

8. Kapitel

Handlungsfehler und Expertise

Jochen Prümper

8.1 Einleitung

Um Dialogsysteme benutzergerecht gestalten zu können, ist es vonnöten, Einblick in die Probleme und Schwierigkeiten zu bekommen, auf die die Benutzer in ihrem alltäglichen Umgang mit Computern stoßen. Der Gedanke liegt dabei nahe, daß die Schwierigkeiten nicht zuletzt von der Art und dem Umfang der Erfahrung abhängen, die der einzelne bereits mit Computern bzw. mit bestimmten Programmen gesammelt hat.

Dieser Überlegung folgen auch die Grundsätze ergonomischer Dialoggestaltung für Bildschirmarbeitsplätze (DIN, 1988), wenn sie die Forderung enthalten, bei der Gestaltung von Dialogsystemen der unterschiedlichen Expertise verschiedenartiger Benutzergruppen Rechnung zu tragen:

Bei der Gestaltung von Dialogsystemen ist zu berücksichtigen, daß sich Benutzer z.B. nach Übungsgrad (im Umgang mit dem Dialogsystem wachsende Kenntnisse und Fertigkeiten), nach Häufigkeit der Benutzung oder nach Vorkenntnissen (Erfahrungen mit bisherigen Arbeitsmitteln, Fachkenntnissen) unterscheiden (DIN 66 234, Teil 8, 1988, S. 1).

In diesem Kapitel soll nun der Versuch unternommen werden, mittels der systematischen Verhaltensbeobachtung (vgl. Mees & Selg, 1977) Handlungsfehler¹ zu untersuchen, die Benutzern mit unterschiedlicher Expertise bei ihrer Arbeit mit dem Computer unterlaufen. Dabei sollen die von DIN aufgestellten Kriterien dazu dienen, Experten und Novizen voneinander abzugrenzen.

¹ Berücksichtigt werden in diesem Kapitel lediglich Nutzungsprobleme - Fehler, die aus einem "mismatch" zwischen Benutzer und Computer resultieren, nicht aber Funktionsprobleme - Fehler, die auf einen "mismatch" zwischen Computer und Arbeitsaufgabe zurückzuführen sind. Vernachlässigt werden auch Interaktionsprobleme - Fehler, die in der Kooperation mehrerer Personen begründet liegen. Außer Fehlern im engeren Sinn werden auch Ineffizienzen einbezogen (bei denen das geplante Ziel zwar erreicht wird, aber nicht auf dem einfachsten und schnellsten Weg).

8.2 Literaturübersicht

8.2.1 Forschung zur Expertise

In der Forschung zur Expertise (Übersicht bei Chi, Glaser & Farr, 1988) werden unterschiedliche Qualitäten angeführt, die Experten gegenüber Novizen auszeichnen. Experten besitzen ein größeres Fachwissen (Johnson et al., 1981; Voss & Post, 1988), Experten verfügen über eine differenziertere Art der Wissensorganisation (Chase & Simon, 1973; deGroot, 1966; Soloway, Adelson & Ehrlich, 1988), Experten betreiben eine ausgeprägtere Problemexploration (Schaub & Strohschneider, 1989), Experten erkennen und repräsentieren fachliche Probleme auf einem höheren Abstraktionsniveau (Chi, Feltovich & Glaser, 1981), Experten besitzen eine höhere Fähigkeit zur Selbstreflexion (Miyake & Norman, 1979; Simon & Simon, 1978), Experten setzen effektivere heuristische Strategien ein (Putz-Osterloh, 1987) und Experten haben realistischere Vorstellungen von den Schwierigkeiten eines Problems (Chi, Glaser & Rees, 1982).

Zu diesen Begründungen, warum Experten die Lösung bestimmter Probleme besser gelingt als Novizen, kommen auch die wenigen Studien, die sich mit Fehlern in der Mensch-Computer Interaktion beschäftigen (z.B. Cuff, 1980; Davis, 1983a, b; Lang, Lang & Auld, 1981a; Wiedenbeck, 1985; Youngs, 1974), obwohl manche von ihnen aus methodischen Gründen nicht unwidersprochen blieben².

Neben dem Ergebnis, daß die Studien zur Expertise implizit oder explizit zu der Schlußfolgerung gelangen, daß Experten besser sind und weniger Fehler machen als Novizen, scheint ihnen noch etwas gemeinsam zu sein: Der Schwerpunkt der Forschung zur Expertise liegt in erster Linie auf recht komplexen (und teilweise dynamischen, vernetzten und intransparenten), häufig künstlich im Experiment geschaffenen Handlungssituationen, in denen z.B. die Rekonstruktion von Schachstellungen (vgl. deGroot, 1965) und Computer-Programmen (vgl. Vessey, 1988) oder das Management diverser computersimulierter Szenarios im Vordergrund standen (vgl. Dörner, 1989; Funke, 1988). Offensichtlich besteht also im Rahmen der Studien zur Expertise bisher lediglich ein geringes Interesse an der Frage, wie Menschen mit alltäglichen Umwelten umgehen, in denen Routinehandlungen dominieren und nicht komplexe und z.T. künstlich angeregte Problemlösungen.

Es gilt nun, die Studien zur Expertise zum einen bezüglich der Anlage der Forschung und zum anderen hinsichtlich der Ergebnisse zu ergänzen bzw. zu relativieren. Wählt man alltägliche Situationen des Arbeitshandelns in seinem natürlichen Kontext als Beobachtungsfeld und wertet man zudem die Beobachtungen aus handlungstheoretischer Sicht aus, dann erscheint es - soviel sei vorweg genommen - nicht unbedingt zwingend, daß Experten zu besseren Resultaten kommen als Novizen, d.h. im vorliegenden Fall, daß sie weniger Fehler machen. Auch die in den einschlägigen Forschungen angeführten Begründungen für die ermittelten Ergebnisse lassen sich aus handlungstheoretischer Sicht ergänzen.

² So kritisieren beispielsweise Litecky und Davis (1976) die unzureichende Stichprobengröße in der Studie von Youngs (1974), oder Allwood (1986) bemängelt die unklare Definition der Expertise in der Untersuchung von Davis (1983a).

8.2.2 Handlungstheoretische Überlegungen zur Expertise

Die Theorie der Handlungsregulation (vgl. Hacker, 1986; Volpert, 1983) geht davon aus, daß weniger gut geübte Handlungen auf der intellektuellen Ebene, besser geübte Handlungen auf der Ebene der flexiblen Handlungsmuster und sehr gut geübte Handlungen auf der sensumotorischen Regulationsebene reguliert werden. Allen diesen handlungssteuernden Ebenen dient als Regulationsgrundlage die Ebene des Wissens, der Vorstellungen, der inneren Modelle (vgl. Kap. 1). Die Zuordnung einer bestimmten Handlung zu einer bestimmten Regulationsebene verändert sich insofern, als Handlungen mit zunehmender Übung von oberen auf untere Regulationsebenen delegiert, also automatisierter werden (vgl. Frese & Stewart, 1984).

Wendet man diese Überlegungen auf unsere handlungsorientierte Fehlertaxonomie an (vgl. auch Zapf et al., 1989), so bedeutet dies, daß bei geringer Expertise in erster Linie Handlungsfehler auf der Regulationsgrundlage, d.h. Wissensfehler im weitesten Sinn (sowie ineffizientes Verhalten aufgrund fehlenden Wissens) auftreten müßten. Bei den ungeübten und damit von der intellektuellen Regulationsebene gesteuerten Handlungen wären außerdem Denk-, Merk-/Vergessens- und Urteilsfehler zu erwarten. Bei hoher Expertise käme es hingegen zu einer Häufung von Handlungsfehlern auf der Ebene der flexiblen Handlungsmuster (Gewohnheits-, Unterlassens- und Erkennensfehlern), zu einem erhöhten Anteil von ineffizientem Verhalten aus Gewohnheit sowie zu mehr Fehlern auf der sensumotorischen Regulationsebene.

Diese Hypothese impliziert, daß nicht per se von einem negativen Zusammenhang zwischen dem Grad der Expertise und einer Gesamtfehlerhäufigkeit ausgegangen werden kann, sondern daß bestimmte Fehler mit einer bestimmten Expertise einhergehen. Experten machen nicht etwa weniger (sondern nur andere) Fehler und kommen damit auch nicht unbedingt zu einem "besseren" Ergebnis als Novizen.

Anhand von zwei Untersuchungen sollen diese Überlegungen empirisch überprüft werden. Zum einen kommt dabei die Beobachtungsstudie zum Tragen, die den meisten Kapiteln dieses Buches zugrunde liegt und im Anhang genauer beschrieben wurde, und zum anderen eine Längsschnittuntersuchung, in der Lernfortschritte unmittelbar meßbar sind.

8.3 Beobachtungsstudie

In der vorliegenden Arbeit werden Unterschiede zwischen Novizen und Experten in ihrem täglichen Umgang mit Dialogsystemen unmittelbar am Arbeitsplatz analysiert. Zur Operationalisierung der Expertise sollen mehrere Dimensionen herangezogen werden, die zunächst einmal a priori als Einzeldimensionen und später a posteriori in ihrem möglichen Zusammenwirken dargestellt und diskutiert werden. Darüberhinaus soll im Rahmen einer kleineren Untersuchung an einer studentischen Stichprobe dem Einfluß der sich verändernden Expertise im Längsschnitt nachgegangen werden.

8.3.1 Methode

8.3.1.1 Operationalisierung der Expertise

Die meisten Studien zur Novizen-Experten Forschung (z.B. Adelson, 1984; Allwood & Eliasson, 1987; Barfield, 1986; Bateson, Alexander & Murphy, 1987; Cooke & Schvaneveldt, 1988; Shneiderman, 1976; Soloway, Adelson & Ehrlich, 1988; Vihmalo & Vihmalo, 1988; Weiser & Shertz, 1983) bedienen sich eines akademischen Klientels im Labor, in denen Professoren, Unterweiser und erfahrene Programmierer mit Studenten oder weniger erfahrenen Programmierern bezüglich ihrer Lösungen von vorgegebenen Aufgaben verglichen werden. So herrscht ein großes Defizit an Untersuchungen, die sich um Beobachtungen von nicht standardisierten Arbeitsaufgaben unter alltäglichen Arbeitsbedingungen bemühen und die den Versuch unternommen haben, der Frage der sich verändernden Expertise im Längsschnitt nachzugehen.

Mit dem beruflichen Status oder dem Ausbildungsstand findet in den einschlägigen Novizen-Experten-Forschungen zudem lediglich eine Dimension der Expertise Berücksichtigung. Die im folgenden beschriebene Querschnittstudie trägt der Überlegung Rechnung, daß es sich bei der Expertise um ein facettenreiches Phänomen handelt, das sich mit einer Dimension sicher nicht hinlänglich erfassen läßt. Deshalb basiert sie auf einer mehrschichtigen Operationalisierung von Expertise, die mit den Variablen Übungsgrad, Vorkenntnisse und Häufigkeit der Benutzung die oben zitierte Anregung des Deutschen Instituts für Normung (DIN, 1988) aufgreift. Entsprechend wird im folgenden von Computer-, Programm- und Arbeitszeit-Expertise die Rede sein, wobei in allen drei Gruppen unterschieden wird zwischen: Novizen mit niedriger Expertise, mittelmäßig erfahrenen Benutzern mit mittlerer Expertise und Experten mit hoher Expertise³.

Computer-Expertise. Der Übungsgrad, d.h. die im Umgang mit dem Dialogsystem gewachsenen Kenntnisse und Fertigkeiten, wurde über die Frage operationalisiert, wie lange die Benutzer schon mit Computern arbeiten. Benutzer, die weniger als ein halbes Jahr Computererfahrung vorweisen können, bezeichnen wir als Benutzer mit niedriger Computer-Expertise, diejenigen, die bis zu drei Jahren mit dem Computer arbeiten, als Benutzer mit mittelmäßiger Computer-Expertise, und diejenigen mit mehr als drei Jahren Computererfahrung als Benutzer mit hoher Computer-Expertise.

Programm-Expertise. Die Vorkenntnisse der Benutzer wurden über die Frage erhoben, wieviele Softwareprogramme überhaupt beherrscht werden - mit den Schnittpunkten ein Softwareprogramm, zwei bis drei Programme und vier und mehr Programme.

Arbeitszeit-Expertise. Um die Häufigkeit der Benutzung in Erfahrung zu bringen, wurde die Frage gestellt, wieviel Prozent der Arbeitszeit im Durchschnitt an einem Computer verbracht wird. Die Variable Arbeitszeit-Expertise erhebt damit die Erfahrung, die ein Benutzer im täglichen Umgang mit Dialogsystemen sammelt. Diejenigen Benutzer, die weniger als 20% ihrer Arbeitszeit am Computer verbringen, bezeichnen wir als Benutzer mit niedriger Arbeitszeit-Expertise, diejenigen, die zwischen 20% und 80% ihrer Arbeitszeit mit dem Computer arbeiten, als Benutzer mit mittelmäßiger Arbeitszeit-Expertise, und diejenigen, mit mehr als 80% am Computer verbrachter Arbeitszeit als Benutzer mit hoher Arbeitszeit-Expertise.

³ Eine vierte Gruppe der naiven Benutzer, ohne jegliche Expertise im Umgang mit Dialogsystemen (vgl. Barfield, 1986; Cooke & Schvaneveldt, 1988; Shneiderman, 1976), muß für die vorliegende Fragestellung außer acht gelassen werden, da diese in der alltäglichen Arbeit nicht auftauchen.

8.3.1.2 Arbeitskomplexität als intervenierende Variable

Im Gegensatz zu Tätigkeiten von niedriger Komplexität verlangen Tätigkeiten von hoher Komplexität bewußte Aufmerksamkeit für Ziel- und Planentwicklung, eine bewußte Behaltensleistung und eine bewußte Verarbeitung der Rückmeldung. Da anzunehmen ist, daß Experten komplexere Arbeiten verrichten als Novizen und im Kapitel 7 gezeigt wurde, daß die Arbeitskomplexität einen Einfluß auf die Häufigkeit von Denkfehlern und Merk- und Vergessensfehlern hat, gilt es, potentielle Einflüsse der Arbeitskomplexität zu kontrollieren. Sie wurde deshalb als Kovariate in den Berechnungen des Zusammenhangs zwischen Handlungsfehler und Expertise berücksichtigt (zur Operationalisierung der Arbeitskomplexität siehe Zapf, 1991).

8.4 Ergebnisse

8.4.1 A priori-Klassifikation der Expertise

Computer-Expertise. Wie aus Tabelle 8.1 ersichtlich ist, gibt es keine überzufälligen Unterschiede zwischen niedriger, mittelmäßiger und hoher Computer-Expertise bezüglich der Gesamtfehlerhäufigkeit.

Erwartungsgemäß unterlaufen aber Benutzern mit niedriger Computer-Expertise signifikant mehr Wissensfehler (bis dreimal so viele) als den beiden anderen Gruppen.

Signifikante Unterschiede ergeben sich auch auf der Ebene der flexiblen Handlungsmuster bezüglich der Gewohnheits- und der Unterlassensfehler. Entsprechend unserer Hypothese ergibt sich mit steigender Computer-Expertise ein Anstieg an Gewohnheitsfehlern. Benutzern mit hoher Computer-Expertise unterlaufen dabei nahezu viermal so viele Gewohnheitsfehler als Benutzern mit niedriger Computer-Expertise. Während das Ergebnis für die Gewohnheitsfehler also recht eindeutig unsere Hypothese belegt, ist der Verlauf für die Unterlassensfehler weniger klar. Hier deutet sich im Gegensatz zu unserer Hypothese ein eher u-förmiger Zusammenhang an.

Alle diejenigen, die einen eindeutigen Zusammenhang zwischen Computer-Expertise und Gesamtfehlerhäufigkeit vermuten, werden enttäuscht sein. Wie lange ein Benutzer schon mit Computern arbeitet, schlägt sich nicht darin nieder, wieviele Fehler er im täglichen Umgang mit ihnen macht.

Während die Hypothesen hinsichtlich der Fehler auf der intellektuellen und der sensorischen Regulationsebene sowie der Ineffizienzen nicht bestätigt werden, erfüllen sich unsere Erwartungen sowohl bezüglich der Wissens- als auch der Gewohnheitsfehler. Offensichtlich benötigen Benutzer ein halbes Jahr Erfahrung im Umgang mit Dialogsystemen, ehe sie gravierendere Wissensprobleme überbrückt haben; erst dann tritt eine gewisse Stabilität ein. Für die Gewohnheitsfehler läßt sich allerdings ein gegenteiliger Effekt beobachten. Mit zunehmender Computer-Expertise kommt es zu immer mehr Gewohnheitsfehlern. Je länger ein Benutzer schon mit Computern Erfahrungen sammeln konnte, desto häufiger werden versehentlich falsche Handlungsgrundmuster aufgerufen. Wie qualitative Aussagen der Beobachter belegen, war diese Problematik besonders in den Fällen augenfällig, in denen die Benutzer kürzlich zum ersten Mal mit einer neuen Programm-

Version konfrontiert worden waren, bei der sich die Softwareentwickler zu einer veränderten Funktionstastenbelegung, Menüorganisation etc. entschieden hatten.

Tabelle 8.1: Durchschnittliche Fehlerhäufigkeit pro Computerarbeitsstunde für Computer-Expertise.

Computer-Expertise Computer-Erfahrung in Jahren:	niedrig <1/2 N=21	mittelmäßig 1/2 - 3 N=118	hoch >3 N=24	F-Wert
Fehler insgesamt	4.75	3.80	3.87	.89
Regulationsgrundlage				
Wissensfehler	.96	.32	.38	7.62***
Intellektuelle Regulationsebene				
Denkfehler	.76	.50	.22	1.90
Merk- und Vergessensfehler	.30	.22	.34	.87
Urteilsfehler	.09	.09	.02	.72
Flexible Handlungsmusterebene				
Gewohnheitsfehler	.19	.45	.73	2.90*
Unterlassensfehler	.69	.40	.84	4.57**
Erkennensfehler	.22	.15	.21	.38
Sensumotorische RE				
Bewegungsfehler	.64	.92	.95	.31
Ineffizienzen				
Ineffizienz Wissen	.95	.36	.14	2.29
Ineffizienz Gewohnheit	.27	.38	.09	.75

Anmerkung: Adjustierte Mittelwerte nach Ausparialisierung der Arbeitskomplexität; Signifikanzniveau der F-Werte: *** $p < .001$; ** $p < .01$; * $p < .05$

Programm-Expertise. Wie aus Tabelle 8.2 ersichtlich ist, bestehen zwischen Benutzern mit niedriger, mittelmäßiger und hoher Programm-Expertise überzufällige Unterschiede bezüglich der Gesamtfehlerhäufigkeit; jedoch nicht in dem Sinne, daß mit steigender Expertise die Fehlerhäufigkeit abnimmt. Im Gegenteil, je mehr Programm-Expertise ein Benutzer vorzuweisen hat, desto mehr Fehler unterlaufen ihm insgesamt.

Auf der intellektuellen Regulationsebene unterscheiden sich die drei Gruppen hinsichtlich der Denkfehler. Benutzern mit niedriger Programm-Expertise unterlaufen dabei die wenigsten und Benutzern mit hoher Programm-Expertise mit Abstand die meisten Denkfehler.

Auf der Ebene der flexiblen Handlungsmuster kommt es zwischen den drei Gruppen lediglich zu Unterschieden bei den Unterlassensfehlern. Entsprechend unseren Erwartungen unterlaufen Benutzern mit niedriger Programm-Expertise die wenigsten und Benutzern mit hoher Programm-Expertise die meisten Unterlassensfehler.

Ebenfalls zu signifikanten Unterschieden kommt es auf der sensumotorischen Regulationsebene. Allerdings zeichnet sich hier ein umgekehrt u-förmiger Zusammenhang ab; den mittelmäßig erfahrenen Programm-Benutzern unterlaufen nahezu doppelt so viele Bewegungsfehler wie den beiden anderen Gruppen.

Die hohe Gesamtfehlerhäufigkeit der Programm-Experten überrascht angesichts der in der Einleitung dargestellten Literatur. Verantwortlich für das Ergebnis, daß es mit steigender Expertise zu einer höheren Gesamtfehlerhäufigkeit kommt, sind vor allem die Denk- und die Unterlassensfehler. Da bei der Operationalisierung der Programm-Expertise die Zeitdauer der Programmkenntnis nicht einfließt, könnte es sich bei den Benutzern mit der geringsten Gesamtfehlerhäufigkeit um diejenigen handeln, die nur mit einem einzigen Programm arbeiten, dieses dafür aber schon sehr lange kennen (wir kommen auf diesen Punkt unter 8.4.2 zurück). Während sich für die Unterlassensfehler unsere Hypothese bestätigt, zeigt das Ergebnis bezüglich der Denkfehler allerdings das genaue Gegenteil. Je höher die Programm-Expertise ist, desto häufiger kommt es zu Denkfehlern. Die Tatsache, daß ein Benutzer mehrere Programme benutzt, die darüber hinaus meist unterschiedlich gestaltet und unterschiedlich funktional sind, wirkt sich bei der Arbeit an einem Programm offensichtlich dahingehend aus, daß für ihn komplexere Operationen einen derartigen Neuigkeitscharakter annehmen, daß falsche oder unrealistische Ziele und Pläne aufgestellt werden.

Tabelle 8.2: Durchschnittliche Fehlerhäufigkeit pro Computerarbeitsstunde für Programm-Expertise.

Programm-Expertise	niedrig	mittel	hoch	F-Wert
Anzahl Programme:	1	2-3	>4	
	N=86	N=66	N=11	
Fehler insgesamt	3.32	4.50	5.27	4.14**
Regulationsgrundlage				
Wissensfehler	.38	.48	.23	.69
Intellektuelle Regulationsebene				
Denkfehler	.38	.51	1.24	4.39**
Merk-und Vergessensfehler	.21	.29	.26	.60
Urteilsfehler	.05	.11	.17	2.12
Flexible Handlungsmusterebene				
Gewohnheitsfehler	.39	.50	.76	1.30
Unterlassensfehler	.41	.51	1.15	5.18**
Erkennensfehler	.17	.15	.31	.73
Sensumotorische RE				
Bewegungsfehler	.64	1.26	.67	2.96*
Ineffizienzen				
Ineffizienz Wissen	.44	.30	.20	.47
Ineffizienz Gewohnheit	.26	.44	.30	.43

Anmerkung: Adjustierte Mittelwerte nach Auspartialisierung der Arbeitskomplexität; Signifikanzniveau der F-Werte: *** $p < .001$; ** $p < .01$; * $p < .05$;

Die Unterlassensfehler weisen darauf hin, daß Benutzer mit hoher Programm-Expertise Handlungsschemata, die ihnen eigentlich zur Verfügung stehen, nicht ausführen, überspringen oder zu einem verspäteten Zeitpunkt aktivieren (z.B. weil die Benutzerführung von Programm zu Programm unterschiedlich ist).

Arbeitszeit-Expertise. Es zeigen sich keinerlei überzufällige Zusammenhänge zwischen Arbeitszeit-Expertise und Fehlern (vgl. Tabelle 8.3).

Das Ergebnis, daß die Erfahrung, die ein Benutzer im täglichen Umgang mit Dialogsystemen sammelt, in keinem Zusammenhang zur Fehlerhäufigkeit steht, verwundert. Anscheinend genügt es bereits, wenn ein Benutzer 20%, d.h. ca. eineinhalb Stunden seiner täglichen Arbeitszeit, am Computer verbringt, um einen Grad an Expertise zu erlangen, der ihn bezüglich der Fehler in keiner Weise von dem Benutzer unterscheidet, der täglich mindestens sechs Stunden mit Computerarbeit verbringt.

Tabelle 8.3: Durchschnittliche Fehlerhäufigkeit pro Computerarbeitsstunde für Arbeitszeit-Expertise.

Arbeitszeit-Expertise tägl. Arbeitszeit in %:	niedrig <20 N=25	mittelmäßig 20-80 N=90	hoch >80 N=47	F-Wert
Fehler insgesamt	3.71	4.27	3.29	1.61
Regulationsgrundlage				
Wissensfehler	.54	.42	.31	.78
Intellektuelle Regulationsebene				
Denkfehler	.68	.48	.40	.70
Merk- und Vergessensfehler	.17	.23	.33	1.12
Urteilsfehler	.05	.09	.08	.25
Flexible Handlungsmusterebene				
Gewohnheitsfehler	.33	.48	.48	.41
Unterlassensfehler	.81	.45	.46	2.55
Erkennensfehler	.11	.19	.17	.35
Sensumotorische RE				
Bewegungsfehler	.76	1.08	.62	1.37
Ineffizienzen				
Ineffizienz Wissen	.10	.47	.32	1.28
Ineffizienz Gewohnheit	.17	.37	.13	1.53

Anmerkung: Adjustierte Mittelwerte nach Auspartialisierung der Arbeitskomplexität; Signifikanzniveau der F-Werte: *** $p < .001$; ** $p < .01$; * $p < .05$;

8.4.2 A posteriori-Klassifikation der Expertise

Während die *a priori* Bestimmung der Expertise einer willkürlichen Setzung unterworfen war, soll bei der *a posteriori* Klassifikation der Expertise der Frage nachgegangen werden, welche genuinen Benutzergruppen sich durch die drei Dimensionen der Computer-, Programm- und Arbeitszeit-Expertise herauskristallisieren. Zur Bestimmung dieser natürlich auftretenden Benutzergruppen wurde mit der Clusteranalyse (vgl. Steinhausen & Langer, 1977) ein Verfahren angewendet, das die Personen nach Maßgabe ihrer Ähnlichkeit in Be-

nutzergruppen klassifiziert, die in sich möglichst homogen sind und sich dabei von den anderen Gruppen möglichst deutlich unterscheiden.

Aufgrund dieser Clusteranalyse bietet es sich an, vier Benutzergruppen zu unterscheiden (s. ausführlicher Prümper, Zapf, Brodbeck & Frese, 1991). Tabelle 8.4 gibt eine Übersicht über die jeweilige Gruppenstruktur.

Tabelle 8.4: Ergebnis der Clusteranalyse.

	Gelegenheits- benutzer N=67	Häufige Benutzer N=63	Anfänger N=26	Generalisten N=6
Computererfahrung in Jahren	2-3	2-3	1/4-1/2	2-3
Anzahl Programme	1.9	1.6	1.2	5.2
tägliche Arbeitszeit in %	20-30	80-90	50-60	40-50

Die zahlenmäßig größte Gruppe stellen die Benutzer dar, die seit zwei bis drei Jahren mit Computern arbeiten, im Durchschnitt fast zwei Programme beherrschen und 20-30% ihrer täglichen Arbeitszeit am Computer verbringen. Vergleicht man diesen Benutzerkreis mit den drei anderen Gruppen, so ist das herausstechende Merkmal darin zu sehen, daß sie einen relativ kleinen Anteil ihrer Arbeit unter Zuhilfenahme eines Computers erledigen. Bei dieser Gruppe handelt es sich also um Gelegenheitsbenutzer. Die zweitgrößte Gruppe konnte ebenfalls schon zwei bis drei Jahre Erfahrungen mit Computern sammeln, beherrscht durchschnittlich 1.6 Programme und verbringt fast den ganzen Tag mit Computerarbeit. Damit arbeitet sie deutlich mehr Stunden am Computer als jede andere Gruppe. Diesen Personenkreis bezeichnen wir als Häufige Benutzer. Die dritte Gruppe ist vor einem viertel bis einem halbem Jahr zum ersten Mal mit Computern in Berührung gekommen, kennt im Durchschnitt etwa 1 Programm und arbeitet etwa die Hälfte des Tages mit ihnen. Diese Benutzer sind die eigentlichen Anfänger. Der vierte und zahlenmäßig kleinste Benutzerkreis hat seit zwei bis drei Jahren Erfahrung mit Computern, kennt recht viele Programme (etwa 5) und verbringt 40-50% der täglichen Arbeitszeit an einem Dialogsystem. Diese Benutzer werden als Generalisten bezeichnet.

In Tabelle 8.5 ist dargestellt, wieviele Fehler bei den vier Gruppen der Clusteranalyse auftraten. Wie ersichtlich wird, existieren überzufällige Unterschiede bezüglich der Gesamtfehlerhäufigkeit zwischen diesen Gruppen. Dabei unterlaufen den Generalisten mit Abstand die meisten, den Anfängern die zweitmeisten und den Häufigen Benutzern sowie den Gelegenheitsbenutzern die wenigsten Fehler.

Unterschiede gibt es auch bei den Wissensfehlern. Die Häufigen Benutzer weisen hier die wenigsten und die Anfänger die meistens Wissensdefizite auf (ca. doppelt soviele Fehler wie die Gelegenheitsbenutzer und die Generalisten und ca. dreimal soviele Fehler wie die Häufigen Benutzer).

Auf der intellektuellen Regulationsebene kommt es bei den Denkfehlern zu Unterschieden. Den Generalisten unterlaufen dabei ca. drei- bis fünfmal soviele Denkfehler wie den anderen drei Gruppen.

Auf der Ebene der flexiblen Handlungsmuster kommt es bei den Gewohnheitsfehlern zu überzufälligen Unterschieden. Wieder sind es die Generalisten, denen deutlich die meisten Gewohnheitsfehler unterlaufen, und wiederum machen sie ca. drei- bis fünfmal soviele Fehler wie die anderen Benutzergruppen.

Auf der sensumotorischen Regulationsebene und bezüglich des ineffizienten Verhaltens zeigten sich keine überzufälligen Unterschiede zwischen den vier Gruppen.

Tabelle 8.5: Durchschnittliche Fehlerhäufigkeit pro Computerarbeitsstunde für Gelegenheitsbenutzer, Häufige Benutzer, Anfänger und Generalisten.

	Gelegenheitsbenutzer N=67	Häufige Benutzer N=63	Anfänger N=26	Generalisten N=6	F
Fehler insgesamt	3.84	3.28	4.87	6.97	3.74**
Regulationsgrundlage					
Wissensfehler	.40	.25	.80	.45	3.56**
Intellektuelle Regulationsebene					
Denkfehler	.37	.42	.64	1.94	5.71***
Merk- und Vergessensfehler	.21	.28	.28	.33	.39
Urteilsfehler	.07	.07	.09	.22	.72
Flexible Handlungsmusterebene					
Gewohnheitsfehler	.47	.46	.24	1.25	2.92*
Unterlassensfehler	.54	.41	.46	.89	.88
Erkennensfehler	.10	.18	.26	.47	2.39
Sensumotorische RE					
Bewegungsfehler	1.03	.80	.86	.69	.27
Ineffizienzen					
Ineffizienz Wissen	.30	.24	.88	.28	2.47
Ineffizienz Gewohnheit	.35	.17	.26	.46	.56

Anmerkung: Adjustierte Mittelwerte nach Auspartialisierung der Arbeitskomplexität; Signifikanzniveau des F-Werts: *** $p < .001$; ** $p < .01$; * $p < .05$;

Zunächst läßt sich als interessantes Ergebnis festhalten, daß die Häufigen Benutzer die geringste und die Generalisten die höchste Gesamtfehlerhäufigkeit aufweisen. Offensichtlich bestätigt sich für die Generalisten, die recht viele Programme benutzen, an dieser Stelle wiederum die Vermutung, daß Benutzer, die mehrere Programme beherrschen, bei der Arbeit mit einem Softwarepaket auf größere Schwierigkeiten stoßen als diejenigen, die sich auf nur wenige Programme konzentrieren.

Den größten Anteil an der hohen Gesamtfehlerhäufigkeit der Generalisten haben die Denk- und Gewohnheitsfehler. In beiden Fällen handelt es sich dabei um Fehler, die bei der Ziel- und Planformulierung auftreten. Offensichtlich stoßen die Generalisten also besonders bei der Zielformulierung und Handlungsplanung auf große Schwierigkeiten. Hinter dem Ergebnis der Generalisten steht allem Anschein nach das Problem, daß Anwender, die aufgrund unterschiedlicher Arbeitsaufgaben mit unterschiedlich gestalteten Softwareprogrammen konfrontiert werden, einerseits ständig vor der Schwierigkeit sich verändernder intellektueller Erfordernisse stehen und es ihnen andererseits immer wieder passiert, daß sie "eigentlich richtige" Handlungen in der falschen Situation anwenden.

Auch das Ergebnis, daß die Häufigen Benutzer, die einen sehr hohen Anteil der täglichen Arbeitszeit mit dem Computer arbeiten, die geringste Gesamtfehlerhäufigkeit aufwei-

sen, ist von Interesse. Anscheinend hat die Häufigkeit der Benutzung also doch einen positiven Einfluß auf die Fehlerhäufigkeit, allerdings nur dann, wenn mit nur wenigen Programmen gearbeitet wird.

Der Vorteil einer mehrdimensionalen gegenüber einer eindimensionalen Herangehensweise tritt auch deutlich im Zusammenhang mit den Wissensfehlern zutage. Während die alleinige Betrachtung der Arbeitszeit-Expertise keine Einflüsse auf die Wissensfehler erkennen ließ, zeigt sich jetzt, daß Benutzer, die wenige Programme, diese aber häufiger benutzen, die geringsten Wissensprobleme haben.

8.4.3 Expertise im Längsschnitt

Um der Frage der sich verändernden Expertise im Längsschnitt nachzugehen, wurde im Abstand von drei Monaten eine Gruppe von 17 Studenten und 6 Studentinnen der Betriebswirtschaftslehre untersucht, die im Rahmen einer einsemestrigen Veranstaltung eine Einführung in ein integriertes, menügesteuertes betriebswirtschaftliches Standard-Softwareprogramm (Stammdatenverwaltung, Kundenauftragsverwaltung, Lagerbestandsführung, Lohn/Gehalt etc.) erhielten. Das Durchschnittsalter lag bei 24 Jahren. 6 Teilnehmer hatten das untersuchte Programm bereits in einem vorhergegangenen Semester kennengelernt und wurden deshalb aus den weiteren Berechnungen ausgeschlossen. Die Beobachtungen fanden jeweils an insgesamt fünfzehn Tagen in einem Zeitraum von drei Wochen statt. Dabei wurden zu jeweils beiden Zeitpunkten Doppelbeobachtungen durchgeführt. Nur die Fehler, die von beiden Beobachtern übereinstimmend geratet wurden (vgl. Kap. 3), wurden in den folgenden Berechnungen berücksichtigt⁴.

Ergebnisse. Wie Tabelle 8.6 zu entnehmen ist, ergeben sich bei der Gesamtfehlerhäufigkeit keine signifikanten Unterschiede.

Dafür reduziert sich jedoch entsprechend unserer Hypothese die Anzahl der Wissensfehler von Zeitpunkt 1 zu Zeitpunkt 2 um mehr als die Hälfte. Ebenfalls erfüllt ineffizientes Verhalten aus fehlendem Wissen heraus die Erwartungen. Zum zweiten Zeitpunkt konnten hier ein Drittel weniger Fehler beobachtet werden als zum ersten Zeitpunkt. Auf der intellektuellen Regulationsebene verringern sich die Denkfehler erwartungsgemäß zum zweiten Zeitpunkt signifikant um die Hälfte; bezüglich der Merk-/Vergessensfehler und Urteilsfehler existieren keine signifikanten Unterschiede.

Auf der Ebene der flexiblen Handlungsmuster nehmen, ebenfalls hypothesengerecht, zum zweiten Zeitpunkt die Unterlassensfehler signifikant zu. Bezüglich der Gewohnheits- und Erkennensfehler konnte dieser Effekt allerdings nicht beobachtet werden.

Auch im Rahmen der Längsschnittuntersuchung läßt sich eine Veränderung der Fehlerquote über die Zeit nur dann beobachten, wenn die Art des Fehlers Berücksichtigung findet. In den Fällen, in denen allerdings eine Veränderung eintritt, ist der Effekt deutlich: innerhalb von nur drei Monaten verringert sich die Anzahl der Wissens- und Denkfehler sowie der Ineffizienzen aufgrund fehlenden Wissens um mehr als die Hälfte. Im gleichen Zeitraum erhöht sich die Anzahl der Unterlassensfehler um fast das Dreifache. Zwar konnte auch im Rahmen der Längsschnittstudie lediglich eine partielle Bestätigung unserer

⁴ Trotz der Untersuchung eines studentischen Klientels ist auch diese Studie weit davon entfernt, den Charakter eines standardisierten Experiments zu haben, da auch hier den untersuchten Personen keine einheitlichen Arbeitsaufgaben vorgegeben wurden. Auf eine etwaige Kontrolle der Arbeitskomplexität mußte allerdings verzichtet werden, da das entsprechende Instrument von Zapf (1991) auf seine Güte unter betrieblichen Bedingungen hin überprüft wurde und von seiner Itemkonstruktion her für diese Stichprobe als nicht geeignet erschien.

Hypothesen geleistet werden, doch in den Fällen, in denen signifikante Unterschiede resultierten, entsprach die Entwicklung den Erwartungen.

Tabelle 8.6: Durchschnittliche Fehlerhäufigkeit pro Computerarbeitsstunde im Längsschnitt.

Zeitpunkt	t1		t2
Fehler insgesamt	5.91		4.77
Regulationsgrundlage			
Wissensfehler	1.65	*	.74
Intellektuelle Regulationsebene			
Denkfehler	1.09	*	.53
Merk- und Vergessensfehler	.18		.29
Urteilsfehler	.82		1.12
Flexible Handlungsmusterebene			
Gewohnheitsfehler	.71		.68
Unterlassensfehler	.21	**	.59
Erkennensfehler	.29		.12
Sensumotorische RE			
Bewegungsfehler	.68		.59
Ineffizienzen			
Ineffizienz Wissen	.27	*	.09
Ineffizienz Gewohnheit	.03		.03

Anmerkung: Signifikanzniveau des einseitigen t-tests für abhängige Stichproben: ** $p < .01$; * $p < .05$
Stichprobengröße: $N = 17$

8.5 Zusammenfassende Diskussion

Entgegen dem Alltagsverständnis und den Ergebnissen der experimentellen Erforschung komplexer Probleme zeigt das Studium alltäglichen Arbeitshandelns: Experten sind nicht durchweg besser als Novizen. Benutzern unterlaufen mit steigender Expertise nicht auch per se weniger Fehler. Im Gegenteil - Benutzer mit hoher Programm-Expertise machen sogar mehr Fehler als Benutzer mit niedriger Programm-Expertise.

Einen tieferen Einblick in den Zusammenhang zwischen Handlungsfehlern und Expertise gewinnt man erst dann, wenn beide Faktoren einer differenzierten Behandlung unterzogen werden. Für den Bereich der Handlungsfehler zahlt sich dabei eine Aufschlüsselung nach unserer Taxonomie aus. Für die Expertise bietet sich ein mehrdimensionaler Ansatz an, bei der verschiedene Dimensionen sowohl einzeln als auch in ihrem Zusammenwirken betrachtet werden.

Betrachtet man die vorgestellten fünf verschiedenen Herangehensweisen, so kommt es bei den Wissens-, Denk- und Unterlassensfehlern in drei Fällen, bei den Gewohnheits-

fehlern in zwei Fällen und bei den Bewegungsfehlern und den Ineffizienzen aufgrund fehlenden Wissens in jeweils einem Fall zu signifikanten Unterschieden. Urteils-, Merk-/Vergessens- und Erkennensfehler sowie Ineffizienzen aus Gewohnheit bleiben von der Expertise hingegen gänzlich unberührt. Betrachtet man diese Ergebnisse unter dem Gesichtspunkt des Handlungsprozesses⁵, so wird deutlich, daß neben der Regulationsgrundlage die Phase der Ziel- und Planformulierung sowohl auf der intellektuellen Regulations-ebene als auch auf der Ebene der flexiblen Handlungsmuster am empfindlichsten auf unterschiedliche Expertise reagiert. Die Phase des Monitorings wird lediglich auf der Ebene der flexiblen Handlungsmuster gestört, und die Phase der Rückmeldung wird von dem Grad der Expertise überhaupt nicht betroffen.

Diese spezifischen Sensibilitäten sowohl bestimmter Fehlerarten als auch bestimmter Benutzergruppen gelten jedoch zunächst lediglich für den Fehler an sich. Betrachtet man in einer weiterführenden Analyse die Fehlerkorrekturzeit als abhängig von Unterschieden in der Benutzerexpertise (zu einer ausführlicheren Darstellung s. Prümper et al., 1991), dann ist das Bild wesentlich einheitlicher. In all den Fällen, in denen es dort zu signifikanten Ergebnissen kommt, beheben die Experten - sowohl unabhängig von der Art des Fehlers als auch der Art der Operationalisierung der Expertise - die Fehler schneller als die entsprechenden Novizen. Dieses Ergebnis spricht dafür, daß Experten bei der Fehlerbewältigung generell Novizen überlegen sind.

In diesem Sinne lassen sich die Ergebnisse zu dem Zusammenhang zwischen Handlungsfehlern und Expertise als Hinweis für spezifische Verbesserungen im Bereich des Softwaretrainings (vgl. Kap. 11), der Softwareberatung (vgl. Kap. 12) und der Softwareentwicklung (vgl. Kap. 13) für einzelne Benutzergruppen heranziehen, wohingegen die Ergebnisse zur Fehlerkorrekturzeit allgemein für das Konzept des Fehlermanagements sprechen (vgl. Kap. 10).

⁵ Der Handlungsprozeß setzt sich so zusammen, daß am Anfang der Handlung ein Ziel aufgestellt und ein entsprechender Plan zur Zielerreichung entworfen wird (Ziel- und Planformulierung). Dieser Plan wird eine zeitlang im Gedächtnis gehalten (Gedächtnis/Monitoring-Phase) und am Ende der Handlung findet die Überprüfung statt, ob das angestrebte Ziel erreicht wurde oder nicht (Rückmeldungs-Phase). Zu einer ausführlichen Diskussion des Handlungsprozesses siehe Kap. 1.

Quelle:

Prümper, J. (1991). Handlungsfehler und Expertise. In M. Frese & D. Zapf (Hrsg.), *Fehler bei der Arbeit mit dem Computer – Ergebnisse von Beobachtungen und Befragungen im Bürobereich* (S. 118-130). Bern: Huber.