



Cornelia Biffi

8.5.2002

### **Evaluation von Bildungssoftware im Spannungsfeld von Objektivität und praktischer Anwendung**

*Das Angebot an Bildungs- bzw. Lernsoftware umfasst eine breite Palette von sehr unterschiedlichen Produkten. Eigentliche Unterrichtsoftware, die speziell für den Unterricht hergestellt und auf Lehrmittel und Curriculum abgestimmt ist, gibt es kaum. Im Projekt «Lernsoftware Evaluation» des Pestalozzianums Zürich ermitteln Lehrpersonen die Qualität von Bildungssoftware, indem sie diese im Unterricht einsetzen und anhand eines Kriterienrasters beurteilen. Eine Analyse der in der Evaluation generierten Daten belegt die Problematik des Anspruchs einer objektiven Qualitätsbeurteilung. Es werden die Probleme bei der Anwendung von Kriterienraster und Beurteilungsmodus aufgezeigt und ermittelt, worin die subjektiven Qualitätsmassstäbe der Evaluierenden begründet sind. Diese Ausführungen sind Teil eines Forschungsprojekts für die Weiterentwicklung einer theoretisch fundierten aber dennoch praxisnahen Evaluation von Bildungssoftware<sup>1</sup>.*

#### **Erwartungen der Lehrkräfte an gute Bildungssoftware**

Die Lehrpersonen der Primarschulstufe schätzen den Computer als vielfältig einsetzbares Hilfsmittel, das die Individualisierung des Unterrichts erleichtert und sich günstig auf bestimmte Fachleistungen auswirkt. Dies

<sup>1</sup> Im Auftrag der Pädagogischen Hochschule Zürich arbeiten wir im Rahmen der Lehrmittel- und Lernmaterialien Forschung an einem Projekt im Bereich «Lernsoftware Evaluation». Das Forschungsdesign sieht in einem ersten Schritt vor, die seit 1997 bestehende «Lernsoftware Evaluation» des Pestalozzianums Zürich heuristisch auszuwerten. Daraus sollen Erkenntnisse für die Weiterentwicklung einer praxisnahen Evaluation und für den Einsatz von Praktiker/innen bei der Beurteilung von Bildungssoftware gewonnen werden.

zeigen die Evaluationsergebnisse des zürcherischen Schulversuchs «Schulprojekt 21» zum computergestützten Lernen. In diesem Sinne wird der Computer als «Lernmedium» und «Arbeitsmittel» zu einem festen Bestandteil des Unterrichts in den Projektklassen (vgl. Büeler et al., 2001). Der Computereinsatz erfolgte dabei in speziellen Unterrichtssituationen. Hauptsächlich wurden die Rechner im Halbklassenunterricht zum Schreiben und Rechnen eingesetzt. Die Schüler/innen arbeiteten und lernten pro Woche durchschnittlich zwanzig Minuten einzeln oder paarweise mit dem Computer. Bezüglich Lernerfolg werden vor allem die Leistungen im Bereich «computer literacy» beschrieben. Nach zweijähriger Projektdauer können laut Evaluationsbericht die meisten Kinder der dritten Jahrgangsstufe den Computer selbstständig bedienen und sie besitzen Grundkenntnisse in den Bereichen Textverarbeitung und Zeichnen. Die Hälfte der Kinder kann auch einfache Suchabfragen im Internet durchführen. Nach Einschätzung der Lehrkräfte erbringen Schülerpaare am Computer generell bessere Leistungen als einzelne Schüler/innen.

Allerdings schätzen die Lehrpersonen auch nach zwei Jahren den Vorbereitungsaufwand für den Computereinsatz im Unterricht noch als hoch ein. Dieser Aufwand wird vor allem auf das Fehlen von geeigneter Bildungssoftware zurückgeführt. Als geeignete Bildungssoftware bezeichnen die Lehrpersonen diejenigen Produkte, welche zu den Lehrmitteln passen. Hilfreich wäre aus Sicht der Lehrkräfte die Möglichkeit, auf einen Fundus an Ideen für Lehr-/Lernaufträge zum computergestützten Unterricht und den Einsatz von Bildungssoftware zurückgreifen zu können.

Diese Erkenntnisse verdeutlichen zum einen, dass ein Bedarf nach Orientierungsmöglichkeiten bezüglich Lernsoftware besteht, zum anderen wird aber auch das Bedürfnis nach Unterstützung bei Vorbereitungsarbeiten für den computerunterstützten Unterricht geltend gemacht.

#### **Qualitätsbeurteilung multimedialer Lern- und Informationssysteme**

Eine Evaluation dient der Bewertung. In diesem Falle geht es um die Bewertung von Bildungssoftware. Ein weites Verständnis von Evaluation umfasst alle Aktivitäten und/oder Ergebnisse, welche die Bedeutung, Verwendbarkeit und Zweckmässigkeit einer Sache zu beurteilen bzw. zu bewerten versuchen. Grundlage der Qualitätsbeurteilung sind so genannte Qualitätskriterien. Diese thematisieren die von den entsprechenden Katalogentwicklern für die Produktqualität als bedeutsam erachteten Merkmale,

wie z. B. Textgestaltung, Anwenderfreundlichkeit, Bildschirmaufbau, Interaktivität.

Es lassen sich vier Phasen in der Evaluation von Bildungssoftware unterscheiden. In einer ersten Phase geht es um die Ausarbeitung des Kriterienkataloges. Es werden die Bewertungskriterien festgelegt und formuliert. In einem zweiten Schritt werden diese Kriterien durch die Formulierung von Leistungsstandards operationalisiert. Die konkrete Durchführung der Evaluation entspricht der dritten Phase. Die vorgegebenen Leistungsstandards werden durch Untersuchung, Messung und Vergleich an die Bildungssoftware angelegt. In einem vierten Schritt geht es um die Bewertung. Diese erfolgt durch die Integration der Ergebnisse zu einem einheitlichen Werturteil.

### **Festlegung von Qualitätskriterien**

Kriterienkataloge zur Qualitätsbeurteilung von Bildungssoftware enthalten Zusammenstellungen von Fragen und Einschätzungsskalen zur standardisierten Beschreibung und Beurteilung von Aspekten der technischen, inhaltlichen und didaktischen Qualität. Zur Formulierung der Wertkriterien und der Leistungsstandards kann nach Scriven (1992, S. 380) eine dreistufige Bedürfnisanalyse verwendet werden.

Die erste Ebene (*necessitata*) dient der Bestimmung unabdingbarer Notwendigkeiten und Erfordernisse. Wird ein Kriterium nicht erfüllt, so hat dies den unweigerlichen Ausschluss des Produktes zur Folge, bzw. es wird als «ungenügend» bewertet. Ein solches *K.-o.-Merkmal* ist zum Beispiel die Hardware-Voraussetzung einer Lernsoftware.

Die zweite Ebene (*desiderata*) beschreibt Funktionen und Eigenschaften, die über das absolute Minimum hinausgehen. Deren Fehlen führt aber nicht zum Ausschluss eines Produktes. Scriven spricht in diesem Fall von Vorzügen. Ein solcher Vorzug stellt zum Beispiel ein Programm dar, das auf mehr als einer Systemarchitektur lauffähig ist.

Der dritten Ebene (*ideals*) werden Ideale zugerechnet. Dabei handelt es sich um kaum erreichbare bzw. realisierbare Zielvorstellungen. Sie dienen als Wegweiser für Verbesserungsansätze.

Kriterienkataloge stellen ein beliebtes Evaluationsinstrument dar. Sie eignen sich für eine differenzierte Beschreibung von Softwaremerkmalen und können damit den Anwendern bei der Entscheidung für den praktischen Einsatz einer Bildungssoftware im Unterricht behilflich sein. Solche Beurteilungsverfahren eignen sich somit insbesondere für eine schnelle Vor-

selektion. Ihre Verwendung bringt aber auch Nachteile. So führt die Anwendung von Kriterienkatalogen oft zu unrealistischen Erwartungen.

Es kann die Erwartung entstehen, die als qualitativ gut beurteilte Bildungssoftware garantiere auch einen entsprechenden Lernerfolg, so dass auf eine empirische Überprüfung der Wirkung der Bildungssoftware verzichtet werden kann. Diese Vorstellung gründet auf der Annahme eines objektiven und prognostisch validen Instrumentariums. Ein weiterer Grund für die Beliebtheit von Kriterienkatalogen besteht darin, dass Kriterienkataloge leicht handhabbar sind und mit ihrer Hilfe eine Qualitätsbeurteilung auch von Nicht-Experten durchgeführt werden kann. Im Übrigen wird der Einsatz von Kriterienkatalogen zur Qualitätsbeurteilung von Bildungssoftware zunehmend kritisiert, da diese Kataloge konzeptuelle Mängel, Schwächen bezüglich Gütekriterien und Anwendungsprobleme aufweisen können (vgl. Tergan, 2001, S. 320).

### **Probleme und Schwächen von Kriterienkatalogen**

In der Literatur werden allgemeine Probleme und Schwächen, die bei der Konstruktion und Anwendung von Fragebogen auftreten sowie die Probleme bei der Auswertung und Interpretation von Skalenwerten als Ergebnisse subjektiver Einschätzungen diskutiert (vgl. Baumgartner 1995; Fricke 2000; Tergan 2001). Tergan zeigt in diesem Zusammenhang drei Problemtypen auf: Konstruktionsmängel, Schwächen bezüglich teststatistischer Gütekriterien sowie Anwendungsprobleme. Im Folgenden werden wir vor allem auf die von Tergan und Fricke beschriebenen konzeptuellen Schwächen von Kriterienkatalogen fokussieren.

Die konzeptuellen Schwächen von Kriterienkatalogen liegen allgemein in der unscharfen Verwendung des Begriffs «Qualitätskriterium». Oft werden Kriterien zusammengenommen, welche auf recht diffuse Weise zu einem Gesamtkatalog aggregiert werden. So stehen bspw. Kriterien zur Motivation, Programmästhetik, Navigation nebeneinander. Aber auch dort, wo es um das Lernen geht, wird der Begriff schon dann verwendet, «wenn lediglich Vermutungen über die Lernwirksamkeit eines Programmmerkmals vorliegen» (Fricke, 2000, S. 75). Qualitätskriterien sind somit Merkmale, von denen man aufgrund der Erfahrung oder plausibler Schlüsse vermutet, dass sie das Lernen positiv beeinflussen oder aber «allgemeine Merkmale einer Lernsoftware, deren Lernwirksamkeit in einer Validitätsstudie wissenschaftlich nachgewiesen wurde» (ebd., S. 75).

Methodisch ist anzumerken, dass die Kriterien der Operationalisierung bedürfen, um ihre standardisierte Anwendung im Rahmen von Kriterienkatalogen zu gewährleisten. Die verwendeten Kriterien sind jedoch selten standardisiert. Es fehlen Angaben darüber, wann bei einer Bildungssoftware ein bestimmtes Kriterium als erfüllt oder in Teilen erfüllt gelten soll. Durch eine Reduktion der Beurteilerübereinstimmung wirkt sich dies auf die Güte der Beurteilung aus.

Inhaltlich zeigt sich, dass in vielen Kriterienkatalogen die technischen Merkmale der Software zu stark gewichtet werden. Letztere sind jedoch bezüglich Lernleistung nicht sehr aussagekräftig. Interessant ist in diesem Zusammenhang die Feststellung von Fricke (2000, S. 78), dass gut operationalisierbare Merkmale zwar eine hohe Messgenauigkeit, aber meist eine geringe Korrelation zum Lernergebnis aufweisen. Allgemein besteht zudem eine Dominanz von Fragen zu Einzelmerkmalen einer Bildungssoftware. Solche eindimensionalen Items beziehen sich meist auf Systemmerkmale. Sie fragen etwa danach, ob ein bestimmtes Systemmerkmal vorhanden ist, der Text verständlich ist, Aufgaben zum Selbsttesten oder eine Fehlerkontrolle vorliegen. Sie beziehen sich aber auch auf pädagogisch-didaktische Softwaremerkmale. Beispielsweise durch eine Fragestellung wie: *«Wird eine bestimmte Lehrerfunktion unterstützt?»*.

Seltener stößt man auf mehrdimensionale Items, welche Zusammenhänge thematisieren. Gefragt wird zum Beispiel: *«Entspricht das Niveau der Darstellung des Sachverhalts den Lernvoraussetzungen der Adressaten?»*. Die isolierte Verwendung von eindimensionalen Kriterien und das Nichtbeachten von Wechselwirkungen zwischen lernrelevanten Variablen hat einen negativen Einfluss auf die prädikative Validität von Kriterienkatalogen. Denn die Wirksamkeit von Merkmalen einer Bildungssoftware äussert sich nicht selten erst in Wechselwirkung mit Lernervariablen und situativen Bedingungen des Softwareeinsatzes (vgl. ebd., S. 79). Hier besteht allerdings die Gefahr einer kombinatorischen Explosion der einzubeziehenden Kriterien.

Aus theoretischer Sicht ist das Fehlen von Annahmen darüber anzumerken, wie der Lehr-Lernprozess beim Arbeiten mit Lernsoftware abläuft. Baumgartner (1995, S. 242) weist auf die theoretische Orientierungslosigkeit von Kriterienkatalogen hin. Besonders Kriterienkataloge ohne theoretisch fundierte Gewichtungsverfahren würden die Frage nach der zugrunde liegenden Lerntheorie vernachlässigen. Dies stellt die pädagogisch-didaktische Angemessenheit der Beurteilung in Frage. Fricke (2000, S. 78) übt massive

Kritik an den bestehenden Kriterienkatalogen und verweist auf die dringende Notwendigkeit, erklärende Lehr-Lerntheorien in die Beurteilung von Lernsoftware einzubeziehen. Bei der Auswahl der zu evaluierenden Software werden zudem vorrangig traditionelle Lernprogramme wie Übungsprogramme und tutorielle Systeme, und seltener Simulationen sowie hyper- oder multimediale Lernumgebungen bewertet. Für die Beurteilung Letzterer sind die entsprechenden Kataloge nur bedingt geeignet.

Aus pädagogisch-didaktischer Sicht geschieht die Bewertung von Bildungssoftware oft unabhängig von den Lerninhalten und dem Kontext ihrer Anwendung. Besondere situative Bedingungen des Softwareeinsatzes wie bspw. die curriculare Integration oder die tutorielle Betreuung der Lernenden bei der Softwareanwendung bleiben unberücksichtigt. Da die Kriterienkataloge scholorientiert sind, kann Software, die ausschliesslich auf das Selbststudium zugeschnitten ist, nicht adäquat beurteilt werden (Korbmacher, 2000, S. 216). Meier (2000, S. 179) weist darauf hin, dass die meisten Fragen und Kriterien auf reine Lernprogramme zugeschnitten sind, die in den herkömmlichen Unterricht zu integrieren sind und bei deren Bearbeitung eine Ausbilderin anwesend ist.

Der Lernerfolg eines Programms hängt nach Fricke (2000, S. 78) von einer Reihe von Kriterien und den Wechselwirkungen dieser Kriterien ab. Der prozentuale Anteil eines einzelnen Kriteriums am Lernerfolg ist somit als sehr gering einzuschätzen.

Für die Ermittlung des Lernerfolgs müssten insgesamt vier Hauptfaktoren berücksichtigt werden: die multimediale Lehr-Lernumgebung, die Lernervariable (Lerner), das Lernthema (Lehrstoff) und das Lernergebnis.

### **Verfahren bei der Beurteilung der Qualität**

Kriterienkataloge verlangen in der Regel die systematische Einschätzung einer Software nach vorgegebenen Kriterien. Über eine Zusammenfassung von zum Teil gewichteten Einzelurteilen gelangen Anwender zu einem Gesamturteil bezüglich der Softwarequalität (vgl. Fricke, S. 75). An eigentlichen Beurteilungsverfahren lassen sich grundsätzlich vier Methoden unterscheiden, die untereinander frei kombinierbar sind:

*Grading (Einstufung)*: Die Produkte werden anhand eines vordefinierten Beurteilungsmassstabes eingestuft.

*Ranking (Reihung)*: Für ein Ranking werden die Produkte miteinander verglichen und beurteilt. So entsteht eine Ordinalskala mit einer Aussage wie

«gut – besser – am Besten», jedoch keine Aussage zu den Abständen untereinander.

*Scoring (Punktevergabe)*: Im Unterschied zum Ranking werden hier Punkte vergeben, deren Abstände untereinander gleich und somit bedeutungsvoll sind. Nur mit dieser Methode sind summative Operationen wie Addition oder Division zulässig.

*Apportioning (Aufteilung, Zuteilung)*: Entsprechend der Wertigkeit eines Produktes werden vorhandene Ressourcen aufgeteilt (Bsp. Aufteilung Bundeshaushalt).

### Objektivität versus Praxisnähe

Zusammenfassend können wir festhalten, dass die Festlegung von Qualität über die Formulierung von Leistungsstandards und Wertekriterien geschieht. Ein beliebtes Evaluationsinstrument dafür sind Kriterienkataloge. Bei der Entwicklung valider Merkmale für eine möglichst objektive Qualitätsbeurteilung besteht die Gefahr, Kriterien zu generieren, die unabhängig vom Lehrstoff, der Personengruppe und der Lernumgebung, zuverlässig die Lernwirksamkeit ermitteln sollen. Dies steht im Widerspruch zum Wunsch der Lehrpersonen nach Orientierungshilfe und unterrichtsbezogenen evaluierten Produkten. Praxisnah gestaltete Evaluationsverfahren weisen dann aber oft Mängel und Schwächen bezüglich Gütekriterien auf. Es stellt sich also die Frage, wie der Anspruch nach objektiver Beurteilung in einer praxisnahen Evaluation umgesetzt werden kann. Im Folgenden gilt das Interesse einem praxisnahen Evaluationsprojekt.

### Lernsoftware Evaluation durch Lehrpersonen

Im Projekt Lernsoftware-Evaluation des Pestalozzianums Zürich wurde angestrebt, mit knappen Ressourcen Lernsoftware möglichst praxisnah zu bewerten und Lehrkräften eine «sachkundige Orientierungshilfe für den Einsatz von Lernsoftware in der Schule anzubieten» (Pfyffer, 2001). Das Evaluationsprojekt beruht auf einem Modell, das Lehrkräfte bezieht, um die Qualität von Bildungssoftware zu beurteilen. Das Sample der Evaluierenden setzte sich hauptsächlich aus Lehrkräften zusammen, welche die zu beurteilende Lernsoftware im Unterricht einsetzten - was die Praxisnähe der Evaluation unterstreicht: Erfahrene Lehrkräfte geben ihr Wissen im Sinne eines „Best-practice“-Ansatzes weiter. Angesichts der eingeschränkten finanziellen Dotierung des Projektes bestand allerdings keine

Möglichkeit, eine Schulung der Evaluator/innen in die Projektumsetzung miteinzubeziehen.

Im Projekt werden seit 1997 im Auftrag der kantonalen Bildungsdirektion<sup>2</sup> unterschiedliche Programme auf ihre Unterrichtstauglichkeit untersucht<sup>3</sup>. So auch Programme, die im weitesten Sinne dem Lernen dienen und auch Informations- sowie Unterhaltungsaspekte beinhalten. Die Beurteilung erfolgt online und sehr benutzerfreundlich mittels Beurteilungsbogen.<sup>4</sup> Nach erfolgter Eingabe sendet die Evaluatorin den Datensatz per Mausklick an die Redaktion, welche die entsprechende Datenbank mit den 100 evaluierten Titeln<sup>5</sup> ausredigiert und aktualisiert. Der Benutzer kann die Datenbank<sup>6</sup> mit den Lernsoftwarebeurteilungen mittels Suchmaske online abfragen. Diese Möglichkeit wird nach wie vor sehr rege genutzt.<sup>7</sup>

### Evaluierte Lern- und Informationssysteme

Insgesamt wurden 97 Programme evaluiert. Aufgeschlüsselt nach Programmtyp sind dies: Lernprogramme (20), Übungsprogramme (44), Informationssysteme (30) und Autorenprogramme/Werkzeuge (3). Nicht in die Evaluation einbezogen wurden Simulationen oder Mikrowelten (0). Die ausgewählte Bildungssoftware deckt eine breite Fächerpalette mit unterrichtsrelevanten Themen für alle Schulstufen ab. Es finden sich Programme aus den Bereichen Sprache (23), Mathematik (26), Mensch und Umwelt (23), Gestaltung und Musik (5) und Fremdsprachen (12).

Unter die *Übungsprogramme* lassen sich vor allem «*drill and practice*»-Programme fassen, die besonders der Aneignung von Faktenwissen dienen. Es gibt Übungsprogramme, die den bereits erarbeiteten Lernstoff mittels

<sup>2</sup> Die beschriebenen Leistungen knüpfen an das Konzept zur «kontinuierlichen Weiterführung und Unterstützung der Schulinformatik» an, welches der Erziehungsrat des Kantons Zürich am 27. August 1996 genehmigt hat. Dieser Entscheid ist als bildungspolitische Weichenstellung aufzufassen und bedeutet eine neue Schwerpunktsetzung in der Weiterbildung von Lehrpersonen und dem Einsatz von Lernsoftware.

<sup>3</sup> Es gibt im deutschsprachigen Raum ähnliche Projekte, so die deutsche Datenbank SODIS («Software Dokumentations- und Informationssystem» über neue Medien). Viele schweizerische Produkte werden hier nicht evaluiert, so dass kein Konkurrenzverhältnis besteht.

<sup>4</sup> <http://www.pestalozzianum.ch:591/evaluationsdb/standard.htm>

<sup>5</sup> CD-Roms und Disketten

<sup>6</sup> <http://www.pestalozzianum.ch:591/lernsoftware/standard.htm>

<sup>7</sup> Neben der Datenbank hat sich ein Beratungs- und Weiterbildungsangebot und die Möglichkeit, Lernsoftware zu visionieren, etabliert.

Übungen vertiefen oder solche, bei denen nach dem Informationsinput entsprechende Fragen beantwortet werden müssen. Vokabel-, Mathematik- oder Rechtschreibtrainer lassen sich dieser Art zuordnen.

Häufig können Thema und/oder Schwierigkeitsgrad gewählt werden. Fehlerprotokolle und individuelle Speichermöglichkeiten des Gelernten sind wichtige Zusatzfunktionen. Adaptive Programme passen sich zudem dem individuellen Wissensstand an.

*Lernprogramme* oder «*tutorials*» vermitteln Basis- oder Vertiefungswissen. Es gibt Programme, die das Lernen helfend und beurteilend begleiten, indem sich ein virtueller Tutor bei Schwierigkeiten automatisch einschaltet<sup>8</sup> oder Programme, denen eher eine Dozentenfunktion zukommt<sup>9</sup>, indem das Programm die Lernenden Schritt für Schritt in ein Wissensgebiet einführt. Der Grad der Interaktivität kann für diesen Programmtyp sehr unterschiedlich ausfallen.

Unter *Informationssysteme* lassen sich vor allem *hypertextbasierte Programme* subsumieren. Hypertext- oder Hypermediaprogramme können als Netzwerke von Informationsknoten und eingebetteten Verweisen beschrieben werden. Es sind Datenbanken, die mit Wort, Bild und Ton vernetzte Wissensinhalte oder Geschichten multimedial darstellen. Es gibt keinen vordefinierten linearen Lernpfad. Der Lernende steuert die Reihenfolge des Lernangebotes und selektiert die entsprechenden Inhalte. Zu diesem Programmtyp gehören *Lexika* oder *Nachschlagewerke* oder auch *Edutainment-Programme* und *interaktive Bücher*.

*Werkzeuge* oder *Autorenprogramme* (*creativity tools*) ermöglichen die Aufarbeitung von Text, Ton und Bildelementen. Das Programm unterstützt die Aufarbeitung oder Umsetzung der Inhalte innerhalb einer vorgegebenen Struktur.

Wie das Sample der evaluierten Programme zeigt, wurden am häufigsten Übungsprogramme evaluiert, gefolgt von Informationssystemen und Lernprogrammen. Vom Programmtyp der Werkzeuge/Autorenprogramme wurden nur sehr wenige in die Evaluation einbezogen. Im Projekt war die Anzahl der Datensätze auf 100 beschränkt. Da statistisch nicht festgelegt wurde, von welchem Programmtyp wie viele Produkte evaluiert werden sollten, wird diese Verteilung wohl vor allem das Marktangebot spiegeln

und kann nicht als Präferenz der Lehrpersonen für den einen oder anderen Programmtyp interpretiert werden. Es erstaunt also nicht, dass vor allem Informationssysteme und Übungsprogramme evaluiert wurden, da diese den Markt beherrschen.

Erstaunlich ist allerdings die grosse Anzahl der Evaluationseinträge unter dem Stichwort der Lernprogramme, da uns nur wenige Programme bekannt sind, welche den oben definierten Kriterien genügen. Eine mögliche Erklärung besteht darin, dass viele Softwareangebote eine Kombination aus Lern- und Übungsprogramm<sup>10</sup> darstellen und dadurch die Zuordnung zum einen oder anderen Programmtyp schwierig ist. Die beachtliche Anzahl an verzeichneten Lernprogrammen zeigt die Tendenz der Lehrpersonen, Programme mit Lern- und Übungsteilen eher als Lernprogramme zu typisieren. Der Umstand, dass nur wenige Werkzeuge bzw. Autorensysteme evaluiert wurden, erstaunt nicht, da diese im Handel nicht so präsent sind. Zudem braucht es für eine sorgfältige Beurteilung und den Unterrichtseinsatz zahlreiche Vorkenntnisse über den Funktionsumfang des Programmes.

### **Kritische Betrachtung der Qualitätsbeurteilung**

Für die Qualitätsbeurteilung wurde ein Verfahren mittels Kriterienkatalog entwickelt und eingesetzt. Der Kriterienkatalog beinhaltet die Dimensionen Inhalt, Didaktik, Technik und Gestaltung der zu beurteilenden Software. Nach Pfyffer (1998, S. 28) sind inhaltlich die fachliche Korrektheit, der Umfang, die Tiefe und der Aktualitätsgrad der Information von Bedeutung. Zudem gelten eine zielgruppengerechte Sprache und die fachgerechte Vernetzung des Inhalts als Qualitätskriterien. Die didaktischen Kriterien umfassen die Aufarbeitung des Lernstoffs, die Motivationserhaltung, die Übungsfunktionen während des Lernprozesses sowie die Möglichkeit, Schwierigkeitsgrade einzustellen, Lernzielkontrollen durchzuführen und Lernhilfen abzurufen. Die technischen Kriterien umfassen die Interaktivität, die intuitive Benutzerführung, den Aufwand für Installation/Deinstallation, die Programmstabilität, die Reaktionsgeschwindigkeit, die Nützlichkeit des Handbuchs und weitere Hilfestellungen. Für die Gestaltungskriterien wird die Qualität von Schrift, Bild und Ton beurteilt.

<sup>8</sup> wie z. B. «Training induktiven Denkens», ein Programm, welches das räumliche Denken fördert.

<sup>9</sup> wie z. B. «Swiss Map Trophy», ein Programm, das schrittweise ins Kartenlesen einführt.

<sup>10</sup> Übungsprogramme beinhalten tutorielle Anteile oder Lernprogramme Übungsteile, wie z. B. Sprachlernprogramme.

Positiv fällt eine pragmatische und direkt auf den Unterrichtseinsatz bezogene Bewertung auf. Die Kehrseite dieses Verfahrens bildet dagegen die Dominanz von Fragen zu Einzelmerkmalen<sup>11</sup> sowie die fehlende Operationalisierung der Kriterien auf. Einschätzskalen zur standardisierten Beschreibung und Beurteilung sowie Wertkriterien bezüglich Leistungsstandards fehlen. Es liegen somit keine Angaben über Indikatoren vor, welche das Vorkommen eines bestimmten Sachverhaltes anzeigen würden, oder Angaben, die zur Erfassung des durch den Begriff bezeichneten Sachverhaltes notwendig sind.

Für die Qualitätsermittlung bedeutet dies, dass die Evaluierenden in eigenem Ermessen darüber entscheiden, ob und ab wann ein Kriterium zutrifft oder nicht. Dies erfolgt durch ankreuzen oder leer lassen der aufgeführten Items. Kritisch zu betrachten ist im Weiteren auch der Beurteilungsmodus. Die evaluierende Person bewertet die Bildungssoftware durch Notengebung für die vier Dimensionen Inhalt, Didaktik, Technik und Gestaltung. Wie im Unterrichtssystem der Schweiz, entspricht die Note 6 der höchsten Beurteilungsstufe und die Note 1 der niedrigsten. Ein Durchschnitt der Noten aus den vier Dimensionen ergibt dann die Endnote für die Gesamtbeurteilung. Die Messung der Qualität erfolgt somit nicht durch systematische Einschätzung nach den vorgegebenen Kriterien, sondern liegt vielmehr im Ermessen der evaluierenden Person. Es ist somit nicht nachvollziehbar, wie die Einschätzung der Qualität über die erhobenen Kriterien in die Beurteilung integriert wird oder welche Variablen wie stark gewichtet werden. Auffällig ist zudem, dass die Bewertungen tendenziell eher zu hoch ausfallen.

Es hat sich gezeigt, dass die Kriterien zwar plausibel, aber nicht operationalisiert sind. Dadurch erscheinen die Ergebnisse im Einzelnen oft reichlich willkürlich. Weder der Kriterienraster noch das Beurteilungsverfahren genügen den wissenschaftlichen Standards nach Objektivität. Somit stellen die Ergebnisse der Evaluation keine nachvollziehbaren objektiven Aussagen über die Qualität der Produkte dar. Positiv ist aber anzumerken, dass in diesem Fragebogen die technischen Merkmale nicht überwiegen, und dass das Beurteilungsverfahren offene, unstrukturierte Antworten ermöglicht, die von den Evaluierenden sehr intensiv genutzt wurden. Vor allem im Bereich Didaktik finden sich Einschätzungen,

<sup>11</sup> Diese Zusammenhänge werden zwar erkannt und in Ansätzen formuliert (Pfyffer, 1998, S. 28), finden aber keine Umsetzung im Kriterienkatalog.

Beschreibungen und Beobachtungen bezüglich Qualität und Einsatz der beurteilten Bildungssoftware. Im Folgenden gilt das Interesse den Problemen bei der Anwendung des Kriterienkatalogs. Dazu werden die durch die Qualitätsbeurteilung generierten Daten analysiert.

### **Analyse des Datenmaterials<sup>12</sup>**

Angestrebt wird im Folgenden eine erste Annäherung ans Material und eine Beschreibung von Tendenzen. Wir möchten feststellen, welche Anwendungsprobleme der Kriterienkatalog erzeugt und welche Annahmen in die Qualitätsbeurteilungen einfließen. Ziel dieser Betrachtung ist es, Schwachpunkte zu erkennen und Erkenntnisse für die Weiterentwicklung der Evaluation auf theoretischer und empirischer Ebene zu generieren.<sup>13</sup>

Dies führt zu folgenden Fragen für die Analyse der Daten: Wie evaluieren Lehrpersonen anhand des vorliegenden Kriterienkatalogs Bildungssoftware im Bereich Didaktik? Wie fallen die Bewertungen aus? Lassen sich Unterschiede für die verschiedenen Programmtypen feststellen? Welche Annahmen werden bezüglich Qualität getroffen?

### **Bewertung nach Programmtyp**

Wie sich gezeigt hat, setzen Lehrpersonen den Computer als ein Hilfsmittel ein, welches die Individualisierung des Unterrichts erleichtert und sich günstig auf bestimmte Fachleistungen auswirkt. In diesem Kontext erwarten sie Bildungssoftware, die auf die Lehrmittel abgestimmt ist. Da diese Erwartung vom bestehenden Bildungssoftwareangebot nicht erfüllt wird und die Lehrpersonen den Curriculumsbezug selbst herstellen müssen, ist anzunehmen, dass die Bewertungen für den Bereich «Didaktik» eher negativ ausfallen.

Im Folgenden zeigt *Tabelle 1* eine Übersicht der Anzahl der evaluierten Programme aufgeschlüsselt nach Programmtyp. Gleichzeitig wird die Verteilung der Bewertung durch Notengebung dargestellt. *Abbildung 1* veranschaulicht die Beurteilungen der verschiedenen Programme für den Bereich «Didaktik». Die schlechten und genügenden Bewertungen (Note 3 und 4) werden den guten bis sehr guten Bewertungen (Note 5 und 6) gegenübergestellt.

<sup>12</sup> Die evaluierten Datensätze können unter folgender Adresse aufgerufen werden: <http://www.pestalozzianum.ch:591/lernsoftware/standard.htm>

<sup>13</sup> Dies geschieht auch im Hinblick auf die Überführung des Projektes «Lernsoftware» in die Pädagogische Hochschule Zürich

Lernprogramme werden mit «Tutor», Übungsprogramme mit «Übg», Informationssysteme mit «Info» und Autorenprogramme mit «Tool» angesprochen.

Programmtyp	Anzahl total	Bewertung							
		3	%	4	%	5	%	6	%
Lernprogramme (tutorielle)	20	1	5	6	30	11	55	2	10
Übungsprogramme	44	1	2	15	34	20	45	8	18
Informationssysteme (Hypertexte)	30	2	7	3	10	20	67	5	17
Werkzeuge (Autorenprogramme)	3	0	0	0	0	3	100	0	0
	97	4	4	24		54		15	

Tabelle 1: Anzahl der evaluierten Programme und Bewertung nach Programmtyp

Auffallend ist, dass bei allen Programmtypen die positiven Bewertungen überwiegen. Ausserordentlich positiv werden Informationssysteme und Autorensysteme/Werkzeuge bewertet.

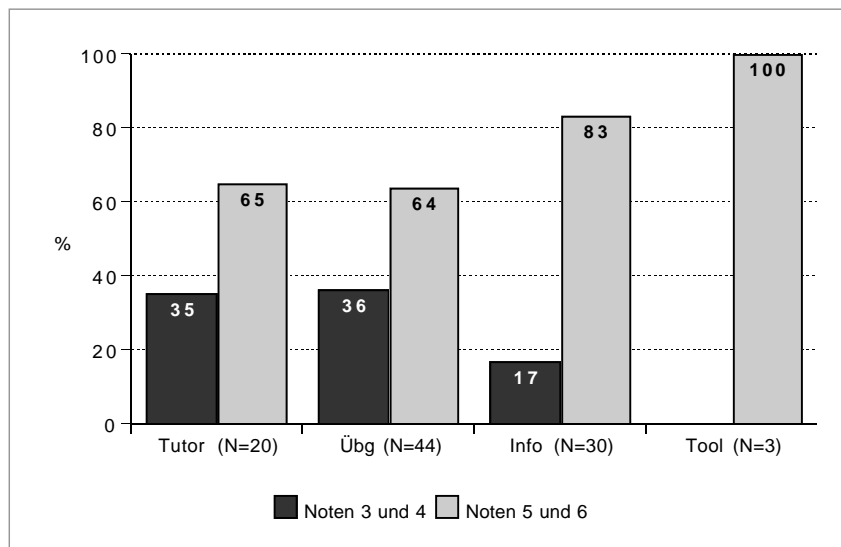


Abbildung 1: Beurteilung der Programmtypen im Bereich Didaktik

Wir können vorerst nur darüber spekulieren, weshalb die Beurteilungen bei allen Programmtypen für den Bereich Didaktik so positiv ausfallen. Die Qualität der Bildungssoftware hat sich in den letzten Jahren zwar verbessert und auch die Schulbuchverlage engagieren sich heute vermehrt in

diesem Bereich. Dennoch scheint die positive Bewertung nicht ausschliesslich mit der Qualität der Produkte zusammenzuhängen. Vielmehr vermuten wir Subjektivitätsaspekte seitens der Beurteilenden oder Wirkungen, die durch beobachtbare Lernaktivitäten oder Motivationseffekte die Bewertungen positiv beeinflussen.

Auch könnte die positive Einstellung dem Medium gegenüber die Bewertungen beeinflusst haben. Die Lehrpersonen beurteilen bereits als positiv, dass der Lerninhalt oder das Wissensgebiet für den Computereinsatz aufbereitet wurde. Weiter erscheint möglich, dass der Einsatz der Bildungssoftware im Unterricht die Entscheidung bezüglich Qualität des Programms bereits grundsätzlich in Richtung positiver Beurteilung beeinflusst. Die gute Bewertung würde dann lediglich bedeuten, dass das Programm im Unterricht integriert werden konnte. Wenn dies zutrifft, würde die Integrationsfähigkeit des Programms ein subjektives Qualitätskriterium darstellen. Natürlich hängt die Integrationsfähigkeit auch von der methodisch-didaktischen Kompetenz der Lehrperson ab. Weiter stellt sich die Frage, ob Lernaktivitäten und Motivationseffekte, die durch den Einsatz von Bildungssoftware im Unterricht beobachtet werden können, unabhängig von der Programmqualität positiv beurteilt werden. Dies wäre ein Hinweis auf implizite Lernauffassungen, die sich über die Produktevaluation in den Beurteilungen widerspiegeln. Wenn dies zutrifft, würde die Aktivitäts- und Motivationserzeugung ebenfalls ein subjektives Qualitätsmerkmal darstellen.

### Kriterien nach Programmtyp

Im Folgenden soll der Frage nachgegangen werden, ob sich unterschiedliche Qualitätskriterien für die einzelnen Programmtypen abzeichnen. Tabelle 2 und die Abbildungen 2, 3 zeigen die Nennungen für die positiven oder negativen Bewertungen auf.

Für die gut bewerteten Programme werden die Kategorien der Benotung 6, für die schlecht bewerteten Programme die Kategorien der Benotung 3 und 4 berücksichtigt. Dies ist darin begründet, dass die Notengebung überdurchschnittlich gut erscheint und die Note 6 somit eine gute und die Note 4 bereits eine schlechte Benotung darstellt. Zudem ist das Sample der Note 3 mit vier Produktbewertungen nicht sehr aussagekräftig. Die nachfolgende Tabelle gibt eine summarische Übersicht für alle drei Programmtypen.

Bereich	Kriterien	Progr: mnty					
		Tutor (%)		Übg. (%)		Info (%)	
		N=7	N=2	N=15	N=8	N=5	N=5
		B3, B4	B6	B3, B4	B6	B3, B4	B6
Funktion im Lernprozess	Eigenständiges Erarbeiten	100	100	73	62	60	100
	Üben (festigen, einprägen)	100	100	100	100	0	20
	Vertiefen (übertragen, nutzen)	100	100	53	100	40	80
Funktion des SW-Einsatzes	Informationssuche	14	0	0	0	80	80
	Kreative Arbeit	0	100	0	25	0	60
	Spielerische Arbeit	29	100	13	50	20	80
	Vielfältige Nutzungsmöglichkeit	29	100	13	88	0	80
Unterstützung ...des Lernens	Lerndiagnose/Fehlerkontrolle	0	100	13	88	0	20
	Anpassung der Schwierigkeit	29	100	27	50	0	0
...des Lehrens	Unterhaltungswert	57	100	13	50	40	80
	Lehrplanbezug	0	0	0	0	20	20

Tabelle 2: Nennungen von Kriterien im Bereich Didaktik in %

Abbildung 2 zeigt sowohl gleiche als auch unterschiedliche Anteile von Kriteriennennungen für gut und schlecht beurteilte Programme. Wir gehen davon aus, dass die bezüglich Häufigkeit übereinstimmenden Nennungen lediglich die Funktion des Programms berücksichtigen und damit programmimmanente Eigenschaften beschreiben. Sie entsprechen somit den Erwartungen an die Erfordernisse für den entsprechenden Programmtyp. Während die Unterschiede bezüglich Häufigkeit in den Kriteriennennungen auf Qualitätskriterien hinweisen.

Die programmimmanenten Eigenschaften für Lernprogramme sind aus Sicht der Lehrpersonen die Kriterien «eigenständiges Erarbeiten», «Vertiefen» und «Üben». Da Lernprogramme tutorielle Funktionen haben, stellt das «eigenständige Erarbeiten» eine notwendige Basisfunktion dieses Programmtyps dar. Die Feststellung, dass die evaluierten Lernprogramme durchwegs auch das «Vertiefen» und «Üben» ermöglichen, lässt sich auf die Übungsteile in tutoriellen Programmen zurückführen.<sup>14</sup>

Die Nennungen für die positiven Bewertungen zeigen, dass die Kriterien «Anpassung der Schwierigkeit» und «Lerndiagnose/Fehlerkontrolle»<sup>15</sup>

<sup>14</sup> Bezüglich Kriterien ist anzumerken, dass «Informationssuche» eigentlich kein didaktisches Kriterium darstellt.

<sup>15</sup> Die Fehlerkontrolle stellt u.E. eine Basisfunktion eines Tutorenprogramms dar und müsste hier als separates Item aufgeführt werden.

Qualitätskriterien darstellen. Das sind Kriterien, welche sich auf die Adaptivität des Programms beziehen.

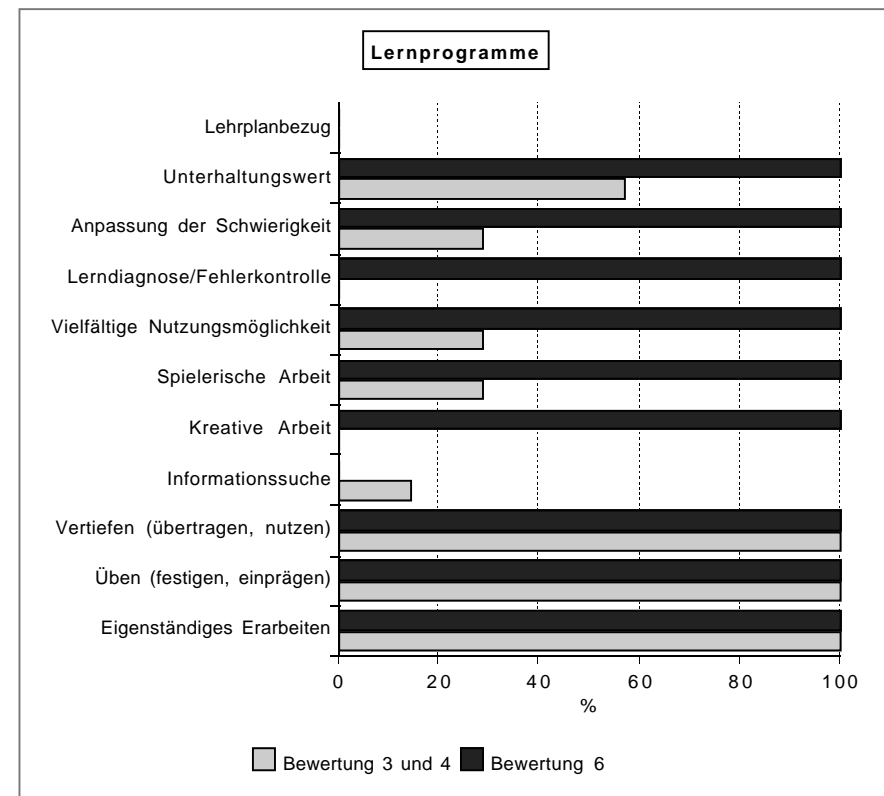


Abbildung 2: Lernprogramme: Nennungen der Kriterien in %

Es ist anzunehmen, dass sich die Kriterien «Unterhaltungswert» und «kreative Arbeit», «spielerische Arbeit» auf beobachtbare Motivations- oder Aktivitätseffekte des Programms beziehen. Diese Kriterien berücksichtigen die Reaktionen der Schüler/innen. Damit wird in der Evaluation eine zweite Ebene erfasst.

Unklar erscheint die Bedeutung des Kriteriums «vielfältige Nutzungsmöglichkeit» in Bezug auf tutorielle Programme, denn diese Programme geben eigentlich eine Lerneinheit vor, die zu erarbeiten ist. In diesem Sinne würde sich die Auslegung des Kriteriums auf den Wissenstransfer beziehen. Werden mit dem Programm bspw. Schlüsselkompetenzen oder kogni-

tive Grundkompetenzen erworben, so könnte man dieses Kriterium im Sinne der Transfermöglichkeit des Gelernten verwenden. Denkbar wäre auch, dass ein Ankreuzen dieses Kriteriums die Kombination von Lern- und Übungsmöglichkeit als Qualitätsmerkmal hervorheben möchte.

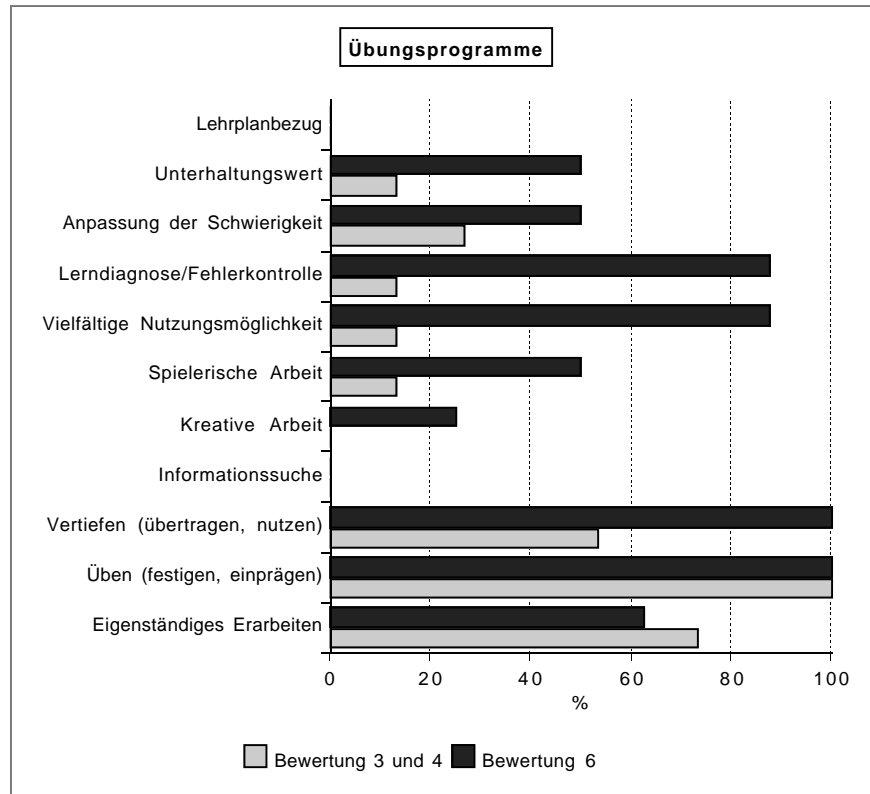


Abbildung 3: Übungsprogramme: Nennungen der Kriterien in %

Wie aus *Abbildung 3* ersichtlich ist, berücksichtigen die bezüglich Häufigkeit übereinstimmenden Nennungen wiederum lediglich die Funktion des Programms. Für die Übungsprogramme wird erwartungsgemäss das Kriterium «Üben» als zentral betrachtet.

Die Qualitätskriterien für Übungsprogramme sollten sich auch hier vor allem in den Unterschieden zwischen guten und schlechten Beurteilungen äussern. So erscheint die Möglichkeit, den Stoff zu «vertiefen» als deutliches Qualitätsmerkmal, während das «eigenständige Erarbeiten» nicht als

solches genannt wird.. Das «eigenständige Erarbeiten» wird nicht als positiv abgegrenztes Qualitätsmerkmal genannt. Das ist nachvollziehbar, da bei Übungsprogrammen der Stoff vorher gelernt werden muss. Die Unterschiede zwischen guten und schlechten Beurteilungen zeichnen sich im Weiteren vor allem dort ab, wo eine «Fehlerkontrolle» gegeben ist und das Programm «vielfältige Nutzungsmöglichkeiten» zulässt. Es ist anzunehmen, dass sich letztere Einschätzung auf die Varianz der Übungsmöglichkeiten bezieht. Das Kriterium «Anpassung der Schwierigkeit» wird sich wohl auf die Vielfalt der Lernaktivitäten beziehen, die das Programm ermöglicht. Auch bei diesem Programmtyp ist die Bedeutung der motivationalen Kriterien schwierig festzustellen.

Auf eine Betrachtung der Informationssysteme werden wir im Folgenden verzichten, da die vorliegenden Kriterien nicht genügend aussagekräftig erscheinen. Es fällt auf, dass der Lehrplanbezug bei den Lern- und Übungsprogrammen gänzlich fehlt und bei den Informationssystemen nur bei 20% der sowohl gut als auch schlecht evaluierten Produkte angegeben wird. Auffallend ist dies umso mehr, da wir ja annehmen können, dass die Lehrpersonen die Übungs- und Lernprogramme nach den curricular festgelegten Inhalten im Unterricht einsetzen.

### Zusammenfassung der Ergebnisse

Angestrebt wurde eine erste Annäherung an die Evaluationsergebnisse einer praxisnahen Qualitätsbeurteilung von Bildungssoftware. Wir wollten für den Bereich Didaktik feststellen, welche Anwendungsprobleme der Kriterienkatalog erzeugt und inwiefern implizite Annahmen in die Qualitätsbeurteilungen einfließen.

Die Analyse der Evaluationsdaten ergab folgende Anwendungsprobleme:

- Unklare Zuordnung der Produkte zur Programmtypologie<sup>16</sup>
- Grosser Interpretationsspielraum in der Anwendung der Kriterien
- Unverständliche oder ungeeignete Kriterien je nach Programmtyp<sup>17</sup>
- Unterschiedliche Auslegung der Kriterien je nach Programmtyp

<sup>16</sup> Sehr deutlich wurde dies bei der Operationalisierung der Begriffe «lernen» und «üben», so dass die Lehrpersonen eine eindeutige Zuordnung von Lern- und Übungsprogrammen gar nicht leisten konnten.

<sup>17</sup> Es hat sich ja auch gezeigt, dass der vorliegende Kriterienkatalog vor allem auf die Evaluation von Übungs- und Lernprogramme ausgerichtet ist und für die Evaluation von Informationssystemen nicht geeignet ist.

Als weiteres Ergebnis hat sich die unerwartet hohe Bewertung für alle Programmtypen im Bereich Didaktik herausgestellt. Diese Tatsache ist wohl nicht ausschliesslich auf die Produktqualität zurückzuführen und wir haben folgende Vermutungen über mögliche Effekte angestellt:

- Positive Einstellung dem Medium Computer gegenüber
- Beobachtbare Lernaktivität und Motivationserzeugung durch den Einsatz der entsprechenden Bildungssoftware im Unterricht
- Integrationsmöglichkeit der Bildungssoftware im Unterricht

Zudem sind wir auf implizite Annahmen in den Qualitätsbeurteilungen gestossen, die wir im Folgenden etwas pointiert darstellen:

- Die Kombination von Lernen und Üben wird bei tutoriellen Programmen vorausgesetzt
- Beobachtbare Aktivitäten und Motivationseffekte garantieren Lern- und Übungserfolge
- Die Programme werden im Unterricht integriert ohne Angabe des Lehrplanbezugs

Diese Ergebnisse zeigen, dass der vorliegende Kriterienraster und das Beurteilungsverfahren subjektive Betrachtungen bezüglich Qualität von Bildungssoftware hervorbringen, die vom Einzelnen ausgehen. Die Lehrperson, welche eine spezielle Bildungssoftware evaluiert, befindet sich in einer individuellen Situation mit spezifischen Zielen vor Augen. Daraus werden die subjektiven Ansprüche an die Qualität der Bildungssoftware abgeleitet. Dieser Umstand ist problematisch, weil diese nicht intendierte Subjektivität im Evaluationsprojekt weder reflektiert noch deklariert wird.

### **Ausblick**

Ziel der Betrachtungen war, Erkenntnisse für die Weiterentwicklung der Evaluation auf theoretischer und empirischer Ebene zu generieren.

Wichtig ist die Erkenntnis, dass die Qualitätskriterien für die verschiedenen Programmtypen unterschiedliche Bedeutungen haben können und somit individuell interpretiert werden müssen. Diese Einsicht zeigt, worin denn nun diese Subjektivität auch begründet ist.

Interpretationsspielraum ergibt sich durch die Notwendigkeit der subjektiven Ermittlung der Bedeutung der Kriterien durch die Evaluierenden. Das liegt einerseits an der fehlenden Reflexion, andererseits

an der fehlenden theoretischen Fundierung und Operationalisierung der Evaluation hinsichtlich Kriterienkatalog und Programmtypologie, sowie an der fehlenden Schulung der Evaluator/innen.

Das Problem der fehlenden Objektivität zeigt sich vor allem auch daran, dass nicht klar wird, worauf sich die ermittelte Bewertung bezieht. Der gewählte Beurteilungsmodus wirkt sich ungünstig auf die Bewertungen aus und muss bezüglich theoretischer Gewichtung der verwendeten Kriterien überarbeitet werden.

Eine Instruktion oder Schulung der evaluierenden Personen ist unumgänglich. Zudem müssen bereits für die Auswahl der zu evaluierenden Produkte Kriterien festgelegt werden. Nur so können auch Programme berücksichtigt werden, die im Marktangebot nicht so präsent sind, oder es können gezielt solche Programmtypen evaluiert werden, deren Evaluation etwas aufwändiger ist, wie z.B. Simulationen, multimediale Lernumgebungen und Autorenprogramme.

Unumgänglich ist auch die Berücksichtigung von Leistungsstandards, wie sie bspw. Scriven in seiner dreistufigen Bedürfnisanalyse vorschlägt. Eine Ermittlung der so genannten Basiskriterien (*necessitata*) würde eine Vorselektion der zu evaluierenden Bildungssoftware ermöglichen.

Für die seriöse Bewertung der verschiedenen Programmtypen müssen differenzierte Leistungsstandards ausgearbeitet werden. Die Qualität der Bildungssoftware würde dann ausschliesslich über die Vorzüge (*desiderata*) eines Programmes ermittelt. Das Entwicklungspotenzial würde sich in den Zielvorstellungen (*ideals*) abzeichnen und nur dann eine sehr hohe Bewertung rechtfertigen. Dabei kann die Festlegung der Leistungsstandards nicht ohne Theoriebezug geschehen, da je nach lerntheoretischem Bezug unterschiedliche Qualitätsansprüche angenommen werden können.. So wäre bspw. die Möglichkeit zur Lerndiagnose bei einem Lernprogramm aus konstruktivistischer Sicht als Basisfunktion aufzufassen, während dies aus behavioristischer Perspektive wohl eher als Vorzug eines Programms gelten würde. Bei einem Übungsprogramm wird dies wohl aus beiden Perspektiven als Vorzug oder gar Zielvorstellung aufgefasst. Für Informationssysteme erscheint diese Funktion als eher irrelevant.

Es wurde deutlich, dass die Kriterien zum Lehrplanbezug sehr sorgfältig ausgearbeitet werden müssen, um überhaupt erhoben werden zu können. Der Nachweis dieses Kriteriums ist wichtig und entspricht u.a. dem ausdrücklichen Wunsch der Lehrkräfte nach Orientierung. Bei der Ausarbeitung des Lehrplanbezugs ist sowohl der fachspezifische Bezug

wichtig, aber auch das Aufzeigen von Bezügen in Richtung formale Bildung, d.h. in Richtung Schlüsselkompetenzen und dem Erlernen von kognitiven Grundkompetenzen.

Im Weiteren hat sich gezeigt, dass die Qualitätsbeurteilung des Produktes eigentlich nicht viel über die Lerneffekte aussagen kann und dass die offenen Antwortfelder im Bereich Didaktik die Lehrpersonen dazu animiert haben, Motivationseffekte, Lernaktivitäten oder den Integrationsanlass der Software zu beschreiben. Es braucht neben der Produktebeurteilung somit eine zweite Ebene der Evaluation. Offene Antwortmöglichkeiten sind methodisch sinnvoll, um die pädagogisch-didaktische Sicht in die Bewertung von Bildungssoftware zu integrieren und dabei den Kontext von Lerninhalten und Anwendung zu ermitteln. Allerdings sollte dies über Leitfragen geschehen, um eine inhaltliche Strukturierung der Antworten zu erreichen.

Gerade in der Kombination von standardisierten und offenen Verfahren liegt das Potenzial einer praxisnahen Evaluation von Bildungssoftware.

### Literatur

- Baumgartner, Peter. (1995). «Didaktische Anforderungen an (multi-mediale) Lernsoftware.» In: L. J. Issing & P. Klimsa (Hrsg.). *Information und Lernen mit Multimedia*. Weinheim: Psychologie Verlags Union. S. 242–252
- Büeler, X; Stebler, R; Stöckli, G; Stolz, D. (2001). *Schulprojekt 21. Lernen für das 21. Jahrhundert? Externe wissenschaftliche Evaluation. Schlussbericht zuhanden der Bildungsdirektion des Kantons Zürich*. (S. 65–111; 242f.). Download unter <[www.schulprojekt21.ch](http://www.schulprojekt21.ch)>
- Fricke, R. (2000). «Qualitätsbeurteilung durch Kriterienkataloge. Auf der Suche nach validen Vorhersagemodellen.» In: Schenkel, P. et al. (Hrsg.). *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme auf dem Prüfstand*. Nürnberg: Bildung und Wissen. S. 75–88
- Korbmacher, K. (2000). «Evaluation von Lernsoftware auf der Basis von SODIS.» In: Schenkel, P. et al. (Hrsg.). *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme auf dem Prüfstand*. Nürnberg: Bildung und Wissen. S. 190–216
- Meier, A. (2000). «MEDA und AKAB: Zwei Kriterienkataloge auf dem Prüfstand.» In: Schenkel, P. et al. (Hrsg.). *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme auf dem Prüfstand*. Nürnberg:

Bildung und Wissen. S. 164–189

- Pfyffer, A. (2001). *Evaluation Lernsoftware im Unterricht. Didaktische Überlegungen und Produkteempfehlungen*. Werkstatt und Projektberichte. Zürich: Pestalozzianum
- Pfyffer, A. (1998). «Lernprogramme: Qualitätskriterien, Einsatzmöglichkeiten.» In: *schweizer schule* 4/98, S. 26–30
- Scriven, M. (1991). *Evaluation Thesaurus*. 4. Aufl. Newbury Park: Sage Publ
- Tergan, S.-O. (2001). «Qualitätsbeurteilung von Bildungssoftware mittels Kriterienkatalogen. Problemaufriss und Perspektiven.» In: *Zeitschrift für Lernforschung*, 29. Jg, Heft 4, S. 319–341