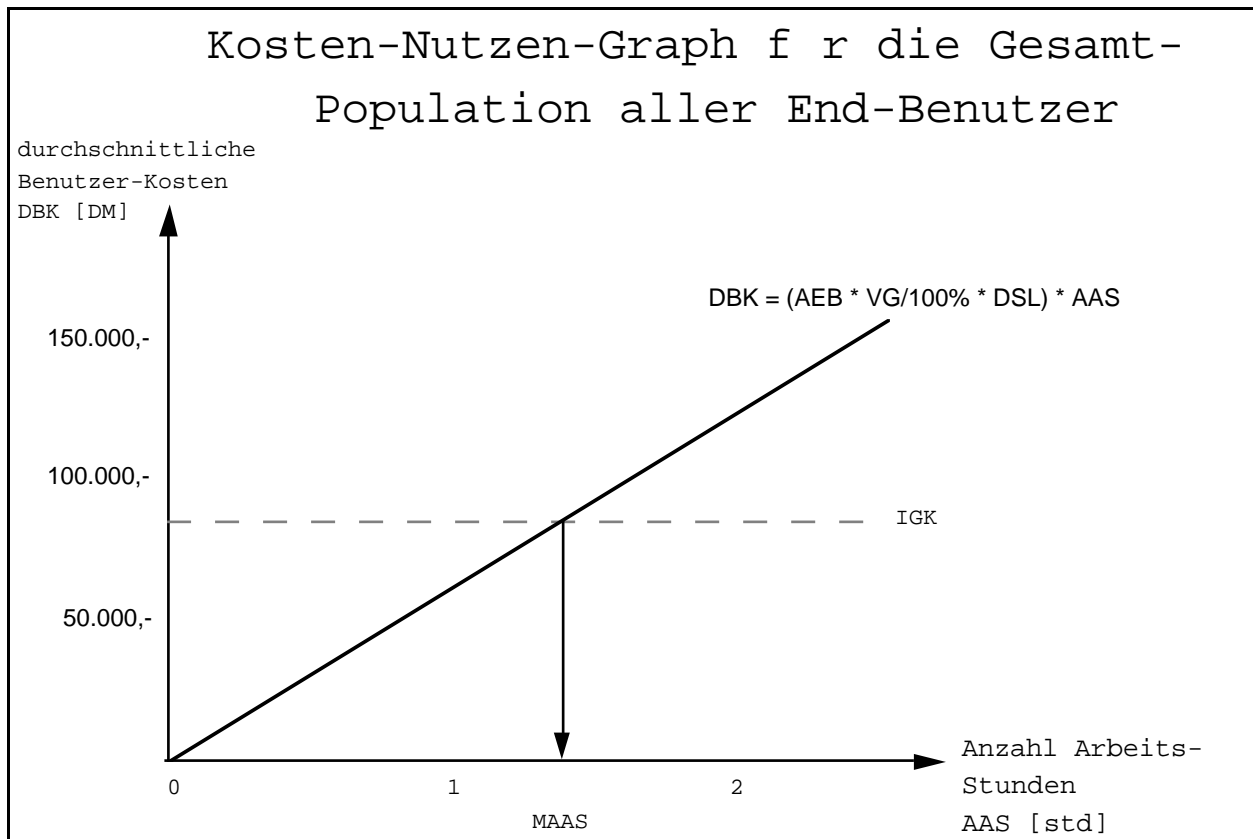
**Formel 3** zur Berechnung der durchschnittlichen Kosten aller End-Benutzer:

$$DBK_{\text{alt}} = (AEB * (PM_{\text{alt}} / PM_{\text{alt}}) * DSL) * AAS$$

$$DBK_{\text{neu}} = (AEB * (PM_{\text{neu}} / PM_{\text{alt}}) * DSL) * AAS$$

Um den "return of investment" abzuschätzen, kann man das folgende Verfahren anwenden. Zuerst berechnet man  $DBK_{\text{alt}}$  (nach Formel 3) mit der alten Version. Dann bestimmt man anhand der gemessenen Performanz-Maße im bBT den Grad der Verbesserung (VG) in Prozent (z.B. der Quotient aus der Bearbeitungszeit (TBZ) mit der neuen und der alten Version; siehe Kapitel 5 im Teil-II). Wird nun der Wert von  $DBK_{\text{alt}}$  mit dem Prozentwert multipliziert, erhält man  $DBK_{\text{neu}}$ . Die Differenz  $EK = (DBK_{\text{alt}} - DBK_{\text{neu}})$  ergibt die eingesparten Kosten (EK). Diesem Wert wird der Wert IGK gegenübergestellt. Da man jedoch die Größe AAS in Formel 3 praktisch nicht genau angeben kann, empfiehlt es sich, einen Kosten-Nutzen-Graphen aufzuzeichnen. Aus dieser Graphik kann man dann unmittelbar die minimale Anzahl Arbeitsstunden (MAAS) der Population aller End-Benutzer ablesen, ab derer sich das Arbeiten mit der neuen Version für die Gruppe aller End-Benutzer "bezahlt" gemacht hat.

Mittels des Kosten-Nutzen-Graphes (Abb. 4.1) kann man lediglich die minimale Anzahl Arbeitsstunden für die gesamte Population der End-Benutzer ermitteln. Für den einzelnen End-Benutzer sieht der zu leistende minimale Arbeitsaufwand (MAAS) etwas anders aus. Der einzelne End-Benutzer muss nämlich erst den Anschaffungs-Preis der Software (APS; bzw. den Upgrade-Preis UPS) "erwirtschaften".



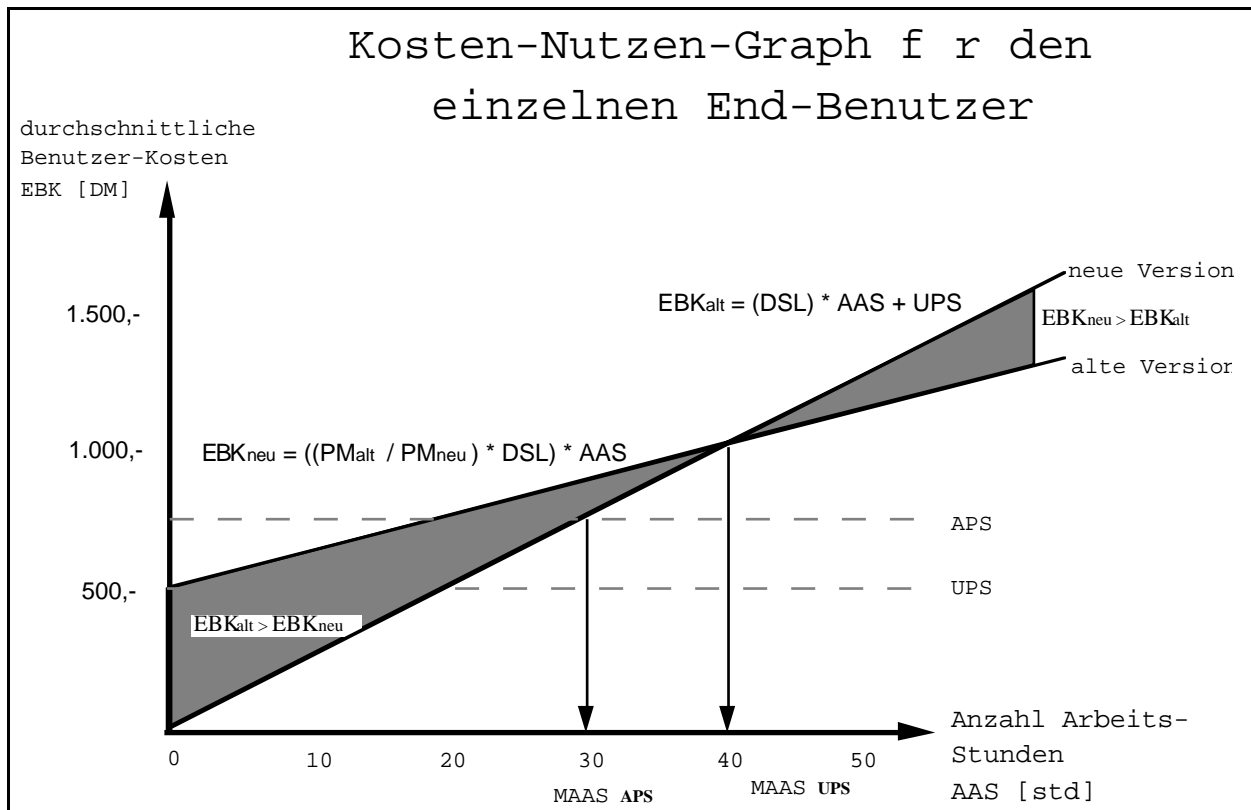
**Abbildung 4.1:** Der Kosten-Nutzen-Graph für die Abschätzung der minimalen Anzahl Arbeitsstunden (MAAS), ab derer sich die investierten Gesamtkosten (IGK) für die gesamte Population aller End-Benutzer (AEB) bei einem durchschnittlichen Stundenlohn (DSL) bezahlt gemacht haben.

**Formel 4** zur Berechnung der durchschnittlichen Kosten eines End-Benutzers:

$$EBK_{\text{alt}} = ((PM_{\text{alt}} / PM_{\text{alt}}) * DSL) * AAS + UPS$$

$$EBK_{\text{neu}} = ((PM_{\text{alt}} / PM_{\text{neu}}) * DSL) * AAS$$

Tragen wir die beiden Geradengleichungen aus Formel 4 in den individuellen Kosten-Nutzen-Graph für den einzelnen End-Benutzer ein, so können wir die minimale Anzahl Arbeitsstunden (MAASAPS; bzw. MAASUPS) ermitteln (Abb. 4.2). Die Steigung der Geraden  $EBK_{\text{neu}}$  setzt sich aus dem geschätzten durchschnittlichen Stundenlohn gewichtet mit dem erreichten Verbesserungsgrad ( $PM_{\text{alt}}/PM_{\text{neu}}$ ) zusammen. Wenn z.B.  $VG = 50\%$  ist, dann ist der Benutzer ( $PM_{\text{alt}} / PM_{\text{neu}}$ ) = 2.0 mal so schnell mit der verbesserten Version. Bei einem durchschnittlichen Stundenlohn von 12,50 [DM/Std] benötigt der Neueinsteiger  $MAAS_{\text{APS}} = 30$  [Std] Arbeitsstunden ( $EBK_{\text{neu}} = APS$ ; d.h.:  $MAAS_{\text{APS}} = (APS / DSL) / (PM_{\text{alt}} / PM_{\text{neu}})$ ), um den Anschaffungs-Preis (z.B.  $APS = 750,-$  DM) zu erwirtschaften. Für den Umsteiger dagegen ist die Gewinnzone [ $EBK_{\text{neu}} > EBK_{\text{alt}}$ ] erst nach  $EBK_{\text{neu}} = EBK_{\text{alt}}$  erreicht ; d.h.:  $MAAS_{\text{UPS}} = (UPS / DSL) / ((PM_{\text{alt}} / PM_{\text{neu}}) - 1) \Rightarrow MAAS_{\text{UPS}} = 40$  [Std] Arbeitsstunden (siehe Abb. 4.2).

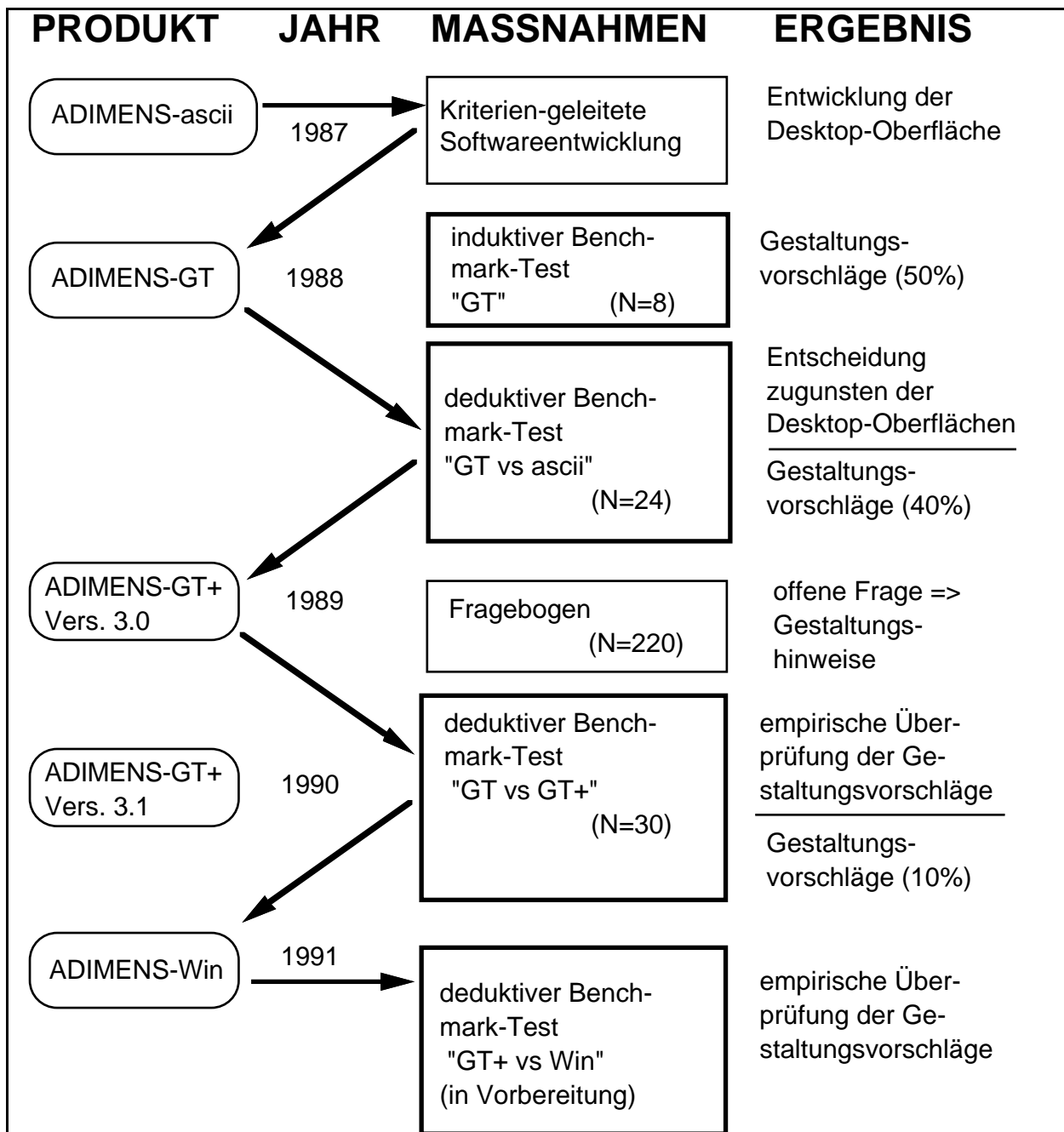


**Abbildung 4.2:** Der Kosten-Nutzen-Graph für die Abschätzung der minimalen Anzahl Arbeitsstunden (MAAS), ab derer sich die investierten Kosten (UPS, bzw. APS) für jeden einzelnen End-Benutzer bei einem durchschnittlichen Stundenlohn (DSL) bezahlt gemacht haben.

Damit nun der Umsteiger nicht mehr investieren muß als der Neueinsteiger ( $MAAS_{UPS} > MAAS_{APS}$ , siehe Abb. 4.2) gilt die folgende Bedingung:  $MAAS_{UPS} = MAAS_{APS}$ ; dies kann erreicht werden, indem der Neupreis  $APS = [1 - (PM_{neu}/PM_{alt})] * UPS$  beträgt. Durch diese Gewichtung des Upgrade-Preises wird erreicht, daß der Neupreis im Verhältnis zu den vom Umsteiger zu investierenden Kosten vergleichbar wird.

## 5. Das Fallbeispiel ADIMENS

Im Rahmen der Benutzerbeteiligung bei der Standardsoftware-Entwicklung des relationalen Datenbank-Programmes ADIMENS wurden Evaluationsstudien (induktive bBTs) und experimentelle Vergleichsstudien mit zwei verschiedenen Oberflächenvarianten (deduktive bBTs) durchgeführt (ADIMENS-ST, ADIMENS-ascii vs. ADIMENS-GT; siehe RAUTERBERG 1989a+b, 1990b). Die konkreten Analyseergebnisse flossen in die Oberflächengestaltung von ADIMENS-GT+ (Version 3.0 und 3.1) und in die Windows-Portierung WIN ein (MOLLENHAUER 1991; siehe auch Abb. 5.1). Die Veränderungen zwischen GT und GT+ wurden bereits experimentell auf ihre ergonomische Relevanz anhand von vier Benchmark-Aufgaben getestet (RAUTERBERG 1991a).



**Abbildung 5.1:** Übersicht über die durchgeführten und geplanten Benchmark-Tests (fett umrandete Maßnahmen) im Rahmen der Standardsoftware-Entwicklung von ADIMENS in den Jahren 1988 bis 1991.

In den empirischen Untersuchungen mit ADIMENS-GT wurde eine Reihe von verbesserungswürdigen Oberflächeneigenschaften gefunden (RAUTERBERG 1988a, 1990b): 1. die Aufteilung der Pull-down Menüs erwies sich als mögliche Quelle von Fehlbedienungen (insbesondere für Anfänger); 2. die Funktionalität der Desktop-Ikonen konnte von Anfängern nicht "entdeckt" werden, weil sie diese in den Pull-down Menü-Optionen vergeblich suchten; 3. die Wirkung einer Reihe von wichtigen Schaltern wurde deshalb oft falsch berücksichtigt, weil ihre Schalterstellung nicht permanent sichtbar waren; 4. die enorme Bedeutung der "Wahl" (permanente Selektion von Datensätzen mittels 'Filter') war durch die sehr häufige Verwendung der Ausgabeform "Anzeigen als

Liste" begründet; hier soll die Möglichkeit zur temporären Selektion vorab Entlastung bringen; dies wird dem Benutzer durch die Implementation des Bearbeitungsmodus "Bearbeiten" ermöglicht; 5. die wichtigste Eigenschaft eines relationalen DBMS - die *Relationalität* - sollte dem Benutzer in einer direkt-manipulativ interaktiven Form durch Verbund-Dateien bei ADIMENS-GT+ zugänglich sein (RAUTERBERG 1991b).

Es wurden, bzw. werden während der Projektlaufzeit des BOSS-Projektes drei vollständige Versions-Entwicklungszyklen durchlaufen und mit benutzungsorientierten Benchmark-Tests begleitet. Ca. 50% dieser Gestaltungsvorschläge -bezogen auf die Oberfläche ADIMENS-WIN - wurden bereits in dem induktiven bBT der ersten Iteration herausgefunden (Aufwand 4-6 Personen-Wochen; siehe RAUTERBERG 1990b). Bemerkenswerterweise werden diese Iterationen jedoch primär durch marketing-strategische Überlegungen getriggert und nicht durch benutzungs-orientierte Optimierungsziele. Die ADI-GmbH hat sich selbst zum Ziel gesetzt, jedes Jahr eine neue Version auf den Markt zu bringen. Die inhaltliche Ausformung dieser Entwicklungen wurde dabei primär auf der Grundlage der Ergebnisse der einzelnen Benchmark-Tests durchgeführt.

## 6. Zusammenfassung

Die Methode der benutzungs-orientierten Benchmark-Tests (bBTs) läßt sich nicht nur zur Gewinnung von Gestaltungsvorschlägen (induktive bBTs), sondern auch zur Entscheidung zwischen Systemalternativen, bzw. zur Überprüfung von getroffenen Designentscheidungen (deduktive bBTs) sinnvoll im Rahmen der partizipativen Entwicklung von Standardsoftware einsetzen. Die primäre Ausrichtung auf den Projekttyp-C der "Standardsoftware" liegt darin begründet, daß für die Durchführung eines bBTs eine lauffähige Version vorliegen muß. Dies ist natürlich auch bei den Projekttypen-A bis -C nach der Fertigstellung der ersten Version gegeben, sodaß im Versionen-<sup>1</sup> bzw. Lebenszyklus dieser Softwareprodukte auch bBTs einsetzbar sind.

Wenn man die zu investierenden Kosten für die Methode des benutzungs-orientierten Benchmark-Testens mit denjenigen Kosten vergleicht, die sich bei der Benutzung der verbesserten Version ergeben, so wird man feststellen müssen, daß sich der Einsatz von Methoden zur Benutzerbeteiligung außerordentlich schnell auszahlt.

Im folgenden zweiten Teil wird ein deduktiver Benchmark-Test beschrieben, welcher mit den beiden Standardsoftware-Produkten ADIMENS-GT und ADIMENS-GTplus durchgeführt wurde. Da ADIMENS-GTplus auf der Grundlage von den Gestaltungsvorschlägen - gewonnen durch Benchmark-Tests mit ADIMENS-GT - entwickelt wurde, dient der folgende deduktive Benchmark-Test primär der Kontrolle der getroffenen Gestaltungsmaßnahmen.

---

<sup>1</sup> zum Versionen-Konzept siehe FLOYD (1984).

## **TEIL-II: Benutzungs-orientierte Benchmark-Tests - Anwendung auf die Gestaltung der Benutzungsoberfläche von ADIMENS-GT hin zu ADIMENS-GTplus<sup>1</sup>**

### **1. Einleitung**

Da im ersten Teil ausführlich die beiden Methoden "benutzungs-orientierter Benchmark-Tests" allgemein dargestellt worden sind, soll nun in diesem zweiten Teil die Anwendung eines deduktiven benutzungs-orientierten Benchmark-Tests an dem Fallbeispiel der Datenbank-Standardsoftware ADIMENS vorgestellt werden. Die Entwicklung der Benutzungsoberfläche von ADIMENS-ST war im Jahre 1987 abgeschlossen (GEISS, 1987). Die Benutzungsoberfläche von ADIMENS-ST ist eine GEM-Entwicklung für den ATARI-Markt. Im Jahre 1988 wurde dann diese ATARI/GEM-Oberfläche auch auf IBM/ GEM-PC's portiert (Markenname ADIMENS-GT). Die ursprüngliche Benutzungsoberfläche von ADIMENS ist Menü- und Funktionstasten-orientiert (Markenname ADIMENS-2.21; wie z.B. ähnlich bei WORD für MSDOS). Die ersten benutzungs-orientierten Benchmark-Tests wurden mit den Oberflächen von ADIMENS-2.21 und ADIMENS-GT<sup>2</sup> durchgeführt (RAUTERBERG, 1988a, 1989a+b, 1990b).

### **2. Stand der Forschung**

Für den anonymen Kundenkreis des Softwareentwicklungstyps D lassen sich verschiedene Stichprobentechniken zur Auswahl und Zusammenstellung von repräsentativen Benutzerbeteiligungsgruppen einsetzen (BORTZ 1984). Mit Hilfe von inferenzstatistischen Methoden können dann die erarbeiteten Gestaltungsvorschläge auf ihre Generalisierbarkeit hin getestet werden.

Im Rahmen der Benutzerbeteiligung bei Standardsoftware-Entwicklungen wurden deshalb experimentelle Vergleichsstudien mit verschiedenen Oberflächenvarianten (ADIMENS-ST, ADIMENS-2.23 vs. ADIMENS-GT; siehe RAUTERBERG 1989a+b, 1990b) durchgeführt. Die konkreten Analyseergebnisse flossen in die Oberflächengestaltung von ADIMENS-GT+ ein. Diese Veränderungen wurden experimentell an einer speziellen Benutzergruppe auf ihre ergonomische Relevanz anhand von ausgewählten Benchmark-Aufgaben getestet. Die Testung erfolgt zwischen den beiden verkaufsfertigen Produkten: ADIMENS-GT und ADIMENS-GT+ (unter GEM auf MSDOS-PC's).

---

<sup>1</sup> Dieser Teil ist eine überarbeitete und ergänzte Fassung des Beitrags "Benutzungsorientierte Benchmark-Tests als Methode zur Kontrolle von Gestaltungsmaßnahmen bei der Veränderung von Desktop-Oberflächen" in RAUTERBERG & ULLICH (1991).

<sup>2</sup> An dieser Stelle möchten wir uns sehr herzlich für die konstruktive und fruchtbare Zusammenarbeit mit der ADI GmbH (Karlsruhe, BRD), insbesondere mit dem Produktmanager Herrn R. Mollenhauer bedanken.

Folgende Punkte sind bei der Methode der benutzer-orientierten Benchmark-Tests zu beachten: 1. Eigenschaften der zu testenden Produkte; 2. die Auswahl der Benchmark-Aufgaben; 3. die Auswahl der Benutzer.

### **3. Eigenschaften des Produktes, Gestaltungsmassnahmen und Test-Fragen**

Um empirische Tests mit einem Software-Produkt in der hier vorgestellten Form durchführen zu können, muß das Produkt mindestens die beiden folgende Anforderungen erfüllen: 1. volle Funktionsfähigkeit der zu testenden Funktionsbereiche und 2. realistisches System-Anwortzeitverhalten. Es müssen mindestens zwei Produktversionen mit diesen Eigenschaften zur Verfügung stehen.

Es wurden in empirischen Untersuchungen mit ADIMENS-GT eine Reihe von verbesserungswürdigen Situationen gefunden. Die kriterien-orientierte Analyse dieser Problemsituationen ergab die folgenden *Gestaltungsmassnahmen* (RAUTERBERG 1988a, 1989a+b):

#### 1. Transparenz der Pull-Down-Menüs:

Die Aufteilung der Pull-down Menüs erwies sich als mögliche Quelle von Fehlbedienungen. Dies wurde dadurch erklärt, daß die Zuordnung der Systemfunktionen zu einem Pull-down-Menü bei der GT-Oberfläche nicht nach den jeweils betroffenen Desktop-Objekten erfolgt war (zB. alle DB-Operationen und alle Datensatz-Operationen waren in dem Menü "Daten" zusammengefaßt). Die entsprechenden Funktionen wurden nun auf die Menüs "Datei" und "Edit" aufgeteilt (siehe Abb. 11.2 und 11.4). Desweiteren gehört sinnvollerweise die Funktion "globales Ändern" zu dem Menü "Rechnen" und nicht in das Menü "Daten". Diese Gestaltungsmassnahme ist ein Resultat des deduktiven Benchmark-Tests (RAUTERBERG, 1990b).

#### 2. Flexibilität & Transparenz der Ikon-Funktionalität:

Die Funktionalität der Desktop-Ikonen (zB. "Mischen", "Klemmbrett", etc.) konnte von Anfängern nicht "entdeckt" werden, weil sie vergeblich die zugehörigen Repräsentationsformen (ein entsprechender Name als Menü-Item) in den Pull-Down Menü-Optionen suchten. Es wurden daher alle von ADIMENS angebotenen Funktionen auch als visuell wahrnehmbare Menü-Items implementiert. Durch die Bereitstellung der Ikon-Funktionalität zusätzlich über entsprechende Menü-Items wurde zusätzlich zur Transparenz die interaktive Flexibilität der Dialog-Komponente vergrößert. Bei primär graphisch-orientierten Oberflächen besteht generell die Gefahr, daß visuell nicht repräsentierte Systemfunktionen ungenutzt bleiben. Die Erschliessung dieser Funktionalität über das Handbuch oder ein Hilfesystem scheint selten stattzufinden. Diese Gestaltungsmassnahme ist ein Resultat des deduktiven Benchmark-Tests (RAUTERBERG, 1990b).

#### 3. Feedback & Flexibilität:

Die Wirkung einer Reihe von wichtigen Schaltern (zB. "Wahl verwenden") wurde oft deshalb falsch berücksichtigt, weil ihre Schalterstellungen nicht permanent sichtbar waren, sondern erst durch das Herunterklappen des entsprechenden Pull-Down-Menüs (zB. "Wahl"; aber auch insbe-



sondere "Schalter") (siehe Abb. 11.1) wahrgenommen werden konnte. Um dem Benutzer die wichtigsten Schalterstellungen permanent auf dem Desktop zur Verfügung zu stellen, wurde die "Schalttafel" implementiert (siehe Abb. 11.3). Einige der über die Schalttafel rückgemeldeten Schalterstellungen können vom Benutzer auch direkt mittels Maus-Klick verändert werden; dies erhöht deswegen auch noch die interaktive Flexibilität, weil so der Benutzer - zusätzlich zu den Pull-Down-Menüs - einzelne Schalter interaktiv setzen kann. Diese Gestaltungsmassnahme ist ein Resultat des induktiven Benchmark-Tests (RAUTERBERG, 1988a, 1990b).

#### 4. Individuelle Auswahlmöglichkeiten:

Die enorme Bedeutung der "Wahl" (permanente Selektion von Datensätzen mittels Filter) war durch die sehr häufige Verwendung der Ausgabeform "Anzeigen als Liste" begründet; da bei ADIMENS-GT vor der Ausgabe als Liste keine temporäre Selektion von Datensätzen über Schlüsselmerkmale möglich war, diese Ausgabeart aber stark bevorzugt wird, mußten die Benutzer die "Wahl" zu Hilfe nehmen. Hier soll nun die Möglichkeit zur temporären Selektion vorab Entlastung bringen. Dies wird dem Benutzer durch die Implementation des Bearbeitungsmodus "Bearbeiten" ermöglicht. Bei ADIMENS-GT+ können nun die Benutzer gleichberechtigt zwischen den beiden Ausgabearten "als Maske" oder "als Liste" je nach individuellen Bedürfnissen auswählen, ohne interaktive Nachteile<sup>1</sup> in Kauf nehmen zu müssen. Diese Gestaltungsmassnahme ist ein Resultat des deduktiven Benchmark-Tests (RAUTERBERG, 1990b).

#### 5. Aufgaben-Orientierung:

Um die wichtigste Eigenschaft eines relationalen DBMS - die *Relationalität* - dem Benutzer in einer direktmanipulativ-interaktiven Form durch Verbund-Dateien zugänglich zu machen, wurde das Konzept des "aktiven Joins" entwickelt und implementiert (siehe zur näheren Beschreibung bei RAUTERBERG, 1991b). Diese Gestaltungsmassnahme ist ein Resultat des deduktiven Benchmark-Tests (RAUTERBERG, 1990b).

Aufgrund der beschriebenen Problembereiche wurden die beschriebenen Verbesserungsvorschläge abgeleitet. Die neu-gestaltete DBMS-Oberfläche von ADIMENS-GT+ soll nun im Rahmen dieser Studie empirisch auf den Grad der erreichten "Verbesserung" hin getestet werden:

*alte desktop-orientierte DBMS-Oberfläche (Version-GT):* die grafik-orientierte, direkt-manipulative DBMS-Oberfläche, bei der die Dialogführung *mit der Maus* durch Anklicken von maus-sensitiven Bereichen vollzogen wird;

*neue desktop-orientierte DBMS-Oberfläche (Version-GT+):* folgende Veränderungen wurden vorgenommen: die Pull-down Menüs wurden neu strukturiert und erweitert; eine "Schalttafel" wird dem Benutzer auf dem Desktop angeboten; es besteht die Möglichkeit zur temporären "Suche" vor der "Ausgabe als Liste" durch den Bearbeitungsmodus "Bearbeiten"; Relationen lassen sich als Join (Verbund) direkt-manipulativ definieren und interaktiv weiterverarbeiten.

---

<sup>1</sup> Die Verwendung der "Wahl" führt zu einer hohen Systemantwortzeit, weil die jeweilige Datei sequentiell und nicht über den schnellen Index-Zugriff durchsucht wird. Dieser Umstand hat dann auch dazu geführt, daß unter den ADIMENS-Kunden die Forderung nach einer "schnellen Wahl" aufgestellt wurde. Diese "schnelle Wahl" kann deshalb nicht implementiert werden, weil nicht gewährleistet werden kann, daß alle Merkmale eines Datensatzes auch als Schlüsselmerkmale deklariert worden sind, eine "Wahl" aber auf allen Merkmalen operieren muß.

Folgende Test-Fragen sollen beantwortet werden:

- 1.) Gibt es einen ergonomisch relevanten Unterschied in der Bearbeitungszeit zwischen diesen beiden Oberflächen ?
- 2.) Gibt es einen Zusammenhang in der Bearbeitungszeit zwischen der Art des Benchmark-Aufgabe und dem Typ der Oberfläche ?
- 3.) Gibt es einen relevanten Unterschied in der subjektiv erlebten Beanspruchung zwischen diesen beiden Oberflächen ?

#### 4. Test-Design

Als Test-Design wurde ein Meßwiederholungsdesign mit permutierter Reihenfolge für den Faktor "Oberfläche" gewählt, um Reihenfolge- und damit Lerneffekte bei der wiederholten Testdurchführung kontrollieren zu können. Bei dem hier vorliegenden Test-Design ergibt sich somit für die Auswertung der Testergebnisse eine drei-faktorielle Varianzanalyse mit Meßwiederholung auf zwei Faktoren (within subjects) und dem Faktor "Reihenfolge" (between subjects):

der 1. *Faktor* ist der "Typ der Benutzungsoberfläche" ("GT" versus "GT+" auf IBM-PCs unter GEM), den 2. *Faktor* bilden die vier "Benchmark-Aufgaben" und den 3. *Faktor* die "Reihenfolge der benutzten Oberfläche". Als *abhängige Variablen* wurden gemessen: die reinen Bearbeitungszeiten gemäß Logfile-Protokoll (bereinigt von den System-Antwortzeiten; USERTIME) pro Aufgabe. Als *intervenierende Variablen* wurden gemessen: die durchschnittliche Entscheidungszeit pro Tastendruck pro Aufgabe (DECISIONTIME)<sup>1</sup>, die Anzahl Tastendrucke insgesamt pro Aufgabe (KEYSTROKES) und die subjektiv erlebte Beanspruchung pro Aufgabenserie (EZ-Skala; APENBURG 1986).

Zwischen den Variablen USERTIME, KEYSTROKES und DECISIONTIME besteht folgende funktionale Abhängigkeit: (USERTIME = KEYSTROKES \* DECISIONTIME).

Um die empirisch vorgefundenen Benutzungsunterschiede auf die Gesamtheit eines anonymen Kundenkreises generalisieren zu können, wird das inferenzstatistische Verfahren der *Varianzanalyse* aus dem Bereich der angewandten Statistik eingesetzt (BORTZ 1984, S. 407ff).

##### 4.1. Auswahl der TestpartnerInnen

Da die Bearbeitung der in diesem Test ausgewählten Benchmark-Aufgaben durch TestpartnerInnen ein Minimum an Systemkenntnissen voraussetzt, muß dieses Wissen vor Beginn der Testung vermittelt werden. Um den zeitlichen Aufwand für diese Instruktionsphase zu minimieren, sind TestpartnerInnen mit EDV-Vorkenntnissen am besten geeignet. Es nahmen daher 31 InformatikstudentInnen aufgeteilt in zwei Gruppen (Faktor "Reihenfolge") an dieser Studie teil. Diese 31

---

<sup>1</sup> Für weitere Erläuterungen der durchschnittlichen Entscheidungszeit pro Tastendruck siehe ACKERMANN & GREUTMANN (1987).

TestpartnerInnen zeichnen sich dadurch aus, daß sie in ihrer täglichen Arbeit schon seit mehreren Semestern (3. bis 9. Semester) mit EDV-Systemen Vorerfahrungen gesammelt haben.

**Gruppe-1 (GT->GT+):**

durchschnittlich 24 ( $\pm 2$ ) Jahre; 14 Männer, 2 Frauen (N=16); 6. ( $\pm 2$ ) Semester; 4.138 ( $\pm 4.022$ ; Range: 450-15.745) Stunden allgemeine EDV-Vorerfahrung, davon 338 ( $\pm 515$ ; Range: 10-1.500) Stunden Vorerfahrung mit DBM-Systemen ohne ADIMENS und 3 ( $\pm 9$ ; Range: 0-35) Stunden mit ADIMENS.

**Gruppe-2 (GT+->GT):**

durchschnittlich 24 ( $\pm 3$ ) Jahre; 14 Männer, 1 Frau (N=15); 5. ( $\pm 2$ ) Semester; 3.706 ( $\pm 4.562$ ; Range: 705-15.535) Stunden allgemeine EDV-Vorerfahrung, davon 67 ( $\pm 119$ ; Range: 0-350) Stunden Vorerfahrung mit DBM-Systemen ohne ADIMENS und 7 ( $\pm 21$ ; Range: 0-75) Stunden mit ADIMENS.

Die Unterschiede hinsichtlich Alter, Semesterzahl und allgemeiner EDV-Vorerfahrung zwischen den beiden Gruppen sind *nicht* signifikant (T-Test, df=30, zweiseitig,  $p>0.05$ ).

## **4.2. Ablauf des Benchmark-Tests**

Die TestpartnerInnen wurden hinsichtlich ihrer Vorerfahrung zur Kontrolle dieser intervenierenden Variable mit einem Vorerfahrungsbogens befragt. Dann erhielten sie eine standardisierte Einweisung in die Handhabung der jeweiligen Desktop-Oberfläche (ca. 1 1/2 Stunden). Direkt vor und nach der ersten Aufgabenserie füllten alle TestpartnerInnen die Eigenzustandsskala von NITSCH (EZ-Skala) aus. Nach der ersten Aufgabenserie wurde die jeweilige Oberfläche mit einem "Handhabungs"-Fragebogen bewertet. Dann erhielten die TestpartnerInnen eine kurze Einweisung in die jeweils alternative Oberfläche. Direkt vor und nach der zweiten Aufgabenserie wurde wiederum die EZ-Skala vorgegeben. Zum Schluß wurde den TestpartnerInnen ein Gesamtbewertungsbogen vorgegeben. Jede TestpartnerIn erhielt 50,-SFR für die Teilnahme. Die gesamte Testdauer pro TestpartnerIn betrug ca. vier Stunden (Einzelsitzungen). Die Reihenfolge der gestellten Aufgaben war für alle gleich. Für die beiden Aufgabenserien gab es zwei Parallel-Versionen mit jeweils vier Benchmark-Aufgaben. Erst wenn eine Aufgabe vollständig bearbeitet worden war, durften die TestpartnerInnen weiterarbeiten. Um einen Testleiter-Effekt weitgehend auszuschalten, wurde ein Testleiter ausgewählt, der in die Entwicklung der GTplus-Oberfläche *nicht* einbezogen war<sup>1</sup>.

## **4.3. Auswahl und Beschreibung der Benchmark-Aufgaben**

Vier Benchmark-Aufgaben wurden so ausgewählt, daß ein jeweils spezifischer Test der vorgenommenen Gestaltungsmaßnahmen durchgeführt werden kann. Aufgabe-1 testet primär die neue Aufteilung der Menüstruktur, Aufgabe-2 den Bearbeitungsmodus "Bearbeiten" mit der Möglichkeit zur temporären Selektion vor Ausgabe der Datensätze in einem Fenster, Aufgabe-3 die Transparenz und die Möglichkeit zur direkten Interaktion mit der "Schalttafel" (siehe das Problem der

---

<sup>1</sup> Wir möchten an dieser Stelle Herrn Dipl. El. Ing. Christian Rolfsen für die Durchführung der Benchmark-Tests sehr herzlich danken.

"Wahl-Falle", RAUTERBERG 1988a, 1990b); Aufgabe-4 testet die direkten Interaktionsmöglichkeiten mit der Darstellung des Inhaltes einer Verbund-Datei als Liste in einem Fenster.

Als Test-Datenbank diente eine Datenbank bestehend aus drei Dateien (PLATZ: 17 Datensätze, ADRESSEN: 280 Datensätze, GRUPPE: 27 Datensätze; sowie bei der GT+-Version die Verbund-Datei "PlatzGruppe") zur 'Verwaltung eines fiktiven Campingplatzes'.

Die Aufgabenserie für die ADIMENS-GT Oberfläche ist wie folgt gegeben:

Aufgabe 1: "Bitte stellen Sie fest, wie viele Datensätze in der Datei 'PLATZ' vorhanden sind."

Aufgabe 2: "Bitte geben Sie in einem einzigen Fenster mindestens alle diejenigen Datensätze der Datei 'ADRESSEN' als Liste auf dem Bildschirm aus, bei denen im Merkmal 'Platz-Nr' eine Platznummer eingetragen ist. In der Liste soll das Merkmal 'Platz-Nr' als erstes Merkmal und 'Land' als zweites Merkmal vorkommen. Korrigieren Sie anschliessend die Merkmalsausprägungen so, dass bei allen Datensätzen mit einer eingetragenen Platznummer, die Platznummer '07' steht."

Aufgabe 3: "Bitte laden Sie die Wahl-Definition 'ANREISE.PIK' für die Datei 'PLATZ' und wenden Sie diese Wahl an, damit alle Fereingäste, die am '02/07/87' angereist sind, ausgewählt werden können. Lassen Sie nur die gefundenen Datensätze als Maske auf dem Bildschirm einzeln ausgeben. Anschliessend lassen Sie die ganze Datei 'PLATZ' vollständig als Liste auf dem Bildschirm ausgeben."

Aufgabe 4: "Erstellen Sie bitte eine Liste aller Gruppenmitglieder mit den Passnummern der zugehörigen Verbindungen, indem Sie für die Datei 'GRUPPE' ein Mischdokument mit dem Texteditor (Parameter: C:\GEMAPPS\EXP\MISCHDOK\VP[Ihre VP-Nr.].DOC) erstellen. Diese Liste soll das Merkmal "Nachname\_Vorname" der Gruppenmitglieder aus der Datei 'GRUPPE' sowie das Merkmal 'Pass-Nr' der zugehörigen Verbindungen aus der Datei 'PLATZ' enthalten. Lassen Sie anschliessend diese Liste durch Verwendung des Mische-Ikons auf dem Drucker ausdrucken."

Um die zu erwartenden Lerneffekte durch die zweimalige Vorgabe dieser ersten Aufgabenserie zu minimieren, wurde eine parallele, zweite Aufgabenserie entwickelt. Da beide Oberflächen (GT und GT+) in unterschiedlicher Reihenfolge getestet werden, wird die parallele Aufgabenserie auf die ADIMENS-GT+ Oberfläche abgestimmt:

Aufgabe 1': "Bitte stellen Sie fest, wie viele Datensätze in der Datei "GRUPPE" vorhanden sind."

Aufgabe 2': "Bitte geben Sie in einem einzigen Fenster mindestens alle diejenigen Datensätze der Datei "ADRESSEN" als Liste auf dem Bildschirm aus, bei denen im Merkmal "Land" eine Nationsangabe eingetragen ist. In der Liste soll das Merkmal "Land" als erstes Merkmal und "Platz-Nr" als zweites Merkmal vorkommen. Korrigieren Sie anschliessend die Merkmalsausprägungen so, dass bei allen Datensätzen die "D.." im Merkmal "Land" haben, stattdessen "BRD" eingetragen wird."

Aufgabe 3': "Bitte laden Sie die Wahl-Definition "TRAINER.PIK" für die Datei "GRUPPE" und wenden Sie diese Wahl an, damit nur die Gruppenmitglieder, die Siegfried Trainer als "Verbindung" haben, ausgewählt werden können. Lassen Sie nur die gefundenen Datensätze als Maske einzeln auf dem Bildschirm ausgeben. Anschliessend lassen Sie die ganze Datei "GRUPPE" vollständig als Liste auf dem Bildschirm ausgeben."

Aufgabe 4': "Erstellen Sie bitte eine Liste aller Gruppenmitglieder mit den Passnummern der zugehörigen Verbindungen, indem Sie die Verbunddatei "PlatzGruppe" verwenden. Diese Liste soll das Merkmal "Nachname\_Vorname" der Gruppenmitglieder, sowie das Merkmal "Pass-Nr" der zugehörigen Verbindungen enthalten. Lassen Sie anschliessend diese Liste auf dem Drucker ausdrucken."

Alle TestpartnerInnen arbeiten mit beiden Aufgabenserien, jedoch je nach Gruppenzugehörigkeit in unterschiedlicher Reihenfolge.

## 5. Darstellung der Ergebnisse zur Bearbeitungszeit

Um zwei sehr wichtige Einflußgrößen auf die Testergebnisse kontrollieren zu können, werden zwei zusätzliche drei-faktorielle Varianzanalysen gerechnet, bei den jeweils eine der beiden intervenierenden Variablen "Anzahl Tastendrucke (KEYSTROKES)", bzw. "durchschnittliche Entscheidungszeit pro Tastendruck (DECISIONTIME)" als Co-Variate in das Modell mit einbezogen wird (Tab. 5.1). Hierdurch wird gewährleistet, daß die gefundenen Effekte für den Faktor "Oberfläche" im Bezug auf die abhängige Variable (USERTIME) *unabhängig* von diesen beiden Co-Variaten getestet werden.

Es sind in allen drei Varianzanalysen die Haupteffekte für den Faktor "Oberfläche" signifikant (Tab. 5.1). Zusätzlich ergibt sich in allen drei Varianzanalysen ein signifikanter Haupteffekte für den Faktor "Aufgabe(1-4)". Dieser Haupteffekt ist durch die bewußt unterschiedlich gestalteten Aufgaben bedingt und kann daher als Bestätigung dieser intendierten Aufgabengestaltung angesehen werden. Wichtig für das gewählte Test-Design ist der *nicht* signifikante Haupteffekt des Faktors "Reihenfolge". Hierdurch ist gewährleistet, daß der Haupteffekt "Oberfläche" nicht auf Lerneffekten durch die zweimalige Vorgabe der einzelnen Aufgaben beruht. Daß es dennoch spezielle Lern- und Transfereffekte ("carry over effects") gibt, wird durch die signifikante zweifache Wechselwirkung "Oberfläche"  $\otimes$  "Reihenfolge" und die signifikante dreifache Wechselwirkung "Oberfläche"  $\otimes$  "Reihenfolge"  $\otimes$  "Test(1-4)" nahegelegt (Tab. 5.1).

**Tabelle 5.1** Ergebnisse der drei-faktoriellen Varianzanalyse für die Bearbeitungszeiten (Variable USERTIME) für die vier Benchmark-Aufgaben. Die zwei Variablen KEYSTROKES "Anzahl Tastendrucke" und DECISIONTIME "mittlere Entscheidungszeit pro Tastendruck" werden als Co-Variate in zwei weiteren Analysen (3. und 4. Spalte) mit einbezogen.

abhängige Variable: USERTIME	Co-Variate: keine			Co-Variate: "KEYSTROKES"			Co-Variate: "DECISIONTIME"		
Quelle der Variation	df	F	signif.	df	F	signif.	df	F	signif.
"Oberfläche"	1	21.19	.001	1	11.16	.002	1	13.51	.001
"Reihenfolge"	1	0.03	.874	1	2.71	.111	1	0.74	.398
"Aufgabe (1-4)"	3	52.02	.001	3	6.97	.001	3	54.43	.001
"Oberfläche" $\otimes$ "Reihenf."	1	31.20	.001	1	16.21	.001	1	6.48	.017
"Oberfläche" $\otimes$ "Aufg(1-4)"	3	10.56	.001	3	.95	.421	3	10.19	.001
"Reihenfolge" $\otimes$ "Aufg(1-4)"	3	1.52	.215	3	2.36	.077	3	1.39	.252
"Oberfl." $\otimes$ "Reihenf." $\otimes$ "Aufg"	3	5.06	.003	3	3.90	.011	3	5.59	.002

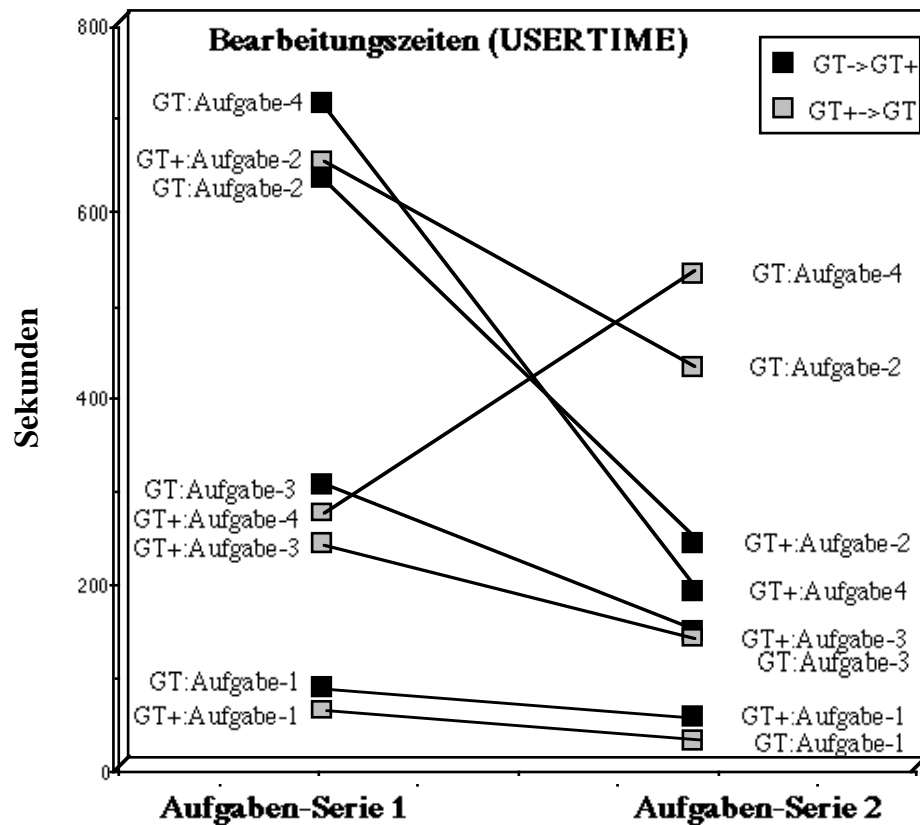
**Tabelle 5.2** Übersicht über die Mittelwerte der Bearbeitungszeiten (USERTIME; Angaben in Sekunden) getrennt nach den verschiedenen Aufgaben. In Klammern hinter den Mittelwerten sind die Standardabweichungen angegeben. Die Anzahl an TestpartnerInnen ist mit N angegeben.

abhängige Variable: USERTIME	<b>Gruppe-1: (GT -&gt; GT+), N = 16</b>		<b>Gruppe-2: (GT+ -&gt; GT), N = 15</b>	
	<b>1.GT</b>	<b>2.GT+</b>	<b>1.GT+</b>	<b>2.GT</b>
<b>Aufgabe-1</b>	71.1 (± 50.7)	41.3 (± 48.1)	68.6 (± 35.8)	34.5 (± 15.3)
<b>Aufgabe-2</b>	638.3 (±386.4)	238.4 (± 85.9)	666.9 (±641.1)	427.8 (±220.0)
<b>Aufgabe-3</b>	298.2 (±280.4)	151.6 (±111.6)	245.5 (± 97.0)	154.2 (± 59.2)
<b>Aufgabe-4</b>	709.0 (±321.2)	209.0 (±169.2)	281.5 (±185.6)	542.3 (±287.9)
<b>Gesamt</b>	429.1 (±383.9)	160.1 (±133.7)	316.4 (±396.0)	289.7 (±272.6)

Der größte Unterschied bei der Bearbeitungszeit (USERTIME) ergibt sich für die Aufgabe-4 (Tab. 5.3). Dieser Unterschied ist im wesentlichen dadurch bedingt, daß bei der GT-Oberfläche unter Verwendung einer einfachen Retrieval-Sprache (siehe ANHANG) ein Listengenerierungsprogramm erstellt werden mußte (größerer interaktiver Aufwand gemessen durch die Variable KEYSTROKES), während bei der GTplus-Oberfläche die Verbunddatei "PlatzGruppe" direkt verwendet werden konnte. Daß dieser höhere interaktive Aufwand bei der GT-Version dennoch nicht für den signifikanten Oberflächeneffekt verantwortlich zu machen ist, läßt sich aus dem signifikanten Haupteffekt des Faktors "Oberfläche" unter Einbezug der Co-Variate "KEYSTROKES" ablesen (Tab. 5.1). Insgesamt benötigten die TestpartnerInnen für die vier Tests mit der GTplus-Oberfläche nur 65% derjenigen Bearbeitungszeit, welche sie mit der GT-Oberfläche aufbringen mußten (Tab. 5.3). Dies entspricht insgesamt einer zeitlichen Verbesserung um 35%.

**Tabelle 5.3** Übersicht über die gewogenen arithmetischen Mittelwerte der Bearbeitungszeiten (USERTIME; Angaben in Sekunden) zusammengefasst über die verschiedenen Aufgaben. Da beide Oberflächen von allen TestpartnerInnen getestet wurden, ist die Anzahl N = 31 für beide Gruppen GT und GT+. In der ganz rechten Spalte ist das prozentuale Verhältnis der beiden Bearbeitungszeiten angegeben.

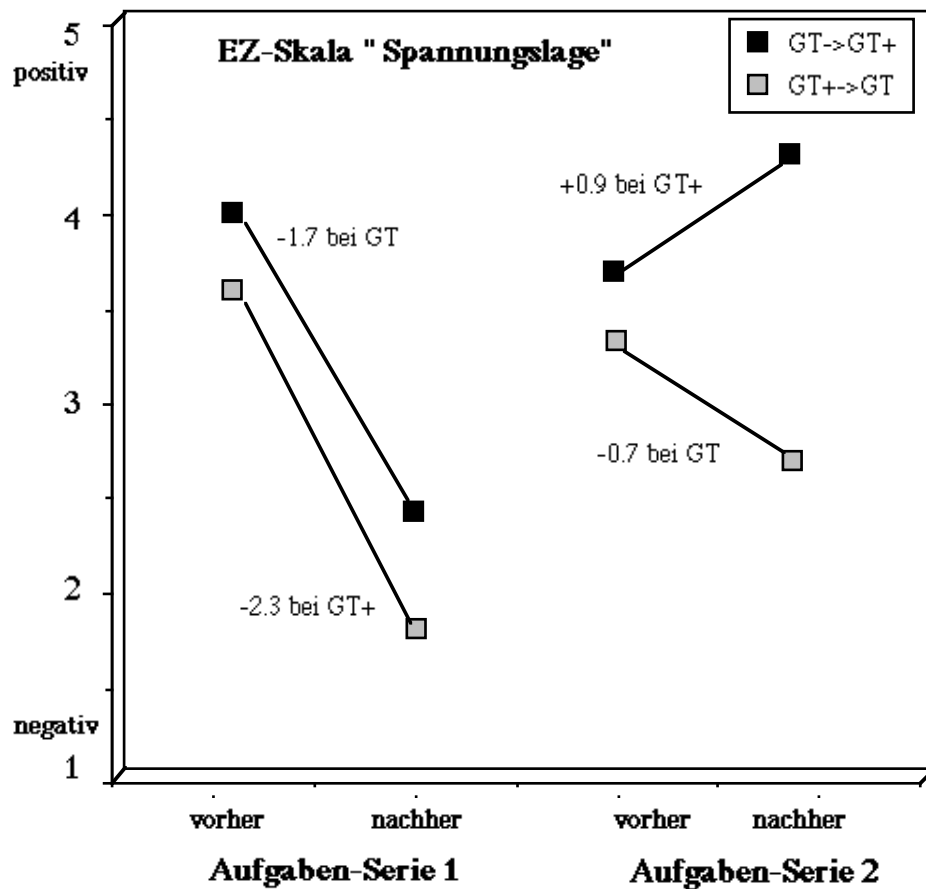
abhängige Variable: USERTIME	<b>Mittelwerte pro Oberfläche</b>		<b>prozentuales Verhältnis</b>
	<b>GT+ (N=31)</b>	<b>GT (N=31)</b>	<b>(GT+/GT)*100%</b>
<b>Aufgabe-1</b>	54.5 (± 44.1)	53.4 (± 41.7)	102%
<b>Aufgabe-2</b>	447.3 (±492.7)	536.4 (±329.7)	83%
<b>Aufgabe-3</b>	197.1 (±113.5)	228.5 (±215.2)	86%
<b>Aufgabe-4</b>	244.1 (±178.2)	628.3 (±312.2)	39%
<b>Gesamt</b>	235.7 (±300.9)	361.7 (±340.6)	65%



**Abbildung 5.1** Die Bearbeitungszeiten für die beiden Gruppen (GT,GT+)-(GT+,GT) aufgliedert nach den vier verschiedenen Benchmark-Aufgaben für die Variable USERTIME ("Bearbeitungszeiten in Sekunden).

## 6. Belastungs- und Beanspruchungs-Parameter

Für die Werte aus der EZ-Skala zur Messung der subjektiv empfundenen Beanspruchungen, welche jeweils für eine komplette Aufgabenserie erhoben wurden, zeigen sich in keiner der acht Skalen bedeutsame Haupteffekte für den Faktor "Oberfläche" (Tab. 6.1). Es zeigen sich jedoch bedeutsame Wechselwirkungen zwischen den Faktoren "Oberfläche"  $\otimes$  "Reihenfolge" für die Skala "Stimmungslage", die Skala "Selbstsicherheit" und die Skala "Spannungslage" (Tab. 6.1). Stellvertretend für diese drei Skalen ist die Skala "Spannungslage" graphisch wiedergegeben (Abb. 6.1).



**Abbildung 6.1** Die subjektiven Beanspruchungswerte der Skala "Spannungslage" gemessen mit dem Eigen-Zustands-Fragebogen nach NITSCH. Je niedriger der Skalenwert ist, desto größer ist die subjektiv empfundene Beanspruchung.

**Tabelle 6.1** Ergebnisse der zwei-faktoriellen Varianzanalyse für die Beanspruchungswerte (EZ-Skala nach NITSCH) für die zwei Aufgabenserien. Es werden nur die Skalen angegeben, bei denen sich signifikante Effekte ergeben haben. Der Faktor "Aufgabenserie" setzt sich zusammen aus den Differenzwerten "vorher-nachher" für die beiden Aufgabenserien.

abhängige Variable: subjektive Beanspruchung	<u>EZ-Skala:</u> "Stimmungslage"			<u>EZ-Skala:</u> "Selbstsicherheit"			<u>EZ-Skala:</u> "Spannungslage"		
	df	F	signif.	df	F	signif.	df	F	signif.
"Oberfläche"	1	1.43	.241	1	0.88	.355	1	0.49	.489
"Aufgabenserie"	1	0.01	.908	1	1.28	.268	1	1.83	.186
"Oberfläche"⊗"Aufgabenserie"	1	4.09	.053	1	5.92	.021	1	7.79	.009

Die Ergebnisse zeigen, daß die Benutzung der GTplus-Oberfläche nicht mit einer Erhöhung von subjektiv erlebten Beanspruchungen einhergeht. Vielmehr ergeben sich bei den EZ-Skalen "Stimmungslage", "Selbstsicherheit" und "Spannungslage" ein entlastender Effekt für die Bedingung, daß die Test-Personen von der GT-Oberfläche zur GTplus-Oberfläche wechseln (signifi-



kante Wechselwirkungen Tab. 6.1). Dies zeigt sehr deutlich, daß durch die vorgenommenen Veränderungen an der GT-Oberfläche nicht nur eine Verkürzung der Aufgabenbearbeitungszeit, sondern auch gleichzeitig damit einhergehend eine nachweisliche Entlastung der subjektiv erlebten Beanspruchung erfolgt ist.

**Tabelle 6.2** Übersicht über die arithmetischen Mittelwerte der Differenzen "nachher-vorher" der acht Skalen des Eigen-Zustands-Fragebogens nach NITSCH; (negative Differenz-Werte bedeuten eine Zunahme der subjektiven Beanspruchung, positive Differenz-Werte bedeuten dagegen eine Abnahme der subjektiven Beanspruchung).

<b>Eigen-Zustands-Fragebogen</b> <b>Differenzwerte</b> <b>"nachher-vorher"</b> <b>(- = neg / + = pos)</b>	<b>Gruppe-1: (GT) -&gt; (GT+),</b> <b>N = 16</b>		<b>Gruppe-2: (GT+) -&gt; (GT),</b> <b>N = 15</b>	
	<b>1.GT</b>	<b>2.GT+</b>	<b>1.GT+</b>	<b>2.GT</b>
Anstrengungsbereit.	0.1 (± 3.4)	-1.4 (± 4.1)	-0.6 (± 4.1)	-2.0 (± 3.6)
Soziale Anerkennung	-0.3 (± 2.0)	-0.1 (± 1.8)	-0.7 (± 1.8)	-0.3 (± 1.8)
Stimmungslage	-1.1 (± 2.3)	0.5 (± 2.0)	-0.4 (± 3.0)	0.0 (± 2.0)
Erholtheit	-0.7 (± 3.3)	0.3 (± 2.9)	-1.3 (± 3.0)	-1.8 (± 2.7)
Kontaktbereitschaft	-0.9 (± 2.5)	0.1 (± 0.9)	-0.1 (± 2.8)	0.2 (± 1.7)
Selbstsicherheit	-0.8 (± 2.8)	0.3 (± 1.3)	-2.0 (± 3.0)	0.4 (± 2.4)
Spannungslage	-1.7 (± 3.9)	0.9 (± 1.6)	-2.3 (± 3.8)	-0.7 (± 2.3)
Schläfrigkeit	-0.4 (± 2.4)	-0.1 (± 1.7)	-0.1 (± 1.6)	0.7 (± 2.3)
<b>Summenwert</b>	<b>-5.8 (± 2.8)</b>	<b>0.7 (± 2.2)</b>	<b>-7.5 (± 2.9)</b>	<b>-3.5 (± 2.4)</b>

## 7. Subjektive Bewertungen der beiden Oberflächen

Die Beurteilungen durch die TestpartnerInnen hinsichtlich der verschiedenen Veränderungen der GTplus-Oberfläche gegenüber der GT-Oberfläche fallen insgesamt sehr positiv aus (Tab. 7.1); die geringste positive Bewertung erhält das Item "Möglichkeit zur direkten Interaktion über die Schalttafel" (Item-Nr. 3; Tab. 7.1) durch die Gruppe-2 (GT+->GT); dieser Unterschied in der Beurteilung zwischen den beiden TestpartnerIn-Gruppen ist signifikant (T-Test, df=30, zwei-seitig).

Auffallend ist, daß die Gruppe-2 "(GT+)->(GT)" insgesamt niedrigere Beurteilungen abgibt als die Gruppe-1.

**Tabelle 7.1** Übersicht über die Mittelwerte des Beurteilungs-Fragebogen (1= unnötig; 2= kaum nötig; 3= weder/noch; 4= nützlich; 5= sehr nützlich) getrennt für die beiden Gruppen (GT->GT+) und (GT+->GT). In Klammern sind die Standardabweichungen angegeben.

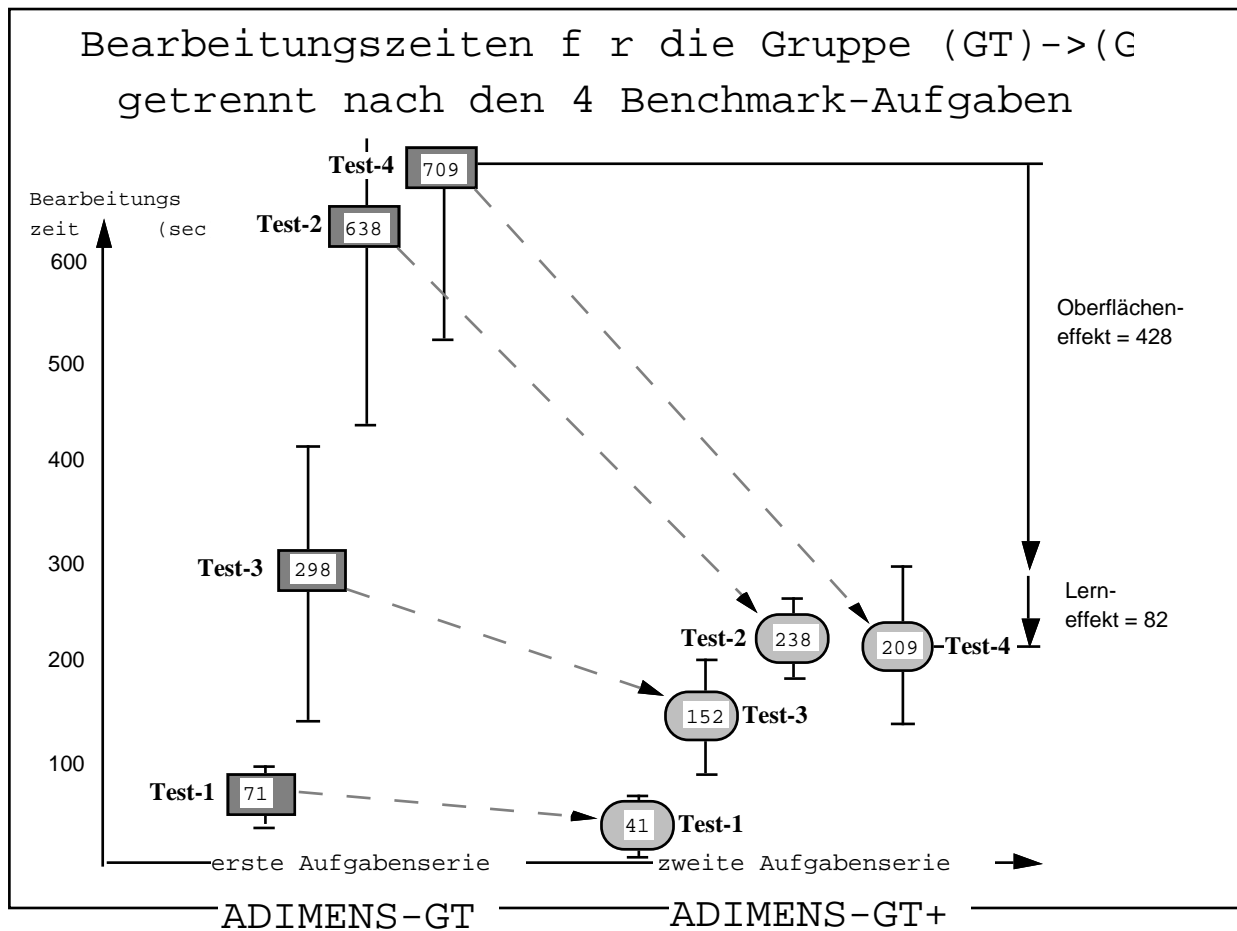
Item Nr.	Fragebogen-Items des Beurteilungsbogens...	Mittelwerte ( $\pm$ Std.)		T-
		(GT->GT+)	(GT+->GT)	Test
1	Die Möglichkeiten, die die Verbunddatei bietet, sind.....	4.75 ( $\pm$ 0.45)	4.60 ( $\pm$ 0.51)	n.s.
2	Durch den Befehl "Bearbeiten" kann eine temporäre Selektion vor der Ausgabe einer Liste gemacht werden. Diese Eigenschaft finde ich.....	4.38 ( $\pm$ 0.96)	4.27 ( $\pm$ 1.03)	n.s.
3	Durch Anklicken von Feldern in der Schalttafel, können verschiedene Schalter aktiviert werden, ohne über die Pull-down-Menüs gehen zu müssen. Diese Eigenschaft ist.....	4.13 ( $\pm$ 1.02)	3.47 ( $\pm$ 1.30)	p<.05
4	Die Schalttafel gibt Auskunft über den aktuellen Zustand der aktivierten Datenbankdatei. Diese Auskunft finde ich.....	4.63 ( $\pm$ 0.81)	4.27 ( $\pm$ 0.70)	n.s.
5	Die Menüleiste ist speziell in die Pull-down-Menüs "Datei", "Edit", "Verbund", etc. aufgeteilt. Die spezielle Art der Aufteilung finde ich....	4.27 ( $\pm$ 0.70)	4.13 ( $\pm$ 0.74)	n.s.

## 8. Berechnung des Oberflächen- und des Lern-Effektes

Als Oberflächen-Effekt bezeichnen wir den Unterschied, welcher sich hinsichtlich einer vorgegebenen Messgröße (z.B. Bearbeitungszeit) zwischen zwei verschiedenen Oberflächen ergibt. Um den Oberflächen-Effekt quantitativ anzugeben, wird die Differenz zwischen den jeweiligen Gruppenmittelwerten berechnet. Hierbei ist zu berücksichtigen, daß diese beiden Mittelwerte unabhängig und nicht durch Lern-Effekte beeinflusst worden sind.

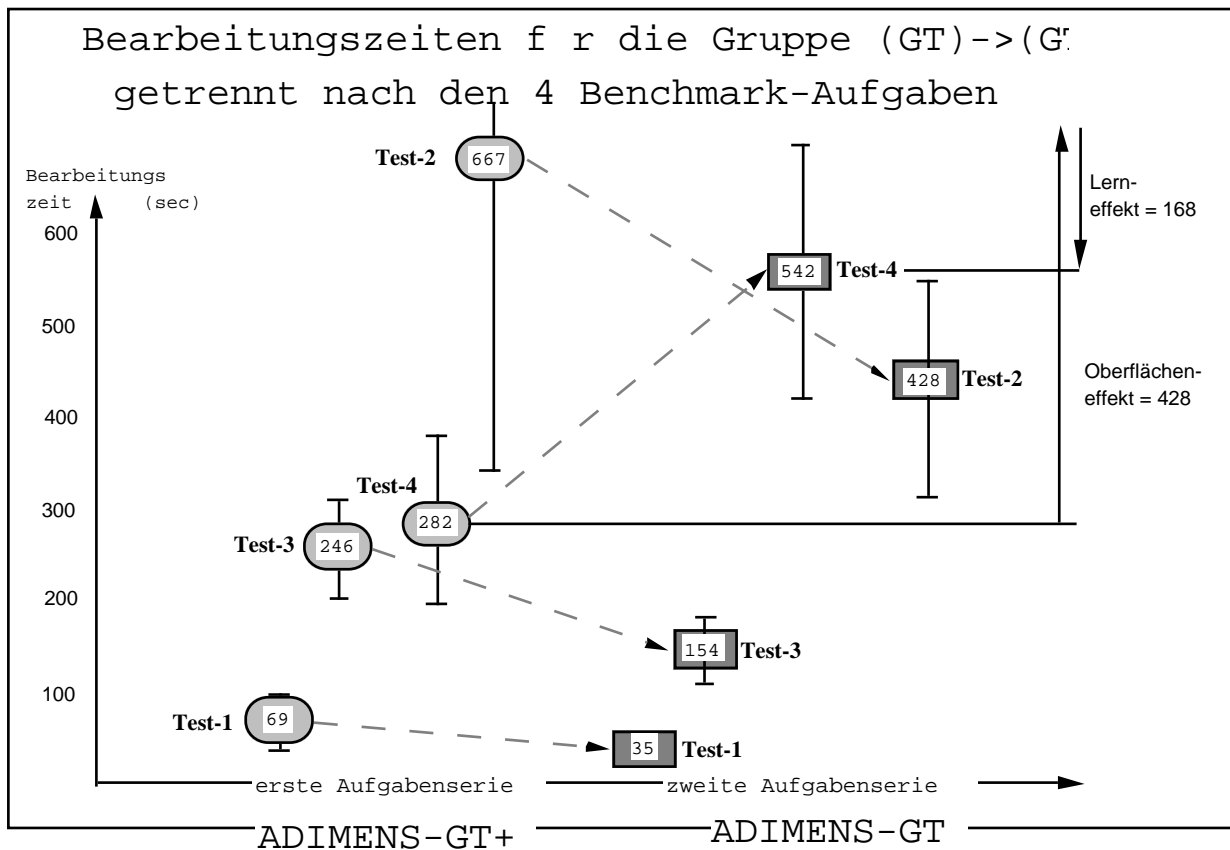
Bei wiederholter Testung einer Benutzungsoberfläche (z.B. mittels parallelisierter Benchmark-Aufgaben) kann man zusätzlich den Lern-Effekt ("carry over effect") abschätzen. Hierzu ist es notwendig, daß die beiden zu vergleichenden Oberflächen in vertauschter Reihenfolge getestet werden. Dieses Test-Design wird "Lateinisches Quadrat" im Rahmen von Meßwiederholungsanalysen genannt (BORTZ 1984, 1989).

An dem Beispiel der Benchmark-Aufgabe-4 wollen wir die einzelnen Berechnungsschritte für die Gruppe (GT)->(GT+) vorführen. Aus der Tabelle 5.2 entnehmen wir den gemessenen Mittelwert für die Oberfläche-GT (1.GT:  $A_4 = 709$  sec.) und den Mittelwert für die Oberfläche-GTplus (1.GT+:  $A'_4 = 282$  sec.). Zunächst berechnen wir den Oberflächen-Effekt (OE), indem wir diese beiden Werten von einander abziehen ( $OE_4 = A_4 - A'_4 = 428$  sec.). Nun können wir den Lern-Effekt (LE) berechnen, indem wir diesmal jedoch die Differenz bezüglich des zweiten Meßzeitpunktes der Gruppe (GT)->(GT+) bestimmen (1.GT:  $A_4 = 709$  sec.; 2.GT+:  $A'_4 = 209$  sec.;  $OE_4 + LE_4 = A_4 - A'_4 = 500$  sec.). Der Lern-Effekt ergibt sich nun durch die Subtraktion des Oberflächen-Effektes aus dieser Differenz ( $LE_4 = 500$  sec. -  $OE_4 = 82$  sec.) (siehe auch Abb. 8.1). Für Aufgabe ergibt sich somit ein Lern-Effekt von 82 sec und ein Oberflächeneffekt von 428 sec.



**Abbildung 8.1** Der Oberflächen- und Lern-Effekt für jede der vier Benchmark-Tests beim Übergang von ADIMENS-GT auf ADIMENS-GT+. Die senkrechten Striche durch die einzelnen Messpunkte symbolisieren die jeweiligen Standardabweichungen.

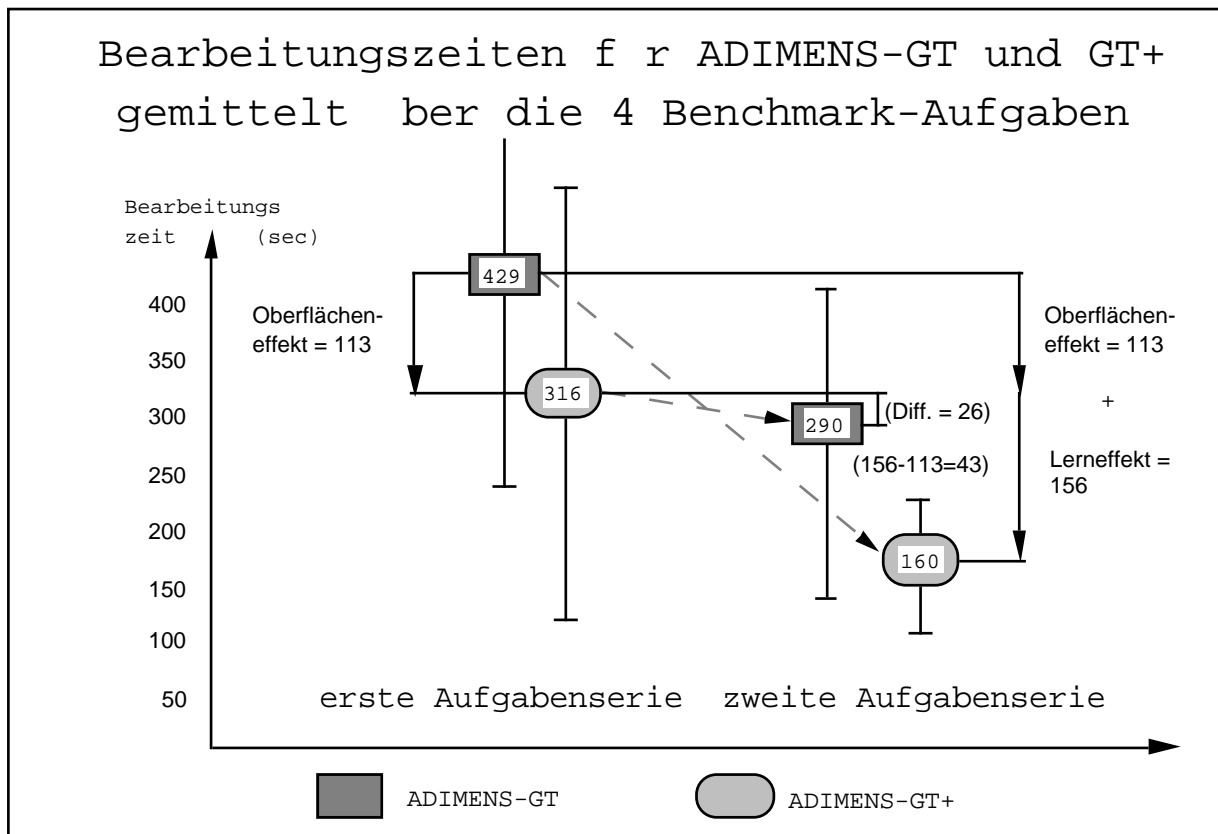
Wenn wir den Lern-Effekte auch für die Gruppe (GT+) -> (GT) errechnen wollen, müssen wir lediglich aus der Differenz (1.GT+:  $A'_4 = 282$  sec.; 2.GT:  $A_4 = 542$  sec;  $OE_4 - LE_4 = A_4 - A'_4 = 260$  sec) den Oberflächen-Effekt subtrahieren ( $LE_4 = OE_4 - 260$  sec = 166 sec). Berechnen wir den Lern-Effekt auch noch für Aufgabe-3 (1.GT+:  $A'_3 = 246$  sec.; 2.GT:  $A_3 = 154$  sec;  $OE_3 - LE_3 = A_3 - A'_3 = -92$  sec), so ergibt sich mit einem Oberflächen-Effekt ( $OE_3 = 52$  sec) folgender Lern-Effekt ( $LE_3 = OE_3 - (-92)$  sec = 144 sec). Man muss lediglich darauf achten, daß die erwarteten Effekte der jeweiligen Oberflächen mit dem entsprechenden Vorzeichen in die Berechnung Eingang finden. So haben wir z.B erwartet, daß die GTplus-Oberfläche eine Verbesserung gegenüber der GT-Oberfläche darstellt und somit der Oberflächen-Effekt beim Übergang von der GT- zur GTplus-Oberfläche zum Lern-Effekt hinzugezählt werden muss (siehe Abb. 8.1), während umgekehrt der Oberflächen-Effekt beim Übergang von der GTplus-Oberfläche zur GT-Oberfläche dem Lern-Effekt entgegenwirkt (siehe Abb. 8.2).



**Abbildung 8.2** Der Oberflächen- und Lern-Effekt für jede der vier Benchmark-Tests beim Übergang von ADIMENS-GT+ auf ADIMENS-GT. Die senkrechten Striche durch die einzelnen Messpunkte symbolisieren die jeweiligen Standardabweichungen.

Da jede der vier Benchmark-Aufgabe einen spezifischen Ausschnitt der beiden Oberflächen testet, ergeben sich auch unterschiedliche Oberflächen-Effekte für jede Aufgabe. Die Lern-Effekte können jedoch für alle Aufgaben annähernd gleich sein; dies kommt immer dann vor, wenn es keine besonderen Transfer-Effekte gibt. Differieren einzelne, aufgaben-spezifische Lern-Effekte deutlich vom Durchschnittswert, so deutet dies auf positive oder negative Transfer-Effekte hin. Negative Transfer-Effekte können dann auftreten, wenn festverankerter Gedächtnisbesitz der TestpartnerIn (z.B. die Funktionstastenbelegung) mit der neuen Oberfläche in Konflikt gerät; dies ist eine spezielle Form von "In-konsistenz".

Der durchschnittliche Lern-Effekt über alle vier Benchmark-Aufgaben hinweg errechnet sich über die globalen Mittelwerte (Tab. 5.2; unterste Zeile) für die jeweiligen Oberflächen in Abhängigkeit von ihrem Testzeitpunkt und beträgt:  $LE_{\text{mean}} = 156 \text{ sec}$ , wenn man einen durchschnittlichen Oberflächen-Effekt  $OE_{\text{mean}} = 113 \text{ sec}$  zugrunde legt (siehe Abb. 8.3). Zur Kontrolle kann man nun die empirisch gefundene Differenz zwischen der GTplus- und der GT-Oberfläche (1.GT+:  $A'_{\text{mean}} = 316 \text{ sec}$ ; 2.GT:  $A_{\text{mean}} = 290 \text{ sec}$ ;  $A'_{\text{mean}} - A_{\text{mean}} = 26 \text{ sec}$ ) mit dem erwarteten Wert ( $OE_{\text{mean}} - LE_{\text{mean}} = 113 \text{ sec} - 156 \text{ sec} = -43 \text{ sec}$ ) vergleichen. Da beide Werte in der gleichen Größenordnung liegen, sehen wir dies als eine Bestätigung für die Berechenbarkeit des Oberflächen- und Lern-Effektes an.



**Abbildung 8.3** Der durchschnittliche Oberflächen- und Lern-Effekt (gemittelt über die vier Benchmark-Aufgaben) beim Übergang von ADIMENS-GT auf ADIMENS-GT+ und vice versa. Die senkrechten Striche durch die einzelnen Messpunkte symbolisieren die jeweiligen Standardabweichungen.

## 9. Kosten-Nutzen-Analyse

Zum Schluß möchten wir noch eine konservative<sup>1</sup> Kosten-Nutzen-Analyse für den hier vorgestellten benutzungsorientierten Benchmark-Tests durchführen. Die investierten Gesamtkosten für die Durchführung des hier im Teil-II beschriebenen deduktiven bBTs belaufen sich auf insgesamt 65.000,- DM. Die einzelnen Positionen sind aus Tabelle 9.1 ersichtlich.

Der erreichte Verbesserungs-Grad der ADIMENS-GT+ Oberfläche gegenüber der ADIMENS-GT Oberfläche beträgt 35% (siehe Kap. 5 Teil-II). Die End-Benutzer sind mit der verbesserten Benutzungsoberfläche ungefähr 1.5 mal schneller als mit der alten Oberfläche. Bei ca. 35.000 End-Benutzern mit vollgraphischen Oberflächen (MOLLENHAUER, 1991, S. 354) ergibt sich bei einem durchschnittlichen Stundenlohn von 15,- DM und einem Verbesserungs-Grad von 35% eine minimale Anzahl Arbeitsstunden (MAAS) von 0,35 [Std]; wenn also alle 35.000 End-Benutzer nur 21 Minuten mit der neuen Version arbeiten würden, haben sich die Investitionskosten in diesen bBT insgesamt, volkswirtschaftlich gesehen ausgezahlt.

<sup>1</sup> "konservativ" deshalb, weil nicht die tatsächlichen Kosten berücksichtigt werden, sondern obere Abschätzungen.

**Tabelle 9.1** Abschätzung der investierten Kosten zur Durchführung des deduktiven bBTs "ADIMENS GT vs. GT+".

Posten	Kosten im einzelnen		Kosten
	Gehälter, etc [DM]	Gewichtung	insgesamt
Testleiter-1	100.000,--	3 Monate	25.000,--
Testleiter-2	60.000,--	6 Monate	30.000,--
Video-Anlage	10.000,--	30 %	3.000,--
Test-Gelder	60,--	* 30	1.800,--
Sekretariat, etc.	5.200,--	* 1	5.200,--
<b>Gesamt</b>			65.000,--

Für den Einsteiger hat sich der Anschaffungspreis (z.B. APS = 800,-DM) bei einem durchschnittlichen Stundenlohn von 15,- DM nach 35 Arbeitsstunden (= MAASAPS) ausgezahlt. Der Umsteiger muß bei einem Upgrade-Preis von ca. 350,- DM für ADIMENS-GT+ höchstens MAASUPS = 47 Stunden arbeiten, um in die Gewinnzone [ $EBZ_{neu} > EBZ_{alt}$ ] hineinzugelangen (siehe Abb. 4.2 in Teil I). In einem Jahr mit einer durchschnittlichen Anzahl Arbeitsstunden von 600 [Std] könnte jeder Umsteiger bei einem Verbesserungs-Grad von 35% mindestens 4.150,-[DM] ( $EBK_{neu} = 13.500,-DM$  und  $EBK_{alt} = 9.000,-DM + 350,-DM$ ) mit der neuen Version ADIMENS-GT+ sparen. Dies entspricht in etwa einem Monatsgehalt.

## 10. Diskussion

Die durchgeführten Benchmark-Tests zeigen, daß sich ein ergonomisch relevanter Unterschied gemessen über die Bearbeitungszeit für die neu konzipierte Oberfläche von ADIMENS-GTplus ergibt (Frage-1, s.o.). Dieser Zeitunterschied in der Bearbeitungszeit bleibt auch bei Berücksichtigung des interaktiven Aufwandes - gemessen über die Anzahl Tastendrucke - bestehen; dieser Unterschied bleibt ebenso bestehen, wenn man die durchschnittliche Entscheidungszeit pro Tastendruck mit in die inferenzstatistische Analyse einbezieht (Tabelle 5.1). Zieht man die Bewertung der TestpartnerInnen zum Fragebogenitem-3 mit in Betracht (Tabelle 7.1), so scheint der Zeitunterschied primär auf einer verbesserten Repräsentation des Dialog- und Anwendungszustandes bei der GTplus-Oberfläche zu beruhen (siehe die Kriterien "Transparenz" und "Feedback" ; SPINAS, WAEBER & STROHM 1990).

Durch den unterschiedlichen interaktiven Aufwand insbesondere für die Bearbeitung des 4. Tests ergibt sich ein signifikanter Zusammenhang zwischen dem Faktor "Aufgabe" und dem Faktor "Oberfläche" (Frage-2, s.o.). Dieser Zusammenhang verschwindet dann auch, sobald der interaktive Aufwand varianzanalytisch kontrolliert wird (Tabelle 5.1; 3. Spalte).

Da die subjektiv erlebte Beanspruchung aller TestpartnerInnen bei der Bearbeitung für beide Oberflächen-Bedingungen nicht signifikant unterschiedlich ist, wird davon ausgegangen, daß mit der neu gestalteten Oberfläche keine zusätzliche Belastung für die TestpartnerIn verbunden ist (Frage-

3, s.o.). Die gefundenen signifikanten Wechselwirkungen zwischen den Faktoren "Oberfläche" und "Reihenfolge" zeigen an, daß der Übergang von der alten GT-Oberfläche auf die neue GTplus-Oberfläche mit einer spezifischen Entlastung verbunden ist (Tabelle 6.1 und 6.2; Abbildung 6.1).

Die TestpartnerInnen beurteilten weitgehend einhellig die Oberflächen-Unterschiede zwischen der alten GT-Oberfläche und der neuen GTplus-Oberfläche positiv zugunsten der GTplus-Oberfläche. Es spiegelt sich somit die Verbesserung - gemessen mit dem objektiven Maß der "Bearbeitungszeit" - auch in der persönlichen Sicht der TestpartnerIn wider.

Insgesamt gesehen, kann die hier vorgestellte *Methode zur Benutzerbeteiligung* bei Standardsoftware-Entwicklungen als ausreichend praktikabel und sinnvoll angesehen werden, welche die geplanten und/oder vorgenommenen Gestaltungsmaßnahmen auf ihre benutzer-orientierte Brauchbarkeit hin überprüfen hilft.

## 11. Literatur

- ACKERMANN D. & GREUTMANN T. (1987): Interaktionsgrammatik und kognitiver Aufwand. in: German Chapter of the ACM Bericht 29: Software-Ergonomie'87; (SCHÖNPFLUG W. & WITTSTOCK M.; Hrsg.); Stuttgart: Teubner; 262-270.
- ACKERMANN D. (1988): Empirie des Softwareentwurfes: Richtlinien und Methoden. in: Einführung in die Software-Ergonomie. Mensch-Computer Kommunikation - Grundwissen 1. (BALZERT H. et al.; Hrsg.), Berlin New York: Walter de Gruyter; 253-276.
- APENBURG E. (1986): Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung. Zeitschrift für Arbeits- und Organisationspsychologie, 30 (N.F. 4); 3-14.
- BEWLEY W. L., ROBERTS T. L., SCHROIT D. & VERPLANK W.L. (1983): Human Factors Testing in the Design of XEROX's 8010 "Star" Office Workstation. in: Proceedings CHI'83: Human Factors in Computing Systems. ACM, New York, pp. 72-77.
- BOEHM B.W. (1981): Software Engineering Economics. Englewood Cliffs, NJ: Prentice-Hall.
- BOOTH P. (1990): An Introduction to Human-Computer Interaction. Hove London Hillsdale: Lawrence Erlbaum.
- BORTZ J. (1984): Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin New York: Springer.
- BORTZ J. (1989): Lehrbuch der Statistik. Berlin Heidelberg New York: Springer.
- CAMPBELL D.T. & STANLEY J.C. (1970): Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung. in: Handbuch der empirischen Unterrichtsforschung I. (INGENKAMP K.H.; Hrsg.), Weinheim: Beltz.
- DZIDA W. (1990): Ergonomische Normenkonformität. in: 5. Herbstschule 'Software-Ergonomie' in Zürich, 18.-21. Spt. 1990, (GI Deutsche Informatik-Akademie, Ahrstr. 45, D-5300 Bonn 2); 1-15.
- DUTKE S. (1988): Lernvorgänge bei der Bedienung eines Textkommunikationssystems. Frankfurt am Main: Peter Lang.
- EKSTROM R.B., FRENCH J.W. & HARMAN H.H. (1976): Manual for Kit of Factor-Referenced Cognitive Tests. Princeton, N.J. ETS.
- FLOYD C (1984): A Systematic Look at Prototyping. in: Approaches to Prototyping. (BUDE R, KUHLENKAMP K, MATHIASSEN L, ZÜLLIGHOVEN H ; eds.) Springer, Berlin, 1-15
- GEISS D. (1987): Ein endbenutzerorientiertes Datenbanksystem auf der grafischen Benutzerschnittstelle GEM. unveröffentlichte Diplomarbeit. Karlsruhe: Universität Karlsruhe - Fachbereich Informatik.
- GOMOLL K. (1990): Some Techniques for Observing Users. in: The Art of Human-Computer Interface Design. (LAUREL B.; ed.) Reading, Mass.: Addison-Wesley; 85-90.

- GREUTMANN T. & ACKERMANN D. (1987): Individual differences in human-computer interaction: how can we measure if the dialog grammar fits the user's needs? in: Human-Computer Interaction: INTERACT'87; (BULLINGER H.-J. & SHACKEL B.; eds.); Amsterdam: North-Holland; 145-149.
- HUNZIKER C. & HÄSSIG S. (1990): ADIMENS-Umfrage: Untersuchung von Datenbankstrukturen. unveröffentlichte Gruppensemesterarbeit WS'89/90. ETH-Zürich: Institut für Arbeitspsychologie.
- KARAT J.. (1988): Software Evaluation Methodologies. in: Handbook of Human-Computer Interaction. (HELANDER M.; ed.), Amsterdam: Elsevier Science; 891-903.
- KARAT C.-M. (1990): Cost-benefit analysis of iterative usability testing. in: Human-Computer Interaction: INTERACT'90; (DIAPER D. et al.; eds.); Amsterdam: North-Holland; 351-356.
- KIRAKOWSKI J & CORBETT M (1990) Effective Methodology for the Study of HCI. in: Human Factors in Information Technology 5. (BULLINGER H-J & POLSON P G; eds.), Amsterdam: North-Holland.
- KLAS H., KUICH D. & LAES T. (1990): Auswertung der Fragebögen zur Anwendung und Beurteilung von ADIMENS ST/GT. unveröffentlichte Gruppensemesterarbeit WS'89/90. ETH-Zürich: Institut für Arbeitspsychologie.
- LANDAUER T.K. (1988): Research Methods in Human-Computer Interaction. in: Handbook of Human-Computer Interaction. (HELANDER M.; ed.), Amsterdam: Elsevier Science; 905-928.
- MACAULAY L., FOWLER C., KIRBY M. & HUTT A. (1990): USTM: a new approach to requirements specification. Interacting with Computers. vol 2, no 1, 92-118.
- MANTEI M. M. & TEOREY T. J. (1988): Cost/benefit analysis for incorporating human factors in the software lifecycle. Communications of the ACM, April, vol. 31, no. 4, 428-439.
- MOLLENHAUER R. (1991): Benutzerorientierte Software-Entwicklung des voll-grafischen Datenbanksystems Adimens: Projektreflexionen aus der Perspektive des beteiligten Softwarehauses. in: Software für die Arbeit von morgen. Bilanz und Perspektiven anwendungsorientierter Forschung. Ergänzung zum Tagungsband. Herausgegeben für DLR, Projektträger Arbeit und Technik. Krefeld: Vennekel & Partner; 353-360.
- MOLL T. (1987): Über Methoden zur Analyse und Evaluation interaktiver Computersysteme. In: Software-Ergonomie. State of the Art 5. (FÄHNRIK K.P.; Hrsg.), München Wien: Oldenbourg, 179-190.
- MÜLLER-HOLZ auf der Heide B., ASCHERSLEBEN G., HACKER S. & BARTSCH T. (1991): Methoden zur empirischen Bewertung der Benutzerfreundlichkeit von Bürosoftware im Rahmen von Prototyping. in: Software für die Arbeit von morgen. Bilanz und Perspektiven anwendungsorientierter Forschung. (FRESE M.; KARSTEN C.; SKARPELIS C. & ZANG-SCHEUCHER B.; Hrsg.), Berlin Heidelberg New York: Springer; 409-420.
- RAUTERBERG M. (1988a): Untersuchung der Benutzerfreundlichkeit einer desktop-orientierten Benutzungsoberfläche am Beispiel eines relationalen Datenbanksystems. unveröffentlichter Forschungsbericht, ETH-Zürich: Institut für Arbeitspsychologie.
- RAUTERBERG M. (1988b): Video-Rating - a reliable and valide evaluation method for the man-computer-interaction (MCI). in: Proceedings of the Tenth Congress of the International Ergonomics Association, 1-5 August 1988, Sydney, Australia; (ADAMS, A.S.; HALL, R.R.; McPHEE, B.J. & OXENBURGH, M.S.; eds.); Vol. II ; 633-635.
- RAUTERBERG M. (1989a): Ein empirischer Vergleich einer desktop- mit einer menü-orientierten Benutzungsoberfläche für ein relationales DBMS. in: Proceedings der GI-Jahrestagung 1989, Band I; Informatik Fachberichte Nr. 222; (PAUL, M.; Hrsg.); Berlin Heidelberg New York: Springer; 243-258
- RAUTERBERG M. (1989b): MAUS versus FUNKTIONSTASTE: ein empirischer Vergleich einer desktop- mit einer ascii-orientierten Benutzungsoberfläche. in: German Chapter of the ACM Bericht 32: Software-Ergonomie'89. (MAASS S. & OBERQUELLE H.; Hrsg.). Stuttgart: Teubner; 313-323.
- RAUTERBERG M. (1990a): Induktive versus deduktive Untersuchungsmethoden: Vor- und Nachteile für die Gestaltung von Benutzungsoberflächen in der Mensch-Computer-Interaktion. in: 32. Tagung experimentell arbeitender Psychologen Regensburg, April 9-12, 1990. (IRTEL H. & DRÖSLER J.; Hrsg.), Fachbereich Psychologie, Regensburg: Universität Regensburg.
- RAUTERBERG M. (1990b): Experimentelle Untersuchungen zur Gestaltung der Benutzungsoberfläche eines relationalen Datenbanksystems. in: Projektberichte zum Forschungsprojekt Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (BOSS), Nr. 3. (SPINAS P.; RAUTERBERG M.; STROHM O.; WAEBER D. & ULICH E. ; Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie.
- RAUTERBERG M. (1991a): Benutzungs-orientierte Benchmark-Tests als Methode zur Kontrolle von Gestaltungsmassnahmen bei der Veränderung von Desktop-Oberflächen. in: Posterband zur Software-Ergonomie'91. (RAUTERBERG M. & ULICH E. ; Hrsg.). ETH-Zürich: Institut für Arbeitspsychologie.



- RAUTERBERG M. (1991b): From passive to active joins - a task oriented desktop interface conception for a relational DataBaseManagementSystem. in: Proceedings of RIAO'91 "Intelligent Text and Information Handling", held in Barcelona (SP), April 2-5; vol. II, 809-821.
- RAUTERBERG M. & ULICH E. (1991; Hrsg.): Posterband zur Software-Ergonomie'91. ETH-Zürich: Institut für Arbeitspsychologie.
- SEIN M.K., BOSTROM R.P. & OLFMAN L. (1987): Conceptual Models in Training Novice Users. in: Human-Computer Interaction: INTERACT'87; (BULLINGER H.-J. & SHACKEL B.; eds.); Amsterdam: North-Holland; 861-867
- SPENCER R H (1985) Computer usability testing and evaluation. Prentice Hall, Englewood Cliffs
- SPINAS P. (1990): Benutzerfreundlichkeit von Dialogsystemen und Benutzerbeteiligung bei der Software-Entwicklung. in: Das Bild der Arbeit. (FREI F. & UDRIS I.; Hrsg.), Bern: Hans Huber; 158-171.
- SPINAS P., WAEBER D. & STROHM O. (1990): Kriterien benutzerorientierter Dialoggestaltung und partizipative Softwareentwicklung; eine Literaturlaufarbeitung. in: Projektberichte zum Forschungsprojekt Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (BOSS), Nr. 1. (SPINAS P.; RAUTERBERG M.; STROHM O.; WAEBER D. & ULICH E. ;Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie.
- STREITZ N. (1986): Cognitive ergonomics - an approach for the design of user-oriented interactive systems. in: Man-Computer Interaction Research MACINTER-I. (KLIX F. & WANDKE H.; eds.). Amsterdam: Elsevier; 21-33.
- STROHM O. (1990): Arbeitsorganisation, Methodik und Benutzerorientierung bei der Softwareentwicklung. in: Projektberichte zum Forschungsprojekt Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (BOSS), Nr. 2. (SPINAS P.; RAUTERBERG M.; STROHM O.; WAEBER D. & ULICH E. ; Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie.
- STROHM O. & SPINAS P. (1991): Ergebnisse der Analyse von Softwareprojekten. unveröffentlichter Forschungsbericht. ETH-Zürich: Institut für Arbeitspsychologie.
- STUTZ R., STAUBL, A., ROHRBACH C. & KRIESL K. (1990): Auswertung der Umfrage über die ADIMENS-Datenbank. unveröffentlichte Gruppensemesterarbeit SS'90. ETH-Zürich: Institut für Arbeitspsychologie.
- ULICH E. (1985): Einige Anmerkungen zur Software-Psychologie. sysdata. Nr. 10, 53-58.
- ULICH E. (1986): Aspekte der Benutzerfreundlichkeit. in: Arbeitsplätze von morgen. Berichte des German Chapter of the ACM, Band 27. (REMMELE W. & SOMMER M.; Hrsg.), Stuttgart: Teubner, 102-121.
- ULICH E. (1991): Arbeitspsychologie. Stuttgart: Poeschel-Verlag.
- VOSSSEN P. (1991): Rechnerunterstützte Verhaltensprotokollierung und Protokollanalyse. in: Posterband zur Software-Ergonomie'91. (RAUTERBERG M. & ULICH E.; Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie; 181-188.
- WAEBER D. (1990): Entwicklung und Umsetzung von Modellen partizipativer Softwareentwicklung. in: Projektberichte zum Forschungsprojekt Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (BOSS), Nr. 4. (SPINAS P.; RAUTERBERG M.; STROHM O.; WAEBER D. & ULICH E. ; Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie.
- WANDKE H. (1990): Zielkonflikte bei der psychologischen Gestaltung von Mensch-Rechner-Schnittstellen. in: Proceedings of 6th International Symposium on Work Psychology Dresden, March 27-29, 1990. Sektion Arbeitswissenschaften Wissenschaftsbereich Psychologie, Dresden: Technische Universität; 132-138.
- WEISER M. & SHNEIDERMAN B. (1987): Human Factors of Computer Programming. in: Handbook of Human-Computer Interaction. (HELANDER, M.; ed.), Amsterdam: Elsevier Science; 1398-1415.
- WHITESIDE J., BENNETT J. & HOLTZBLATT K. (1988): Usability Engineering: Our Experience and Evolution. in: Handbook of Human-Computer Interaction. (HELANDER M.; ed.), Amsterdam: Elsevier Science; 791-817.
- WILLIGES R.C., WILLIGES B. & ELKERTON J. (1987): Software interface design. in: Handbook of Human Factors. (SALVENDY G.; ed.), New York: Wiley & Sons.
- YAUVERBAUM G.J. & CULPAN O. (1990): Exploring the Dynamics of the End-User Environment: The Impact of Education and Task Differences on Change. Human Relations. vol 43 no 5, pp. 439-454.
- ZAPF D. & FRESE M. (1989): Benutzerfehler im Kontext von Arbeitsaufgabe und Arbeitsorganisation. in: Software-Ergonomie'89. Aufgabenorientierte Systemgestaltung. (MAASS S. & OBERQUELLE H.; Hrsg.), Stuttgart: Teubner; 213-222.

12. Anhang

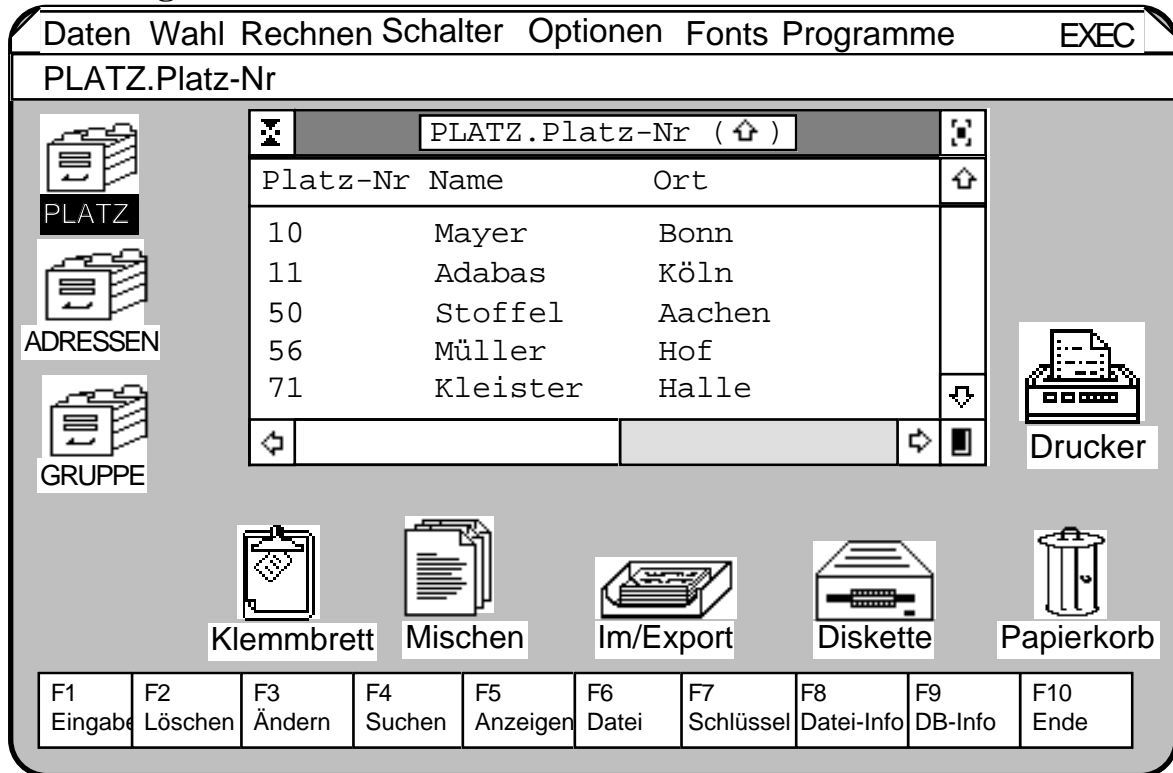


Abbildung 11.1 Die Desktop-Oberfläche von ADIMENS-GT. Der TestpartnerIn steht lediglich die Möglichkeit zur Verfügung, pro Datei ein Fenster mit den einzelnen Datensätzen zu öffnen (Darstellung "als Liste"). Eine "inter"-aktive Liste mit Merkmalen aus mehreren Dateien gleichzeitig ist nicht möglich.

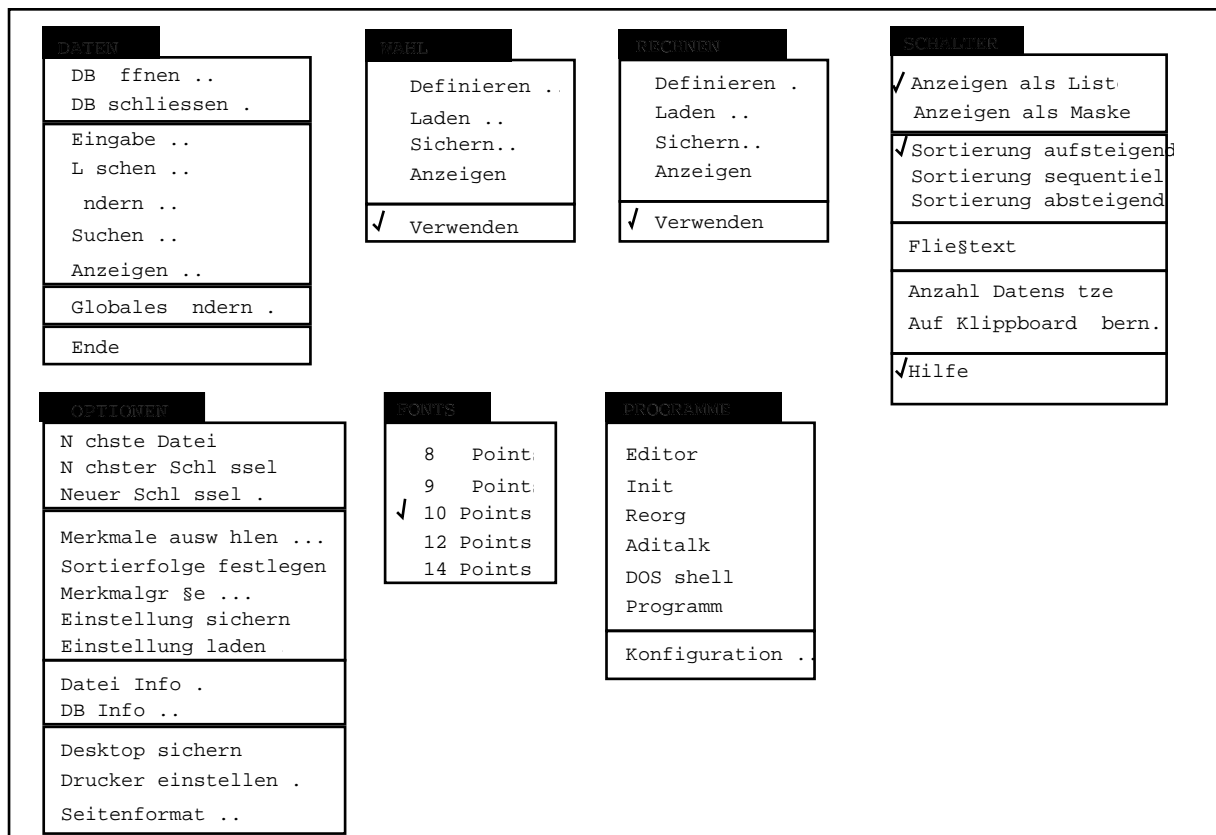
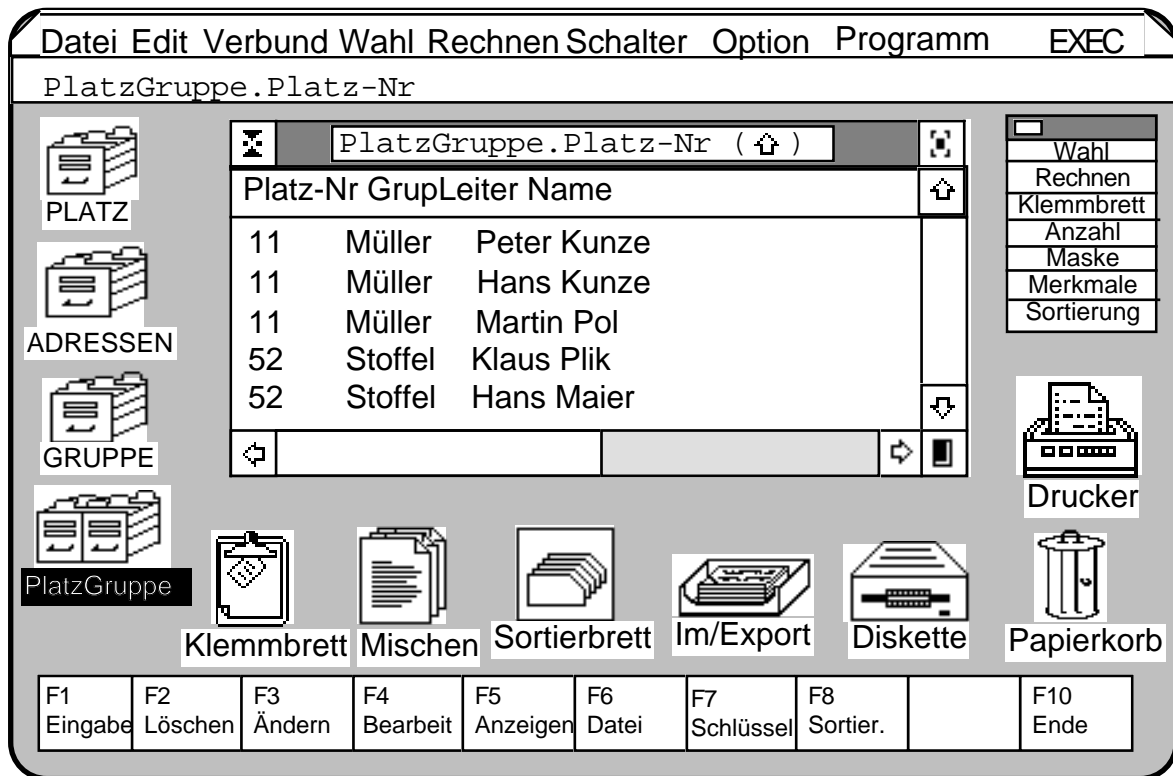
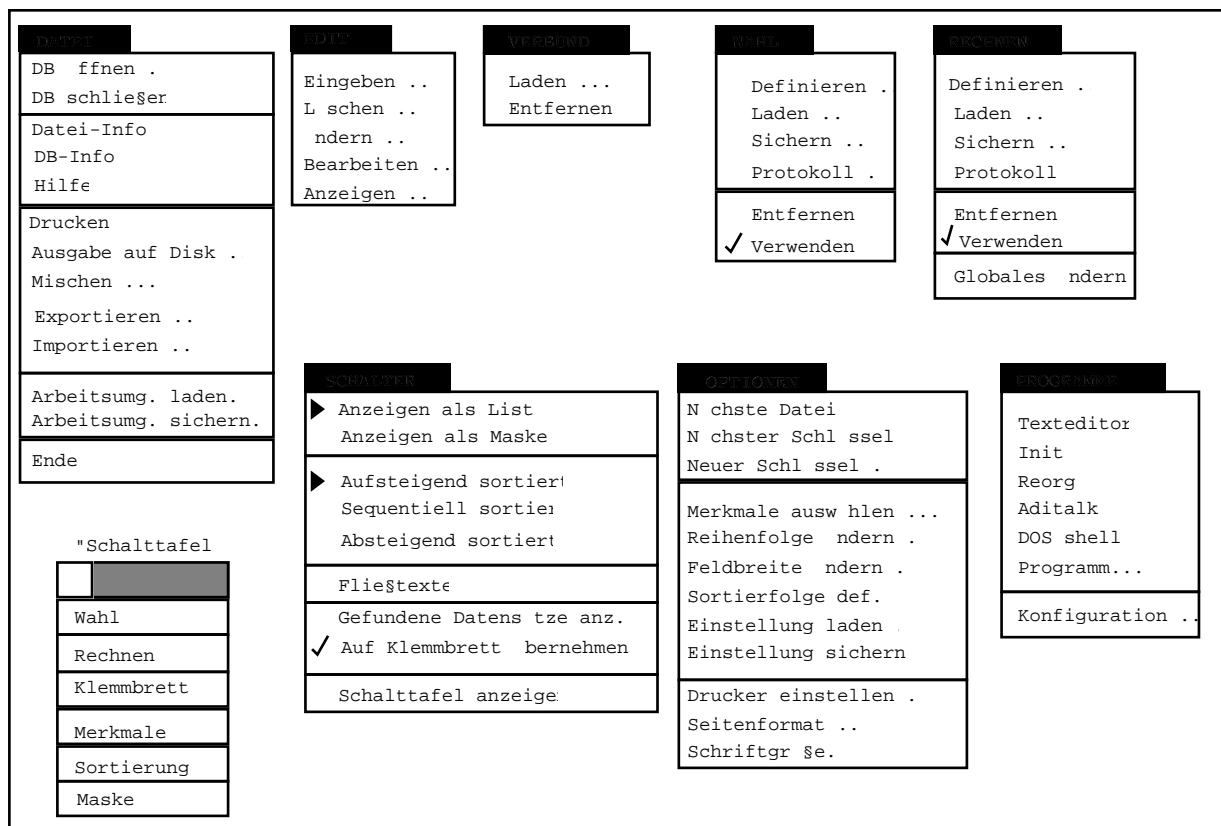


Abbildung 11.2 Die aufgeklappten Pull-down Menüs der Desktop-Oberfläche von ADIMENS-GT.

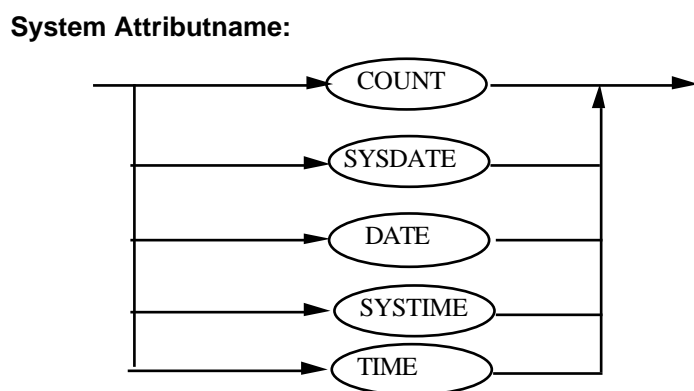
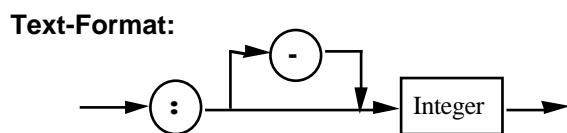
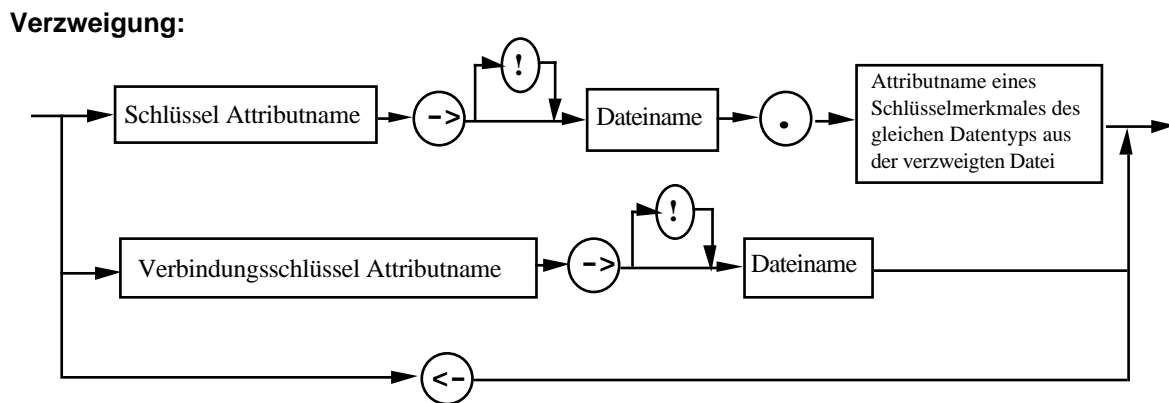
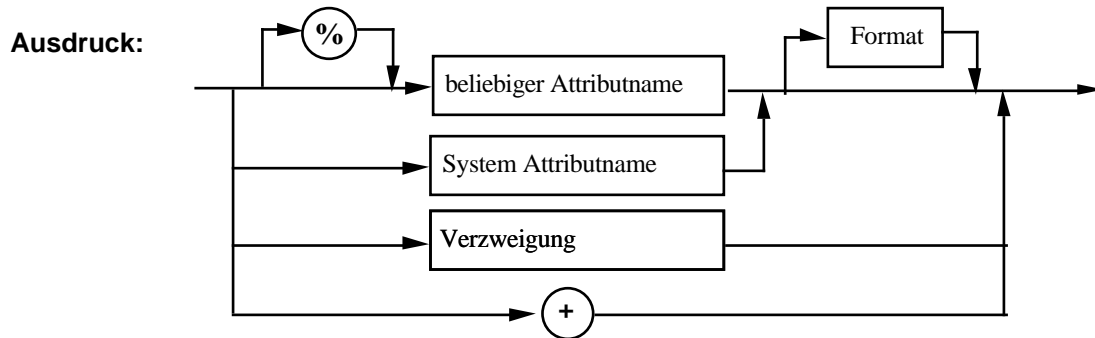
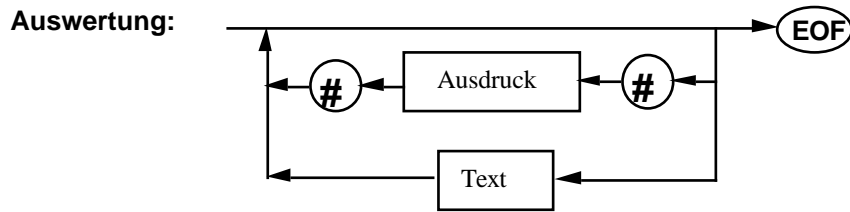


**Abbildung 11.3** Die Desktop-Oberfläche von ADIMENS-GT+. Neu ist: die Aufteilung der Pull-down-Menüs, die Schalttafel rechts oben im Bild, das Sortierbrett, der Dialogmodus "Bearbeiten" und die Verbund-Dateien (z.B. "PlatzGruppe").



**Abbildung 11.4** Die aufgeklappten Pull-down Menüs der Desktop-Oberfläche von ADIMENS-GT+. Zusätzlich ist links unten die "Schalttafel" als weiteres, bewegliches Menü zu sehen.

Die Syntax der Retrieval-Sprache zur Erstellung von Listen mit Merkmalen aus mehreren Dateien:



- # : linke, bzw. rechte Klammer
- % : erzeugt Ausgabe nur wenn Attributwert ungleich leer
- + : Aufaddieren von Attributwerten
- > : Verzweigungsoperator
- ! : Verzweigung erfolgt nur zu genau einem passenden Datensatz
- : linksbündige Ausgabe, sonst rechtsbündig
- :
- : Trennsymbol zwischen Dateiname und Attributname

**Attributname:** { Auswahl-Liste der Attributnamen der aktuell aktiven Datei }  
 oder  
 { Auswahl-Liste der vom System beim DB-Entwurf entsprechend der eingegebenen Reihenfolge vergebenen Nummern für die Attributnamen der aktuell aktiven Datei }

EIDGENÖSSISCHE TECHNISCHE  
HOCHSCHULE (ETH)

**© Institut für Arbeitspsychologie**

**Instruktion zum Ablauf  
des Benchmark-Tests**

**1990**

**INSTRUKTIONEN:**

Herzlich willkommen bei der Teilnahme an dem benutzungs-orientierten Benchmark-Test zur Evaluation des ADIMENS-Datenbankprogrammes.

Dieses Datenbankprogramm dient dazu, eine Reihe von verschiedenen Datensätzen ("Karteikarten") in der jeweils zugehörigen Datei ("Karteikasten") zu verwalten und in verschiedenster Art und Weise zu bearbeiten.

Sie werden in dem ersten Teil dieses Tests zunächst systematisch in die Handhabung dieses Datenbankprogrammes eingewiesen. Danach werden eine Reihe von Aufgaben gestellt, die Sie bearbeiten sollen.

Es werden Ihnen zwischendurch ausreichend Pausen zur Verfügung stehen.

Es wäre für uns sehr hilfreich, wenn Sie während der gesamten Aufgabenbearbeitung laut "denken" würden. Bitte sprechen Sie alles, was Ihnen dann durch den Kopf geht, laut und deutlich aus. Wir sind insbesondere an Ihren Zielen, Absichten, Wünschen, etc. interessiert.

Das Ziel dieser Untersuchung ist die Analyse von Problembereichen und Schwachstellen im Umgang mit diesem Programm. Daher sind wir insbesondere an denjenigen Situationen interessiert, bei denen Sie Schwierigkeiten im Umgang mit diesem Programm haben werden.

**BEARBEITEN SIE DIE AUFGABEN BITTE ZÜGIG UND OHNE UNNÖTIGE ZEITVERLUSTE.**

Falls Sie jedoch gar nicht weiter wissen sollten, steht Ihnen der Testleiter gerne mit einer kleinen Hilfestellung zur Seite. Der Testleiter reagiert jedoch nur dann, wenn Sie ihn direkt ansprechen.

Falls Sie überhaupt nicht mehr weiter wissen sollten und auch die Hilfen des Testleiters Ihnen nichts nützen, können Sie jederzeit die Aufgabenbearbeitung vorzeitig abbrechen.

Bei der Durchführung der einzelnen Aufgaben werden Sie die Rolle eines/r Camping-Platzwartes/in einnehmen, der einen Teil seiner Büroarbeit mit Hilfe dieses Datenbankprogrammes erledigt. Wie das im einzelnen aussieht, wird Ihnen vom Testleiter noch genau erklärt.

Vor und nach der Durchführung der einzelnen Benchmark-Aufgaben bitten wir Sie, einige Fragebögen auszufüllen.

Alle Ihre Angaben werden anonym und streng vertraulich behandelt !

# Ablauf des Benchmark-Tests:

## **1. EDV-Vorerfahrung:**

Zunächst wird Sie der Testleiter nach Ihrer Vorerfahrung im Umgang mit EDV-Systemen befragen.

20 Minuten

## **2. Übungsteil der 1. Version:**

Der Testleiter wird Sie systematisch in die Bedienung des Datenbankprogrammes einführen. Wenn Sie etwas nicht verstehen, fragen Sie bitte sofort nach.

60 Minuten

**PAUSE**

5 Minuten

## **3. Aufgabenbearbeitung der 1. Version:**

Vor der Aufgabenbearbeitung erhalten Sie einen Eigenzustandsbogen zum ausfüllen. Anschliessend erhalten Sie eine Mappe mit 4 verschiedenen Aufgaben. Bitte bearbeiten Sie die Aufgaben in der vorgelegten Reihenfolge. Arbeiten Sie dabei bitte zügig und ohne unnötigen Pausen. Falls Schwierigkeiten auftreten, benutzen Sie zuerst die im Programm vorgesehene Hilfe-Funktion, falls diese Ihnen nicht weiter hilft, lesen Sie bitte in der beigelegten Dokumentation nach. Falls dies auch nicht ausreicht, können Sie sich direkt an den Untersuchungsleiter wenden. Nach der Aufgabenbearbeitung erhalten Sie einen Eigenzustandsfragebogen sowie einen Handhabungsfragebogen zum ausfüllen.

55 Minuten

**PAUSE**

10 Minuten

## **4. Übungsteil der 2. Version:**

Der Testleiter wird Ihnen die Unterschiede der zwei Versionen des Datenbankprogrammes erklären.

15 Minuten

## **5. Aufgabenbearbeitung der 2. Version:**

Vor der Aufgabenbearbeitung erhalten Sie einen Eigenzustandsbogen zum ausfüllen. Anschliessend erhalten Sie eine Mappe mit 4 verschiedenen Aufgaben. Bitte bearbeiten Sie die Aufgaben in der vorgelegten Reihenfolge. Arbeiten Sie dabei bitte zügig und ohne unnötigen Pausen. Falls Schwierigkeiten auftreten, benutzen Sie zuerst die im Programm vorgesehene Hilfe-Funktion, falls diese Ihnen nicht weiter hilft, lesen Sie bitte in der beigelegten Dokumentation nach. Falls dies auch nicht ausreicht, können Sie sich direkt an den Untersuchungsleiter wenden. Nach der Aufgabenbearbeitung erhalten Sie einen Eigenzustandsfragebogen sowie einen Nachbefragungsfragebogen zum ausfüllen.

55 Minuten

**PAUSE**

5 Minuten

## **6. Interview:**

Sie werden die Gelegenheit haben, sich mit dem Testleiter über Ihre persönlichen Erfahrungen zu unterhalten.

10 Minuten

**Insgesamt dauert der Benchmark-Test ungefähr 4 Stunden.**

EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE  
(ETH)

© Institut für Arbeitspsychologie

**Vorerfahrungsbogen über EDV-Wissen  
im Umgang mit  
interaktiver Software**

1990

TestpartnerIn- Codierung:	
Datum :	
Startzeitpunkt :	
Endzeitpunkt :	





	1. Zeitraum: von - bis (Datum - Datum)	durchschnittliche Intensität (Stunden/Zeiteinheit <sup>1</sup> )	2. Zeitraum: von - bis (Datum - Datum)	durchschnittliche Intensität (Stunden/Zeiteinheit)
Praktische Vorerfahrung mit Programmiersprachen				
Vorerfahrung mit einem Betriebssystem oder mit Oberflächen von Betriebssystemen				
Vorerfahrung mit Textverarbeitungssystemen				
Vorerfahrung mit Statistikprogrammen				
Vorerfahrung mit Graphikprogrammen und CAD				
Vorerfahrung mit Datenbanken ausser ADIMENS				
Vorerfahrung mit ADIMENS				
Vorerfahrung mit Computerspielen				
Interaktive Arbeit an einem Rechner				
Interaktive Arbeit an einem Rechner mit einer direkt- manipulativen Oberfläche (Desktop- Oberfläche)				
Interaktive Arbeit an einem Rechner mit Maus				

<sup>1</sup> Zeiteinheit = Jahr, Monat, Woche oder Tag. Bitte wahlweise angeben.



EIDGENÖSSISCHE TECHNISCHE  
HOCHSCHULE (ETH)

**Institut für Arbeitspsychologie**

**Eigenzustandsfragebogen**  
© Apenburg 1986

1990

TestpartnerIn- Codierung:
Datum :
Startzeitpunkt :
Endzeitpunkt :

Im folgenden finden Sie eine Liste mit verschiedenen Eigenschaftswörtern. Fragen Sie sich bitte bei jedem Eigenschaftswort:

In welchem Grade trifft dieses Wort auf meinen augenblicklichen Zustand - so wie ich ihn jetzt erlebe - zu?

Jedem Eigenschaftswort ist ein Masstab zugeordnet. Dieser hat 6 Stufen von 1 = "kaum zutreffend" bis 6 = "völlig zutreffend".

Kreuzen Sie bitte jeweils diejenige Stufe an, die Ihrem augenblicklichen Zustand am ehesten entspricht.

Es gibt keine richtigen oder falschen Beantwortungen. Es kommt nur auf Ihren persönlichen Eindruck an. Scheuen Sie sich nicht, jede Skalenstufe von 1 bis 6 anzukreuzen, die Ihnen angemessen erscheint !

Bitte beachten Sie folgende Punkte:

- Sollte Ihnen die Antwort einmal schwerfallen, so entscheiden Sie sich bitte für die Antwortmöglichkeit, die Ihnen am ehesten sinnvoll und zutreffend erscheint.
  
- Überlegen Sie sorgfältig, aber nicht zu lange.
  
- Lassen Sie keine Zeile aus und machen Sie in jeder Zeile nur ein Kreuz.

Auf meinen augenblicklichen Zustand zutreffend:

	kaum 1	etwas 2	einiger- massen 3	ziem- lich 4	über- wiegend 5	völlig 6
Gutgelaunt						
Schläfrig						
Ausgeglichen						
Erfahren						
Anstrengungsbereit						
Gesprächig						
Energiegeladen						
Geachtet						
Heiter						
Ausdauernd						
Selbstsicher						
Ausgeruht						
Überlegen						
Erholungsbedürftig						
Locker						
Unternehmungslustig						
Abgehetzt						
Kontaktbereit						

Auf meinen augenblicklichen Zustand zutreffend:

	kaum 1	etwas 2	einiger- massen 3	ziem- lich 4	über- wiegend 5	völlig 6
Ruhig						
Anerkannt						
Müde						
Fröhlich						
Nervös						
Konzentrationsfähig						
Angesehen						
Sicher						
Matt						
Mitteilsam						
Gelassen						
Aktiv						
Gefestigt						
Erholt						
Geschätzt						
Arbeitsfreudig						
Vernügt						
Verausgabt						

Berechnung der Skalenwerte erfolgt durch die Summenbildung der jeweiligen Antwortwerte pro Fragebogen-Item. Die einzelnen SKALEN setzen sich aus den jeweiligen Fragebogen-Items wie folgt zusammen:

ANSTRENGUNGSBEREITSCHAFT:	aktiv, anstrengungsbereit, arbeitsfreudig, ausdauernd, energiegeladen, konzentrationsfähig;
ERHOLTHEIT:	abgehetzt <sup>U</sup> , ausgeruht, erholt, erholungsbedürftig <sup>U</sup> , verausgabt <sup>U</sup> ;
KONTAKTBEREITSCHAFT:	gesprächig, kontaktbereit, mitteilsam, unternehmungslustig;
SCHLÄFRIGKEIT:	matt, müde, schläfrig.
SELBSTSICHERHEIT:	erfahren, gefestigt, selbstsicher, sicher, überlegen;
SOZIALE ANERKENNUNG:	anerkannt, angesehen, geachtet, geschätzt;
SPANNUNGS-LAGE:	ausgeglichen, gelassen, locker, nervös <sup>U</sup> , ruhig;
STIMMUNGS-LAGE:	fröhlich, gutgelaunt, heiter, vergnügt.

---

<sup>U</sup> der aktuelle Wert dieses Fragebogen-Items muß vor der Summenbildung umgepolt werden. Dies wird dadurch erreicht, daß der jeweilige Antwortwert von dem höchsten "Skalenwert + 1" (hier: "6 + 1 = 7") abgezogen wird.



EIDGENÖSSISCHE TECHNISCHE  
HOCHSCHULE (ETH)

**© Institut für Arbeitspsychologie**

**Gruppe 1: Aufgabenserie 1 für Version "A"**

1990

**Warten Sie, bis der Testleiter Sie auffordert, anzufangen...**

**Aufg.1**

Bitte stellen Sie fest, wie viele Datensätze in der Datei "PLATZ" vorhanden sind.

Sobald Sie mit dieser Aufgabe fertig sind, verlassen Sie das Datenbankprogramm über die Menü-Option "DOS-Shell" in dem Menü "PROGRAMME" und warten auf die Anweisungen des Testleiters.

## **Aufg. 2**

Bitte geben Sie in einem einzigen Fenster mindestens alle diejenigen Datensätze der Datei "ADRESSEN" als Liste auf dem Bildschirm aus, bei denen im Merkmal "Platz-Nr" eine Platznummer eingetragen ist. In der Liste soll das Merkmal "Platz-Nr" als erstes Merkmal und "Land" als zweites Merkmal vorkommen. Korrigieren Sie anschliessend die Merkmalsausprägungen so, dass bei allen Datensätzen mit einer eingetragenen Platznummer, die Platznummer "07" steht.

Sobald Sie mit dieser Aufgabe fertig sind, verlassen Sie das Datenbankprogramm über die Menü-Option "DOS-Shell" in dem Menü "PROGRAMME" und warten auf die Anweisungen des Testleiters.

**Aufg.3**

Bitte laden Sie die Wahl-Definition "ANREISE.PIK" für die Datei "PLATZ" und wenden Sie diese Wahl an, damit alle Fereingäste, die am "02/07/87" angereist sind, ausgewählt werden können. Lassen Sie nur die gefundenen Datensätze als Maske auf dem Bildschirm einzeln ausgeben. Anschliessend lassen Sie die ganze Datei "PLATZ" vollständig als Liste auf dem Bildschirm ausgeben.

Sobald Sie mit dieser Aufgabe fertig sind, verlassen Sie das Datenbankprogramm über die Menü-Option "DOS-Shell" in dem Menü "PROGRAMME" und warten auf die Anweisungen des Testleiters.

#### **Aufg. 4**

Erstellen Sie bitte eine Liste aller Gruppenmitglieder mit den Passnummern der zugehörigen Verbindungen, indem Sie für die Datei "GRUPPE" ein Mischdokument mit dem Texteditor

(Parameter: C:\GEMAPPS\EXP\MISCHDOK\VP[Ihre VP-Nr.].DOC) erstellen. Diese Liste soll das Merkmal "Nachname\_Vorname" der Gruppenmitglieder aus der Datei "GRUPPE" sowie das Merkmal "Pass-Nr" der zugehörigen Verbindungen aus der Datei "PLATZ" enthalten. Lassen Sie anschliessend diese Liste durch Verwendung des Mische-Ikons auf dem Drucker ausdrucken.

Sobald Sie mit dieser Aufgabe fertig sind, verlassen Sie das Datenbankprogramm über die Menü-Option "DOS-Shell" in dem Menü "PROGRAMME" und warten auf die Anweisungen des Testleiters.

EIDGENÖSSISCHE TECHNISCHE  
HOCHSCHULE (ETH)

**© Institut für Arbeitspsychologie**

**Nachbefragungsbogen: Handhabung**

1990

TestpartnerIn- Codierung:	
Datum :	
Startzeitpunkt :	
Endzeitpunkt :	

Beschreiben Sie bitte anhand der nachfolgenden Gegensatzpaare von Eigenschaftswörtern, wie Sie die Handhabung des Computerprogrammes erlebt haben.

Es gibt keine richtigen oder falschen Einschätzungen. Es kommt nur auf Ihren persönlichen Eindruck an. Scheuen Sie sich nicht, jede Skalenstufe von -2 bis +2 anzukreuzen, die Ihnen angemessen erscheint !

Bitte machen Sie nur ein Kreuz in jeder Zeile.

	ziemlich (-2)	eher (-1)	teils/ teils (0)	eher (+1)	ziemlich (+2)		
zeitraubend						zeitsparend	ZEITSPAR
starr						flexibel	FLEXIBEL
einfach						umständlich	EINFACH
durchschaubar						undurchschaubar	DURCHSCH
verwirrend						eindeutig	EINDEUTG
übersichtlich						unübersichtlich	UEBERSIC
kompliziert						unkompliziert	UNKOMPLI
frei						vorgeschrieben	FREI
vorhersehbar						unvorhersehbar	VORHERSEH
beeinflussbar						unbeeinflussbar	BEEINFLU
ärgerlich						erfreulich	ERFREULI
	ziemlich (-2)	eher (-1)	teils/ teils (0)	eher (+1)	ziemlich (+2)		

EIDGENÖSSISCHE TECHNISCHE  
HOCHSCHULE (ETH)

**© Institut für Arbeitspsychologie**

**Beurteilungsfragebogen für das  
ADIMENS-Datenbankprogramm**

**1990**

TestpartnerIn- Codierung:
Datum :
Startzeitpunkt :
Endzeitpunkt :



**Beurteilungsfragebogen:**

Dieser Fragebogen enthält einige Fragen, die Ihre Arbeitstätigkeit mit dem Datenbankprogramm ADIMENS-Version "A" (die neutrale Test-Bezeichnung) betreffen. Dabei geht es uns um Ihre persönliche Meinung, also darum, wie Sie Ihre Arbeit mit dem Datenbankprogramm erleben. Alle Angaben, die Sie machen, werden vertraulich behandelt. Die Auswertung des Fragebogens erfolgt absolut anonym, das heisst ohne Namensnennung.

Wir bitten Sie, den Fragebogen möglichst vollständig auszufüllen. Selbstverständlich steht es Ihnen aber auch zu, eine Frage offen zu lassen, deren Beantwortung Ihnen grosse Mühe bereitet.

Für zusätzliche Bemerkungen und Kritik zur Befragung sind wir Ihnen dankbar. Bitte benötigen Sie dazu die letzte Seite dieses Fragebogens.

**Anleitung zum Ausfüllen des Fragebogens:**

Bitte kreuzen Sie eine der vorgegebenen Antwortmöglichkeiten an, wobei Sie bitte nur ein Kreuz in eines der vorgesehenen Felder eintragen.

Beispiel:

	Unnötig - 1 -	Kaum nötig - 2 -	Weder / noch - 3 -	Nützlich - 4 -	Sehr nützlich - 5 -	Keine Angabe möglich
Die Datensätzen können auch als Liste angezeigt werden. Diese Möglichkeit finde ich....			X			

Es gibt keine richtigen oder falschen Einschätzungen. Es kommt nur auf Ihren persönlichen Eindruck an. Scheuen Sie sich nicht, jede Skalenstufe von 1 bis 5 anzukreuzen, die Ihnen angemessen erscheint !

Nur, wenn Sie wirklich nichts mit der Frage anfangen können, kreuzen Sie die Kategorie "keine Angabe möglich" an.

Vielen Dank für Ihre Mitarbeit !

	Unnötig - 1 -	Kaum nötig - 2 -	Weder / noch - 3 -	Nützlich - 4 -	Sehr nützlich - 5 -	Keine Angabe möglich
Die Möglichkeiten, die die Verbunddatei bietet, sind.....						
Durch den Befehl "Bearbeiten" kann eine temporäre Selektion vor der Ausgabe einer Liste gemacht werden. Diese Eigenschaft finde ich.....						
Wenn der Schalter "Anzeigen als Maske" gesetzt ist, können durch den Befehl "Bearbeiten" die angezeigten Datensatzmasken mit Funktionstasten weiter bearbeitet werden. Diese Möglichkeit finde ich.....						
Durch Anklicken von Feldern in der Schalttafel, können verschiedene Schalter aktiviert werden, ohne über die Pulldown-Menüs gehen zu müssen. Diese Eigenschaft ist....						
Die Schalttafel gibt Auskunft über den aktuellen Zustand der aktivierten Datenbankdatei. Diese Auskunft finde ich.....						
Die Menüleiste ist speziell in die Pulldown-Menüs "Datei", "Edit", "Verbund", etc. aufgeteilt. Die spezielle Art der Aufteilung finde ich....						

**Falls Sie weitere Anmerkungen, Kritik, etc. haben, schreiben Sie diese hier bitte auf:**