

Proseminar hypermediale Lernsysteme
Kapitel 12
Im Land der Nullhypothesen

Ausarbeitung des Vortrags von:

Peter Keller

Matr.Nr.:711003

27. Februar 2003

Inhaltsverzeichnis

1	Bemerkungen	3
2	Pädagogische Mythen des Computerlernens	4
3	Historische Bedingtheit von Evaluationen	5
3.1	Beispiele für überholte Studien	5
3.1.1	Qualität und Auswirkungen von Bildschirmen mit Bildröhre	5
3.1.2	Lesen am Bildschirm versus Lesen auf Papier	6
3.1.3	Forschungen zu Eigenschaften von Bildschirmdarstellungen	6
3.1.4	Antwortzeiten des Rechners	6
3.1.5	Standbildsequenz versus Video	6
3.2	Gründe für die laufende Veränderung der Parameter	7
3.2.1	Erfindung immer neuer Interaktionsformen und Oberflächen.	7
3.2.2	Leistungsfähigere Hardware und mächtigere Grafikwerkzeuge erhöhen Qualität und Performanz der Anwendungen	7
3.2.3	Entwicklung der Oberflächen von einfachen textbasierten Anwendungen zu multimedialen, multi-Window, multi-Desktop-Anwendungen	8
3.2.4	Sinkende Kosten für Hardware	8
3.2.5	Einholen der Hardwareprobleme durch Software	8
3.2.6	Nivellierung des Novitätseffektes	8
3.3	Schlussfolgerung	8
4	Methodenvergleich und Kontrolldesign	9
4.1	Ergebnisse	9
4.1.1	Lernzielabhängigkeit	9
4.1.2	Eignungsabhängigkeit	9
4.1.3	Vorbereitung der Umgebung	9
4.1.4	Motivation durch Novität	10
4.1.5	Nivellierung durch Berücksichtigung des Hawthorne-Effekts	10
4.2	Geringe Typisierungsfähigkeit	10
4.3	Falsche Ettikettierung	11
4.4	Differenzierung der Variablen	11
4.5	Auskunfts-fähigkeit der Versuchspersonen	12

5	Meta-Analysen zum Computerlernen	13
5.1	Die Rolle des Lehrers	14
5.2	Interaktivität	14
5.3	Methodologische Genauigkeit	15
5.4	Grundlegende methodologische Einwände	15
5.5	formative Evaluation	17
5.6	Nichtuntersuchte Faktoren	17
6	Quellen	19

1 Bemerkungen

Dieser Text ist die Ausarbeitung des Vortrags zum Thema „Im Land der Null-Hypothesen“. Die Ausarbeitung und der Vortrag beschäftigen sich mit der wichtigen und eindringlichen Frage, ob Multimedia überhaupt einen verstärkten Lerneffekt erzielen kann. Desweiteren werden Fragen nach wesentlichen Faktoren gestellt, die das Lerngeschehen am Computer beeinflussen und welche Ergebnisse bisherige Analysen ergeben haben.

Der Vortrag wurde im Rahmen des Proseminars „hypermediale Lernsysteme“ der Universität Potsdam gehalten und die Ausarbeitung orientiert sich stark an den Folien des Vortrags. Änderungen der Reihenfolge des Inhaltes der Folien ergaben sich nur dann, wenn es aus Sicht des Autors notwendig erschien. Sie dienen im wesentlichen nur der besseren Verständlichkeit des Textes und des besseren Leseflusses.

Peter Keller, 27. Februar 2003

2 Pädagogische Mythen des Computerlernens

Im Land der Null-Hypothesen

Der Begriff „Im Land der Null-Hypothesen“ entstammt einem Kommentar von Elshout(1992) zu einer vergleichenden Analyse zum Paarlernen. Bei den insgesamt 22 durchgeführten Studien ergaben sich nur in zwei Fällen signifikante Unterschiede. Die eine bevorzugte das Paarlernen, die andere fand das Gegenteil heraus. Daraus kann man schließen, dass die meisten experimentellen Vergleiche von Unterrichtsmethoden keine Unterschiede zwischen den Methoden feststellen können, oder dass sich die wenigen relevanten Ergebnisse auch noch gegenseitig widersprechen.

Auch Clark(1983) kommt zu einem ähnlichen Ergebnis, als er Studien zum Lernen mit dem Radio aus den 50'ern, Lernen mit dem Fernseher aus den 60'ern und Studien über das Computerlernen danach untersucht. „Ähnliche Fragestellungen bei den Untersuchungen der Medien resultierten in ähnlichen Ergebnissen und Daten. Mediavergleichsstudien, unabhängig vom untersuchten Medium, tendieren zu Ergebnissen, bei denen kein signifikanter Unterschied feststellbar ist.“

Die Gründe hierfür sind vielfältig:

- Viele der Studien sind stark abhängig vom aktuellen Stand der Technik. Durch die rasante Entwicklung der Computertechnologie in den letzten Jahren, sind die meisten Studien bereits nach sehr kurzer Zeit veraltet oder sogar hinfällig. Aus diesem Grunde sind bereits Studien, die einen Zeitraum von wenigen Monaten überblicken, als Longitudinal-Studien zu bezeichnen. D.h. diese Studien erheben gar nicht mehr den Anspruch überhaupt für einen längeren Zeitraum gültig zu sein. Dies führt natürlich bei Untersuchungen von Studien, die mehrere Jahre oder Jahrzehnte überdecken, immer wieder zu widersprüchlichen Ergebnissen.
- Zum Teil werden Lehrmethoden verglichen, die nicht oder nur sehr schwierig vergleichbar sind, wodurch es zum Teil zu einer sehr unfairen Behandlung der Einzel-Faktoren kommen kann. Zudem sind die Kontrolldesigns von sehr vielen Variablen abhängig und mit sehr schwierigen Problemen behaftet, so dass eine solide und realistische Auswertung der Ergebnisse beinahe unmöglich erscheint.
- Die untersuchten Studien basieren zum Teil auf veralteten Theorien und sind daher eher von geringer Relevanz für die Analysen. Trotzdem

werden immer wieder auch solche Studien einbezogen, weswegen es auch hier immer wieder zu Widersprüchen und Null-Ergebnissen kommt.

3 Historische Bedingtheit von Evaluationen

Evaluationen sind nicht losgelöst vom zeitlichen Kontext betrachtbar. Das heißt, dass Studien, die vor zwanzig Jahren durchgeführt wurden, andere Mittel zur Analyse benutzen müssen, als heutige. Das liegt zum großen Teil an der rasanten Entwicklung der Technologie und der sich ständig ändernden Bedingungen. Daraus ergibt sich auch eine notwendige dynamische Entwicklung der Untersuchungsmethoden. So ist es kaum möglich Untersuchungsmethoden aus den 70'ern auf Problemfragen der 90'er anzuwenden. Aus diesem Grunde wird auch der Vergleich zwischen zeitlich weit auseinanderliegenden Studien immer komplizierter.

Ebenso wie die ständig wechselnde Technologie ist auch der Faktor Mensch und Gesellschaft wesentlich für die Untersuchung von Lerneffekten. Es ist bis heute nicht hundertprozentig geklärt, welchen Einfluss der Computer auf die Sozialisierung des Menschen, speziell der Jugend, nimmt bzw. genommen hat. Parallel zur ständigen Entwicklung der Technologie werden auch die Einflüsse des Computers auf den Menschen einer immer rasanteren Entwicklung unterworfen.

Weiterhin ist nicht bekannt, welchen Einfluss der Computer auf die Entwicklung der Gesellschaft genommen hat und welche Faktoren dabei eine Rolle spielen bzw. wie diese zusammenhängen. Sicher ist nur, dass diese Faktoren Einfluss auf die auftretenden Effekte beim Lernen mit hypermedialen Systemen nehmen und einer eingehenden Untersuchung bedürfen.

3.1 Beispiele für überholte Studien

3.1.1 Qualität und Auswirkungen von Bildschirmen mit Bildröhre

Untersuchungen zu diesem Thema sind fast vollständig überflüssig geworden, denn hochauflösende und strahlungsarme Bildschirme sind schon seit langem Standard und werden ständig verbessert und optimiert. Zudem werden immer neue Technologien entwickelt, man denke nur an die neuen Flachbildschirme, die sicher schon bald Standard sein werden, oder an ehrgeizige Projekte, bei denen man versucht Bildschirme in Glasscheiben einzubetten oder ähnliche noch in der Entwicklung befindliche Visionen.

3.1.2 Lesen am Bildschirm versus Lesen auf Papier

Auch diese Untersuchung zeigt ein typisches Verhalten im zeitlichen Kontext. So wurde noch vor wenigen Jahren untersucht, in wie weit sich die Lese/Erinnerungsqualität beim Lesen vom Bildschirm vom Lesen auf Papier unterscheidet. Damals fand man heraus, dass das Lesen auf Papier deutliche Vorteile bringe. Zu beachten ist dabei, dass damals noch Bildschirme mit einer geringen Auflösung, Pixelfonts und monochromen Bildschirmen benutzt wurden und zudem die Programme noch nicht über Scroll- und Sprungfunktionen verfügten.

Würde man eine solche Untersuchung mit aktuellen Techniken wiederholen, würde sich wahrscheinlich ein anderes Ergebnis zeigen. Im Zeitalter von antialiased Fonts, Vektorfonts und hochauflösenden Bildschirmen bringt das Lesen auf dem Bildschirm sicherlich Vorteile gegenüber schlecht ausgeleuchteten und unter Umständen schlecht gedruckten Texten auf Papier.

3.1.3 Forschungen zu Eigenschaften von Bildschirmdarstellungen

Wie schon beim Vergleich zwischen dem Lesen auf Papier und dem Lesen am Bildschirm, so ergibt sich auch bei diesem Beispiel die Irrelevanz der Untersuchung durch die rasante Entwicklung der Technik. Denn nicht nur die Hardware entwickelt sich immer weiter und ermöglicht damit immer aufwendigere Darstellungen in immer kürzerer Zeit, sondern auch die Software. Immer bessere und schnellere Algorithmen ermöglichen immer bessere Performanz. So wurde etwa vor kurzem ein Algorithmus entwickelt, der die dreidimensionale Darstellung von Wasser in Echtzeit erlaubt, d.h. der die Manipulation von und Interaktion mit wasser-ähnlichen Oberflächen erlaubt.

3.1.4 Antwortzeiten des Rechners

Untersuchungen von Antwortzeiten von Computers sind heute nur noch in verteilten Systemen oder in Bezug auf das Internet relevant, denn die rasante Entwicklung der Technologie wirkt auch hier als entscheidender Faktor mit. Bei Prozessorgeschwindigkeit von mittlerweile über zwei Gigahertz und RAM-Speichern von einem Gigabyte sind selbst aufwendigere Darstellungen und Berechnungen fast ohne Zeitverlust möglich.

3.1.5 Standbildsequenz versus Video

Christel beschäftigte sich 1994 mit der Frage, wie die Erinnerungsfähigkeit von Versuchspersonen beschaffen ist, die eine Videosequenz mit Text bzw. eine Standbildsequenz mit Text betrachteten. Er fand dabei heraus, dass 89%

der Probanden sich nach dem Betrachten einer Videosequenz an die Informationen erinnern konnten, bei denjenigen, die die Standbildsequenz gesehen hatten nur 71%. Christel kommt zu dem Ergebnis, dass höherwertige Grafik, dass Lernen unterstützt. Schulmeister merkt dazu an, dass dieses Beispiel kontextabhängig ist und man mit geeigneten Mitteln auch das Gegenteil beweisen könne. Legt man den Gedanken zugrunde, dass das gezeigte Video auch vom Inhalt des Textes ablenken kann, versteht man Schulmeisters Einwände.

Desweiteren ist auch der Einsatz von Standbildsequenzen beim Design von Lernprogrammen mittlerweile unüblich, da Videosequenzen einfach herzustellen und abzuspielen sind und das in guter bis sehr guter Qualität ohne Ruckeln oder Hängen (natürlich nur bei modernen Rechnern).

3.2 Gründe für die laufende Veränderung der Parameter

Die Gründe für die laufende Veränderung der Parameter ist ebenso vielfältig, wie die Beispiele, die sich für überholte Studien finden lassen.

3.2.1 Erfindung immer neuer Interaktionsformen und Oberflächen.

Als Beispiel sei hier nur die Erfindung des window-orientierten Betriebssystems Windows erwähnt. Dieses Programm bot, bei seiner Einführung, einen völlig neuartigen Umgang mit den Daten und mit der Erfindung der Maus einen völlig neuen Umgang mit dem Computer, der die nachfolgende Entwicklung mit Sicherheit stark geprägt hat.

3.2.2 Leistungsfähigere Hardware und mächtigere Grafikwerkzeuge erhöhen Qualität und Performanz der Anwendungen

Grafikkarten mit eigenem Speicher, spezielle Hardwarelösungen für Berechnungen von Grafiken und eigene Grafikprozessoren erhöhen die Performanz von Bildschirmdarstellungen. Softwarealgorithmen, die Darstellungen optimieren und Rechenzeit sparen erhöhen ebenfalls die Performanz und verändern damit die Möglichkeiten zum Einsatz von Multi- und Hypermedia.

3.2.3 Entwicklung der Oberflächen von einfachen textbasierten Anwendungen zu multimedialen, multi-Window, multi-Desktop-Anwendungen

Die Veränderung der Programmoberflächen von einfachen text-basierten schwierig zu bedienenden Programmen, hin zu bunten multimedialen und einfach mit der Maus zu bedienenden Programmen, hat die Einstellung zum Computer geändert und immer mehr User unterschiedlichsten Alters interessiert und animiert. Daher ist auch dieser Punkt ein entscheidender Faktor, wenn man Studien über einen längeren Zeitraum untersuchen will.

3.2.4 Sinkende Kosten für Hardware

Durch die rasante Entwicklung der Technologie wird die Anschaffung neuer Hard- und Software immer attraktiver. Zeitgleich sinken, durch die erhöhte Nachfrage und ständige Aktualisierung, die Preise und ermöglichen so selbst finanzschwachen Mitgliedern der Gesellschaft den Zugang zu Computern. Auch hier ändert sich die Einstellung der Menschen zum Medium Computer und beobachtete Effekte beim Lernen können schnell ihre Gültigkeit verlieren.

3.2.5 Einholen der Hardwareprobleme durch Software

Durch neue und die Optimierung bestehender Algorithmen können sich Hardwareprobleme softwaretechnisch lösen. So ist zum Beispiel eine niedrige Auflösung mit geringer Farbtiefe mit Hilfe von Dithering zu vermindern, andererseits kann man mit Hilfe von Antialiasing Treppeneffekte bei niedriger Auflösung entgegenwirken, indem man zwischen Farben, die sich stark unterscheiden, mit Hilfe von Zwischenfarben überblendet.

3.2.6 Nivellierung des Novitätseffektes

Aufgrund der immer stärker werdenden Einbeziehung des Computers in das alltägliche Leben und der ständigen Konfrontation schleifen sich Novitätseffekte schnell ab. Anfängliche Begeisterung für das neue Medium wird dementsprechend schnell von einer kritischeren Einstellung abgelöst, die wiederum die Ergebnisse der Studien stark beeinflusst.

3.3 Schlussfolgerung

Negroponte kommt anhand der eben genannten Gründe selbst zu dem Schluss, dass es kein universelles Bestes geben könne, weil Menschen unterschiedlich seien, Situationen sich verändern und die Umstände der Interaktion vom

verfügbaren Medium geleitet würden. Daher könne es auch kein ultimatives Design von Schnittstellen zur Evaluationen geben.

Die von Negroponte aufgestellte Behauptung erscheint logisch. Die genauen Gründe werden in den nächsten Kapiteln näher beleuchtet und erleichtern den Zugang zum Verständnis der Meinung Negropontes bzw. untermauern diese.

4 Methodenvergleich und Kontrolldesign

Die Untersuchung der Methoden und der Kontrolldesigns deckte einige interessante Aspekte auf, die beim Umgang mit Lernsystemen eine Rolle spielen.

4.1 Ergebnisse

4.1.1 Lernzielabhängigkeit

Der Effekt des Lernens hängt deutlich vom Lernziel ab. 1993 untersuchten Shiu und Smaldino ein Computerprogramm zum Chinesisch-Lernen mit einem Tonband. Sie fanden heraus, dass die Probanden deutlich bessere Ergebnisse in den Wochen erzielten, in denen das Computerprogramm eingesetzt wurden. Allerdings zeigten sich bei Zuhören und Verstehen keine Unterschiede. Unterschiede zeigten sich nur dann, wenn es um das Schreiben chinesischer Schriftzeichen ging.

Je nach Lernziel-Setzung unterscheidet sich das Ergebnis also wesentlich und eine genaue Ursachenanalyse bzw. -differenzierung ist schwierig und zum Teil nicht einmal eindeutig.

4.1.2 Eignungsabhängigkeit

Die Effektivität eines Lernprogramms ist auch davon abhängig, inwieweit sich überhaupt das angestrebte Lernziel für ein interaktives Programm eignet.

4.1.3 Vorbereitung der Umgebung

Die Umgebungen, in denen die zu untersuchenden Verfahren/Methoden getestet werden sollen, müssen optimal vorbereitet werden, damit die jeweiligen Methoden ihre Vorteile voll ausreizen können. Ist dies nicht der Fall oder sind sich die Umgebungen zu ähnlich, so ergeben sich keine signifikanten Unterschiede mehr.

4.1.4 Motivation durch Novität

Der Novitäts- oder Hawthorne-Effekt erzeugt eine außergewöhnliche Motivation der Probanden, die die Untersuchungsergebnisse stark verfälscht. Dieser wird durch die Konfrontation mit einem neuen Medium hervorgerufen und ist wohl auf die allgemein positive Einstellung zu Computerprogrammen zurückzuführen. Zu diesem Thema müssen noch weitere Studien durchgeführt werden, um die tatsächlichen Ursachen für diesen Effekt herauszufinden.

4.1.5 Nivellierung durch Berücksichtigung des Hawthorne-Effekts

Unter Berücksichtigung des Hawthorne-Effektes ergeben sich dann keine Unterschiede, wenn man starke pädagogische Mittel einsetzt. Zu Evaluation dieser Behauptung hat man einer Gruppe von Schüler ein Abenteuerspiel für den PC vorgestellt und mit einer anderen Gruppe führte man im Klassenverband ein interaktives Spiel zur Geographie durch. Bei der Untersuchung der Ergebnisse stellte man keine Unterschiede bezüglich des Lerneffektes fest. Weder die eine noch die andere Gruppe zeigte bessere Ergebnisse.

Atkins und Yildiz meint dazu, dass man damit aufhören müsse, die Effektivität von Multi-Media Simulationen im Vergleich zu anderen Technologien zu testen, um sich stattdessen darauf zu konzentrieren die Relationen und Lern-Eigenschaften zu erforschen.

4.2 Geringe Typisierungsfähigkeit

Die in den Studien untersuchten Programme sind sehr unterschiedlich designed und konzipiert. Aus dieser breiten Fächerung des Designs ergibt sich zwangsläufig eine geringe Typisierungsfähigkeit. Es ist bei den meisten Programme nicht eindeutig möglich einen genauen Typ festzustellen. So gibt es Programme, die, je nach Sichtweise, dem einen oder anderen Programmtyp angehören. Auch die ständige Weiterentwicklung bestehender Programmtypen oder die Erfindung neuer Gattungen ist der geringen Typisierungsfähigkeit nicht förderlich. Daraus ergeben sich weitgehende Schwierigkeiten.

Auch die breite Varianz der Lehrmethoden und deren Vermischung verhindert eine genaue Diskretisierung und somit eine genaue Untersuchung. Das führt sogar soweit, dass generelle Aussagen nicht mehr möglich sind.

Green meint 1991 dazu dass man die Untersuchung in zwei Gebiete unterteilen müsse. Zum einen solle man das Design evaluieren, d.h. die tatsächliche Implementierung müsse untersucht und eingeordnet werden. Zum anderen müsse der Designtyp und dessen Eigenschaften evaluiert werden um, das Programm in einen größeren Kontext einzuordnen.

4.3 Falsche Ettikettierung

Leider werden immer wieder Fragestellungen untersucht, die keinerlei Bezug mehr zur aktuellen Situation und dem Stand der Technik haben. So wird ernsthaft untersucht wie lange man braucht die Maus zu einem bestimmten Menüpunkt zu bewegen oder wieviel Zeit zum Ausfüllen eines Formulars benötigt wird. Diese Fragestellungen sind zwar zweifellos interessant, bringen der Forschung aber eher wenig. Green schlägt 1991 vor, man solle doch eher untersuchen, welche Strategien benutzt werden, um in einer Lernumgebung zu navigieren oder sich anderen zentralen kognitiven Konzepten zuwenden, anstatt sich mit unsinnigen Untersuchungen aufzuhalten.

4.4 Differenzierung der Variablen

Wie schon in den vorangegangenen Ausführungen bemerkt wurde, ist das Kontrolldesign mit sehr schwierigen und teilweise schwer zu verifizierenden Variablen und Faktoren behaftet. Alle Faktoren zu beachten und in ihrer vollen Wertigkeit einzubeziehen ist schwierig und sogar teilweise nicht möglich, aufgrund ihrer Komplexität. Schulmeister bemerkt sogar, dass die Anzahl der Variablen gegen Unendlich gehe.

Einige der wichtigsten Faktoren hierbei sind:

- Länge der linearen bzw. nicht-linearen Lernmaterialien
- Art der Lerninhalte
- Lernziele
- Lernvoraussetzungen
- Organisation der Hypertextbasis
- Ausmaß der semantischen Spezifikationen der Verknüpfungen*
- Bereitstellen von Orientierungshilfen und Metainformationen
- Bereitstellen von Navigationshilfen
- Ausmaß der Multimedialität
- Flexibilität im Design der Benutzerschnittstellen

*Beschreibung der Funktion einer Verknüpfung

Das Fazit aus diesen Betrachtungen ist, dass kein sinnvolles Kontrolldesign möglich ist, wenn man die hochgradige Differenzierung im Variablenbereich beachten möchte. Eine Einschränkung der Variablen muss zwangsläufig zu einer Verfälschung der Ergebnisse führen. Trotzdem erscheint eine solche Eingrenzung in den meisten Fällen sinnvoll, da sonst einige Studien überhaupt nicht durchführbar wären oder sehr viel Zeit in Anspruch nähmen. Außerdem ist die Einschränkung auch stark abhängig vom Ziel, das man mit den Studien erreichen will.

Ein großes Problem solcher Studien ist sicherlich auch die, sich ständig erweiternde, Palette an Faktoren in angemessenem Maße zu berücksichtigen. Das wirft gerade dann Probleme auf, wenn man versucht Studien zu vergleichen oder wenn man versucht Studien, die einen längeren Zeitrahmen überspannen, miteinander in Einklang zu bringen. Eine weitere Untersuchung der Verteilung der Variablen ist sicherlich sehr schwierig, aber dennoch wichtig, um Fehler oder falsche Ergebnisse bei den Evaluationen zu vermeiden.

4.5 Auskunftsfähigkeit der Versuchspersonen

Die Auskunftsfähigkeit der Versuchspersonen ist einer der wichtigsten und zugleich am schwierigsten zu kontrollierenden Faktoren. Es muss bei den Untersuchungen nicht nur sichergestellt werden, dass die Versuchspersonen geeignet sind für die Untersuchung, sondern auch dass die Probanden verbindliche Selbstauskünfte über sich selbst und ihrer Einstellung und Leistung geben können. Doch gerade hier sind schwerwiegende Mängel zu beobachten. So wurden 1988 Versuchspersonen jeweils mit der elektronischen Version des Lexikons „Academic American Encyclopedia“ und mit der gedruckten Version konfrontiert. Marchionini, der diese Untersuchung führte, musste feststellen, dass seine Versuchspersonen angaben die elektronische Version hätte nicht nur mehr Inhalt, sondern sie wäre auch noch schneller als die Buchform. Nach einer Untersuchung der tatsächlichen Gegebenheiten stellte sich jedoch heraus, dass die Schüler sehr viel länger benötigten, um eine Information in dem elektronischen Lexikon zu finden, als in der gedruckten Form. Auch die Aussage, dass die elektronische Form mehr Inhalt enthalte stellte sich als Trugschluss heraus, denn die elektronische Version wies in keinem Punkt Abweichungen gegenüber dem Buch auf. Worauf aber ist dieser Effekt zurückzuführen?

In erster Linie ist wohl die anfängliche Begeisterung für das neue Medium (Hawthorne-Effekt) und für den Computer an sich für die erstaunlichen Ergebnisse verantwortlich. Zweitens ist wohl auch die mangelnde Fähigkeit der Probanden sich selbst einzuschätzen schuld an den Ergebnissen.

Sicher ist zudem, dass auch Vorkenntnisse und die Umgebung eine entschei-

dende Rolle spielen. MacCormac untersuchte zu diesem Thema 1992 ein Computerprogramm, das Krankenpfleger bei der Ausbildung unterstützen sollte. Er untersuchte die Einstellung und Lernergebnisse von Erstsemestlern, die eine durchgehend sehr gute Qualität des Programms bestätigten. Ein Jahr später wiederholte er den Versuch mit Drittsemestlern. Das Ergebnis des Vorjahres musste deutlich nach unten korrigiert werden, denn die Probanden hatten bereits Erfahrungen mit dem überprüften Programm und konnten so wesentlich kritischer bewerten, ohne durch anfängliche Begeisterung Fehler und Unstimmigkeiten zu übersehen. Hinzu kam, dass die meisten der Probanden schon Erfahrungen mit anderen Programmen besaßen und daher auch Faktoren bewerteten, die eigentlich nicht vorgesehen waren und zu einer genaueren Wertung verhalfen.

Aus den genannten Gründen müssen neue Untersuchungen gemacht werden, die feststellen, in wie weit sich die Meinung allgemein im zeitlichen Kontext ändert und wie die Ergebnisse davon beeinflusst werden.

5 Meta-Analysen zum Computerlernen

Definition :

Meta-Analysen sind Sekundär-Analysen, die eine exakte Beurteilung der Größe des intendierten Effektes erlauben, weil sie die Originaldaten in Prozentanteile der Standardabweichungen des Punktgewinnes bei Abschlußtests umwandeln

Diese Methode, die auf den ersten Blick wahrscheinlich erst einmal verwirrt, ermöglicht aber recht einfach den Vergleich zwischen Lehrmethoden. Sie berechnet also im Grunde wieviel mehr Punkte die Probanden bei getesteten Programmen erreichen, als bei den Vergleichsmethoden.// Kulik und Kulik et al kommen 1979, bei der Meta-Analyse über das Keller-Plan-Modell, zu dem Ergebnis, dass diese eine recht hohe Erfolgsquote aufwiesen (nämlich 50%). Diese Ergebnisse sind ähnlich hoch wie Analysen zum Computergestützten Lernen, dahingegen sind die Ergebnisse zum individualisierten Unterricht deutlich höher.

Clark meint zu dieser Problematik 1983 noch, dass die ähnlichen Ergebnisse zwischen Keller-Plan-Modellen und computergestützten Lehrmethoden darin begründet seien, dass letztlich die computergestützten Lehrmethoden eine Simulation der Keller-Plan-Modelle seien. Schulmeister dahingegen ist ganz anderer Meinung, er schließt daraus, dass letztlich jede Veränderung des Unterrichts etwas bringe, solange Anstrengungen erkennbar würden, die traditionellen Sozialisierungsformen zu modifizieren und den Unterricht zu

reformieren.

Bei der Untersuchung verschiedener Studien hat man einige wesentliche Faktoren und Probleme herausgestellt, die im Folgenden kurz erläutert werden sollen.

5.1 Die Rolle des Lehrers

Nach der Untersuchung von Studien zum Computer-Lernen, mußte man feststellen, dass der Lehrer eine zentrale Rolle einnimmt. So wurde in verschiedenen Studien deutlich, dass Lerneffekte abnehmen, wenn der selbe Lehrer die Versuchs- und Kontrollgruppen unterrichtet. Dies wird besonders in den Studien von Kulik und Kulik et al von 1980 klar. Sie untersuchten mehr als 500 Studien und stellten bei den meisten einen Lerneffekt von 0.5 Standardabweichungen fest. Bei den Studien aber bei denen der selbe Lehrer beide Versuchsgruppen unterrichtet hatte nivellierte sich der Effekt auf 0.13 Standardabweichungen. In den Analysen von Bangert und Kulik et al von 1983 ist der Effekt sogar noch geringer, mit 0.10 Standardabweichungen. Dies scheint laut Schulmeister ein deutlicher Hinweis auf die gewichtige Rolle des Lehrers als Variable zu sein. Weiterhin scheint das auch ein Indiz dafür zu sein, wie gering die Unterschiede zwischen den Lehrmethoden tatsächlich sind.

Daraus resultiert auch die Schwierigkeit überhaupt zu entscheiden, ob die Vorteile eines Mediums oder die Unterschiede zwischen Inhalt, Methode und Medium verglichen wurden, wenn die Methoden von unterschiedlichen Lehrkräften eingeführt wurden. Es kommt zu einer Verwechslung von Medium und Methode. Demnach ist die Effektivität von Instruktionen entscheidend von der Lehrerrolle abhängig.

5.2 Interaktivität

Interaktivität ist ein Schlagwort, das in der Entwicklung von Lehrprogrammen immer wieder fällt und einen hohen Grad an Effektivität verspricht. Verschieden Studien und Meta-Analysen haben aber gezeigt, dass Interaktivität allein nicht ausschlaggebend ist für den Lerneffekt. Interaktivität im eigentlichen Sinne ist mit den meisten anderen Medien in vergleichbarer Weise durchaus möglich und wird zum Teil auch effektiv genutzt, insbesondere mit dem Medium Lehrer.

Außerordentlich positive Ergebnisse mit Computerprogrammen sind auch

hier wieder hauptsächlich auf den Novitäts- oder Hawthorne-Effekt zurückzuführen.

5.3 Methodologische Genauigkeit

Bei den Untersuchungen der Studien stieß man auf einen interessanten Effekt. Wurden die Studien mit einer recht geringen methodologischen Strenge durchgeführt, konnte man dem computergestützten Lernen noch einen hohen Wirkungsgrad anrechnen. Sobald aber die methodologische Genauigkeit zunimmt, desto mehr werden die positiven Effekte der Lehrmethoden am Computer nivelliert.

Problematisch werden solche Analyse allerdings dann, wenn die Strenge der Analyse zu hoch ist. Im Grunde eine vernünftige Vorgehensweise, doch die Studie von Joliquet und Berger von 1986 zeigt Gegenteiliges. Sie wählten aus einer Unzahl von Studien 47 aus, die in Frage kommen für ihre Untersuchung. Nach einer etwas genaueren Untersuchung blieben von den 47 gerade mal zwei übrig. Der Grund für die Herausnahme von so vielen potentiell in Frage kommenden Studien liegt in der Tatsache, dass in den meisten Studien mehr als ein Programm Einsatz gefunden hatte. Aufgrund dieser Auswahl war keine vernünftige und repräsentative Untersuchung möglich. Berger und Joliquet kommen daher zu dem Schluss, dass, mit der bestehenden Datenmenge, keine zufriedenstellenden Meta-Analysen möglich sind.

Allgemein ist aus der derzeitigen Situation kaum eine vernünftige Aussage über die prinzipielle Lernwirkung von Multimedia möglich, dennoch ist wohl unverkennbar, dass Multimedia Potenziale hat. Wie sich die Multimedia in Zukunft entwickelt ist ungewiß, sicher ist nur, dass noch zahlreiche Studien notwendig sind, um wirklich heraus zu finden welche Rolle sie beim Lernen spielt.

5.4 Grundlegende methodologische Einwände

Mit der Betrachtung der weitreichenden Faktoren und Einflüsse kommt Schulmeister zu einer grundlegenden Kritik an bisher durchgeführten Studien und Evaluationsmethoden:

- Typisierungsfähigkeit der Anwendungen gering
- Experimentalsituationen häufig artifiziell

- Experimente abhängig vom institutionellen und interaktionellen Kontext
- Die Ergebnisse unterschiedlicher Methoden nicht mit dem selben Posttest vergleichbar
- Zahl beeinflussender Faktoren geht gegen unendlich, nur ein kleiner Teil in die Untersuchungen einbezogen
- Wenn die entscheidende Variable für den Lernerfolg der Lehrer ist, sorgt das Engagement des Lehrers für den Lernerfolg
- Positive Erfolge sind nicht in erster Linie auf die verwendete Technologie zurück zu führen, sondern eher auf die realisierten Methoden
- fehlende Differenzierung des Lernenden
- starker Einfluss des Hawthorne-Effektes

Die aufgeführten Einwände zeigen ganz deutlich, dass in der Forschung zur Wirkung von Multimedia und Hypermedia noch einiges passieren muss, gerade in Hinblick auf die Fehler, die gemacht worden sind, und die auch weiterhin gemacht werden. Es reicht nicht aus vorhandene Studien zu überprüfen, stattdessen sollte man anfangen Ursachenforschung zu betreiben und bestehende Studien kritischer zu betrachten.

Die geforderten Punkte sind überdies nicht nur wichtig, sondern leider auch sehr schwer umzusetzen. Gerade wenn man bedenkt, dass es eine Unzahl von Faktoren gibt, die kontrolliert werden müssen. Es scheint sehr schwierig ein passendes Kontrolldesign zu erstellen und die Tatsache, dass Kontrolldesigns nicht universell einsetzbar sind, fördert nur im geringen Maße die Motivation weitere Forschungen zu betreiben. Die angesprochene ständige Weiterentwicklung verhindert zudem die Endgültigkeit der Ergebnisse. Was heute noch bewiesen scheint, kann sich morgen als komplett unsinnig und nichtig herausstellen.

Der Versuch Kontrolldesigns zu optimieren und die Zahl der zu untersuchenden Variablen gering zu halten resultiert zumeist nur in absolut artifiziellen, d.h. künstlichen, Umgebungen, die der Realität in keinsten Weise entsprechen. Außerdem spielen scheinbar auch die Umgebung und die vorhandenen Mittel eine zentrale Rolle. Stellt man sich einen Test zu Lerneffekten mit Videosequenzen am PC's von 1990 vor, so wird schnell klar, dass ein vergleichbarer Test in angenehmer und technisch aktuellerer Umgebung andere Ergebnisse bringen wird.

5.5 formative Evaluation

Formative Evaluation ist eigentlich ein Hilfsmittel bei der Entwicklung von Programmen und weist eine gewisse Ähnlichkeit zum Konzept der Qualitätssicherung auf. Sie befasst sich mit der Erfassung von Daten, die bei der Entwicklung direkt Aufschlüsse über die ablaufenden Prozesse und Einflußfaktoren gibt. Bei der Untersuchung von Studien gibt sie insofern Einsichten, als das sie eine Prozesserfassung ermöglicht. Das heißt man kann gezielt die Entwicklung von Prozessen und die Einflussnahme von Faktoren im Zeitverlauf feststellen und auswerten.

Trotz dieser offensichtlichen Vorteile gegenüber den Evaluationen, ist diese Methode nicht dazu geeignet die bestehenden Probleme zu lösen. Sie bietet allerdings eine bessere Einsicht in bestehende Probleme und kann zu deren Aufklärung und der Verbesserung der Evaluationsmethoden beitragen. Die formalen Evaluationen können jedoch nicht das Experiment ersetzen, sondern höchstens unterstützen.

5.6 Nichtuntersuchte Faktoren

Das Gebiet der zu untersuchenden Faktoren ist umfangreich und kann ohne weiteres nicht in ausreichender Genauigkeit evaluiert werden. Aus diesem Grunde sehen sich viele Wissenschaftler gezwungen die Anzahl der zu untersuchenden Faktoren stark einzugrenzen, nicht zuletzt um die Kontrolldesigns zu vereinfachen und Ergebnisse deutlicher ablesen zu können.

Unterzieht man den bisher durchgeführten Studien, so sind zahlreiche Mängel nicht verkennbar. Viele wichtige Faktoren, die die Ergebnisse entscheidend verfälschen, wie etwa der Hawthorne-Effekt, werden nicht berücksichtigt bzw. nicht aus den Ergebnissen herausgefiltert.

Die Frage ist aber, welche Faktoren noch nicht berücksichtigt wurden. Schulmeister gibt selbst zu, dass einige dieser Fragen einer dringenden Klärung, zum besseren Verständnis der Lerneffekte, bedürfen.

Warum gehen Jugendliche überhaupt gerne mit dem Computer um? Gibt es einen speziellen Grund dafür oder liegt es an der allgemein veränderten Sozialisierung und der sich rasant entwickelnden Technik?

Warum steigt das Interesse an interaktiven Umgebungen? Ist die natürliche Umgebung zu uninteressant geworden, welche Faktoren spielen beim Interesse der Jugendlichen eine Rolle?

Warum sind Spiele und Animationen am Computer so aufregend?

Welchen Grund hat die Unlust am formalen Unterricht? Sind traditionelle Lehrmethoden zwecklos?

Dies sind nur einige Fragen, die in den bisherigen Untersuchungen keine Beachtung fanden. Aber nichts desto trotz sind sie wichtig und bedürfen einer genaueren Klärung, nur so kann man Lehrmethoden und Programme effektiv verbessern und bewerten.

6 Quellen

Als Quelle des Vortrages und der Ausarbeitung sind folgende Bücher und Websites verwendet worden:

- Grundlagen hypermedialer Lernsysteme, Schulmeister, 1997
- www.jura.uni-sb.de/-fritz/infwiss/dist/bibl/Referate/stutz.htm – Thema Evaluation