

Kontrollierte Evaluationsstudie von MicroPat — Interaktiver Atlas der Histopathologie

S. Schulz, T. Auhuber, R. Klar

Abteilung Medizinische Informatik
Universitätsklinikum Freiburg i. Br.
Stefan-Meier-Str. 26, D-79104 Freiburg i. Br.
http://www.imbi.uni-freiburg.de/medinf/mi_home.htm
schulz@ami1.ukl.uni-freiburg.de

Für die Weiterentwicklung von Bildungsstandards ist es notwendig, die Qualität neuer Lehr- und Lernmittel zu prüfen. Im Rahmen einer kontrollierten Studie wurde mit Hilfe eines Crossover-Versuchsplans der an der Universität Freiburg entwickelte Atlas für Histopathologie MicroPat auf die Verbesserung der Lernleistung, auf die inhaltliche Leistungsfähigkeit, auf Erweiterungsmöglichkeiten und auf seine Akzeptanz hin überprüft. Die Erfolgskontrolle am Mikroskop ergab eine Verbesserung der Lernleistung. Deren Ausprägung und Signifikanz hing jedoch stark davon ab, ob man die Zahl der Richtig-Antworten allein oder in Zusammenhang mit einem vom Benutzer erfragten Sicherheits-Score der Analyse zugrunde legt. Hinweise für einen Vorteil computererfahrener Probanden bezüglich des Abschneidens mit dem Lernsystem fanden sich nicht. Das Programm hinterließ einen durchweg positiven Gesamteindruck.

Keywords: CBT, Evaluation, Histopathologie

1 Grundlagen

In der Medizin ist die Verantwortung der Lehre für das zukünftige ärztliche Handeln besonders wichtig [4]. Deshalb ist es notwendig, sich über die Qualitätsmessung und -prüfung von Lehr- und Lernmitteln Gedanken zu machen. Bei dieser Studie wurde ein Modell zur Qualitätsprüfung für medizinische CBT-Programme entwickelt, das am Beispiel des elektronischen Histopathologie-Atlanten MicroPat getestet wurde.

MicroPat ist ein histopathologisches Lernsystem zur Unterstützung des Kurses „Allgemeine Pathologie“ für Medizinstudenten der Universität Freiburg. Das System ist in zwei verschiedene Module aufgeteilt: ein Atlas für Histopathologie (Abb. 1) und ein Lernprogramm mit Quizmöglichkeit [2].

Enthalten sind 900 digitalisierte Diapositive aus dem Bildarchiv des Pathologischen Instituts der Universität Freiburg, die die wesentliche Grundlage des Systems bilden. Zwischen den einzelnen Vergrößerungsstufen läßt sich über eine Zoomfunktion navigieren. Die Texte sind thematisch untergliedert in allgemeine Abschnitte wie Definition, Pathogenese und Klinik, die unabhängig vom gerade angezeigten Bild angesehen werden können. Die morphologischen Bildbeschreibungen sind für jede Abbildung spezifisch erstellt worden. Anmerkungen, Bildverweise

und Erklärungen zu Färbungen sind ebenso wie Querverweise als Hyperlinks realisiert. Eine Indexfunktion erlaubt eine Schlagwortsuche. Gemäß dem Aufbau des Mikroskopierkurses wurde das Text- und Bildmaterial topographisch angeordnet. Die Benutzerführung orientiert sich an gängigen GUI (graphical user interface)-Standards.

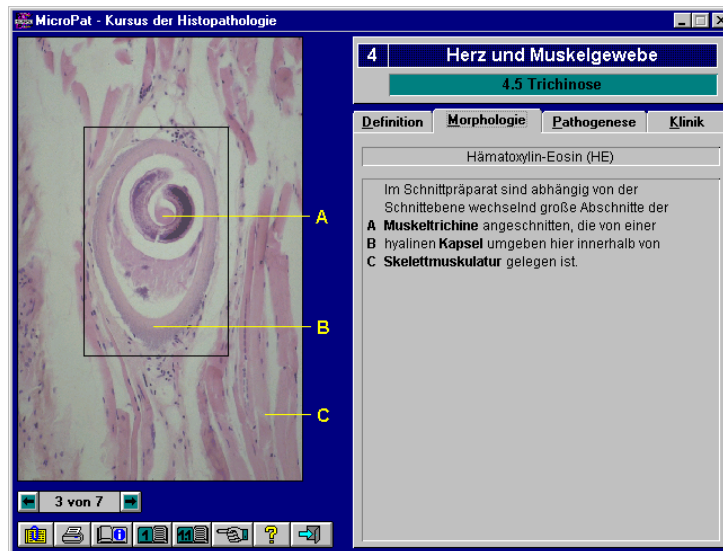


Abb. 1: Beispiel einer Bildschirmseite aus MicroPat

2 Studiendesign

Ziel der Studie war es, den Lernerfolg des Programms im Vergleich zu den zur Vorbereitung auf die Histopathologie-Testate üblicherweise verwendeten Lehrbücher zu messen. Dazu wurde anhand einer realen Prüfungssituation (Präparate-Erkennen am Mikroskop) ein Testat durchgeführt. Im Gegensatz zu der üblichen Praxis wurden gänzlich unbekannte Präparate vorgelegt, d. h. die Probanden mußten sich den Stoff selbständig mit Hilfe von Büchern bzw. mit dem Programm ohne tutorielle Hilfe erarbeiten.

Die Evaluation wurde an 72 freiwilligen Probanden aus dem ersten klinischen Studienabschnitt durchgeführt. Während der Studie war nur der Atlasteil des Systems aktiviert, und die Navigation war nur innerhalb desjenigen Themenbereichs möglich, welcher im Sinne dieser Evaluationsstudie bearbeitet werden sollte.

Die Studenten wurden randomisiert zwei Gruppen zugeordnet. Eine Gruppe sollte den Themenbereich A (Prostatakarzinom, invasiv lobuläres Mammakarzinom, invasiv duktales Mammakarzinom und Komedokarzinom) ohne Computer ausschließlich mit Büchern und den Themenbereich B (Adenomyomatose, Mastopathie, Fibroadenom und Gallertkarzinom) mit Computer und nur fakultativ mit Büchern bearbeiten. Die andere Gruppe verfuhr umgekehrt (Crossover-Versuchsplan [3]). Nach der Vorbereitung mußten die Studenten aus den beiden Themenbereichen je drei unbekannte Präparate mit Hilfe des Mikroskops diagnostizieren. Neben der Diagnose wurden die Probanden nach der Sicherheit ihrer Aussage auf einer dreistufigen Skala (von „Ich bin bei meiner Diagnose ganz sicher.“ = 3 Punkte, bis „Ich habe bei meiner Diagnose geraten.“ = 1 Punkt) gefragt. Weiterhin sollten sie 17 Fragen über die eigenen Com-

puterkenntnisse, die Beurteilung der Bilder, der Texte und des Programms allgemein per Fragebogen beantworten. Die Bewertung erfolgte hier auf einer Ordinalskala von eins (= „sehr gut / absolut richtig“) bis fünf (= „sehr schlecht / absolut falsch“).

3 Ergebnisse und Diskussion

Deutlicher Vorteil für MicroPat nur bei Analyse der Scores

Unter Berücksichtigung des Sicherheits-Scores läßt sich errechnen, daß 58.3% der Probanden mit dem Computer bessere Resultate erzielen konnten, 31.9% mit Büchern besser zurechtkamen und 9.7% der Studenten in beiden Prüfungsabschnitten gleich gut abschnitten. Nach dem Vorzeichentest bzw. McNemar-Test sind die Ergebnisse auf einem Niveau von $\alpha=0.05$ signifikant. Der Wilcoxon-Test konnte die Signifikanz ebenfalls bestätigen. Bildet man die Differenz des Computerergebnisses mit dem Buchresultat unter Berücksichtigung der unveränderten Lernleistungen, ergibt sich für das Gesamtkollektiv ein globaler Vorteil von annähernd 17%.

Betrachtet man dagegen nur die Richtig- bzw. Falsch-Antworten (Tab. 1, untere Hälfte), so ist nur ein geringer Unterschied zwischen Computer- und konventioneller Testatvorbereitung erkennbar, welcher nicht signifikant ist.

Die Entscheidung für ein Score-System hat sich dennoch als sinnvoll herausgestellt, da die Ergebnisse differenzierter betrachtet werden können und Tendenzen der Fragenbeantwortung besser klar werden. Wie diese Studie jedoch deutlich zeigt, muß der Score-Betrachtung eine rein objektive Lernerfolgskontrolle im Sinne einer Richtig-Falsch-Analyse gleichberechtigt gegenübergestellt werden.

	arithmetisches Mittel	Median
<i>Score [-9;+9] beschreibt die Summe der Einzelscores von drei Präparaten nach folgendem Schema: 1=„geraten“, 2=„relativ sicher“, 3=„sicher“ (Angabe des Probanden), multipliziert mit 1 für eine richtige oder -1 für eine falsche Diagnose (nach Auswertung)</i>		
Score Themenbereich A	0,6	0
Score Themenbereich B	3,5	4
Score Computer	2,6	3
Score Buch	1,5	2
<i>korrekte Diagnosen [0;+3] beschreibt die Anzahl der richtigen Diagnosen von drei Präparaten</i>		
korrekte Diagnosen Themenbereich A	1,6	1
korrekte Diagnosen Themenbereich B	2,2	2
korrekte Diagnosen Computer	2,0	2
korrekte Diagnosen Buch	1,8	2

Tab. 1: Ergebnisse nach Themenbereichen und nach Lernarten

Es gab leichte und schwere Themenbereiche.

Nach der exakten Kontrolle der erreichten Punktzahlen stellte sich heraus, daß die Themenbereiche unterschiedlich schwer waren. Die Probanden schnitten im Themenbereich B (Adenomyomatose, Mastopathie, Fibroadenom und Gallertkarzinom) wesentlich besser ab (Tab. 1). A priori waren die Prüfungsgebiete so ausgewählt, daß es möglichst zusammenhängende Themen sind, daß alle Organsysteme paritätisch verteilt sind und Verwechslungsmöglichkeiten gegeben sind. Aus der Sicht des Dozenten waren beide Themenbereiche als gleich schwer beurteilt worden. Für die schlechteren Resultate bei der Bearbeitung von Themenbereich A liegt der Grund vor allem im Vertauschen vom Präparat „Invasiv lobuläres Mammakarzinom“ mit dem Präparat „Invasiv duktales Mammakarzinom“. Die besten Ergebnisse wurden beim „Fibroadenom“, die schlechtesten beim „invasiv duktales Mammakarzinom“ erreicht.

Varianzanalyse zum Ausschluß von Wechselwirkungen

Da keine Normalverteilungen vorlagen – die summierte Score-Variable ist am ehesten als binomialverteilt auf einem Träger von -9 bis +9 zu beschreiben –, wurde eine Winkeltransformation (Arcus-Sinus-Transformation) [5] vorgenommen, um die erhaltenen Werte dann einer Varianzanalyse zu unterziehen. Hierbei konnten Wechselwirkungen mit folgenden Variablen ausgeschlossen werden: Prüfungstag, Reihenfolge der Bearbeitung, Auswahl der Präparate. Dagegen fiel auf, daß die Aufteilung in einen leichten und einen schweren Themenbereich nur bei den Präparaten der Computergruppe hochsignifikant ist ($\alpha=0.001$); für die Buchgruppe läßt sich diese Wechselwirkung nicht nachweisen.

Zusammen mit dem Befund, daß für das Computer-Lernen signifikant höhere Sicherheitsangaben (Wilcoxon, $\alpha=0.05$) vergeben wurden, unabhängig davon, ob die Antworten sich danach als richtig oder als falsch erwiesen), wird deutlich, daß die höheren Scores im wesentlichen durch diese höhere subjektive Sicherheit bedingt sind. Damit ist auch erklärt, warum dieser Vorteil nur bei der „leichteren“ Präparategruppe auftrat. Bei der „schwereren“, mit einem Median von Null (Tab. 1, obere Hälfte) halten sich positive und negative Effekte die Waage. Denkbare Gründe sind einerseits die Vergabe höherer Scores für das „Computer-Lernen“ im Sinne einer Einflußnahme auf die Studie – die Probanden wußten ja, daß das Lernprogramm evaluiert werden sollte – und andererseits eine mögliche Beeinflussung der Studenten durch das Medium Computer, das eine trügerische Sicherheit vermitteln könnte.

Da die Karzinome im Vergleich zu anderen Präparaten immer weniger gut diagnostiziert wurden, kann man vermuten, daß das Verständnis für diese Präparate einen höheren zeitlichen und didaktischen Mehraufwand erfordert. Inhaltliche Fehler in den Büchern und im Programm konnten ausgeschlossen werden

Das Crossover-Modell hat sich als Modell zur Evaluation dieses CBT-Programms bewährt.

Da kontrollierte Evaluationsstudien oft fehlen oder schlecht vergleichbar sind [1], sollte mit dieser Crossover-Versuchsplanung mit anschließender, den realen Verhältnissen entsprechenden Prüfung, eine leicht durchführbare und dennoch aussagekräftige Validierungsmethode entwickelt werden. Der Crossover-Plan hat einige Vorteile. Jeder Proband ist gleichzeitig Test- und Kontrollperson. So kann die Anzahl der benötigten Teilnehmer verringert werden. Außerdem können einige Effekte, wie z. B. ein Motivationsverlust bei der Kontrollgruppe, vermieden werden, da jeder Proband an allen Elementen der Studie teilnimmt. Der Nachteil eines solchen Verfahrens liegt in Perioden- und Residualeffekten [3]. Periodeneffekte, wie z. B. ein Nachlas-

sen der Motivation, eine Veränderung der Prüfungsbedingungen oder Gewöhnungseffekte während der Tage der Durchführung, konnten nicht nachgewiesen werden. Durch die Variation der Lernfolge und der Testformen wurden darüber hinaus Residualeffekte vermieden. Wechselwirkungen innerhalb der unterschiedlichen Kollektive wurden in der statistischen Analyse gut sichtbar.

Es konnten aus einer Gesamtheit von 237 Studenten 72 gewonnen werden, was unter den statistischen Bedingungen eines nicht normalverteilten Kollektivs bezüglich der erreichten Punktzahl eine eher geringe Zahl ist. Da die Teilnahme freiwillig war, kann die Stichprobe nicht als repräsentativ betrachtet und eine mögliche Überrepräsentation PC-erfahrener Studenten nicht ausgeschlossen werden. Allerdings muß man auch die zeitliche Terminierung (letzte Semesterwoche) als wesentlichen Grund der Nicht-Teilnahme und die Verlosung von Lehrbüchern unter den Probanden als zusätzlichen Anreiz für die Teilnahme in Betracht ziehen. Allerdings ist eine Gruppe von 72 Personen aus einer Gesamtheit von 237 so groß, daß damit unter Verwendung eines Crossover-Versuchsplans charakteristische Aussagen über eine bedeutende Untermenge von 30% getroffen werden können.

Fragebogen als Ergänzung zum histopathologischen Testat

Als Ergänzung zur Hauptfrage nach dem Erfolg in einer histopathologischen Prüfung wurden die Probanden aufgefordert einen Fragebogen (Abb. 2) auszufüllen.

Die Auswertung des Fragebogens ergab, daß die Bedienbarkeit und die Schnelligkeit des Programmes durchschnittlich mit der Note 1.5 bewertet wurde. Ebenfalls gut fiel die Benotung der Bilder mit 1.8 aus; mit 2.4 Punkten schnitten die Texte deutlich besser als befriedigend ab. Mit Werten von durchschnittlich 1.6 hinterließ das Programm einen durchweg positiven Gesamteindruck.

Die Probanden halten ein zusätzliches Begleitheft zum Programm im großen und ganzen für nicht notwendig. Ob das Programm ein Buch für Histopathologie ersetzen kann, wurde sehr different beantwortet. Im Gegensatz dazu sind sich die Testpersonen weitgehend einig, daß ein solches Lernprogramm für eine Ergänzung durch ein Lehrbuch für Pathologie weniger geeignet ist.

Eine Korrelation der „Einschätzung der eigenen Computerkenntnisse“ (aus dem Fragebogen) mit der Differenz des Lernerfolgs („Computer-Lernen“ minus „Buch-Lernen“) findet sich nicht (Spearman'scher Korrelationskoeffizient 0.04). Dasselbe Ergebnis erhält man auch unter Vernachlässigung der Scores (Spearman'scher Korrelationskoeffizient -0.03). Die Freude am Arbeiten mit dem PC hat keinen Einfluß auf die Lernleistung. Dieses Bild wird vervollständigt durch die Überzeugung der Studenten, daß sie nach ihrer eigenen Einschätzung nur teilweise schneller und effizienter lernen. Allerdings kann durch die Angst vor dem neuen System oder durch die Begeisterung beim ersten Kontakt eine Beeinflussung der Lernleistung nicht ausgeschlossen werden [7]. Mit Ausnahme einiger Werte, die jedoch für eine sinnvolle Beurteilung nicht geeignet sind, konnten Abhängigkeiten vom Prüfungstag und der Prüfungsform nicht nachgewiesen werden.

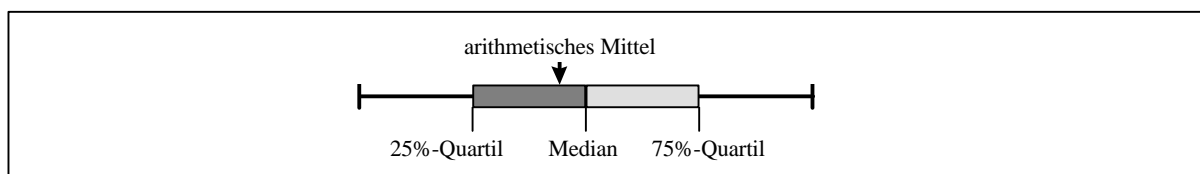
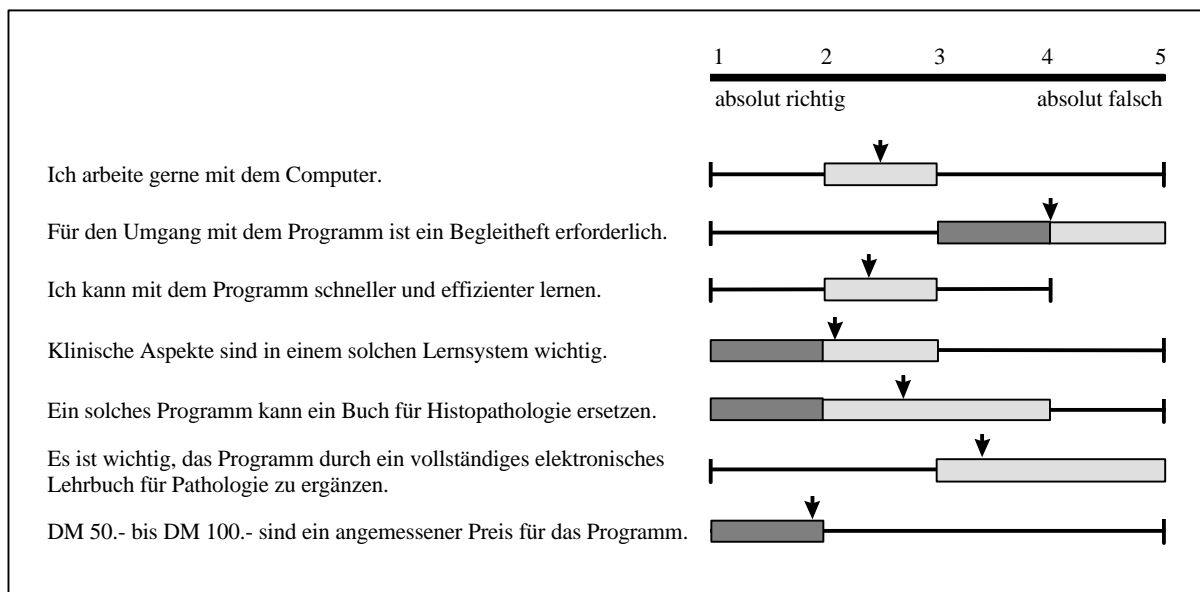
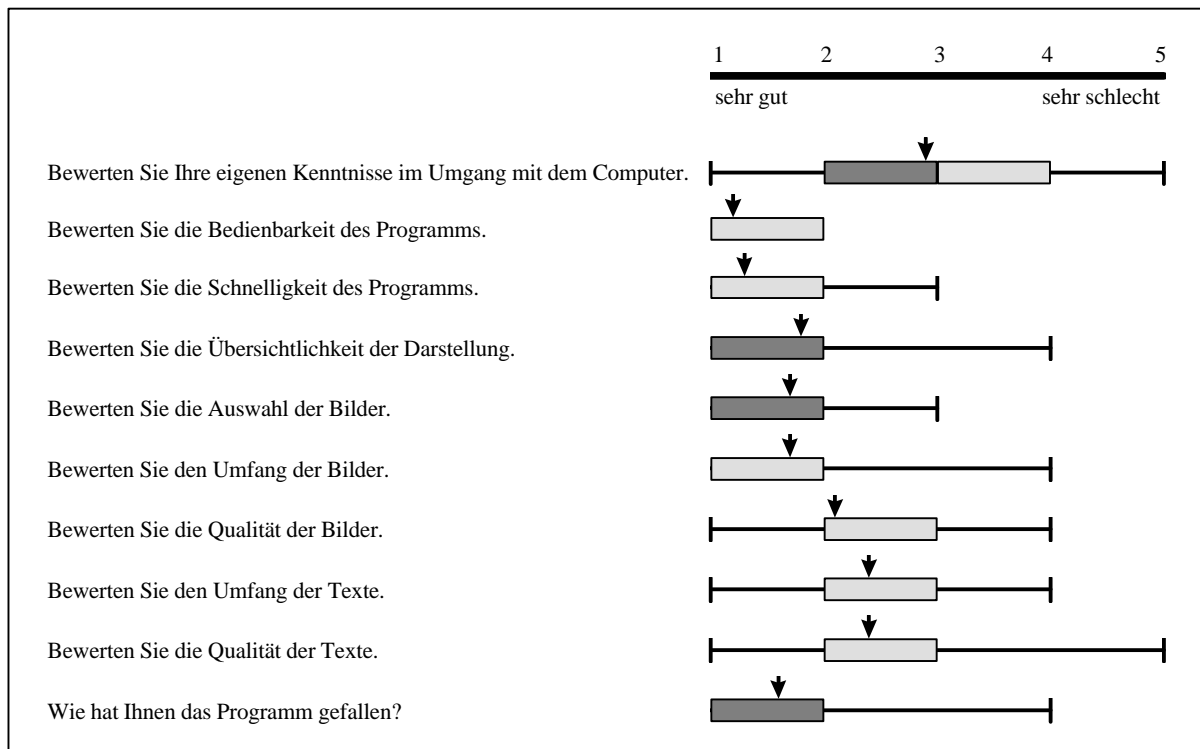


Abb. 2: Übersicht über die Ergebnisse des Fragebogens als Boxplot

Eine weitergehende Analyse der Daten sowie zusätzliche Studien zur Messung des Lernerfolgs von CBT-Systemen sind notwendig. Das hier vorgestellte Evaluation zeigt hierbei interessante Ergebnisse. Unter Berücksichtigung der Akzeptanz bei den Studenten und unabhängig von der didaktischen Beurteilung von Lernsystemen ist die Tendenz zur CD-ROM-basierten Publikation unverkennbar, können doch mit diesem Medium große Bilddatenmengen in guter Farbqualität und mit einem, im Gegensatz zum Buch, geringen finanziellen Aufwand dargestellt werden [4][6].

Literatur

- [1] Fieschi, D.; Fieschi, M.; Soula, G.; Degoulet, P.: Evaluation of computer-assisted instruction methods. Apropos of 26 comparative studies published between 1989 and 1992. *Pathologie Biologie*. 42 (2). 1994. 183-190.
- [2] Klar, R.; Auhuber, T.; Schulz, S.: MicroPat – An Example for Integrating an Atlas of Digital Images into a Computer Based Training System. in: *Health Telematics Education*. 1997 (in print).
- [3] Lehmacher, W.: Zwei Stichproben von Verlaufskurven und Crossover-Versuche. *Medizinische Informatik und Statistik* 67. Heidelberg. Springer. 1986. 77-85.
- [4] Leven, F. J.; Schulz, S.; Alle, W.; Klar, R.: Rechnergestützte Lehr- und Lernsysteme in den Klinika: Stand und zukünftige Entwicklungen. in: Buchholz, W.; Haux, R. (Hrsg.). *Informationsverarbeitung in den Universitätsklinika Baden-Württembergs*. Heidelberg. 1995. 187-193.
- [5] Sachs, L.: *Angewandte Statistik. Anwendung statistischer Methoden*. 7., völlig neubearbeitete Auflage. Berlin. Springer. 1992. 354-355.
- [6] Schulz, S.; Schrader, U.; Klar, R.: Computer-based Training and Electronic Publishing in the Health Sector: Tools and Trends. in: *Methods of Information in Medicine* (in print).
- [7] Selbmann, H.-K.: Validierung von Software und Teachware. in: Baur, M.; Michaelis, J. (Hrsg.). *Computer in der Ärzteausbildung*. München. Oldenbourg. 1990. 165-173.