

# Benutzer-orientierte Benchmark-Tests als Methode zur Kontrolle von Gestaltungsmassnahmen bei der Veränderung von Desktop-Oberflächen\*

Matthias Rauterberg

## Zusammenfassung

Es wird die Methode der benutzungsorientierten Benchmark-Tests im Rahmen einer Standardsoftwareentwicklung vorgestellt. Unter Verwendung von vier Benchmark-Aufgaben wurde das zu bewertende interaktive System ADIMENS durch 31 ausgewählte Benutzer empirisch getestet. Die Ergebnisse aus dem Testvergleich der beiden unterschiedlichen Benutzungsoberflächen - ADIMENS-GT und ADIMENS-GT+ - werden vorgestellt und auf dem Hintergrund von benutzer-orientierter Softwaregestaltung diskutiert. Es zeigt sich, daß die vorgenommenen Veränderungen bei der ADIMENS-GT+ Oberfläche eine signifikant schnellere Bearbeitung der Benchmark-Aufgaben ermöglichen, ohne gleichzeitig dabei die subjektiv erlebte Beanspruchung zu erhöhen.

## 1. Einleitung

Es lassen sich vier Arten von Software-Entwicklungsprojekten unterscheiden (WAEBER 1990): spezifische Anwendungen für firmeninterne Fachabteilungen (*Typ-A*); spezifische Anwendungen für externe Kunden (*Typ-B*); Standardbranchenlösungen für externe Kunden (*Typ-C*); Standardsoftware für anonymen Kundenkreis (*Typ-D*). Während bei Typ-A bis C die Gruppe der potentiellen Benutzer weitgehend bekannt ist, müssen bei Typ-D neue Wege der Benutzerbeteiligung besprochen werden. Es wird nun die Möglichkeit vorgestellt, durch experimentelle Studien generalisierbare Aussagen in Form von benutzer-orientierten Gestaltungsvorschlägen auf der Grundlage des Kriterienkonzeptes von ULICH (1991) ableiten zu können (RAUTERBERG 1988, 1990).

## 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; RAUTERBERG 1989, 1990) 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 zwei verkaufsfertigen Produkten: ADIMENS-GT und ADIMENS-GT+.

---

\* 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.

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

### 3. Eigenschaften des Produktes 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 (RAUTERBERG 1988, 1989): 1. die Aufteilung der Pull-down Menüs erwies sich als mögliche Quelle von Fehlbedienungen; 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 oft deshalb 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. sollte die wichtigste Eigenschaft eines relationalen DBMS - die *Relationalität* - dem Benutzer in einer direkt-manipulativ interaktiven Form durch Verbund-Dateien zugänglich sein; dies ist bei ADIMENS-GT nicht der Fall.

Aufgrund der beschriebenen Problembereiche wurden verschiedene Verbesserungsvorschläge abgeleitet. Die neu-gestaltete DBMS-Oberfläche von ADIMENS-GT+ soll im Rahmen dieser Studie empirisch auf den Grad der "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 maussensitiven Bereichen vollzogen wird (ANHANG: Abb. 3);

*neue desktop-orientierte DBMS-Oberfläche (Version-GT+):* folgende Veränderungen wurden vorgenommen: die Pull-down Menüs wurden neu strukturiert; 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 (ANHANG: Abb. 4).

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 BenutzerInnen

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

---

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

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 der Untersuchung

Die BenutzerInnen 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 BenutzerInnen die Eigenzustandsskala von NITSCH (EZ-Skala) aus. Nach der ersten Aufgabenserie wurde die jeweilige Oberfläche mit einem "Handhabungs"-Fragebogen bewertet. Dann erhielten die BenutzerInnen 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 BenutzerInnen ein Gesamt-Bewertungsbogen vorgegeben. Der gesamte Test dauerte pro BenutzerIn 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 BenutzerInnen weiterarbeiten. Um einen Beobachter-Effekt weitgehend auszuschalten, wurde ein Beobachter ausgewählt, der in die Entwicklung der GT+-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 "Wahl-Falle", RAUTERBERG 1988, 1990); 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'.

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."

---

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

**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 (siehe RAUTERBERG & ROLFSEN 1991).

## 5. Darstellung der Ergebnisse

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. 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 sind.

**Tabelle 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:			Co-Variate:			Co-Variate:		
	keine			"KEYSTROKES"			"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"⊗"Reihenf."	1	31.20	.001	1	16.21	.001	1	6.48	.017
"Oberfläche"⊗"Aufg(1-4)"	3	10.56	.001	3	.95	.421	3	10.19	.001
"Reihenfolge"⊗"Aufg(1-4)"	3	1.52	.215	3	2.36	.077	3	1.39	.252
"Oberfl."⊗"Reihenf."⊗"Aufg"	3	5.06	.003	3	3.90	.011	3	5.59	.002

Es sind in allen drei Varianzanalysen die Haupteffekte für den Faktor "Oberfläche" signifikant (Tab.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 Auf-

gaben 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 Tests beruht. Daß es dennoch spezielle Lern- und Transfereffekte gibt, wird durch die signifikante zweifache Wechselwirkung "Oberfläche"  $\otimes$  "Reihenfolge" und die signifikante dreifache Wechselwirkung "Oberfläche"  $\otimes$  "Reihenfolge"  $\otimes$  "Test(1-4)" nahegelegt.

**Tabelle 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 BenutzerInnen ist mit N angegeben.

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

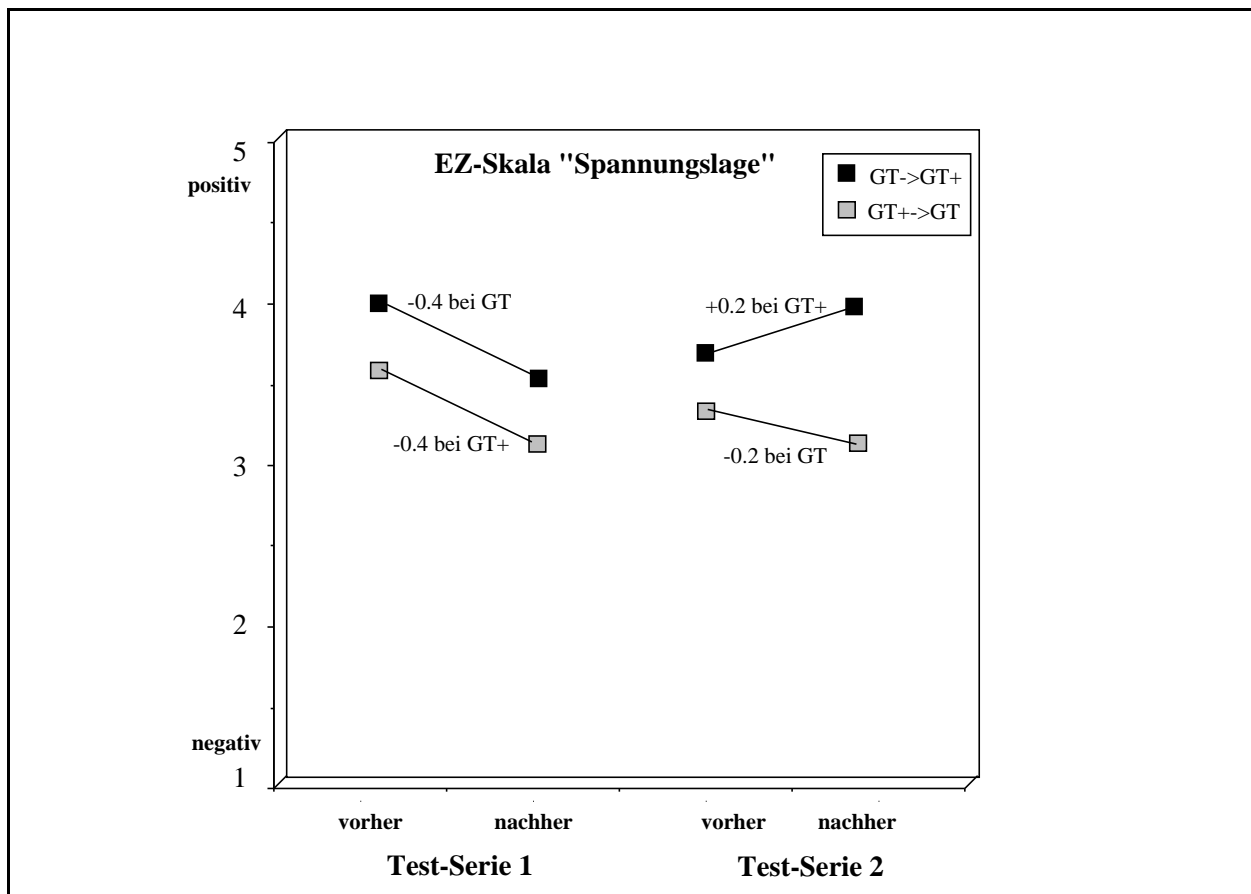
**Tabelle 3** Übersicht über die gewogenen arithmetischen Mittelwerte der Bearbeitungszeiten (USERTIME; Angaben in Sekunden) zusammengefasst über die verschiedenen Aufgaben. Da beide Oberflächen von allen Benutzern getestet wurden, ist die Anzahl N = 31 für beide Gruppen GT und GT+. In der ganz rechten Spalte ist das prozentuale Verhältnis der beiden Bearbeitungszeiten angegeben.

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

Der größte Unterschied bei der Bearbeitungszeit (USERTIME) ergibt sich für den Test-4 (Tab. 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 KEY-STROKES), während bei der GT+-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



"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



**Abbildung 2** Die subjektiven Beanspruchungswerte, gemessen mit der Eigen-Zustands-Skala nach NITSCH. Der niedrigste Skalenwert beträgt 0 ("negativ") und der höchste 5 ("positiv").

Die Beurteilungen durch die BenutzerInnen hinsichtlich der verschiedenen Veränderungen der GT+-Oberfläche gegenüber der GT-Oberfläche fallen insgesamt sehr positiv aus (Tab. 5); die am wenigsten positive Bewertung erhält das Item "Möglichkeit zur direkten Interaktion über die Schalttafel" (Item-Nr. 3; Tab. 5) durch die Gruppe-2 (GT+>GT); dieser Unterschied in der Beurteilung zwischen den beiden Benutzer-Gruppen ist signifikant (T-Test,  $df=30$ , zwei-seitig).



**Tabelle 5** 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- Test
		(GT->GT+)	(GT+->GT)	
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.

## 6. Diskussion

Die durchgeführten Benchmark-Tests zeigen, daß sich ein ergonomisch relevanter Unterschied gemessen über die Bearbeitungszeit für die neu konzipierte Oberfläche von ADIMENS-GT+ 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 (Tab. 1). Zieht man die Bewertung der BenutzerInnen zum Fragebogenitem-3 mit in Betracht (Tab. 4), so scheint der Zeitunterschied primär auf einer verbesserten Repräsentation des Dialog- und Anwendungszustandes bei der GT+-Oberfläche zu beruhen (siehe das Kriterium "Transparenz"; SPINAS, WAEBER & STROHM 1990).

Durch den unterschiedlichen interaktiven Aufwand insbesondere für die Bearbeitung des 4. Tests ergibt sich ein signifikanter Zusammenhang zwischen dem Faktor "Test" und dem Faktor "Oberfläche" (Frage-2, s.o.). Dieser Zusammenhang verschwindet dann auch, sobald der interaktive Aufwand varianzanalytisch kontrolliert wird (Tab.1; 3. Spalte).

Da die subjektiv erlebte Beanspruchung aller BenutzerInnen 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 Benutzer 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 GT+-Oberfläche mit einer spezifischen Entlastung verbunden ist (Tab. 4; Abb. 2).

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

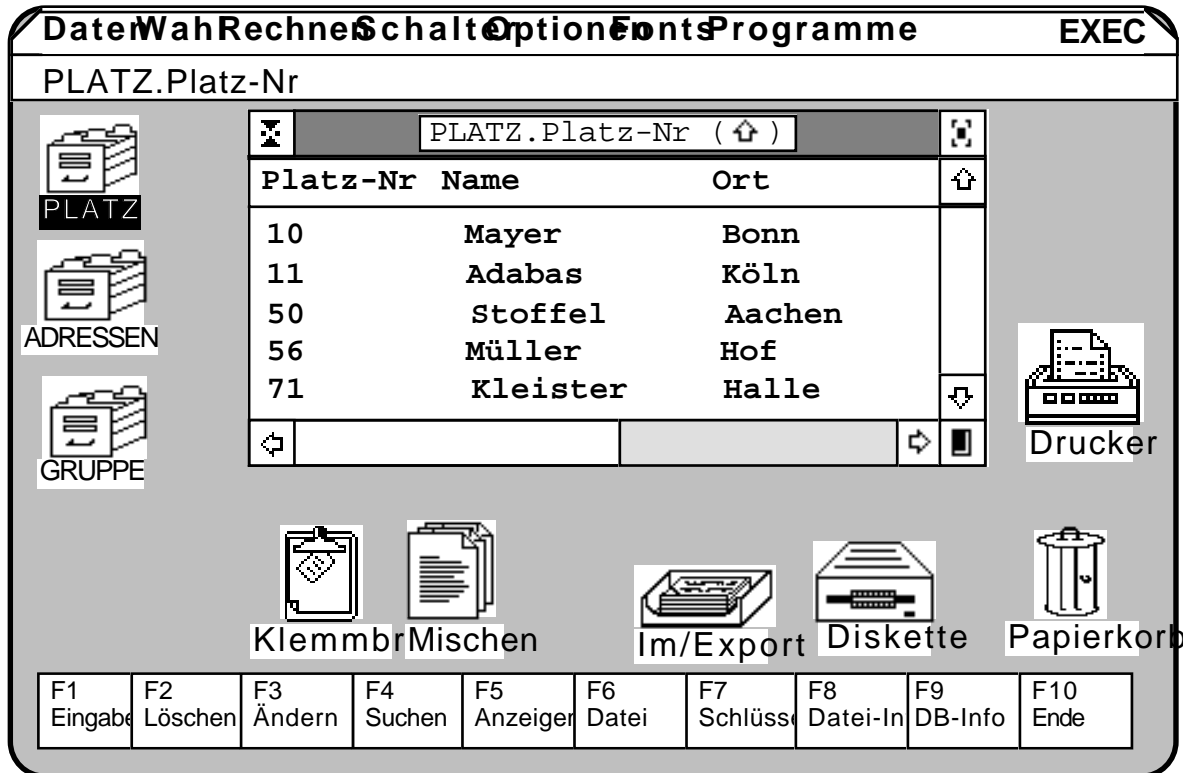
Insgesamt gesehen, kann die hier vorgestellte *Methode zur Benutzerbeteiligung* bei Standardsoftware-Entwicklungen als ausreichend praktikabel und sinnvoll angesehen werden, welche die geplanten und/oder vorgenommenen Gestaltungsmaßnahmen auf ihre benutzer-orientierte Brauchbarkeit hin überprüfen hilft.

## 7. Literatur

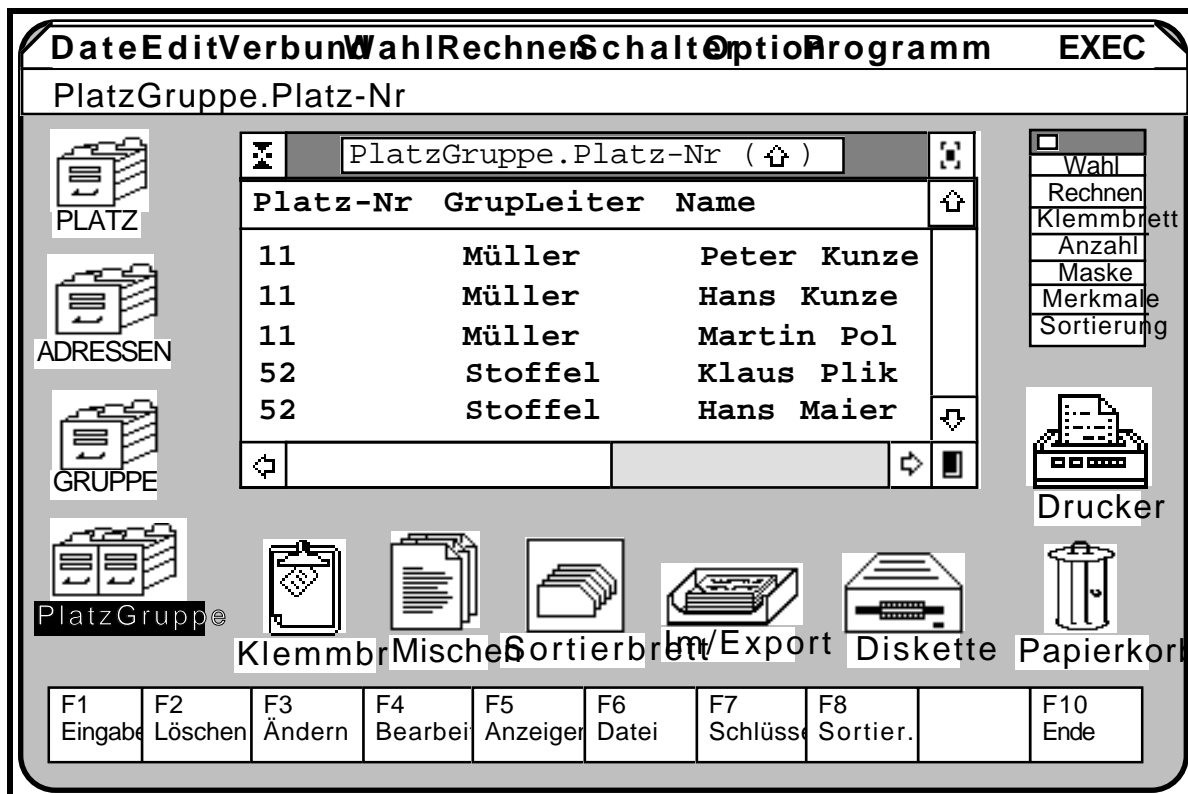
- ACKERMANN, D. & GREUTMANN, T. (1987): Interaktionsgrammatik und kognitiver Aufwand. In: German Chapter of the ACM Bericht 29: Software-Ergonomie'87; (SCHÖNPFLUG, W. & WITTSTOCK, M.; Hrsg.); Stuttgart: Teubner; 262-270.
- APENBURG, E. (1986): Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung. Zeitschrift für Arbeits- und Organisationspsychologie, 30 (N.F. 4); 3-14.
- BORTZ, J. (1984): Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin New York: Springer.
- RAUTERBERG, M. (1988): Untersuchung der Benutzerfreundlichkeit einer desktop-orientierten Benutzungsoberfläche am Beispiel eines relationalen Datenbanksystems. unveröffentlichter Forschungsbericht, Zürich: ETH-Institut für Arbeitspsychologie.
- RAUTERBERG, M. (1989): 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; S. 243-258
- RAUTERBERG, M. (1990): 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. & ROLFSEN, C. (1991): Benutzerorientierte Benchmark-Tests: eine Methode zur Benutzerbeteiligung bei der Entwicklung von Standardsoftware. In: Projektberichte zum Forschungsprojekt Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (BOSS), Nr. 6. (SPINAS, P.; RAUTERBERG, M.; STROHM, O.; WAEBER, D. & ULICH, E. ;Hrsg.), ETH-Zürich: Institut für Arbeitspsychologie.
- 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.
- 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.
- 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.
- ULICH, E. (1991): Arbeitspsychologie. Stuttgart: Poeschel-Verlag.

Matthias Rauterberg  
 Institut für Arbeitspsychologie  
 Eidgenössische Technische Hochschule  
 CH-8092 ZÜRICH

### 8. Anhang



**Abbildung 3** Die Desktop-Oberfläche von ADIMENS-GT. Dem Benutzer 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 4** 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. "Bestellung").

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

