

# Qualitätskriterienbasierte Analyse von Lernsystemen

Dissertation zur Erlangung des akademischen Grades eines  
Doktors der Ingenieurwissenschaften (Dr.-Ing.)  
im Fachbereich Maschinenbau  
der Universität Kassel

vorgelegt von: Dipl.-Ing. Ingo Wagner  
geboren am 7.11.1972  
in Konstanz

Gutachter:  
Prof. Dr.-Ing. Dr. h.c. Gunnar Johannsen  
Prof. Dipl.-Ing. Adolf Reinhardt

Disputation: 01.02.2006



## 0 Vorwort

Als rohstoffarmes, exportorientiertes Hochlohnland ist Deutschland in besonderem Maße auf ein leistungsfähiges universitäres Bildungssystem angewiesen. Qualifizierte Wissenschaftler, Forscher, Studenten und Absolventen von Universitäten stellen hierfür eine wichtige Zukunftsressource dar. Sie tragen zu einer wirtschaftlich stabilen Überlebenschance der Nation bei durch Wissen, Kompetenz, Forschungsgeist, Kreativität und Innovation. Dies sind Ziele und Ergebnisse von Lernprozessen, die eine inhalts- und methodenreiche, multimediale, interaktive Bildungslandschaft bedingen (BmBF 2002).

Die Begriffe Neue Medien, E-Learning und virtuelle Hochschule stehen für eine visionäre Bildungslandschaft, in der sich durch den gezielten Einsatz von Informations- und Kommunikationstechnologie bemüht wird, eine neue Qualität des Lehrens und Lernens zu installieren. Dies ist kein neuer Gedanke, wenn man sich die geschichtliche Entwicklung technikbasierter Lehrens und Lernens anschaut. Bei Niegemann *et al.* (2004, S.3ff.) findet sich hierzu eine historische Beschreibung: angefangen vom Leserad von Ramelli 1588, über Skinner, H. 1866, der das erste Patent auf eine Lernmaschine erhielt, über die Buchstabiermaschine von Aikins 1911, über Presseys Test- und Lernmaschine von 1926, über die linearen Lehrprogramme von Skinner, B. und Holland, über Crowdes verzweigte Lehrprogramme 1959, hinzu neueren Entwicklungen des computerunterstützten Unterrichts (CUU) und der aktuell populären Bezeichnung E-Learning für Lernsysteme, die die Anwendung einer Webtechnologie ermöglichen.

Mit dem Einsatz von Lernsystemen kann eine zeitliche und örtliche Unabhängigkeit des Lehrens und Lernens, sowie die Ansprache von unterschiedlichen Wahrnehmungskanälen des Menschen ermöglicht werden. Die traditionellen Bildungsmöglichkeiten sind bereits in manchen Lehrveranstaltungen an Universitäten durch alternative Lernwege und effizientere Lernprozesse erweitert worden. Die Studierenden können von beinahe überall aus über ein Lernsystem zeitlich synchron mit dem Lehrenden verbunden sein. Vielversprechend sind Lehrprozesse, die die Stärken der Lernsysteme mit personalen Vermittlungsformen kombinieren (Blended Learning) (Wagner 2005, S.26). Die Wissensvermittlung kann jedoch auch asynchron von der gleichzeitigen Präsenz des Lehrenden entkoppelt verlaufen.

Der Technikeinsatz zur Vermittlung von Lerninhalten ist in Übereinstimmung mit Schenkel (2000a, S.7) ein notwendiger Fortschritt und Baustein für unsere Lern- und Wissensgesellschaft: „Wenn Wissen zum kritischen Erfolgsfaktor von Gesellschaften, Unternehmen und Individuen wird – und dies ist in der Wissensgesellschaft unabweisbar – dann sind Formen des Lehrens und Lernens erforderlich, die sich nicht allein auf personale Vermittlungsprozesse stützen. Die ständige Verfügbarkeit eines Lehrers, Ausbilders, Trainers kann nicht die Grundlage lebenslangen Lernens bilden“.

Schulmeister (1998, S.39) nennt acht Strategien, die sich auf den Computereinsatz in Lehr- und Lernprozessen vorteilhaft auswirken:

1. Förderung der Studiertechnik;
2. Qualitätsverbesserung bei fachwissenschaftlichen Arbeiten;
3. Nutzung kognitiver und konstruktiver Werkzeuge;
4. Individuelles Lernen mit Lernprogrammen;
5. Das Prinzip der Virtualität;
6. Medienunterstützung in konventioneller Lehre;
7. Kooperatives Lernen in telematischen Umgebungen;
8. Bereicherung des Lehrangebots.

Viele Universitäten, so auch die Universität Kassel, richten Multimedia-Koordinationszentren zur Unterstützung der E-Learning-Aktivitäten der Lehrenden ein. Für die Studierenden werden gut ausgestattete Lernräume angeboten. Trotzdem, so zeigen eigene Erfahrungen aus dem Lehr- und Lernbereich Maschinenbau, wird diese Unterstützung von den Dozenten und den Studierenden überwiegend nicht genutzt. Woran liegt das? Wie lässt sich bereits während der Entwicklung, die Eignung bzw. die Qualität von Lernsystemen feststellen, bewerten und sichern?

Im Rahmen meiner Dissertation ging ich diesen und anderen Fragen nach. Dies wurde begünstigt durch die Lehraufgaben<sup>1</sup> im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter an der Universität Kassel im Fachbereich Maschinenbau. Mit den vielfältigen Lehrgelegenheiten und dem Kontakt zu Studenten, Lehrenden, Entwicklern und Administratoren wurde von mir die Wissensvermittlung mit Neuen Medien unvoreingenommen in einer authentischen Lehrumgebung erprobt, evaluiert und weiterentwickelt.

In der vorliegenden Arbeit wird die Evaluation von unterschiedlichen Lernsystemen fokussiert. Die durchgeführten kriterien- und webbasierten Expertenbefragungen zur Qualitätssicherung der Lernsysteme förderten zahlreiche Erkenntnisse zum Potenzial der Evaluationsmethodik zutage.

Mit dieser sowohl theorie- als auch praxisorientierten Forschungsarbeit wird den Lesern, die sich mit der schwierigen Aufgabe der Installation einer effektiven und effizienten Qualitätssicherung von Lernsystemen befassen, mit geeigneten Systematisierungen, neuen Methoden, Hypothesen und Studien zur Qualitätssicherung, sowie mit praktischen Werkzeugen zur Umsetzung der Qualitätssicherung von Lernsystemen gedient.

---

<sup>1</sup>Systemtechnik-Übung, Mensch-Maschine-Systeme-Praktikum, Mensch-Rechner-Interaktion-Übung, Seminar Mess- und Automatisierungstechnik (Vorträge) und Vorlesung Wissensbasierte Systeme (Vortrag Wissensmanagement und künstliche neuronale Netze).

Gedankt sei an dieser Stelle Herrn Professor Dr.-Ing. Dr. h.c. Gunnar Johannsen, der mir als geschäftsführender Direktor des Instituts Mess- und Automatisierungstechnik Raum und Zeit für die Durchführung der Forschungsarbeit bot und der mir stets durch seine Diskussionsbereitschaft ein wertvoller Begutachter meiner Arbeit war. Herrn Prof. Dipl.-Ing. A. Reinhardt danke ich für seine eingehende Durchsicht der vorliegenden Arbeit.

Ich danke allen Kollegen Dipl.-Ing. Gerd Strätz und Natascha Feder für ihre Hilfsbereitschaft, Dr.-Ing. Burkhard Borys für seine intensive Beratung und Dr.-Ing. Andreas Völkel, der mir die Erstellung dieser Arbeit mit dem Textsatzsystem L<sup>A</sup>T<sub>E</sub>X empfohlen hat.

Allen Versuchsteilnehmern sei gedankt, da durch die erhobenen Untersuchungsdaten interessante wissenschaftliche Erkenntnisse gewonnen werden konnten.

Ich danke meinen Eltern für die wertvolle Unterstützung und insbesondere meiner Frau Tanja Wagner, die mich während der Promotionszeit mehr mit Notebook und Fachbüchern gesehen hat als ohne.

### **Erklärung nach Paragraph 5 der Promotionsordnung**

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

# Inhaltsverzeichnis

<b>0</b>	<b>Vorwort</b>	<b>i</b>
	<b>Inhaltsverzeichnis</b>	<b>iv</b>
<b>I</b>	<b>Überblick und Ausgangssituation</b>	<b>1</b>
<b>1</b>	<b>Kurzfassung</b>	<b>1</b>
<b>2</b>	<b>Zusammenfassung</b>	<b>1</b>
<b>3</b>	<b>Ziele</b>	<b>3</b>
<b>4</b>	<b>Aufbau der Arbeit</b>	<b>3</b>
<b>5</b>	<b>Qualitätssicherung der Lehre</b>	<b>4</b>
<b>6</b>	<b>Qualitätssicherung von Lernsystemen</b>	<b>6</b>
6.1	Qualitätsbegriff . . . . .	6
6.2	Lernsystem-Evaluation . . . . .	14
6.3	Qualitätskriterien . . . . .	16
6.4	Qualitätskriterienkataloge . . . . .	17
6.5	Schwächen von Qualitätskriterien . . . . .	19
6.6	Vorzüge von Qualitätskriterien . . . . .	21
6.7	Güte von Qualitätskriterien . . . . .	22
<b>II</b>	<b>Qualitätssicherungsmethodik</b>	<b>24</b>
<b>7</b>	<b>Qualitätsverbesserungsprozess</b>	<b>24</b>
7.1	Wirkungskreise . . . . .	25
7.2	Vorentwicklung . . . . .	27
7.3	Anforderungsanalyse . . . . .	28
7.4	Qualitätsregelteam . . . . .	29
7.5	Lernsystem-Prototyping . . . . .	30
7.6	Formative Evaluation . . . . .	31
7.7	Summative Evaluation . . . . .	32
7.8	Nachbesserungsevaluation . . . . .	32
7.9	Produktevaluation . . . . .	33

<b>8</b>	<b>Expertenbefragung</b>	<b>33</b>
8.1	Anreizsysteme . . . . .	34
8.2	Webbasierte Befragung . . . . .	35
8.2.1	Layout des LSA-Fragebogens . . . . .	38
8.2.2	Metainformationen . . . . .	41
8.2.3	Fragefolge . . . . .	43
8.2.4	Vortest des LSA-Fragebogens . . . . .	44
8.3	Lernsystem-Konfiguration . . . . .	47
<b>9</b>	<b>Die Software LernSystemAnalyse</b>	<b>49</b>
9.1	Projektverwaltung . . . . .	50
9.2	Anforderungen . . . . .	50
9.3	Personenverwaltung . . . . .	51
9.4	Qualitätskriterienkatalog . . . . .	52
9.5	Evaluation . . . . .	60
<b>III</b>	<b>Lernsysteme</b>	<b>62</b>
<b>10</b>	<b>Autorensysteme</b>	<b>63</b>
<b>11</b>	<b>Lernsystem-Klassifikation</b>	<b>65</b>
<b>12</b>	<b>Entwickelte Lernsysteme zur Systemtechnik</b>	<b>68</b>
12.1	Fuzzy-Logik CBT . . . . .	70
12.2	Fuzzy-Logik WBT . . . . .	71
12.3	Fuzzy-Logik Kransteuerung . . . . .	72
12.4	Magnat Schwebebahn . . . . .	74
12.5	Planspiel-Autorensystem . . . . .	75
<b>IV</b>	<b>Studie zur Expertenbefragung</b>	<b>77</b>
<b>13</b>	<b>Versuchsplanung</b>	<b>77</b>
13.1	Expertenanzahl . . . . .	78
13.2	Fragestellungen . . . . .	82
13.3	Versuchsvariablen . . . . .	83
<b>14</b>	<b>Durchführung</b>	<b>85</b>

<b>15 Auswertung</b>	<b>90</b>
15.1 Qualitätsindikatoren . . . . .	91
15.2 Zeitaufwand . . . . .	95
15.3 Hinweisklassen . . . . .	97
15.3.1 Begründungstexte . . . . .	98
15.3.2 Interaktionsgrad . . . . .	99
15.3.3 Fragenanzahl . . . . .	101
15.3.4 Expertengrad . . . . .	103
15.3.5 Anforderungserfüllung . . . . .	107
15.4 Hinweisklassen – Expertenanzahl . . . . .	109
15.5 Monopol-Hinweisklassen – Expertenanzahl . . . . .	112
15.6 Standardabweichung der Bepunktung . . . . .	117
15.7 Freie Antwortmöglichkeit . . . . .	122
15.8 Qualitätskriterien . . . . .	123
15.9 Metainformationen . . . . .	125
<b>16 Übertragbarkeit der Ergebnisse</b>	<b>127</b>
<b>17 Ausblick</b>	<b>128</b>
<b>V Verzeichnisse</b>	<b>131</b>
<b>Abbildungsverzeichnis</b>	<b>131</b>
<b>Tabellenverzeichnis</b>	<b>133</b>
<b>18 Glossar</b>	<b>134</b>
<b>Literaturverzeichnis</b>	<b>138</b>

## Teil I

# Überblick und Ausgangssituation

## 1 Kurzfassung

Drei Faktoren sind für die Qualität von Lernsystemen entscheidend: qualifiziertes Personal, eine adäquate unterstützende Technologie und eine die beiden Faktoren umfassende Qualitätssicherungsmethodik.

Diese Arbeit stellt keine Enzyklopädie der Qualitätssicherung von Lernsystemen dar, sondern es wird ein spezielles Qualitätssicherungsverfahren vorgestellt und untersucht. Der Untersuchungsschwerpunkt betrifft das Potenzial einer Expertenbefragung für den formativen (entwicklungsbegleitenden) Evaluationskontext von Lernsystemen.

Die Experten haben die Aufgabe, ein selbst getestetes Lernsystem anhand von Qualitätskriterien zu bewerten. Durch Punktvergabe erfolgt die Gewichtung und Bewertung der Umsetzung der Qualitätskriterien im Lernsystem. Zusätzlich ist der Experte aufgefordert, seine Bewertung schriftlich zu begründen.

Im Rahmen einer Studie wird untersucht, welche Einflussgrößen sich wie auf die Anzahl der für die Qualitätssicherung nutzbaren Expertenhinweise auswirken. Mit diesem Wissen kann das Potenzial des Verfahrens auch bei sich ändernden Rahmenbedingungen angemessen abgeschätzt werden.

## 2 Zusammenfassung

Die Entwicklung von Lernsystemen ist eine zeit- und kostenintensive Aufgabe, die das Zusammenwirken vielfältiger Kompetenzen<sup>2</sup> während der gesamten Wertschöpfungskette bedingt. Neben der Kürzung der benötigten Zeit und der Kosten der gemeinsamen Arbeit, rückt zunehmend die Qualität der Lernsysteme in den Vordergrund von Zielformulierungen.

Viele Lernsysteme werden mit dem Anspruch entwickelt, das bestehende Lehrangebot aufzuwerten. Doch wie überprüft man diesen Anspruch bei Lernsystem-Prototypen frühzeitig? Gibt es Kriterien, die sich eignen, um die Qualität eines Lernsystems zu messen und das Lernsystem in der mit dem Kunden vereinbarten und von ihm erwarteten Qualität zu fertigen? Ist die Expertenbefragung ein probates und effizientes Verfahren?

---

<sup>2</sup>Projektmanager, Programmierer, Autoren, Designer, Lerninhalteexperten, Pädagogen u.a.

Die Qualitätssicherung eines Lernsystems erfordert ein in den Entwicklungsprozess verankertes Vorgehensmodell, sowie Qualitätskriterien und Software, durch die die Qualität eines Lernsystem-Prototyps ganzheitlich und effizient bewertet werden kann. Das Vorgehensmodell besteht im Kern aus der entwicklungsbegleitenden sogenannten formativen Evaluation (siehe Kap. 7 auf Seite 24). Ihr Potenzial ist, frühzeitig kreative Problemlösungsprozesse zur Verbesserung von Lernsystem-Prototypen anzustoßen.

In einem Zeitraum von vier Jahren wurden mehrere Lernsystem-Projekte mit dem Vorgehensmodell entwickelt. Untersucht wurde eine Auswahl von Lernsystem-Prototypen (siehe Kap. 12 auf Seite 68), die sich in wesentlichen technischen und didaktischen Aspekten unterscheiden. Anhand dieser Prototypen wird das Potenzial der kriterienbasierten Analyse zur Qualitätssicherung von Lernsystemen in einer Studie aufgedeckt (siehe Teil IV auf Seite 77). Bei den entwickelten Lernsystemen handelt es sich um ein Planspiel, ein Planspiel-Autorensystem und drei computer- und webbasierte Trainings- und Simulationssysteme zur Fuzzy-Logik, die alle in der Systemtechnik-Lehre zum Einsatz kommen. Diese Lernsysteme wurden mit Expertenbefragungen entwicklungsbegleitend bewertet und kontinuierlich verbessert.

Bei allen Lernsystemen wurde für den Bewertungsprozess auf denselben Qualitätskriterienkatalog zurückgegriffen (Lernsystem-Konfiguration, siehe Kap. 8.3 auf Seite 47). Die Lernsystem-Konfiguration deckt multiple Qualitätsaspekte ab. Sie vereint und bewahrt fächerübergreifend allgemeines und projektspezifisches Systemtechnik-Wissen, technisches Design-Knowhow und pädagogisch-didaktische Expertise zur Evaluation von Lernsystemen.

Neben dem Vorgehensmodell zur Qualitätssicherung von Lernsystemen wird die methoden-angebundene webbasierte LernSystemAnalyse-Software (siehe Kap. 9 auf Seite 49) vorgestellt. Durch den Einsatz von Informations- und Kommunikationstechnologie kann der Aufwand für die formative Evaluation in engen Grenzen gehalten werden.

Die qualitätskriterienbasierte Analyse entspricht dem Bedarf der Entwicklungspraxis an einer schnellen und handhabbaren Qualitätssicherung. Die Analyse berücksichtigt den Anspruch der Wissenschaft an eine fundierte und korrekte Durchführung der Evaluation von Lernsystemen.

### 3 Ziele

„In der Wissenschaft und in der Praxis fehlt die dringend notwendige Auseinandersetzung über die Optimierung von Lernarchitekturen und Lernprozessen“ (Schenkel 2000b, S.54). Mit dieser Arbeit soll ein wissenschaftlicher Beitrag zur Qualitätssicherung von Lernsystemen durch formative Evaluation geliefert werden. Eine wichtige Nebenbedingung für diese Evaluation ist, dass sie die Entwicklung von Lernsystemen nicht aufhält, sondern sie fördert. Dies erfordert eine einfache Integration und effiziente Durchführung der Evaluation in den Entwicklungsprozess eines Lernsystems. Für diesen Kontext ist das Potenzial von Expertenbefragungen durch eine eigene Studie gründlich zu untersuchen. Die Organisation, Durchführung und Auswertung der Expertenbefragung soll durch ein geeignetes softwaretechnisches Werkzeug unterstützt werden.

### 4 Aufbau der Arbeit

In dieser Arbeit wird eine Qualitätssicherungsmethodik vorgestellt, die schwerpunktmäßig über Expertenbefragungen Ansatzpunkte zur Verbesserung von untersuchten Lernsystemen generiert. Die Arbeit ist in fünf aufeinander aufbauende Teile gegliedert:

Teil I erläutert die Notwendigkeit von Qualitätssicherung der Lehre und der eingesetzten Lernsysteme. Nach der Klärung von Begrifflichkeiten zur Qualitätssicherung von Lernsystemen wird die Qualitätsbewertung anhand von Qualitätskriterien diskutiert.

In Teil II wird die Qualitätssicherungsmethodik für die Entwicklung und Evaluation von Lernsystemen beschrieben. Diese ist auch für kleine und mittlere universitäre Projekte geeignet. Die Umsetzung der Qualitätssicherungsmethodik wird durch die hier vorgestellte LernSystemAnalyse-Software begünstigt.

Teil III führt die für den Einsatz in der Systemtechnik-Lehre entwickelten und evaluierten Lernsysteme auf.

Der Teil IV enthält eine Studie zur Expertenbefragung zur Potenzialabschätzung dieser Methode zur formativen Qualitätssicherung von Lernsystemen.

Eine Auswahl verwendeter Begrifflichkeiten kann alternativ zu den Erläuterungen des Kontextes im Glossar (siehe Kap. 18 auf Seite 134 im Teil V) nachgeschlagen werden.

## 5 Qualitätssicherung der Lehre

Das Hochschulrahmengesetz von 1998 machte Qualitätssicherung der Lehre in §6 ausdrücklich zur Hochschulaufgabe. Der Rat der EU hatte im gleichen Jahr eine Empfehlung zur europäischen Zusammenarbeit zur Qualitätssicherung in der Hochschulbildung verabschiedet. Das daraufhin im März 2000 gegründete europäische Qualitätsnetzwerk European Network for Quality Assurance in Higher Education (ENQA) unterstreicht die Bedeutung der Qualitätssicherung an Hochschulen. Die Mitgliedstaaten sollen transparente Qualitätsbewertungssysteme an Hochschulen fördern oder schaffen, falls noch keine bestehen.

In fast allen Landeshochschulgesetzen wurden die Hochschulen verpflichtet, Rechenschaftsberichte bzw. Lehrberichte zu verfassen. Diese dienen der Kontrolle der Studentenzahlen, der Qualitätssicherungsmaßnahmen und des verantwortungsbewussten Einsatzes von Steuergeldern. Allerdings kann nur vorgeschrieben werden, dass eine Einrichtung interne Qualitätsentwicklung betreibt, aber nicht wie.

Der Wille zur Qualitätssicherung seitens der Hochschulen ist auch vorhanden, denn gerade im Angesicht von sinkenden Mittelzuweisungen gilt es, wettbewerbsfähige und effiziente Studiensysteme zu realisieren. Dazu sind „die Hochschulen an ihrem wissenschaftlichen und an ihrem Ausbildungserfolg zu messen. Die Konkurrenzfähigkeit im internationalen Rahmen ist eine wichtige Nebenbedingung“ (Lömker 1998, S.20). In Übereinstimmung mit Blom (2000, S.21) empfiehlt es sich, eine Studiensituation zu schaffen, die Studierenden hilft, in der Berufswelt zu bestehen.

An der Universität Kassel richtet sich der prüfende Blick der Hochschulleitung und einzelner Fachbereiche bzw. Fachgebiete unlängst auf die Qualität ihrer Präsenzlehre und den möglichen bzw. schon realisierten Einsatz von Lernsystemen zur Unterstützung der Aus- und Weiterbildung. Der Antrieb liegt im Verständnis der Universität als eine Art von Dienstleister, der den Studierenden durch sein akademisches Lehrangebot eine gute Berufsausbildung anbietet.

Im Wintersemester 2004/2005 wurden die Studenten des FB15 Maschinenbau der Universität Kassel gebeten, die zutreffenden Aussagen (siehe Tab. 1 auf der nächsten Seite) in Bezug auf die besuchte Lehrveranstaltung per Antwortskala 1 „trifft völlig zu“ bis 5 „trifft überhaupt nicht zu“ zu beantworten.

Die Ergebnisse aus den Mitteilungen vom Dekanat Maschinenbau (2005, S.7) zeigen, dass die Dozenten über alle Lehrveranstaltungen gut vorbereitet wirken, freundlich und aufgeschlossen gegenüber den Studenten auftreten. Das Verbesserungspotenzial hierzu liegt bei der Anregung, der Anforderung und dem Interesse der Lernenden. Dies könnte in Zukunft beispielsweise durch den Einsatz von interessanten Medien (Lernsystemen), aktivierenden Lehrmethoden und anwendungsorientierten Lerninhalten geschehen. Zur Überprüfung dieser Hypothese sind detailliertere Befragungen zur Qualität der in der Maschinenbau-Lehre eingesetzten Lernsysteme nötig.

## 5 QUALITÄTSSICHERUNG DER LEHRE

---

<b>Aussage</b>	<b>Zutreffen</b>
Die Lehrinhalte werden verständlich vermittelt.	2,2
Der Aufbau und die Ziele der Lehrveranstaltung wurden klar vorgestellt.	2,1
Auf Fragen, Anregungen und Einwände von Studenten wird sorgfältig eingegangen.	2,0
Der Dozent ist im Umgang mit Studierenden freundlich und aufgeschlossen.	1,7
Der Dozent ist meist gut vorbereitet.	1,7
Die eingesetzten <i>Medien</i> (Folien, Tafelbild etc.) sind hilfreich.	2,2
Die Mitstudierenden haben ein großes <i>Interesse</i> an dieser Lehrveranstaltung.	2,8
Die Arbeitsatmosphäre unter den Studierenden ist kooperativ.	2,4
Das Thema der Lehrveranstaltung interessiert mich sehr.	2,4
Ich lerne in dieser Lehrveranstaltung viel.	2,5
Die Lehrveranstaltung ist <i>anregend</i> .	2,7
Die <i>Anforderungen</i> in der Lehrveranstaltung sind nicht zu hoch.	2,8
Im Allgemeinen bin ich mit der Lehrveranstaltung zufrieden.	2,3

---

Tab. 1: Lehrveranstaltungsevaluierung im WS2004/2005: FB15

Im nachfolgendem Teil II dieser Arbeit wird eine Qualitätssicherungsmethodik vorgestellt. Die Methodik basiert hauptsächlich auf Evaluationen, bei denen potenzielle Benutzer der Lernsysteme befragt werden. Die Befragung hilft, konkrete Ansätze zur Verbesserung von Lernsystem-Prototypen zu generieren.

## 6 Qualitätssicherung von Lernsystemen

Nachdem im Bildungswesen beträchtliche Geldsummen für Lernsystem- bzw. E-Learning-Projekte investiert wurden und letztlich sich der Erfolg gemessen am Aufwand bescheidener dargestellt hatte als ursprünglich geplant, verstärkt sich nun das Interesse der Allgemeinheit, die Qualität eines Lernsystems frühzeitig bestimmen zu wollen. Als Ausschreibungskriterium wurde beispielsweise vom BmBF (2005) im Rahmen des Förderprogramms „Neue Medien in der Bildung“ die Evaluation vorgeschrieben.

Es gibt aber auch Gegenstimmen zur Evaluation. In diesem Zusammenhang stehen die Argumente, dass systematische Qualitätsbewertung überflüssig sei, denn der Markt würde die Filterfunktion zwischen „guten“ und „schlechten“ Lernsystemen einnehmen. Ähnlich geschehe es schon seit Jahren bei Literatur, bei denen die Qualität auf den ersten Blick auch nicht ersichtlich ist und oft nur durch deren Nachfrage indiziert wird.

Lernsysteme sind allerdings keine Bücher, sondern es sind interaktive Medien. Der Umgang und die Bewertung von Literatur ist bei fast jedem Menschen durch jahrelange Erfahrung fest verankert. Dies trifft bei Lernsystemen überwiegend nicht zu.

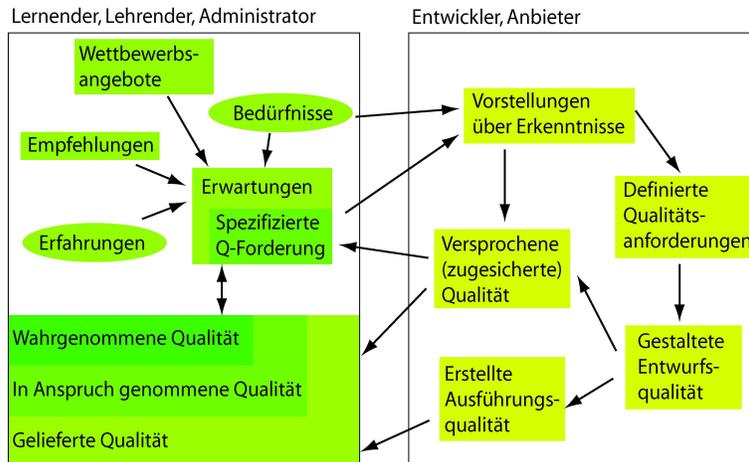
Bevor im zweiten Teil der Arbeit ein Weg zur Qualitätsbestimmung von Lernsystemen aufgezeigt wird, sollen in diesem Kapitel zunächst die Begriffe „Qualität“ und „Qualitätssicherung“ erörtert werden.

Der Begriff Qualität hat einen lateinischen Ursprung: *qualis* = wie beschaffen, *qualitas* = Beschaffenheit, Verhältnis, Eigenschaft. In dieser ursprünglichen Bedeutung schwingt mit dem Begriff Qualität entgegen dem allgemeinen Verständnis nicht etwas „Positives“ oder „Erstrebenswertes“ mit. Wenn wir Qualität neutral als Beschreibung der Eigenschaften und Merkmale eines Lernsystems sehen, bedeutet der Begriff „Qualitätssicherung“ die Wahrung der Eigenschaften und Merkmale. Aber um welche Eigenschaften und Merkmale geht es konkret, und wie können sie wahrgenommen werden? Das folgende Unterkapitel liefert hierzu eine Antwort.

### 6.1 Qualitätsbegriff

Die populäre Auffassung „Qualität ist das, was der Kunde wünscht“ fordert die Erfüllung der Kundenbedürfnisse durch Ausrichtung des Produkts auf die vom Benutzer eines Lernsystems benötigten Eigenschaften und Merkmale. Diese Kundenorientierung ist tatsächlich auch bei Lernsystemen eine wichtige und notwendige Denkrichtung für das Qualitätsverständnis. Kunden können dabei Lernende, Lehrende, Administratoren oder Entwickler sein. Lernende möchten durch die Verwendung des Lernsystems neues Wissen erlangen. Lehrende wollen über das Lernsystem Lehrmöglichkeiten nutzen, die ohne das Lernsystem nicht möglich wären. Administratoren brauchen gut dokumentierte und wartbare Lernsysteme. Entwickler

achten auf Erleichterungen bei der Kodierungsarbeit des Lernsystems.



Wallmüller stellt die Zusammenhänge der Qualitätsbeurteilung in Abb. 1 sehr differenziert dar (Wallmüller 1995, S.13). Die wahrgenommene Qualität ist in der Regel immer geringer als die in Anspruch genommene Qualität. Als effizient erscheint es, ein Lernsystem in der in Anspruch genommenen Qualität zu liefern – nicht mehr und nicht weniger.

Abb. 1: Qualitätsbeurteilung nach Wallmüller

In Anlehnung an die Darwin'sche These „Survival of the fittest“ ist das Lernsystem erfolgreich, welches eine gewisse intelligente Anpassungsfähigkeit an die gegebenen Anordnungen aufweist: „It is not the strongest species who survives or the most intelligent but the most responsive to change“ (Darwin 1859).

Die oft zitierte ISO-Norm 8402 erweitert den ursprünglichen Qualitätsbegriff durch den wichtigen Nebensatz, der die Relation der Eigenschaften und Merkmale eines Produkts auf deren Eignung zur Erfüllung von Anforderungen betont: „Qualität ist die Gesamtheit von Eigenschaften und Merkmalen eines Produkts oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen“ (Deutsches Institut für Normung 1992). Nach DIN EN ISO 9000:2000 wird Qualität als „Vermögen einer Gesamtheit inhärenter Merkmale eines Produkts, eines Systems oder eines Prozesses zur Erfüllung von Forderungen von Kunden oder anderen interessierten Parteien“ verstanden (Deutsches Institut für Normung 2000).

Aus diesen genannten Definitionen der Qualität lässt sich die folgende prägnante Gleichung für die subjektive Qualität eines Lernsystems ableiten:

$$Q_{\text{subjektiv}} = \frac{\sum_{i=0}^{n-1} Ae_i}{n} \quad (1)$$

$Q_{\text{subjektiv}}$  = Subjektive Qualität,  $Ae_i$  = Anforderungserfüllung für Anforderung mit Index  $i$ ,  $n$  = Anzahl der Anforderungen.

Mit Annahme der Gültigkeit von Gleichung (1) ist klar, dass die Qualität kein statisches, feststehendes Merkmal eines Produktes ist, sondern ein dynamisch veränderbares und nie sicher gegebenes Merkmal. So kommt es, dass gegenwärtig überzeugende Produkte mit der Zeit durch noch bessere ersetzt werden. Mit sich verändernden Anforderungen bei einem

Produkt wechselt zumeist auch dessen Qualität zur Erfüllung der Anforderungen. Aufgrund dieser Umfeldabhängigkeit wird die Qualität in Gleichung (1) als subjektiv gekennzeichnet. Subjektivität entsteht zudem durch das Qualitätsmessverfahren, bei dem die Anforderungserfüllung weitestgehend von Menschen bewertet werden muss. Qualität lässt sich nicht ausschließlich quantitativ messen, sondern (kann) nur kommunikativ durch Verständigung ermittelt werden (Lüthje 1997, S.7). Rössler hält Qualität daher für ein „violdimensionales Konstrukt, das je nach Beurteilungsgegenstand und je nach Perspektive des Beurteilenden unterschiedlich gemessen werden kann“ (Rössler 2004, S.128).

Es bedarf einiger Zwischenschritte bis die subjektive Qualität eines Lernsystems messbar wird (Kaufmanns 2002, S.81):

- Wenn etwas besser ist, dann ist es in einem relevanten Aspekt anders.
- Wenn etwas in einem relevanten Aspekt anders ist, dann ist dies wahrnehmbar.
- Wenn etwas wahrnehmbar ist, dann ist es zählbar.
- Wenn etwas zählbar ist, dann ist es auch messbar.
- Wenn etwas messbar ist, kann man auch den möglichen Nutzwert ableiten.

Der springende Punkt in dieser Aufzählung betrifft die Wahrnehmung der unterschiedlichen Merkmale und Eigenschaften eines Lernsystems. Nicht alle Menschen haben die gleiche Wahrnehmung.

Es wird also ein Medium benötigt, welches die Wahrnehmung des Menschen auf bestimmte Qualitätsaspekte lenkt. Da dies nur zum Teil gelingen kann, reicht es für eine ganzheitliche Bewertung des Lernsystems nicht aus, eine Person einzuplanen. Erst mit mehreren Personen gelingt es, dass viele Qualitätsaspekte eines Lernsystems wahrgenommen und bewertet werden. Abb. 2 auf der nächsten Seite verdeutlicht dies durch ein erklärendes Modell.

These 1: Die objektive Qualität eines Lernsystems lässt sich nicht direkt messen. Sie kann nur über die subjektive Qualität nach Gleichung (1) auf der vorherigen Seite als arithmetischer Mittelwert von Anforderungserfüllungen annähernd bestimmt werden.

Objektivierte Qualität ist nicht mit der objektiven Qualität gleichzusetzen, sondern sie ist als eine Näherung zur objektiven Qualität zu verstehen. Eine angemessene Vielzahl von Qualitätsbewertungen, die sich teilweise auch ähneln können (siehe überlappende Ellipsen in der Abb. 2 auf der nächsten Seite), führen zu einer Objektivierung von Qualität. Die sternartige Fläche stellt die objektive Qualität dar. Diese wird mit zunehmender Anzahl von Bewertungen abgedeckt. Eine annähernde Abdeckung der bildlichen Sternfläche durch die in der Darstellung kenntlich gemachten Bewertungsellipsen, gilt als Ziel jeder wissenschaftlichen Qualitätsmessung. Die Bewertungsellipsen unterscheiden sich in Größe und Form, da die Bewerter quantitativ und qualitativ individuelle zum Ausdruck gebrachte Wahrnehmungen zur Qualität des Lernsystems haben.

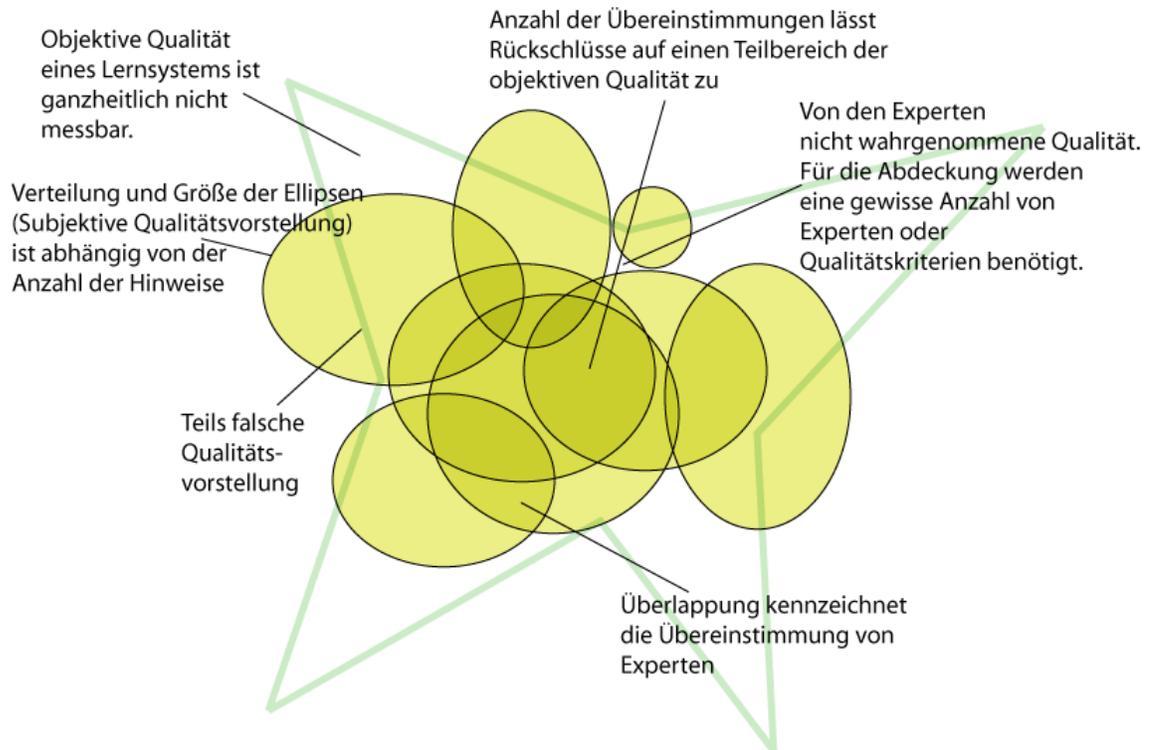


Abb. 2: Subjektive und objektive Qualität

Es drängt sich die Frage auf, wieviel Bewerter bzw. Experten nötig sind, um die subjektive Qualität zu objektivieren. Im vierten Teil der Arbeit wird in diesem Zusammenhang eine Möglichkeit aufgezeigt, die für ein gutes Aufwand-Nutzen-Verhältnis der Evaluation optimale Anzahl von Experten anhand von bestimmten Einflussgrößen abzuschätzen. Zudem wird dargelegt, wie durch die Auswertung von subjektiven Qualitätshinweisen und -bewertungen von Experten auf die Anforderungserfüllung und damit auf die objektivierbare Qualität eines Lernsystems geschlossen werden kann.

Nach der Gleichung (1) auf Seite 7 stellt die Bestimmung der Anforderungserfüllung eine wichtige Grundlage zur Qualitätsbewertung dar. Diese basiert letztlich immer auf der Bewertung eines Soll- und Istvergleichs von Eigenschaften und Merkmalen des Lernsystems.

Für die Aufstellung von Anforderungen bietet sich das weit verbreitete Verfahren Quality Function Deployment (QFD) an. Die Methode wurde von Akao (1990, S.1ff.) in Japan im Jahr 1966 eingeführt. Das sogenannte „House of Quality“ (HoQ) ist eine Methode innerhalb des Quality Function Deployment. Durch das HoQ können systematisch die geplanten Lernsystem-Eigenschaften und -merkmale begründet und Anforderungen validiert werden.

Das QFD-Verfahren gliedert sich in die folgenden Schritte:

- Erfassen der Kundenanforderungen,
- Gewichten der Kundenanforderungen,
- Wettbewerbsanalyse zur Erfüllung der Kundenanforderungen,
- Ableiten technischer Merkmale,
- Interdependenzen der technischen Merkmale,
- House of Quality HoQ (Beziehungsmatrix),
- Qualifizieren der technischen Spezifikationen,
- Wettbewerbsanalyse von Ausprägungen der technischen Merkmale aus Herstellersicht,
- Bewertung technischer Merkmale zur Erfüllung der Kundenwünsche.

Erst nachdem die Anforderungen an das Lernsystem klar definiert worden sind, kann die Qualität eines verfügbaren Prototyps durch Wahrnehmungs- und Bewertungsprozesse gemessen werden. Die Initiierung solcher Prozesse ist die Aufgabe der Qualitätssicherung von Lernsystemen. Für den Bereich Qualitätssicherung können je nach Fokus folgende Verfahrensklassen unterschieden werden:

- Prozessverbesserung,
- Vergleich bzw. Benchmarking,
- Ergebniskontrolle.

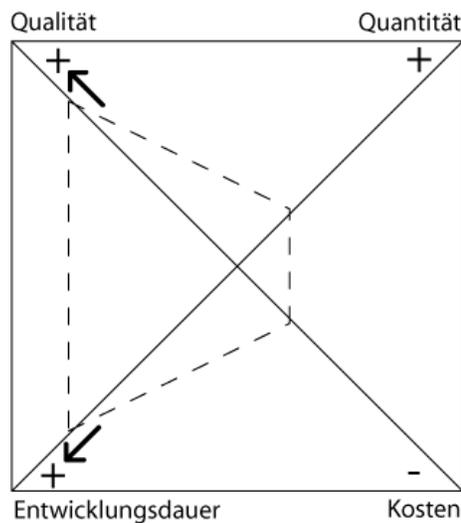
Die vorliegende Arbeit konzentriert sich schwerpunktmäßig auf die zuletzt genannten Verfahrensklassen: Vergleich, Benchmarking und Ergebniskontrolle von Lernsystem-Prototypen.

**Prozessverbesserung** Zu dieser Gruppe können diejenigen Qualitätssicherungsmethoden gerechnet werden, die darauf abzielen, die Prozesse zur Entstehung des Lernsystems zu verbessern. „Ein sinnvolles Systemmanagement wird daher nicht versuchen, einen Schaden nach dem anderen dort, wo er gerade auftaucht zu reparieren und damit den Ereignissen bildlich hinterher zuhinken, sondern durch eine systemorientierte Planung und Steuerung die Weichen für eine andere Konstellation des Systems zu stellen, in der solche Schäden weniger Chancen haben aufzutreten“ (Vester 1999, S.62). Hier wird also nicht direkt das Lernsystem verbessert, sondern ausschließlich die Arbeitsabläufe zum fertigen Produkt. Indirekt soll eine Verbesserung des Lernsystems erzielt werden. Dieses Vorgehen ist insbesondere dann nachvollziehbar, wenn Qualität in Abhängigkeit von personellen Ressourcen, Entwicklungszeit und -kosten definiert wird.

Prozessverbesserung wird in der Wirtschaft oft und in Universitäten vereinzelt von einer Zertifizierungseinrichtung begleitet. Die Zertifizierungen nach EFQM (European Foundation for Quality Management), DIN EN ISO 9000ff. (International Standardization Organization), BAOL (British Association for Open Learning) etc. sind weniger für kleine und mittlere Projekte an Universitäten geeignet. Zertifizierungen sind vor allem für durch Arbeitsteilung geprägte Produktionsprozesse interessant, bei denen beispielsweise das Qualitätsbewusstsein und die Qualifikation der Beteiligten sowie

die Transparenz der Arbeitsschritte gefördert werden muss. Es werden lediglich in einem Qualitätsmanagementhandbuch festgelegte Prozesse wie Planung, Bereitstellung und Durchführung eines Lernsystems auf deren Einhaltung geprüft und nicht ob sich das Produkt dieser Prozesse tatsächlich bewährt. Zudem hat das Zertifikat nur durch Kenntnis des Qualitätsmanagementhandbuchs eine Aussagefähigkeit.

Das Ziel ist eine Qualitätssicherung durch Verfahrenssicherung. Häufig wird dieses Ziel zu dem Tätigkeitsfeld von Qualitätsmanagement gerechnet, wie bereits an dem Begriff Qualitätsmanagementhandbuch deutlich wird. Die Integration der in Teil II der Arbeit beschriebenen speziell für die Entwicklung von Lernsystemen abgestimmten Qualitätssicherungsmethodik kann vom Tätigkeitsfeld des Qualitätsmanagements nicht losgelöst behandelt werden. Die eher prozess- und projektrelevanten Aspekte des Qualitätsmanagements werden hier nicht künstlich von der Begrifflichkeit Qualitätssicherung getrennt, da der Übergang von Qualitätssicherung und Qualitätsmanagement fließend verläuft. Dies wird an Themen wie die Berücksichtigung der Kosten, der Dauer und des Zeitpunkts der Qualitätssicherung deutlich.



Sneed verdeutlicht den Verbesserungsversuch von Qualität, Quantität, Kosten und Zeit mit dem Teufelsquadrat. Die Verbesserung der einen Kenngröße führt in der Regel zur Verschlechterung der anderen Kenngrößen (Sneed 1987). Die Einstellung der Qualität ist ein Kompromiss aus den übrigen Kenngrößen und damit Aufgabe einer Prozessverbesserung.

Abb. 3: Teufelsquadrat nach Sneed

Die Einhaltung der standardisierten Prozesse ist allerdings kein Garant für die Qualität der Produkte. Der Verfahrenssicherung wird häufig zugeschrieben, dass der Dokumentationsaufwand zu hoch sei. Daneben entstünden mit der Zertifizierung Kosten, die mit der nicht kommerziellen universitären Entwicklungspraxis unvereinbar sei. Mit Spannung wird daher der nächste noch unvollendete Schritt der Arbeitsgruppe „Aus- und Weiterbildung unter besonderer Berücksichtigung von E-Learning“ erwartet (Deutsches Institut für Normung 2004). Es wird sich zeigen, ob der geplante Qualitätsmanagement-Ansatz sich für kleine und mittlere universitäre Lernsystem-Projekte eignet.

**Vergleich bzw. Benchmarking** Qualitätssicherungsverfahren können auch zum Vergleich von Lernsystemen eingesetzt werden. Der Markt für Lernsysteme gestaltet sich oft sehr intransparent und die Einsatzwilligen von Lernsystemen benötigen Rankings, um leichter eine überlegte Auswahl treffen zu können. Allerdings sind solche vergleichenden Evaluationen nicht nur als Verbraucherschutz und -ratgeber sinnvoll, sondern auch intern für das Prototyping. Im Rahmen eines kontinuierlichen Verbesserungsprozesses kann überprüft werden, ob ein Lernsystem-Prototyp gegenüber dem Vorgängerprototyp tatsächlich besser abschneidet. Das Bestreben sollte bei einer Prototyp-Entwicklung darin liegen, möglichst frühzeitig den Grad der Anforderungserfüllung festzustellen, um rechtzeitig Verbesserungsmaßnahmen einzuleiten. Dazu wird nicht nur das Produkt am Ende des gesamten Entwicklungszyklusses getestet, sondern die Prototypen, die fortlaufend während der Entwicklung entstehen. „Eine Möglichkeit der Reaktion auf vorhandene Qualitätsdefizite besteht nicht in der – oft versuchten – Nachbesserung der Programme, sondern in der konsequenten Qualitätssicherung während aller Produktionsphasen der Softwareherstellung“ (Gräber 1990, S.4).

Mit dem Begriff „kontinuierliche Verbesserung“ verbindet man die Ansichten von Deming (1986), die das Qualitätsmanagement weltweit stark geprägt haben. Qualität kann nach Demings Meinung nicht erpüft, sondern sie muss erzeugt werden. Dies bedeutet, dass mit kontinuierlicher Verbesserung nicht allein das ständige Prüfen der Qualität wichtig ist, sondern ergänzend eine nachhaltige und ganzheitliche Verbesserung erfolgt. Dies erfordere eine Unternehmensphilosophie, die nach den 14-Management-Punkten<sup>3</sup> organisiert ist (Deming 1982).

Das Konzept „Kaizen“<sup>4</sup> wird häufig als Hauptgrund für die dramatischen Qualitätssteigerungen in der japanischen Wirtschaft gesehen. Die Botschaft von Kaizen beinhaltet, dass kein Tag ohne irgendeine Verbesserung im Unternehmen vergehen soll. Kaizen ist somit die Philosophie der ewigen Veränderung und der Flexibilität, um auf die Veränderungen des Umfeldes zu reagieren.

Das Vorgehen wurde vom Fachgebiet Systemtechnik und Mensch-Maschine-Systeme an der Universität Kassel im Projekt DIAMANTA 1997 für Software-Entwicklungsprozesse umgesetzt (siehe Abb. 4 auf der nächsten Seite). DIAMANTA war ein Projekt mit drei industriellen Partnern, das im Rahmen des ESPRIT-Programms der Europäischen Gemeinschaft unterstützt wurde (Johannsen *et al.* 1997).

Die periodischen (formativen) Evaluationen des DIAMANTA-Projektlebenszyklus dienen der kontinuierlichen Dokumentation der Projektaktivitäten und des Projektfortschritts, sowie der Überprüfung der Änderung von Zielen und Situationen. Zusätzlich

---

<sup>3</sup>1) Nachhaltige Geschäftspolitik, 2) Neue Denkweise, 3) Lückenlose Kontrolle, 4) Beschaffung auf Grund des Preises, 5) Andauernde Verbesserung des Systems, 6) Training on the Job, 7) Motivierende Führung, 8) Furchtfreies Arbeitsklima, 9) Interne Schranken, 10) Schlagwörter, 11) Quoten und Leistungsziele, 12) Erfolgsergebnisse, 13) Mitarbeiterförderung, 14) Aufbruch zu neuen Horizonten

<sup>4</sup>Kai = Veränderung, Wandel; Zen = zum Besseren

wird beim Abschluss einer Entwicklungsphase eine Evaluation des entwickelten Prototyps initiiert. Diese periodischen Evaluationen dienen der Überprüfung der Anforderungen. Durch die finale Evaluation wird die subjektive Zufriedenheit mit dem Produkt zum Ende des Entwicklungsprozesses erhoben.

Obwohl die formative Evaluation in vielen Entwicklungszyklus-Modellen theoretisch verankert ist, findet sie bei kleinen und mittleren Projekten kaum Anwendung. Dies liegt größtenteils daran, dass nur ein geringer Anteil der bekannten Evaluationsverfahren kosten- und zeitgünstig durchgeführt und ausgewertet werden kann.

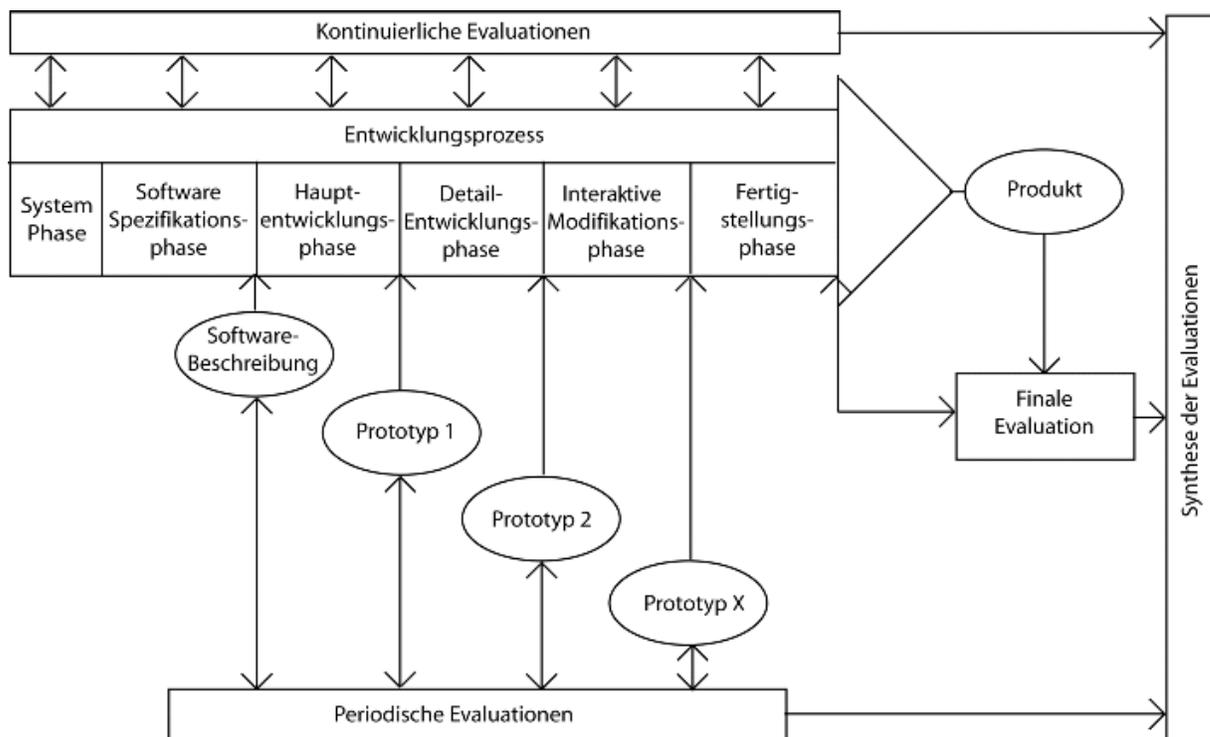


Abb. 4: Evaluationschritte des DIAMANTA Projektlebenszyklus (Tiemann *et al.* 1996)

**Ergebniskontrolle** Die Ergebniskontrolle verfolgt das Ziel der Verbesserung der Produktqualität durch Evaluation des Lernsystems. Bei der Evaluation werden bestimmte Merkmale und Eigenschaften eines Lernsystems analysiert. Gängige Evaluationsmethoden stellen die qualitätskriterienbasierte Befragung und die Beobachtung von Zielgruppen bei der Nutzung des Lernsystems dar.

In Bezug auf die Messung des Lernerfolgs und die Transferfähigkeit des Lernenden wird häufig ein anderer Weg beschritten. Die Lernenden bekommen vor der Arbeit mit den Lernsystemen Aufgaben präsentiert, die sie aufgrund ihres Vorwissens selbstständig lösen sollen. Nach dem Studium des Lernsystems werden weitere Aufgaben gestellt. Die Lösungen und Herangehensweisen zu den zu unterschiedlichen Zeitpunkten gestellten Aufgaben der Prä- und Posttests werden dann miteinander verglichen

und gegenüber gestellt. Derartige Untersuchungen sind sehr fehleranfällig bezüglich des Versuchsdesigns. Auftretende Fehler sind hier besonders kritisch, denn sie führen dazu, dass die Versuchsergebnisse (z.B. mit anderen Personen) nicht wiederholbar sind.

Eine weitere Methode ist die Protokollierung des Nutzungsverhaltens der Anwender des Lernsystems. Nach der Analyse der Nutzungsdaten erfolgt ein Rückschluss auf den Aufwand und den Nutzen des Lernsystems. Darüber hinaus kann nach Mandl (2001, S.2) die Qualitätsbewertung von Inhalten teilweise automatisiert werden: „Experimentelle Systeme behandeln zahlreiche Aspekte und nutzen unterschiedliche Definitionen von Qualität“.

## 6.2 Lernsystem-Evaluation

Jede Lernsystem-Evaluation erfordert zunächst Ressourceneinsatz und ein zusätzliches Engagement der Projektmitglieder, ohne bliebe die Qualität des Lernsystems unklar und der Kunde würde unfreiwillig zur Versuchsperson. Qualitätssicherung von Lernsystemen ist ohne eine Evaluation nicht möglich (Schott 2000, S.107).

Evaluationen schaffen Orientierung und Argumentationssicherheit. Sie geben Impulse, um Handlungen zur Verbesserung der Qualität abzuleiten. Für Stufflebeam (1972, S.124) bedeutet Evaluation die Gewinnung von Informationen durch formale Mittel wie Kriterien, Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällen von Urteilen in Entscheidungssituationen zu erhalten. Nach Will *et al.* (1987) hat die Evaluation eine

- Steuerungs- und Verbesserungsfunktion;
- Bewertungs- und Beurteilungsfunktion;
- Kontroll- und Disziplinierungsfunktion.

Generell können nach Stockmann (2004, S.25) mit Evaluationen folgende Ziele angestrebt werden:

- Gewinnung von Erkenntnissen,
- Kontrolle,
- Schaffung von Transparenz, um einen Dialog zu ermöglichen,
- Dokumentation des Erfolgs (Legitimation).

Die Zielerreichung ist im Wesentlichen abhängig von

- dem Untersuchungsgegenstand (Lernsystem),
- den Evaluatoren,
- dem Budget,
- der zeitlichen Restriktion und
- der/den Evaluationsmethode(n).

## 6 QUALITÄTSSICHERUNG VON LERNSYSTEMEN

Die Tab. 2 führt die zur Evaluationsebene (siehe Abb. 5) und die zur Lebensphase (siehe Kap. 7 auf Seite 24) des Lernsystems passenden Evaluationsmethoden auf.

Evaluationsebene	Lebensphase	Evaluationsmethode
1. Eigenschaften und Merkmale 2. Reaktion der Benutzer	} Alle	Inhaltsanalyse, qualitätskriterienbasierte mündliche und schriftliche Benutzer-/Expertenbefragung, Beobachtung
3. Lernerfolg		
4. Transferebene	Lernsystemnutzung	
5. Effizienzebene	Lernsystemnutzung	Betriebsdaten, Log-File- Analyse
6. Kosten-Nutzen-Relation	Vorentwicklung, Lernsystem- nutzung	Kosten-Nutzen-Kalkulation

Tab. 2: Evaluationsmethoden

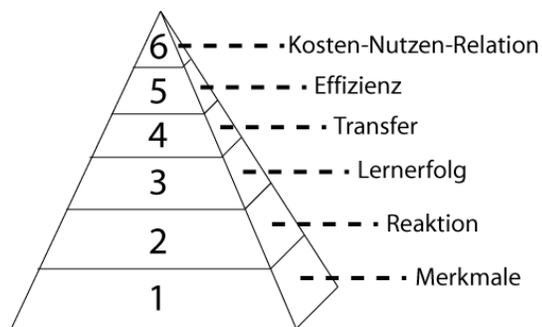


Abb. 5: Evaluationsebenen Evaluationsnetz (<http://www.evaluationsnetz.com/>)

Bei der Auflistung in Tab. 2 fällt auf, dass die zu den ersten beiden Evaluationsebenen zugeordneten Evaluationsverfahren ein weites Lebensphasen-Einsatzspektrum aufweisen. Selbst wenn nur Prototypen bereit stehen, können die Verfahren angewendet werden. Nicht zuletzt wegen der vergleichsweise einfachen Durchführbarkeit eignen sie sich deshalb für den formativen Entwicklungskontext. Aufgrund dieser Bedeutung und der Anzahl der Qualitätsaspekte, die untersucht werden können, bilden die Merkmalsebene und die Reaktionsebene in Abb. 5 daher die Basis des Pyramidenmodells.

Für die Abdeckung anderer Evaluationsebenen sind hingegen die weiteren genannten Evaluationsmethoden zu empfehlen. Die Evaluationsebenen Lernerfolg, Transfer, Effizienz, Kosten-

Nutzen-Relation zielen eher auf einen summativen Evaluationskontext, bei dem ein vollständiges Produkt bereit steht (siehe UKap. 7.7 auf Seite 32).

Auf jeder Evaluationsebene wird auf bestimmte Qualitätsaspekte geachtet. Bei der in Teil IV auf Seite 77 durchgeführten Studie zur Expertenbefragung kommen Qualitätskriterien zum Einsatz, die bestimmte Qualitätsaspekte für den Bewerter aufbereitet beschreiben und anhand derer eine zielgerichtete Bewertung stattfindet. Im nachfolgenden Unterkapitel wird erläutert, was unter Qualitätskriterien zu verstehen ist.

### 6.3 Qualitätskriterien

Die Unendlichkeit des Lösungsraums zur Gestaltung eines Lernsystems wird begrenzt durch die Notwendigkeit, bestimmte Anforderungen zu erfüllen. Zur Überprüfung der Anforderungserfüllung werden Qualitätskriterien formuliert. Qualitätskriterien kommunizieren das zugrunde liegende Qualitätsmodell und sorgen für eine klare Grenzziehung zwischen qualitätsrelevanten und -irrelevanten Aspekten. „Eine gute Möglichkeit, die Brauchbarkeit von Lernsoftware zu untersuchen, ist die Erarbeitung von Kriterien über die Güte von Lernprogrammen“ (Thomé 1989, S.47).

In der Literatur werden die Möglichkeiten von Qualitätskriterien häufig überschätzt: „Qualität im Bereich der Weiterbildung ist ein Konstrukt, welches bislang nicht eindeutig definiert und in Kriterien aufgeschlüsselt ist ... Dies gilt sowohl für klassische Formen der Weiterbildung als auch für onlinegestützte Formen der Weiterbildung“ (Ehlers 2004, S.51). Derartige allgemeingültige Qualitätskriterien sind Wunschenken, denn unter Verwendung der bereits diskutierten Gleichung (1) auf Seite 7 misst sich Qualität an der Erfüllung der projektspezifischen Anforderungen. Ein bewährtes Merkmal eines Lernsystems muss in einem anderen Anforderungskontext nicht unbedingt die gleiche Wirkung erzielen. „Es können keine allgemeinen Aussagen über E-Learning gemacht werden, da E-Learning-Angebote sich gravierend in Zielen, Szenarien, Lernumgebungen und Lernobjekten unterscheiden“ (Schulmeister 2005, S.477). „Es lassen sich keine allgemeinen Urteile über die Qualität von E-Learning-Umgebungen (Design, Usability) fällen, da die Urteile abhängig sind von unterschiedlichen Einstellungen und Fachkulturen der Studierenden und der Lehrenden“ (Schulmeister 2005, S.480).

Qualitätskriterien sind daher nicht als Checklisten zu verstehen, die eine bestimmte Gestaltung von Lernsystemen vorschreiben. Gestaltungsvorschriften sollten vielmehr in Anforderungen formuliert werden. Durch dieses Verständnis unterscheidet sich der in der Studie des Teils IV auf Seite 77 verwendete Qualitätskriterienkatalog von den meisten im nachfolgenden Unterkapitel aufgelisteten Qualitätskriterienkatalogen.

In dieser Arbeit werden Qualitätskriterien als multipler Fokus einer zielgerichteten Befragung und differenzierten Bewertung verstanden. Bei einer ganzheitlichen Bewertung von Lernsystemen ist es wichtig, multiple Qualitätskriterien zu definieren, die helfen, die Komplexität

der Bewertung durch die Abfrage von abgegrenzten Teilaspekten der Qualität zu reduzieren. Dabei wird die Wahrnehmung von Bewertern auf die relevanten Qualitätsaspekte eines Lernsystems mit dem Ziel gelenkt, umfassende und verdichtete Informationen aus der Sicht der Bewerter zu erhalten. Anders ausgedrückt tragen die mit den Qualitätskriterien verbundenen Fragestellungen dazu bei, besser nachvollziehbare und einordbare Bewertungshinweise von den antwortenden Personen zu erheben.

### 6.4 Qualitätskriterienkataloge

Eine Sammlung über nationale und internationale Qualitätskriterienkataloge und eine vergleichende Diskussion dieser finden sich auf dem deutschen Bildungsserver,<sup>5</sup> bei Gräber (1996, S.26ff.) und bei Jelitto (2005). Vielen der dort genannten Qualitätskriterienkatalogen ist gemein, dass die Lernsysteme bezüglich der Erfüllung bzw. Nicht-Erfüllung von bestimmten Kriterien geprüft werden.

Die bekanntesten Qualitätskriterienkataloge sind nachfolgend beschrieben:

- **MEDA'97:** Meda wurde mit finanzieller Unterstützung der Kommission der Europäischen Gemeinschaft durch ein internationales Wissenschaftlerteam geschaffen. MEDA ist ein hierarchisch strukturiertes Instrument für die Beobachtung, Analyse und Bewertung von Lernsystemen. Den Arbeitsbereichen Entwicklung, Anwendung und Vertrieb sind sogenannte Intentionen zugeordnet, die Kriterienvorschläge beinhalten.
- **AKAB:** Der Arbeitskreis der deutschen Automobilbranche (AKAB) hat einen gleichnamigen Kriterienkatalog entworfen, um computerbasierte Trainingsprogramme (CBT) mit Hilfe festgelegter Qualitätskriterien zu beurteilen. Der gesamte Katalog ist bei Kubicek *et al.* (1998, S.236ff.) abgedruckt.
- **SODIS:** Das Software Dokumentations- und Informationssystem<sup>6</sup> (SODIS) ist ein gemeinsames Angebot der deutschen Länder und Österreichs. Bewertungen und Erfahrungsberichte aus allen kooperierenden Ländern werden in der SODIS-Datenbank erfasst. In Hessen gibt es den Hessischen Kompass für Bildungssoftware. Die Bewertungen werden von Lehrern anhand eines Leitfadens vorgenommen. Es wird nicht nur das Produkt Lernsystem, sondern auch die Verwendung im Unterricht bewertet.<sup>7</sup>
- **ELISE:** Das theorie-, adressaten- und anwenderorientierte ELISE-Verfahren steht für eine „Effiziente Lern- und InformationsSystem Evaluation“ und beansprucht unter Einbeziehung der Universal Constructive Instructional Theory (UCIT)<sup>8</sup> eine ganzheitliche,

---

<sup>5</sup><http://www.bildungsserver.de/zeigen.html?seite=1564>

<sup>6</sup><http://www.sodis.de>

<sup>7</sup>[http://medien.bildung.hessen.de/kompass/Hinweise/Leitfaden\\_fur\\_Evaluatoren.pdf](http://medien.bildung.hessen.de/kompass/Hinweise/Leitfaden_fur_Evaluatoren.pdf)

<sup>8</sup>[http://www.cmr.fu-berlin.de/lehre/ws02/lssing/Lernen\\_Multimedia/Powerpoint-Pr%20entationen/Thesenpapier031202.pdf](http://www.cmr.fu-berlin.de/lehre/ws02/lssing/Lernen_Multimedia/Powerpoint-Pr%20entationen/Thesenpapier031202.pdf)

systemische und präskriptive Sichtweise unter Berücksichtigung psychischer Prozesse und Kosten-Nutzen-Relationen.

- **Evit@:** Evit@ stellt ein Bewertungsinstrument zur vergleichenden Produktanalyse und zur Qualitätssicherung dar, mit dessen Hilfe qualitative Aussagen über elektronische Informationsmittel möglich werden sollen.<sup>9</sup>
- **Basic Clear:** Im Projekt clear2b wurde ein Konzept zur Qualitätsanalyse von Lernsystemen entwickelt. Ein wesentlicher Aspekt hierbei ist, dass die Qualitätsanalyse eines Lernsystems immer in Bezug zu dem konkreten Lernumfeld steht. Das bedeutet, dass ein Lernsystem hinsichtlich der jeweiligen Lernziele, Zielgruppen und des Lernumfeldes analysiert und bewertet wird. Es wurde ein Kriterienpool auf der Basis bestehender Kriterienkataloge (z.B. AKAB, MEDA) zusammengestellt.<sup>10</sup>
- **Evaluationsnetz:** Das Evaluationsnetz (2005) ist ein webbasiertes Informations- und Beratungsinstrument für E-Learning-Angebote. Das Evaluationsnetz ist damit ein wachsendes System, das in seiner Grundstruktur von Betrieben übernommen und dann betriebspezifisch angepasst werden kann.
- **QuIT-L:** Zielsetzung des Projektes QuIT-L (Qualitätskriterien für IT-basierte Lernsysteme) ist es, inhaltsunabhängige Qualitätskriterien für E-Learning-Produkte zusammenzustellen und im Hinblick auf ihre Aussagekraft zu evaluieren.<sup>11</sup>
- **dmmv:** Im Sommer 2003 erfolgte die Definition von Qualitätskriterien für E-Learning-Angebote durch den Arbeitskreis Aus- und Weiterbildung des Deutschen Multimedia Verbandes e.V. (dmmv).<sup>12</sup>
- **Q.E.D.:** Die Qualitätsinitiative E-Learning in Deutschland (Q.E.D.) entwickelt neue Standards und erschließt neue Märkte und Geschäftsmodelle. Für die innovativen Anwendungsfelder Mobile Learning und Rich Media Content werden Referenzmodelle aufgebaut und entsprechende Werkzeuge für die sofortige Umsetzung und Nutzung erstellt.<sup>13</sup>
- **Criteria for Evaluating Internet Resources:** Auch im Ausland werden Qualitätskriterien formuliert. Hierfür steht exemplarisch der Katalog zur Evaluation von Internet Ressourcen der Universität von British Columbia.<sup>14</sup>

Die qualitätskriterienbasierte Evaluation hat Schwächen, aber auch Vorzüge. Die Schwächen können durch Gestaltungsmöglichkeiten (siehe UKap. 8.1 auf Seite 35.) abgeschwächt werden, sodass sich das Verfahren als unverzichtbarer Bestandteil einer formativen Quali-

---

<sup>9</sup><http://www.fbi.fh-koeln.de/institut/projekte/evit@/evit@001.htm>

<sup>10</sup><http://www.lernet-info.de>

<sup>11</sup><http://www.uni-oldenburg.de/sport/bww/quitl/Texte/kritforsch.html>

<sup>12</sup>[http://www.bvdw.org/shared/data/doc/a\\_dmmv\\_9\\_qualitaetskriterien\\_elearning\\_040512.doc](http://www.bvdw.org/shared/data/doc/a_dmmv_9_qualitaetskriterien_elearning_040512.doc)

<sup>13</sup><http://www.qed-info.de>

<sup>14</sup><http://www.library.ubc.ca/home/evaluating/>

tätssicherung darbietet. Die nachfolgenden zwei Unterkapitel diskutieren die Schwächen und Vorzüge der qualitätskriterienbasierten Evaluation.

### 6.5 Schwächen von Qualitätskriterien

Die Anwendung eines Qualitätskriterienkataloges ist ein beliebtes, kostengünstiges, zeitsparendes, einfaches und methodisch korrektes Verfahren zur Bewertung von Lernsystemen. Doch wie jedes Verfahren ist es nur bedingt reliabel. Fricke (2000, S.75ff.) und Tergan (2001, S.319ff.) benennen einige Schwächen der kriterienbasierten Evaluation. Diese lassen sich jedoch aufgrund der in UKap. 6.2 auf Seite 14 aufgezeigten Abhängigkeiten einer Evaluation nicht verallgemeinern. Wenn „die ausgewählten Kriterien nicht den Inhalten angemessen und die Vorgehensweisen nicht den Situationen adäquat sind, braucht es nicht zu verwundern, dass viele Experten von den Resultaten enttäuscht sind“ (Zimmer und Psaralidis 2000, S.263). Folgende Problembereiche gilt es bei jedem Kriterienkatalog individuell zu überprüfen:

**Transparenz:** Qualitätskriterien beinhalten häufig wohl formulierte Qualitätsaspekte. Verborgenen bleiben jedoch die Hintergründe, die lerntheoretischen Annahmen oder die empirischen Untersuchungen, die zu der Aufstellung der Qualitätskriterien führten (Fricke 1995, S.409f.). Bei vielen Qualitätskriterienkatalogen ist weder die Vorgehensweise, noch die Auswahl und die Gewichtung der Qualitätskriterien offensichtlich. Ohne diese Informationen ist eine Auswertung schwer nachzuvollziehen und das Ergebnis beliebig manipulierbar. Durch die scheinbare Seriosität der Anbieter solcher Qualitätskriterien werden die Evaluatoren zu leicht von der Verantwortung entbunden, sich kritisch mit den Hintergründen der Qualitätskriterien auseinander zusetzen. Qualitätskriterienkataloge erscheinen in dieser Hinsicht oft als eine ad hoc zusammengetragene Ansammlung von Einzelaspekten.

Ein weiterer Kritikpunkt betrifft die Interrelation zwischen den Qualitätskriterien. Die durch Qualitätskriterien beschriebenen Qualitätsaspekte überschneiden sich vielfach inhaltlich mit anderen Qualitätskriterien. „The fact that all components of an instructional situation interact with one another is a central problem for evaluating educational software. The existing interrelations between the components of learning systems, the fact, that the usefulness and effectivity of a particular system type may be dependent on the context of use and the particular educational goals make it impossible to follow a direct checking approach in software evaluation on the basis of a simple software evaluation criteria list“ (Tergan 1996, S.47).

Qualitätskriterien müssen von den Bewertern einheitlich verstanden werden. Fehlen entsprechende Erläuterungen zu den Qualitätskriterien, können Gutachter die Kriterien im eigenen Ermessen interpretieren. Dies führt letztlich ungewollt zu einer Reduktion der Bewerterübereinstimmung.

**Ganzheitlichkeit:** Der Detaillierungsgrad von Qualitätskriterienkatalogen beschränkt sich zumeist auf Kriterien zur Bildschirm- und Dialoggestaltung. Selten finden sich auch pädagogisch-didaktische Qualitätskriterien oder welche zum Lernumfeld. Formative Qualitätssicherung von Lernsystemen kann jedoch nur mit ganzheitlichem Denken beherrscht werden. „Eine ganzheitliche Sicht ist notwendig, weil die Güte einer Bildungsmaßnahme, etwa die Effizienz der Zielerreichung, ... von dem Gesamtzusammenhang der sie beeinflussenden Faktoren abhängt. Solche Faktoren sind unterschiedliche Eigenschaften der Lernenden, des Lernstoffes, der Lehrmethoden und Medien sowie des Umfeldes, in dem die Bildungsmaßnahme stattfindet“ (Schott 2000, S.107). Ein Qualitätskriterienkatalog sollte die wesentlichen Qualitätsaspekte eines Lernsystems behandeln. Viele Qualitätskriterienkataloge versuchen daher, durch eine Vielzahl von Qualitätskriterien oder weitreichende Fragestellungen möglichst viele Qualitätsaspekte abzudecken. Jedoch befindet sich ein Qualitätskriterienkatalog in dem Dilemma, dass er nie den Anspruch auf Vollständigkeit erfüllen kann, um noch mit möglichst wenigen und eindeutig abgegrenzten Kriterien praktikabel zu sein. Die wenigstens Qualitätskriterienkataloge sind als offenes und skalierbares System angelegt, welches zu gegebener Zeit mit neuen Erkenntnissen zu aktualisieren ist.

**Empirie:** Qualitätskriterien sind in der Regel wenig empirisch validiert. Die Erforschung der Neuen Medien befindet sich noch am Anfang. Erschwerend kommt hinzu, dass sich die Lerntechnologien ständig weiter entwickeln und empirische Evaluation Zeit braucht. Daher hinken die Forschungsergebnisse immer den aktuellen Entwicklungen hinterher. Zudem sind viele Studienergebnisse nur bedingt auf ein anderes als das im Versuchsdesign untersuchte Lernsystem oder Lernumfeld zu übertragen.

Dies ist jedoch kein Grund zu resignieren, wie die Untersuchungen von Gigerenzer (1997, S.55) vom Max-Planck-Institut für Psychologische Forschung zeigen. Er stellt fest, dass sich mit einfachen Heuristiken schnell und gut auf Unbekanntes schließen lässt. Dort wo Studien fehlen, können Heuristiken (plausible Vermutungen, Schlüsse oder Erfahrungswerte) Übergangslösungen bilden. Zum Beispiel nennt Nielsen (1993, S.155ff.) folgende typische Heuristiken: Erkennbarkeit des Systemstatus, Passung des Systems mit der Realität, Benutzerkontrolle und -freiheit, Konsistenz und Standards, Fehlervermeidung, Wiedererkennung anstatt Erinnerung, Flexibilität und Effizienz beim Gebrauch, Ästhetik und Minimalistisches Design, Benutzer helfen, Fehlerdiagnose und -beseitigung, Hilfe und Dokumentation.

**Situation:** Wie bei jedem Werkzeug kommt es bei einem Qualitätskriterienkatalog auf die situationsgerechte Anwendung an. Bei einigen Qualitätskriterienkatalogen wird die situative Dimension bzw. der Verwertungszusammenhang vernachlässigt.

Aufgrund dieser Schwächen zu schließen, dass die kriterienbasierte Evaluation kein Mittel zur Evaluation sei, wäre falsch. „Perfektion ist nicht erreichbar, aber prozentuale Verbesserungen sind möglich und es ist wert, danach zu streben“ (Shneiderman 2002, S.185). Viele der

genannten Schwächen können durch die Gestaltung der Fragebögen (siehe UKap. 8.1 auf Seite 35) oder durch das Hinzuziehen von weiteren Evaluationsmethoden beseitigt werden (siehe Kap. 7 auf Seite 24). Natürlich hat die kriterienbasierte Evaluation nicht nur die aufgezeigten Schwächen, sondern auch Vorzüge, die im folgendem Unterkapitel diskutiert werden.

### 6.6 Vorzüge von Qualitätskriterien

Für die Entwicklung von Lernsystemen im universitären Bereich verlangt es nach einer praktikabel durchzuführenden, aussagekräftigen und anerkannten Evaluationsmethode.

Folgende Vorzüge sind mit der qualitätskriterienbasierten Evaluation verbunden:

- **Aufwand:** Durch die einfache, schnelle und kostengünstige Durchführung ist die qualitätskriterienbasierte Evaluation trotz der im vorherigen Unterkapitel benannten Schwächen am weitesten verbreitet und besonders geeignet für die formative Qualitätssicherung. Die qualitätskriterienbasierte Evaluation hält den Entwicklungsfortschritt nicht auf, sondern unterstützt ihn und schafft Klarheit über die Qualität der bewerteten Lernsysteme.
- **Methodik:** Die Verwendung von Qualitätskriterienkatalogen ist methodisch sauber und wissenschaftlich anerkannt, da eine Vergleichbarkeit von Bewertungen unterschiedlicher Personen gegeben ist. Die Anwendung von Qualitätskriterienkatalogen bietet gegenüber einer intuitiven Bewertung eine objektiviertere, systematische und strukturierte Beurteilung. Die Qualitätskriterien dienen als Strukturierungs-, Wahrnehmungs-, Argumentations- und Verständigungshilfe in Bezug auf die Modellvorstellung von Qualität. Evaluatoren werden durch die Qualitätskriterienkataloge sensibilisiert hinsichtlich der Komplexität der Bewertungsaspekte. Der Qualitätskriterienkatalog verhindert, dass wichtige Gesichtspunkte bei der Beurteilung von Software vergessen werden (Dick 2000, S.233). Qualitätskriterien sind kein Denkersatz. Ein davon abweichender höherer Anspruch ist aufgrund des unendlichen Lösungsraums der Entwicklung von Lernsystemen nicht möglich.
- **Wissensmanagement:** Der Qualitätskriterienkatalog bietet eine Struktur, in der Evaluationswissen, Projekterfahrungen (Lessons Learned), Literaturquellen, Ansprechpartner uvm. dokumentiert, gesichert und verbreitet werden können.
- **Einsatzmöglichkeiten:** Der Qualitätskriterienkatalog dient
  - als Checkliste für bestimmte Gestaltungsentscheidungen,
  - als Ideenpool, Orientierungs- und Nachschlagewerk,
  - zum Vergleich/Benchmarking,
  - zur Prognose,
  - zur Publikation/Dokumentation,
  - zur Qualitätssicherung und Selbstreflektion.

Weitere realisierbare Vorteile der computergestützten kriterienbasierten Evaluation zählt Hippler *et al.* (1988, 79f.) auf:

- Vorherige Antworten sind zu verschiedenen Zwecken in späteren Fragen verwendbar;
- Eingabeüberprüfung ist möglich;
- Komplexe Filtersprünge sind realisierbar;
- Die Fragen- oder Itemreihenfolge kann randomisiert werden;
- Antwortzeiten sind erfassbar.

Die Bewerter können Wörter und Sätze aus *vorherigen Antworten* kopieren. Dies kann Tipparbeit einsparen.

Zusätzlich zur *Eingabeüberprüfung* können Vorkehrungen gegen unbefugtes Bearbeiten des Fragebogens durch verschlüsselte Datenübertragung, Passwortschutz und Begrenzung des Befragungszeitraums getroffen werden.

*Filtersprünge*: Durch Auswertung der Antworten zur Laufzeit der Befragung können zusätzliche kontextspezifische Qualitätskriterien eingereiht werden, wodurch ein Nachfragen ähnlich wie bei Interviews realisiert werden kann.

Eine zufällige *Fragen- oder Itemreihenfolge* ist eher eine Störgröße beim Vergleich von Bewertungsdaten.

Die *Antwortzeiten* hängen auch von der Geschwindigkeit der verfügbaren Netzanbindung ab. Pausenzeiten können nur bedingt ermittelt werden (Hoepner 1994, S.93ff.). Doch nicht nur Antwortzeiten sind für die Auswerter interessant und aufschlussreich, sondern auch die Häufigkeiten der aufgerufenen Bedienelemente des Fragebogens, die protokolliert werden können.

Auf der Grundlage der vorangegangenen Diskussion der Schwächen und Vorzüge von Qualitätskriterien können generelle Anforderungen an deren Güte gestellt werden. Das nachfolgende Unterkapitel beschreibt wichtige zu beachtende Anforderungen.

## 6.7 Güte von Qualitätskriterien

Für die Abschätzung der Güte eines Qualitätskriterienkatalogs können die Standards der DeGEval (2003) angewendet werden:

- Nützlichkeit,
- Durchführbarkeit,
- Fairness,
- Genauigkeit.

Qualitätskriterienkataloge sollen handhabbar, effizient, aussagefähig, umfassend und möglichst wissenschaftlich fundiert sein. Durch deren Verwendung im Rahmen einer Evaluation

sollen nicht nur die genauen Vorzüge und Schwächen eines Lernsystems erhoben werden, sondern auch Gestaltungsempfehlungen für die Qualitätssicherung geliefert werden.

Qualitätskriterien sollten für alle an der Evaluation beteiligten Personen glaubwürdig, verständlich, nachvollziehbar sein und eine situationsgerechte, ganzheitliche Evaluation ermöglichen.

Detaillierte Anforderungen in Bezug auf einen „optimalen“ Qualitätskriterienkatalog nennt Meier (1995, S.189f.):

1. Die Beurteilung erfolgt durch einen – im Idealfall geschulten – Fachmann.
2. Die Kriterien sind vollständig, valide und reliabel.
3. Die Kriterien liegen in strukturierter Form vor und sind in Kriteriumskategorien (mit Überschriften) unterteilt.
4. Diese Kriteriumskategorien sind einzeln als Prüfinstrumente anwendbar.
5. Im Idealfall ist ein umfangreicher Katalog zu einer „Kurzprüfliste“ extrahiert.
6. Die Kriterien sind sprachlich korrekt, verständlich, treffend und knapp formuliert.
7. Der Katalog ist einfach und im Idealfall elektronisch auswertbar. Voraussetzung hierfür ist u.a., dass die Kriterien in Sachzusammenhängen vorstrukturiert und die Bewertungsskalen einfach konstruiert sind (z.B. Ja/Nein-Antworten).
8. Der Katalog verfügt über sogenannte Filter, d.h. Mechanismen, die eingangs den betreffenden Anwendungsbereich, Unterrichtszusammenhang, Zielgruppe u.ä. erfragen und so den Anwender von vornherein zu den für ihn individuell relevanten Kriterien führen. Diese Filter machen eine Qualitätsprüfung außerordentlich effektiv.
9. Der Kriterienkatalog ist sprachlich und grafisch einwandfrei und übersichtlich zu bearbeiten.

Eine weitere Anforderung betrifft die Flexibilität von Qualitätskriterienkatalogen. Diese „müssen flexibel genug sein, um verschiedenen Softwareformen und Inhalten gerecht zu werden, sowie exakt und komplex sein, um alle Besonderheiten eines Programms erfassen zu können“ (Thomé 1989, S.47).

Nach den genannten Anforderungen zur Güte wurde ein Qualitätskriterienkatalog entwickelt. Dieser wird bei der Expertenbefragung eingesetzt. Die Expertenbefragung ist Teil der nachfolgend erläuterten Qualitätssicherungsmethodik.

## Teil II

# Qualitätssicherungsmethodik

Die in diesem Teil vorgestellte Qualitätssicherungsmethodik beschreibt nicht, welche Informatiktechniken bei der Codierungsarbeit von Lernsystemen (beispielsweise Syntax-Kontrolle, Unit-, Integrations-, Last-, Performance-Tests, Debugging, Build-Prozess, Versionsverwaltung etc.) verwendet werden sollten. Diese werden von Liggesmeyer (2002) ausführlich beschrieben. Ebenso geht sie nicht dezidiert auf die Messung des Lernerfolgs infolge der Benutzung des Lernsystems ein. Sie ist kein Ersatz für Begabung, Fähigkeiten, Intuition, Kreativität und Situationskenntnis der Projektmitglieder. Sie ist primär auch nicht eine Checkliste für das Design von Lernsystemen, sondern ein nutzbringendes Leitmodell zur Organisation der Qualitätssicherung von Lernsystemen.

In diesem Teil der Arbeit werden die wesentlichen organisatorischen Abläufe und Instanzen der entwickelten Qualitätssicherungsmethodik behandelt. Diese besteht im Kern aus

- dem **Qualitätsverbesserungsprozess** für die Organisation von lebensphasenorientierten Evaluationen,
- der **Expertenbefragung** als formative Evaluationsmethode,
- der **Lernsystem-Konfiguration** als ordnende Struktur für Qualitätskriterien
- und der unterstützenden **LernSystemAnalyse**-Software zur effizienten Handhabung der Qualitätssicherungsmethodik.

Das vorgestellte Verfahren kann ergänzend zu bestehenden Qualitätssicherungsverfahren oder ausschließlich eingesetzt werden. Im nachfolgenden Kapitel wird zuerst auf den alle Lebensphasen eines Lernsystems umspannenden Qualitätsverbesserungsprozess (QVP) eingegangen.

## 7 Qualitätsverbesserungsprozess

Die Notwendigkeit einer Qualitätssicherung steht häufig in Zielkonflikt mit den Entwicklungskosten und der verfügbaren Entwicklungszeit. In vielen Lernsystem-Projekten läuft die Qualitätssicherung insgesamt Gefahr, zum Projektbaustein zu werden, den es in der letzten Projektphase einer Entwicklung abzuheben gilt. Wenn ein Lernsystem den Produktstatus erreicht, wird allzu oft die letzte Phase „Qualitätssicherung“ beschnitten, da die vorgelagerten Phasen bereits ihren Zeitrahmen überzogen haben.

Tritt dieser Fall ein, wird der Kunde unfreiwillig zum Produkttester, dem früher oder später klar wird, dass das Lernsystem nicht die vom Anbieter bescheinigte Qualität aufweist. Das

Lernsystem wird unvollkommen ausgeliefert und nach mehreren Reklamationen im Laufe der Zeit zur vollen Reife ausgefeilt. Diese korrektive Qualitätssicherung gilt in der IT-Branche als zumutbar.

Universitäre Lernsysteme können sich dieses mangelnde Qualitätsbewusstsein langfristig nicht erlauben. Bildungsprodukte für eine potenziell große Zielgruppe dürfen nicht unausgereift veröffentlicht werden. Studenten bringen dem Lernangebot in der Regel ein großes Vertrauen entgegen. Damit stehen die Lernsystem-Entwickler in einer gewissen Verantwortung, Qualität zu erzeugen.

Der nutzbringende Einsatz von Lernsystemen ist an die ganzheitliche Qualität des Lernsystems gebunden und nicht etwa durch den Einsatz von Neuen Medien vorprogrammiert. Bevor ein Lernsystem freigegeben wird, sollte die Qualität eines Lernsystems erwiesen sein. Dies kann durch die Verankerung effizienter Qualitätssicherungsmaßnahmen geschehen, die das Lernsystem kontinuierlich überprüfen und verbessern.

In einem solchen Qualitätsverbesserungsprozess (QVP) gilt es, möglichst frühzeitig die Zielgruppen Lernende, Lehrende und Administratoren bzw. Entwickler einzubinden. Die iterativ durch Prüfung der Lernsystem-Prototypen gewonnenen Erkenntnisse müssen dann in eine aufgabenangemessene Gestaltung umgesetzt werden. Bei einer effizienten Lernsystem-Entwicklung bilden Design, Implementierungsarbeit und Qualitätssicherung eine Einheit. Für diese Einheit sorgt die Anwendung des QVP. Dieser umfasst alle Lebensphasen eines Lernsystems – von der Idee über die Erstellung des Entwurfs, die anschließende Implementierung, bis hin zur Lernsystembenutzung. In diesem Prozess findet eine auf die Lebensphase des Lernsystems zugeschnittene Evaluation zu Terminen statt, an denen ein Meilenstein der Projektentwicklung erreicht wurde. Die Evaluationen dienen dem Zweck, die Qualität der Lernsystem-Prototypen bzw. -Produkte frühzeitig zu überprüfen und rechtzeitig neue Ansatzpunkte zur kontinuierlichen Verbesserung zu generieren.

### 7.1 Wirkungskreise

Qualitätssicherung sollte den Entwicklungsprozess über alle Lebensphasen eines Lernsystems begleiten. Hierzu sind regelmäßig Evaluationen durchzuführen, die Auskunft über die Qualität des Lernsystems, den Projektstand bzw. den Projektfortschritt geben können.

In der Abb. 6 auf der nächsten Seite sind neben den lebensphasensorientierten Evaluationen zwei unterschiedliche Wirkungskreise zu erkennen. Sie deuten an, dass im Unterschied zu sequentiellen (wasserfallartigen) Entwicklungsmodellen hier ein zirkulärer QVP-Ansatz mit vorwiegend formativer und summativer Evaluation favorisiert wird. Der in Abb. 6 dargestellte innere Wirkungskreis stellt die formative Evaluation dar (siehe UKap. 7.6 auf Seite 31). Der ebenfalls abgebildete äußere Wirkungskreis schließt weitere Evaluationen ein, insbesondere die summative Evaluation (siehe UKap. 7.7 auf Seite 32). Die Terminologie „formativ und summativ“ wurde von Scriven (1967, S.39f.) eingeführt. Formativ bedeutet, dass eine

Evaluation des Lernsystems noch während dessen Entwicklung stattfindet. Die summative Evaluation untersucht hingegen das bereits fertig entwickelte Lernsystem-Produkt. Hierbei werden vorwiegend Benutzertests durchgeführt.

Die Lernsystem-Entwicklung sollte nicht in einmal zu durchlaufenden abgeschlossenen Lebensphasen organisiert werden. In späteren Phasen zunehmendes Systemwissen sollte durchaus dazu führen, dass die vorgelagerten Phasen erneut zyklisch durchlaufen werden können, um notwendige Korrekturen einfließen zu lassen. Den Zeitpunkt, wann aus diesen Wirkungskreisen ausgebrochen wird, bestimmen die Ergebnisse der durchgeführten lebensphasenorientierten Evaluationen. Diese lassen einen Rückschluss auf die momentane Qualität des Prototypen zu und eröffnen die Möglichkeit zur frühzeitigen Verbesserung des Prototypen.

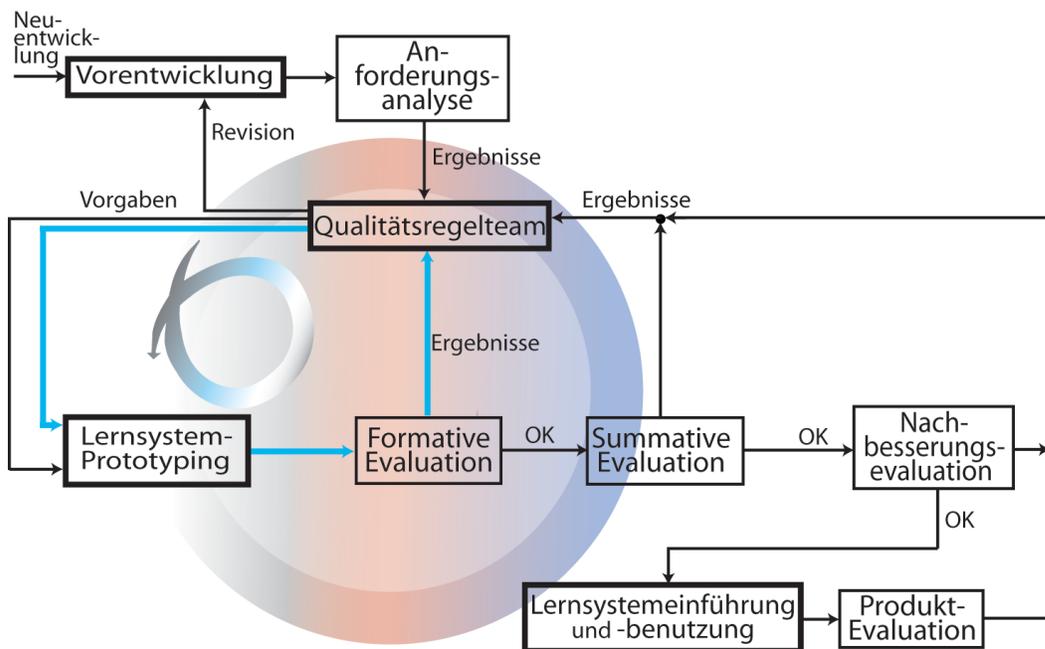


Abb. 6: Wirkungskreise des Qualitätsverbesserungsprozesses

In Abb. 7 auf der nächsten Seite ist ein typischer Entwicklungsprozess abgebildet. Die Kosten der Implementierung steigen erfahrungsgemäß im Laufe eines Projekts und damit auch die Kosten für Korrekturen. Zugleich nimmt das Systemwissen über das zu fertigende Produkt zu. In frühen Lebensphasen sind Qualitätsmängel noch relativ einfach zu lokalisieren aufgrund der geringeren Softwarekomplexität der Prototypen. Fehler von Software resultieren nach einer Einschätzung von Walraet (1991, S.7ff.) überwiegend aus der Phase der Vorentwicklung (56%), zu 10% aus der Designphase und nur zu 7% aus Fehlern der Kodierungs- bzw. Programmierungsphase. Daher sollte in frühen Entwicklungsphasen der Qualitätssicherungsaufwand besonders hoch angesetzt werden.

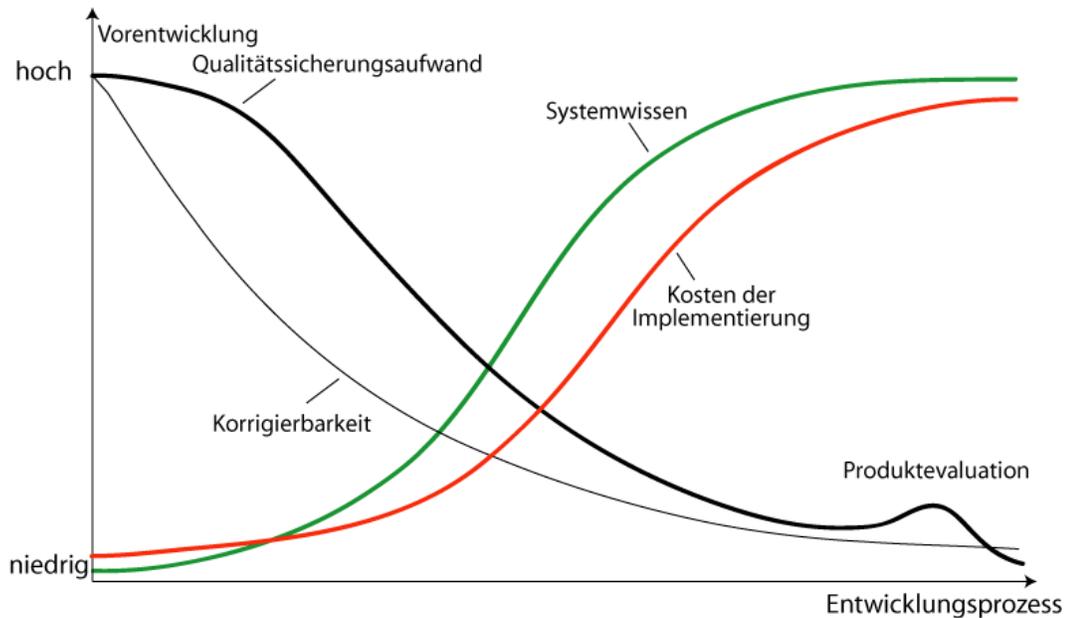


Abb. 7: Entwicklungsprozess

Für das weiterführende Verständnis der Abb. 6 auf der vorherigen Seite werden die Lebensphasen und die zugeordneten Evaluationen in den nächsten Unterkapiteln beschrieben.

### 7.2 Vorentwicklung

Der Vorentwicklung eines Lernsystems kommt für das Gelingen eines Lernsystem-Projektes eine bedeutende Rolle zu. In dieser Planungsphase soll sich schließlich aus einem Auftrag, einer Idee, einer Inspiration, einem festgestellten Handlungsbedarf oder einem visionären Weitblick konzeptionelle Klarheit für eine Lernsystem-Lösung entwickeln.

Viele Lernsystem-Projekte sind gescheitert oder wurden mit Verspätung fertig gestellt, da die Entwickler zu früh mit der programmiertechnischen Umsetzung eines Vorhabens begonnen haben. Der zu geringe Wirkungsgrad dieses unsystematischen Vorgehens wird zumeist zu spät gegen Ende des Projektes erkannt. Nur relativ kleine und einfache Projekte können aufgrund der Überschaubarkeit der Problemstellungen und der fehlenden Komplexität, schneller zur Phase der Realisierung übergehen.

Ansätze des Extreme Programming (XP)<sup>15</sup> sind da schon zielführender. Hierbei wird das Produkt gemeinsam mit den identifizierten Benutzergruppen besprochen (User stories). Gemeinsam mit den Entwicklern wird das Lernsystem gedanklich konfiguriert. Bedürfnisse und

<sup>15</sup><http://www.extremeprogramming.org/rules.html>

die Eigenschaften<sup>16</sup> der Benutzer, sowie das aktuelle Lernumfeld, die ausgewählte Lerntechnologie, die Lerninhalte und die Lehrmethode werden bei der Vorentwicklung angemessen berücksichtigt. Der Austausch ist die Grundlage für eine angemessene Festlegung von Zeitplänen, Ressourcen, Anforderungen, Qualitätskriterien und Evaluationen.

Die Konzepte, die in der Vorentwicklung zu erarbeiten sind, werden im Projektverlauf immer feiner ausgearbeitet. Sie müssen allerdings nicht unbedingt vollständig sein. Nicht alles lässt sich sinnvoll planen, denn manche Ideen oder Probleme treten erst bei der Realisierung auf. Ein häufig vorliegendes Problem ist, dass die Kosten und die Zeit fixiert werden, noch bevor durch konkrete Anforderungen klar ist, was überhaupt realisiert werden soll. Natürlich sollte nach der Vorentwicklung eines Lernsystems weiterhin genug gestalterischer Raum und Zeit für kreatives Design sein. Gerade bei einem frühen Prototyp sollten noch unvorhergesehene und notwendige Anpassungen möglich sein. Die Vorentwicklung ist im weiteren Verlauf des Projekts stets zu überprüfen, nicht nur ob der geplante Weg zur Realisierung des Lernsystems eingeschlagen wurde, sondern auch ob ein anderer besserer Weg favorisiert werden müsste. Möglichst viele Schwierigkeiten und Probleme sollten im Vorfeld erkundet und beispielsweise durch einen Problemlösungszyklus<sup>17</sup> ausgeräumt werden.

Die an der Entwicklung beteiligten Personen können durch die Vorentwicklung ein klares Bild von ihren Aufgaben und den für diese Aufgaben vorgesehenen Ressourcen gewinnen. Die Ergebnisse der Vorentwicklung werden in Gestalt von nachprüfbaren Anforderungen festgehalten, an denen sich die subjektive Qualität<sup>18</sup> eines Lernsystems messen lassen kann. Auf der Grundlage der Anforderungen werden Qualitätskriterien ausgewählt oder neu definiert. Qualitätskriterien sollen alle in Bezug auf die Anforderungen wichtigen Qualitätsaspekte und Fragestellungen umfassen. Auf diese Weise helfen sie den Bewertern, ein Verständnis zu entwickeln, bis zu welchem Grad ein Qualitätskriterium erfüllt ist. Die Beantwortung der Fragen durch Experten trägt zur Bewertung der Anforderungserfüllung und Ableitung von Verbesserungsmaßnahmen zum Lernsystem bei.

Die Gültigkeit, Vollständigkeit und Verständlichkeit von Anforderungen kontrolliert die Anforderungsanalyse. Diese wird im folgenden Unterkapitel erläutert.

### 7.3 Anforderungsanalyse

Requirements-Engineering ist die Antwort auf die Softwarekrise in der Mitte der sechziger Jahre. Identifizierte Probleme dieser Krise waren fehlerhafte, nicht wartbare Software, Projekte, die nicht termingerecht fertig wurden und unzureichend eingehaltene Anforderun-

---

<sup>16</sup>Vorerfahrungen, Wissen, Fertigkeiten, Lernmotivation, Fähigkeiten, Interessen, Einstellungen, kognitive Strategien, mentale Modelle etc.

<sup>17</sup>Siehe beispielsweise das allgemeine Problemlösungsverfahren von Klauer (1985, S.5ff.).

<sup>18</sup>Siehe Gleichung (1) auf Seite 7 .

gen. Die Definition von Anforderungen sollte daher möglichst zu Beginn eines Lernsystem-Projekts vollzogen werden.

Eine Anforderungsanalyse sorgt für die Richtigkeit und Vollständigkeit der Anforderungen während der gesamten Projektlaufzeit. Konkret bedeutet dies, dass nachgeprüft wird, ob die richtigen Benutzer des Lernsystems und deren Eigenschaften und Aufgaben identifiziert wurden, die Projektziele vollständig, verständlich und umsetzbar sind etc.. Weiterhin wird analysiert, ob absehbare Probleme erkannt und voraussichtlich gelöst werden können, die Priorisierung angemessen ist, Qualitätskriterien zur Überprüfung des Erfüllungsgrades den Anforderungen zugewiesen sind uvm.. Fällt die Analyse zufriedenstellend aus, erfolgt das Prototyping und die begleitenden formativen Evaluationen.

Das Qualitätsregelteam hat im weiteren Projektverlauf nach dieser Analyse mit den verifizierten und validierten Anforderungen eine Grundlage, die Qualität des Lernsystems anhand der Anforderungserfüllung zu bestimmen.

### 7.4 Qualitätsregelteam

Die Qualitätssicherungsmethodik sieht die Einrichtung eines Qualitätsregelteams vor. Das Qualitätsregelteam sorgt für die konsequente Umsetzung der systematischen und ganzheitlichen Qualitätssicherung von Lernsystemen.

Eine Hauptaufgabe betrifft die Initiierung und technisch-organisatorische Betreuung der lebensphasenorientierten Evaluationen. Alle erhobenen Evaluationsdaten wertet das Qualitätsregelteam aus. Beispielsweise werden aus den freien Antworttexten aus den Expertenbefragungen inhaltlich zusammenfassende Hinweisklassen formuliert. Diese dienen als Entscheidungsgrundlage für die Umsetzung von Handlungsmaßnahmen zur Verbesserung des untersuchten Lernsystems. Nach der Information und Umsetzung der Verbesserungsmaßnahmen durch die Entwickler überprüft das Qualitätsregelteam erneut die Qualität des Lernsystems durch eine Evaluation. Daher nimmt das Qualitätsregelteam in allen Lebensphasen eine Schlüsselrolle für den qualitätsrelevanten Informationsaustausch innerhalb des Projektes ein.

Interne Qualitätssicherungsmaßnahmen sind auf ein hohes Maß an Ehrlichkeit und Offenheit aller Beteiligten angewiesen. Das Qualitätsregelteam sollte für eine kontinuierliche Ansammlung und Aktualisierung von gewonnenen Projekterfahrungen sorgen. Diese können im Qualitätskriterienkatalog hinterlegt werden. Auf diese Weise profitieren alle derzeitigen und zukünftigen Projektmitglieder von diesen Erfahrungen und ein kollektives Lernen kann stattfinden.

Mit ihrem durch die Qualitätssicherungsaufgabe erworbenen projektinternen Wissen beurteilt das Qualitätsregelteam das Lernsystem. Diese begründete Bewertung geht als Quali-

tätsindikator in die Berechnung der Anforderungserfüllung mit ein (siehe Kap. 15 auf Seite 90).

Das Qualitätsregelteam sollte ein tiefes, d.h. nicht nur theoretisches, sondern auch praktisches Verständnis über diejenigen Prozesse besitzen, die zu einer nachhaltigen Qualitätsverbesserung führen. Das Qualitätsregelteam ist im Idealfall ein interdisziplinäres Team von Lernsystem- und Qualitätssicherungs-Experten, deren Mitglieder zusammen eine Vielzahl für die Lernsystembeurteilung notwendiger Fach- und Methodenkompetenzen abdecken. In der Praxis sind Experten allerdings eine knappe und teure Ressource. In kleineren Projekten kann das Team daher nur von einer Person besetzt werden. Diese Person sollte zumindest mediendidaktische Kenntnisse aufweisen können.

### 7.5 Lernsystem-Prototyping

Wenn Evaluationen zu den verschiedenen Lebensphasen eines Lernsystems durchgeführt werden, dann steht zumeist kein voll entwickeltes Lernsystem für die Evaluation zur Verfügung. Es handelt sich dann zumindest um einen mehr oder weniger weit entwickelten Prototypen. In vielen Bereichen werden Prototypen eingesetzt, um möglichst schnell und kostengünstig ein Objekt zu schaffen, dessen Eigenschaften und Merkmale man testen kann. An Prototypen können in Verbindung mit Evaluationen generell die Software-Architektur, die Funktionalität, die Integrierbarkeit in ein bestehendes übergeordnetes System, die Eignung von Entwicklungswerkzeugen, der Lerninhalt, die Lehrmethodik, die Akzeptanz, der Aufbau und die Performanz des Lernsystems etc. erprobt werden.

Prototypen können in verschiedenen Detaillierungsstufen implementiert werden. Wird nur ein bestimmter Aspekt untersucht, so kann man in der Regel davon unabhängige Funktionsbereiche der späteren Software vernachlässigen. Es wird daher nur der Code implementiert, der für einen bestimmten zu untersuchenden Aspekt nötig ist. Ein Prototyp ist somit nicht in allen Bereichen funktionsfähig. Dies unterscheidet ihn von dem zu schaffenden Produkt.

Prototypen können dabei horizontal oder vertikal angelegt sein. Horizontal bedeutet, dass die Prototypen die wesentlichen Funktionen einer Software andeuten. Dies kann beispielsweise durch ein Navigationsmenü zum Aufruf der Funktionen umgesetzt werden. Eingeschränkt ist hingegen die Detailimplementierung. Diese muss nicht vollständig und für den Anwender benutzbar sein. Bei vertikalen Prototypen sind nicht alle Funktionen verfügbar, dafür sind die vorhandenen Funktionen für die Erfüllung bestimmter Aufgaben funktionsfähig. Zuletzt unterscheidet man noch Szenarien-Prototypen. Diese sind für ein bestimmtes (Benutzer-) Szenario in der notwendigen Tiefe und Breite der Funktionalität ausgearbeitet.

Beim Prototyping gibt es verschiedene Ansätze bezüglich der Wiederverwendung des Programmcodes, die je nach Ziel und Aufgabe Sinn machen. Bei sogenannten Wegwerf-Prototypen soll die implementierte Software nach der Gewinnung von Erkenntnissen durch das Testen des Prototypen nicht weiter verwendet werden. Dies hat den Vorteil, dass sich die

Programmierer nicht viel Gedanken um die Effizienz, Sicherheit, Robustheit, Vollständigkeit, Strukturiertheit und Modularität des Programmcodes machen müssen, was letztlich Zeit kostet. In vielen Fällen kann man auf die Programmierarbeit auch völlig verzichten, wenn ein Papierprototyp für einen bestimmten Zweck ausreicht. Geschaffen wird eine sogenannte schnelle und unsaubere Lösung (Quick and Dirty). Wenn hingegen von vornherein geplant ist, den Prototyp weiter auszubauen, muss in Richtung Programmierarbeit mehr Aufwand betrieben werden. In diesem Fall spricht man vom „Rapid Prototyping“.

Die Reihenfolge der in den nachfolgenden Unterkapiteln beschriebenen Evaluationen entspricht der zeitlichen Abfolge und dem wachsenden Entwicklungsgrad des Lernsystems. In der Vorentwicklung steht unter Umständen nur ein Papierprototyp zur Verfügung, bei dem die Benutzerschnittstelle des Lernsystems auf Papier skizziert ist. In den nachfolgenden Evaluationen wird hingegen dann ein mehr oder weniger komplexer vertikaler oder horizontaler Softwareprototyp in einer formativen Evaluation bewertet. Das nächste Unterkapitel erläutert, was unter einer formativen Evaluation zu verstehen ist.

### 7.6 Formative Evaluation

Die formative (:=entwicklungsbegleitende) Evaluation ist ein wichtiger Bestandteil iterativer Softwareentwicklung (ISO 13407 1999). Mit ihr werden häufig folgende Evaluationsverfahren wie Inspektion, Review, Gutachten, Rezension, Walkthrough und Interview in Verbindung gebracht. Bei den genannten Verfahren können Fragebögen, Leitfäden, Kriterienkataloge, Checklisten und Kennzahlen zum Einsatz kommen. Hinsichtlich der Eignung für den formativen Evaluationskontext, d.h. unter besonderer Berücksichtigung der Ökonomie ist die kriterienbasierte Evaluation prädestiniert. Zum einen steht der Durchführungsaufwand in einem guten Verhältnis zum Nutzen und zum anderen liefert sie aussagekräftige Ergebnisse, noch bevor der aktuelle Lernsystem-Prototyp wesentlich weiter entwickelt wurde.

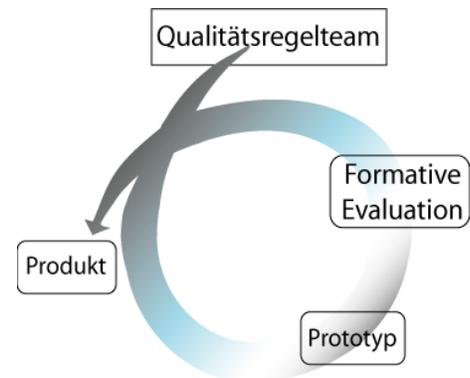


Abb. 8: Formativer Wirkungskreis

Die frühzeitige Entdeckung von Schwachstellen des Lernsystems ist bedeutend, da in frühen Phasen der Softwareentwicklung Mängel noch relativ einfach zu beseitigen sind aufgrund der geringeren Softwarekomplexität der Lernsystem-Prototypen. Die Bereitschaft für Anpassungen ist hoch, und die Kosten der Änderung fallen geringer aus. Die formative Evaluation soll bei Lernsystem-Prototypen Vorzüge und Schwächen aufdecken, noch bevor die Software den Produktstatus erreicht. Auf diese Weise erhalten die Projektmitglieder frühzeitig eine Rückmeldung zu ihrer Entwicklung. Notwendige Verbesserungsmaßnahmen können mit dieser Erkenntnis ökonomisch in den Entwicklungsprozess integriert werden.

## 7.7 Summative Evaluation

Die Ergebnisse von Karat *et al.* (1992, S.397ff.) sprechen dafür, dass die formative Evaluation den Produkttest mit End-Nutzern nicht ersetzen kann. Eine beträchtliche Menge an schweren Usability-Fehlern wurde nur durch den Test mit End-Nutzern aufgedeckt. Die formative Experten-Evaluation und die summative Produkt-Evaluation sind komplementär, wie Studien belegen (Jeffries *et al.* 1991, S.119ff., Karat *et al.* 1992, S.397ff.). Daher ist zu empfehlen, dass sobald ein Lernsystem die Produktreife erlangt, eine summative Evaluation durchgeführt wird. Eine finaler Prototyp erlangt die Produktreife, wenn er alle Muss-Anforderungen laut den formativen Evaluationsergebnissen vereinbarungsgemäß erfüllt.

Bei einer summativen Evaluation sind benutzerorientierte (empirische) Evaluationsverfahren einsetzbar, die in der Regel mehr Zeit und Vorbereitung als die formative Evaluation brauchen:

- Labortests als Produkttests, Inhalts- und Medienanalyse, visuelle und auditive Verhaltensbeobachtung/-analyse (direkt, einseitig durchsichtige Spiegel, Video,<sup>19</sup> Protokollierung über Computer-Logfiles,<sup>20</sup> Eyetracking,<sup>21</sup> Polygraphmessung,<sup>22</sup> Methode des lauten Denkens<sup>23</sup> etc.);
- Lernerfolgsmessung (Transferanalyse, Leistungsmessung, Einsatz von Testaufgaben etc.).

Aufgrund der Schwerpunktsetzung der vorliegenden Arbeit auf die kriterienbasierte formative Analyse von Lernsystemen wird auf die stichpunktartig genannten summativen Evaluationsverfahren hier nicht weiter eingegangen. Mehr Information findet sich bei Shneiderman (2002).

## 7.8 Nachbesserungsevaluation

Die Nachbesserungsevaluation (Final Check) soll den Prototyp auf die Stufe des auslieferbereiten Produkts verhelfen. Die Freigabe des Lernsystems erfolgt, nachdem es auf verschiedenen Computersystemen erfolgreich getestet wurde. Gegebenenfalls vorhandene Inkompatibilität bei der Darstellung und der Funktionsweise wird in der Nachbesserungsevaluation entdeckt und behoben.

---

<sup>19</sup>Videokonfrontation: Beobachteter soll im Anschluss seinen Film kommentieren.

<sup>20</sup>Aufzeichnung der Interaktivität zwischen Benutzer und Lernsystem durch eine Software.

<sup>21</sup>Augenbewegung verfolgen, auch Blickaufzeichnung genannt.

Siehe beispielsweise die Website <http://www.poynterextra.org/eyetrack2004/photos.htm>.

<sup>22</sup>Stressmessung, Herzaktivität, Leitfähigkeit der Haut, Muskelaktivität, etc.

<sup>23</sup>Die Testperson soll ihre Gedanken bei der Anwendung laut aussprechen.

## 7.9 Produktevaluation

Die Produktevaluation erfolgt nach der Lernsystemeinführung mit den Zielgruppen am Einsatzort. Nach der Auslieferung des fertigen Produktes ist zu gewährleisten, dass die Reaktionen der Benutzer erhoben werden und das Feedback an das Qualitätsregelteam weitergeleitet wird. Für die Projektmitglieder ist es wichtig zu erfahren, ob ihr Lernsystem-Projekt erfolgreich war und das Lernsystem von den Benutzern akzeptiert und planmäßig genutzt wird.

Zwar werden durch die vorangegangenen Evaluationen der Großteil des Verbesserungspotenzials bereits festgestellt. Es kann jedoch vorkommen, dass diese allein nicht ausreichen. Die Effektivität der vorangegangenen Evaluationen hängt von verschiedenen Einflussgrößen ab. Beispielsweise sind bei der formativen Befragung die Expertenanzahl und der Erfahrungsgrad der Experten entscheidend (siehe UKap. 15.3.4 auf Seite 103 und UKap. 15.4 auf Seite 109). Experten mit einem hohem Erfahrungsgrad stehen jedoch oft nicht in ausreichender Zahl zur Verfügung oder sind bei einer internen Evaluation möglicherweise betriebsblind.

Die Produktevaluation wird mit realen (und nicht mit potenziellen) Benutzern des Lernsystems durchgeführt. Die Auswertung von Nutzungsdaten (Web-Tracking), sowie Beanstandungen, Reklamationen, Kritik und Anregungen der Benutzer können zu einer Qualitätssicherung des Lernsystems beitragen. Die abgeleiteten Verbesserungsmaßnahmen können allerdings erst bei einer Revision des Lernsystems umgesetzt werden.

## 8 Expertenbefragung

Der Begriff Expertenbefragung impliziert, dass Experten befragt werden. Experten sind Personen, die zu dem relevanten Lehrbereich oder/und zur Qualitätssicherung von Lernsystemen über ein solides Wissen verfügen. Sie stellen die Quelle für Hinweise zur Qualität des Lernsystems dar. Die Hinweise werden über einen Fragebogen erhoben und vom Qualitätsregelteam ausgewertet.

Expertenbefragungen können intern durchgeführt oder extern in Auftrag gegeben werden. Für die erste Alternative sprechen zumeist die Kostenfrage und die selbstbestimmte Verfügbarkeit von Ergebnissen, die zweite Alternative ist für die Glaubwürdigkeit von Daten dienlich. Externe Evaluationen verursachen zusätzliche Kosten und können Angstgefühle und Abwehrreaktionen bei den Projektmitgliedern auslösen. Für die formative Evaluation, bei der die Erzielung einer Außenwirkung der Qualitätssicherung noch nicht gewünscht wird, erweist sich die interne Evaluation als zweckmäßig.

In Übereinstimmung mit Hampe-Neteler (1994, S.130) ist die Verfahrensökonomie ein wichtiges Bewertungskriterium für die Praxistauglichkeit der Evaluationsmethode. Die Expertenbefragung stellt eine schnell, einfach und kostengünstig durchzuführende Evaluationsmethode

de dar. Zu diesem Schluss kommen auch andere Untersuchungen (Hamborg 2002, S.303ff., Nielsen 1994, S.25ff., Virzi *et al.* 1993, S.309ff., Jeffries *et al.* 1991, S.119ff.). Einschränkend muss allerdings angemerkt werden, dass die Kosten von der Anzahl und Verfügbarkeit der befragten Experten abhängen.

Bei der Befragung von Experten soll nach Judd *et al.* (1991, S.485f.) folgendes vermieden werden:

- Einbeziehung von Personen ohne deren Zustimmung bzw. Information;
- Zwang zur Teilnahme;
- Verschleiern des wahren Zwecks einer Untersuchung;
- Täuschung der Teilnehmer;
- Verführung der Teilnehmer zu entwürdigenden Verhaltensweisen;
- Verletzung des Rechts auf Selbstbestimmung der Teilnehmer;
- Verursachung von physischem oder mentalem Stress bei den Teilnehmern;
- Eindringen in die Privatsphäre der Probanden;
- Vorenthaltung von versprochenen Belohnungen;
- Unfaire Behandlung der Teilnehmer sowie mangelnde Achtung und Respekt.

Eine freiwillige Teilnahme der Experten kann über Anreizsysteme erzielt werden. Mit diesem Thema setzt sich das folgende Unterkapitel auseinander.

## 8.1 Anreizsysteme

Befragungen werden von vielen Teilnehmern als langweilig empfunden. Zudem ist der Nutzen für den Teilnehmer oft nicht erkennbar. Trifft bereits eine dieser Empfindungen zu, besteht die Gefahr, dass die Fragen halbherzig beantwortet werden.

Experten als Quelle der für die Qualitätssicherung notwendigen Bewertungsdaten müssen zunächst für die Befragung gewonnen werden. Die Motivation der Experten zur Teilnahme an einer Befragung hängt in der Regel von folgenden Faktoren ab:

- (monetäre, nicht monetäre) Anreize;
- Beantwortungszeit: Je länger der Fragebogen, desto höher wird die Belastung an Zeit und Aufwand der Bewerter.
- Design des Fragebogens;
- Art und Verständlichkeit der Fragestellungen;
- Antwortmöglichkeiten;
- Thema der Befragung.

Durch Anreizsysteme bzw. Incentives sollen die Motivation für die Teilnahme an der Expertenbefragung gesteigert werden.

Bei der in Teil IV auf Seite 77 diskutierten Studie wurden folgende nicht-monetäre Anreize für die Teilnehmer geschaffen:

- Es fand eine Aufklärung der Bewerber über Sinn und Zweck der Befragung statt;
- Die Experten wurden persönlich eingeladen und in die Benutzung des Fragebogens eingewiesen;
- Es wurde ein Bonussystem angeboten. Für jede Teilnahme an einer Befragung erhielt der Experte sogenannte Evalits. Diese Evalits konnte dieser für die lizenzierte Nutzung der LernSystemAnalyse-Software einsetzen.
- Der Aus- und Wiedereinstieg in den Fragebogen waren jederzeit möglich;
- Der Experte hatte die Möglichkeit, durch die Metainformationen neues Wissen zu erwerben (siehe UKap. 8.2.2 auf Seite 41);
- Der webbasierte Fragebogen wurde vor der Hauptbefragung im Rahmen von Voruntersuchungen ausgiebig getestet, sodass die Experten von einer stabilen Software ausgehen konnten.

Viele der genannten Anreize hängen mit der Gestaltung des bei der Expertenbefragung verwendeten Fragebogens zusammen. Wer über einen webbasierten Fragebogen Informationen erheben will, muss die Anforderungen, Wünsche und Möglichkeiten der Experten ernst nehmen. Vor allem ist die Benutzerschnittstelle des Fragebogens so zu gestalten, dass sie die Experten stimuliert, möglichst viel von ihren Erfahrungen mit dem zu bewertenden Lernsystem preiszugeben.

Die Fragebogengestaltung ist das erste Glied einer Kette von Handlungen (Datenerhebung, Aggregation, Analyse, Auswertung, Präsentation), die für eine effiziente Befragung verantwortlich sind. Eine aufgabenangemessene Fragebogengestaltung unterstützt den Befragten durch die Bereitstellung und klare Anordnung von effizient zu bedienenden und aufgabenangemessenen Gestaltungselementen des Fragebogens. Umgekehrt kann eine unangemessene Gestaltung zu einem Mehraufwand und schlimmstenfalls zu falschen, ungültigen oder gar keinen Angaben des Befragten führen.

Vor allem für ungeübte Computernutzer kann das Ausfüllen eines Webformulars in einer webbasierten Befragung eine Hürde darstellen. Eine kurze Einweisung zur Erklärung ist in solchen Fällen unentbehrlich. Im folgenden Unterkapitel werden weitere Voraussetzungen der webbasierten Befragung beschrieben.

### **8.2 Webbasierte Befragung**

Das Internet ist bereits in viele Lebensbereiche des Menschen vorgedrungen. Es ist naheliegend, diesen Vorteil der Erreichbarkeit für die Expertenbefragung zu nutzen.

1986 erkannte Sproull (1986, S.160) E-Mail als eine Form zur Befragung in Unternehmensnetzwerken. Zum Einsatz sind seiner Meinung nach folgende Voraussetzungen zu erfüllen:

- Zugang der Probanden zur notwendigen Technologie;
- Bereitschaft an einer Umfrage über das Medium teilzunehmen;
- Kompatibilität mit konventionell erhobenen Daten.

Diese Voraussetzungen für den Einsatz gelten auch bei einer Befragung über das Internet mittels Webformularen.

Bei einem persönlich-mündlichen Interview ergibt sich nach Hippler und Beckenbach (1992, S.44) folgende Situation, die sich auch auf die webbasierte Expertenbefragung übertragen lässt:

- Sinkende Ausschöpfungsquoten in repräsentativen Bevölkerungsumfragen;
- Zunehmende Tendenz zur Befragung von Spezialpopulationen;
- Gestiegene Anforderungen an Design und Komplexität von Umfragen;
- Wachsende Ansprüche an die Qualität der erhobenen Daten;
- Wunsch nach Aktualität und schneller Verfügbarkeit der Daten.

Mit dem Einsatz von webbasierten Befragungstechniken kann eine größere Reichweite von Befragungen erzielt werden. Dies ist vor allem dann förderlich, wenn die für die Befragung wichtigen Ressourcen – die Experten – zeitlich und örtlich beschränkt zur Verfügung stehen. Durch webbasierte Fragebögen ist es möglich, dass Bewerber am eigenen Computer zu einer ihnen angenehmen Zeit befragt werden können. Aufgrund der Medienunterstützung und des sich daraus ergebenden Geschwindigkeitsvorteils, können mehr Experten beispielsweise zur gleichen Zeit über dasselbe Medium in ihrer bevorzugten Sprache befragt werden. Auf der Grundlage einer eigenen Einschätzung wurden vier mögliche Befragungen hinsichtlich der in Tab. 3 aufgestellten Kriterien bewertet.

<b>Kriterien</b>	<b>WWW-Befragung</b>	<b>mündliche Befragung</b>	<b>telefonische Befragung</b>	<b>schriftliche Befragung</b>
Kosten	sehr gering	sehr hoch	mittel	mittel
Rücklaufgeschwindigkeit	schnell	unmittelbar	unmittelbar	langsam
Rücklaufquote	unterschiedlich	hoch	mittel	mittel
Geographische Reichweite	hoch	sehr niedrig	hoch	sehr hoch
Verfügbarkeit der Befragungsmedien	unterschiedlich	direkt	unterschiedlich	hoch
Erreichbarkeit verschiedener Zielgruppen	viele	viele	viele	beinahe alle

Tab. 3: WWW-, mündliche, telefonische und schriftliche Befragung

Zur Einschätzung der Präferenzen bezüglich der Computerbefragung und der papiergebundenen Befragung existiert die Untersuchung von Fuhrmann Verdaguer (2005, S.118). Um die Präferenzen herauszufinden, wurde von Fuhrmann Verdaguer ein papiergebundener Fragebogen zum Ankreuzen von zutreffenden Antworten und ein computerbasierter Fragebogen mit freier Antwortmöglichkeit in einer Befragung verwendet. Im Anschluss an die Befragung wurden die 10 Personen gebeten, die unterschiedlichen Fragebögen zu beurteilen. 60% der Personen ziehen die papiergebundene der computergestützten Befragung vor. Allerdings haben diejenigen Experten die Computerbefragung bevorzugt, die weniger Zeit zur Beantwortung der Fragen benötigten. Das Ergebnis der Untersuchung deutet darauf hin, dass je nach Computererfahrung unterschiedliche Präferenzen bestehen.

Fuchslocher und Fitz (1992, S.221) haben die Vorteile und Nachteile der Computerbefragung in Tab. 4 zusammengefasst.

Vorteile	Nachteile
<ul style="list-style-type: none"> <li>● Kostenersparnis;</li> <li>● Kein Interviewerbias;</li> <li>● Schnelle Verarbeitung, Koordination und Auswertung;</li> <li>● Keine intensive Interviewschulung notwendig;</li> <li>● Spieltrieb, Spaß an der Befragung bzw. Bedienung;</li> <li>● Autonome Entscheidungsfreiheit.</li> </ul>	<ul style="list-style-type: none"> <li>● Übertriebene, mehr oder weniger unkontrollierte Befragung;</li> <li>● Komplexität der Befragung;</li> <li>● Befragungsdauer;</li> <li>● Befragungsabbruch bei Unverständnis und Zeitproblemen;</li> <li>● Unbeweglichkeit, ortsgebunden, fest installiert;</li> <li>● Größe der Geräte;</li> <li>● Psychologische Konflikte.</li> </ul>

Tab. 4: Vor- und Nachteile der Computerbefragung nach Fuchslocher und Fitz (1992)

Eine *Kostenersparnis* ist erst dann gegeben, wenn über den Fragebogen eine große Personenanzahl befragt wird und die Software für andere Befragungen wiederverwendet werden kann. Der Kostenvergleich von Watt (1997) zwischen WWW-Befragung, postalischer Befragung und Telefoninterview zeigt, dass sich die internetbasierte Befragung erst ab ca. 1000 Teilnehmern als deutlich günstiger erweist.

Die Änderung des Antwortverhaltens (*Interviewerbias*) äußert sich vor allem in deutlichen Ergebnisdifferenzen bei Fragen mit starker Erwünschtheit (Hippler und Beckenbach 1992, S.50f., Ewen und Gelszus 1990, S.120). Laube (1994, S.794) hält die Befragung ohne Interviewer für vorteilhaft: „Die Qualitätssicherungs-Beauftragten können mehr Bereitschaft zur Antwort und ehrliche anstatt sozial erwünschte Bewertungen erwarten, als wenn sie persönlich das Interview führen“.

*Schnelle Verarbeitung, Koordination und Auswertung* sind Randbedingungen der formativen Evaluation. „Am liebsten hätte man die Ergebnisse seiner Fragen noch während der laufenden Sitzung. Dies wird bei ausgewählten Zielgruppen mit Hilfe des Internets in absehbarer Zeit möglich sein“ (Haupt 1996, S.63).

Das von Fuchslocher und Fitz (1992) genannte Problem bezüglich der *Größe der Geräte* ist mit den heute verfügbaren portablen Computersystemen nicht mehr gegeben.

Im nächsten Unterkapitel wird das Layout des webbasierten LernSystemAnalyse-Fragebogens beschrieben. Der Fragebogen wird in der Studie, die in Teil IV auf Seite 77 diskutiert wird, verwendet.

### **8.2.1 Layout des LSA-Fragebogens**

Mit der Textauszeichnungssprache HTML und der Programmiersprache Javascript lassen sich alle aus papiergebundenen Befragungen bekannten Antwortfelder auch in webbasierten Fragebögen realisieren. Darüber hinaus sind mit der Programmiersprache Java weitere Eingabeobjekte denkbar, wie beispielsweise ein Schieberegler, bei dem der Benutzer durch Mausbewegungen anstelle einer Tastatureingabe einen Wert einstellen kann.

Der LernSystemAnalyse-Fragebogen enthält vier Eingabeobjekte (siehe Abb. 9 auf der nächsten Seite). Zwei Schieberegler sind für die Einstellung einer Punktzahl zur Gewichtung des Qualitätskriteriums und zur Bewertung des Lernsystems im Hinblick auf den nachgefragten Qualitätsaspekt vorgesehen. Eine hohe Punktzahl bei der Gewichtung bedeutet, dass man die Fragestellung für entsprechend bedeutsam hält. Hingegen steht eine hohe Punktzahl für eine Einschätzung des Bewerter, dass er das Qualitätskriterium für entsprechend gut im Lernsystem umgesetzt sieht. Es stehen 5-Punkt-Skalen für eine differenzierte Einstufung der Gewichtung und Bewertung zur Verfügung. Falls ein Experte die Fragestellung für nicht bewertbar hält, kann er dies mit Options-Schaltflächen erklären. Das letzte und wichtigste Eingabeobjekt ist das Begründungstextfeld. Hier sollen sämtliche Begründungen, Anregungen, Kritik etc. zur Bewertung notiert werden. Da die zu erwartende Art und der Inhalt der Hinweise im Vorfeld nicht vorausgesehen werden können, ist eine standardisierte anstelle einer freien Antwortmöglichkeit nicht zielführend. Die frei formulierten Begründungstexte sollen nicht nur die von den Projektmitgliedern vermuteten Defizite und Vorzüge des Lernsystems bestätigen, sondern darüber hinaus unbekanntes Verbesserungspotenzial aufdecken.

Qualitätskriterium →

Leitfragen →

Hauptfrage →

Bewertbarkeitsoption,  
Gewichtungs- und  
Bewertungsskala →

Begründung →

**Mensch**

1. Welche benötigten **Funktionalitäten** bzw. Möglichkeiten enthält das Lernsystem für welche Zielgruppe (nicht)? Welche aktuellen Chancen eröffnet das Lernsystem für die Zielgruppen?
2. An welcher Stelle im Lernsystem erkennen Sie, dass der erforderliche **Aufwand für Ihr Arbeits-** bzw. **Lernergebnis** (nicht) angemessen ist?
3. Inwieweit passt das Lernsystem, insbesondere die Erzähl- und Darstellungsform zu Ihrer **Arbeits-** bzw. **Lerngewohnheit**?
4. Sehen Sie eine (**Feedback-**) Möglichkeit, um auf die Entwicklung bzw. Verbesserung des Lernsystems z.B. durch Meinungs- und Erfahrungsäußerung einzuwirken?

Informationen zu den Leitfragen...

**Gewichtung und Bewertung der Aufgabenangemessenheit des Lernsystems:**  
**In wieweit halten Sie persönlich das Lernsystem für aufgabenangemessen?**

**Bewertbarkeit**

Gegeben  Nicht gegeben

**Gewichtung** **Bewertung**

1 2 3 4 5 1 2 3 4 5

**Begründung**

Einfach und schnell einstellbar.  
Leicht zu verstehen.  
Komplexe Darstellungen möglich.  
Sicher gut in verschiedenen Bereichen einsetzbar.

25% Seite 1/4

Abb. 9: Fragebogen: Qualitätskriterium Mensch

Weniger motivationsfördernd ist es, wenn die Experten viel Text am Bildschirm lesen müssen. Allzu umfangreiche Texte, insbesondere seitenlange Ausfüll- und Teilnahmehinweise zu Beginn einer Befragung, sind zu vermeiden, da im allgemeinen von einem flüchtigen Leseverhalten am Bildschirm ausgegangen werden kann (Will 1997, S.209, Naumann *et al.* 1999, S.3f.). Daher sind eine mehrkanalige Informationsaufnahme, Spracherkennung und -synthese zu überdenken, um möglichst viele Informationen zu transferieren. Allerdings sind hier die Grenzen der Informationsaufnahmefähigkeit des Menschen und die der zumutbaren Lade- und Rechenzeit, aufgrund des datenintensiven Materials, zu überprüfen. Derzeit bietet das Open-Source Projekt Sphinx<sup>24</sup> unter Sourceforge.net technische Vorstöße zur Spracherkennung, -analyse und -synthese. Es existiert hierzu noch keine portable webbasierte Lösung. Insbesondere die Spracherkennung ist aufgrund der derzeit begrenzten Bandbreite des Internets und der im Durchschnitt zur Verfügung stehenden Rechnerleistung technisch noch utopisch. Es ist wichtig, dass die Werkzeuge zur Befragung plattformunabhängig und webbasiert arbeiten, da die teilnehmenden Experten in der Praxis unterschiedliche Hardwareausstattungen besitzen und einen zeit- und ortsunabhängigen Zugriff benötigen.

Ein häufiges Argument für die Unterstützung mündlicher Sprache ist, dass computerunerfahrene Personen bei Tastatureingaben sehr langsam seien. Diese sind aber nicht unbedingt die Nutzer von Lernsystemen.

Gegen die Nutzung mündlicher Sprache sprechen die in der Praxis mangelnden Möglichkeiten, das Gesprochene zu korrigieren oder zu erweitern. Zudem ist die Auswertung von gesprochenem Text aufwendiger als bei geschriebenem Text.

Der Fragebogen enthält neben den Hauptfragestellungen eine Reihe zusätzlicher Informati-

<sup>24</sup><http://cmusphinx.sourceforge.net/sphinx4/>

onstexte und Leitfragen (siehe Abb. 9 auf der vorherigen Seite und Abb. 10). Diese sollen den Experten bei seiner Bewertungsaufgabe unterstützen.

[▶Beschreibung](#)
[▶Erläuterung](#)
[▶Anleitung](#)
[▶Fundierung](#)
[Zurück](#)

## Mensch

Wirksame Lernsysteme können nicht unabhängig davon konzipiert und bewertet werden, wer sie wie und wofür benutzt. In der Praxis werden dennoch viele Lernsysteme an der Zielgruppe vorbei konstruiert. Dies ist kein böser Wille, sondern viel mehr eine logische Konsequenz, denn Lernsysteme werden zu weiten Teilen auf der Code-Ebene entworfen. Dabei gerät der Mensch aus dem Blickfeld.

Bei einer benutzergerechten Gestaltung steht die Mensch-Lernsystem-Interaktion im Mittelpunkt. Für eine solche Gestaltung ist es wichtig, den Benutzerkreis genau zu definieren, denn ein Lernsystem zur Unterstützung der Lehre richtet sich in Regel nicht an die breite Öffentlichkeit. Durch eine Zielgruppen- bzw. Aufgabenanalyse gilt es, die unterschiedlichen Anwenderbedürfnisse (Ziele, Anforderungen, Aufgaben usw.) und -merkmale (Rollen, Nutzerklassen, Vorwissen, Einstellungen usw.) zu identifizieren und insbesondere für die Gestaltung der Benutzerschnittstelle angemessen zu berücksichtigen. In der Literatur findet man für dieses Vorgehen die Begriffe **User-Centered Design**, **Benutzerzentrierung** und **Benutzerpartizipation**.

Die Lernerorientierung wird oft als maßgeblich für den Erfolg eines Lernsystems ausgewiesen. Dabei wird häufig vergessen, dass es neben **Lernenden** noch die Anwendergruppe der **Lehrenden** und der **Administratoren/Entwickler** gibt. Haben die Entwickler des Lernsystems eine dieser Zielgruppen nicht angemessen in ihrem Design berücksichtigt, wird es auf Akzeptanzprobleme stoßen. Die rollenabhängige Aufgabenbewältigung in den drei Nutzerklassen sollte vom Lernsystem ermöglicht und angemessen unterstützt werden. Angemessen heißt, dass Werkzeuge, Dialoge und Bedienabläufe so auf die Aufgabe zugeschnitten sein sollen, dass das Arbeitsziel möglichst gut erreicht werden kann. Hierbei sollte der Zeitaufwand, die mentale Anstrengung und die Anzahl der benötigten Arbeitsschritte gering sein.

Das Qualitätskriterium "Mensch" fragt nach Ihrer persönlichen Bewertung der Aufgabenangemessenheit des Lernsystems.

Abb. 10: Beschreibung des Qualitätskriteriums Mensch

Es wurde beim Design des Fragebogens darauf geachtet, dass die zusätzlichen Metainformationen nur bei Bedarf abgerufen werden können, um einen Informationsüberfluss zu vermeiden. Der Abruf erfolgt, indem der Bewerter auf im Kontext stehende Hyperlinks klickt. Diese Hypertext-Lösung der Informationsanordnung auf mehreren Ebenen priorisiert gleichzeitig die Inhalte. Die Inhalte, die unbedingt für die Bewertungsaufgabe gelesen werden sollten, befinden sich direkt auf der Seite. Zusatzinformationen erfordern eine Interaktion durch Überfahren eines Begriffs mit der Maus, um beispielsweise eine Begriffsdefinition anzuzeigen oder durch einen Mausklick, der zu kontextbezogenen Metainformationen (Beschreibung, Erläuterung, Anleitung, Fundierung) führt.

Das nachfolgende Unterkapitel erklärt die Bedeutung der Metainformationen für den Bewertungsprozess.

### 8.2.2 Metainformationen

Der Einfluss der Erfahrung bei Befragungen wurde in vielen Studien belegt (Nielsen 1992, S.373ff., Desurvivre 1994, S.173ff., Karat *et al.* 1992, S.397ff.).

Personen	aufgetretene Probleme	potenzielle Probleme
Experten:	44%	31%
Software-Entwickler:	16%	24%
Nicht-Experten:	8%	3%

Tab. 5: Vorhersagerate für Endbenutzer-Probleme (Desurvivre 1994, S.185)

*Aufgetretene Probleme* sind diejenigen Probleme, die bei der Benutzung durch den Evaluanden der Software auftraten.

*Potenzielle Probleme* sind vorhergesehene Probleme, die auftreten könnten.

Die jeweilige Prozentzahl gibt an, welcher Anteil von der Gesamtanzahl bekannter Probleme gefunden wurde.

Der in der Tab. 5 erkennbare Unterschied zwischen Experten und Nicht-Experten überrascht nicht. Experten sind Individuen, und sie verfügen somit über einen „unterschiedlichen theoretisch-empirischen Wissensstand und Erfahrungshintergrund sowie über implizite Theorien bezüglich der Bedeutsamkeit von Softwaremerkmalen und Bedingungen des Softwareeinsatzes“ (Tergan 2000, S.138). Dies beeinflusst die Interpretation einer Fragestellung und das individuelle Urteilsvermögen.

Der Interpretationsspielraum einer Fragestellung kann sowohl als Stärke wie auch als Schwäche gesehen werden. Einerseits ermöglicht genau dieser Interpretationsspielraum ein breites Anwendungsfeld der Methode, sowie die Entwicklung von individuellen Problemlisten auf deren Basis Handlungsempfehlungen zur Verbesserung des Lernsystems abgeleitet werden können. Andererseits bedeutet dieser Interpretationsspielraum, dass die befragten Personen individuelle, nicht nachvollziehbare und ungenannte Bewertungskriterien zur Einschätzung des Lernsystems verwenden. Die Bewerterübereinstimmung und damit die Durchführungsobjektivität wäre stark eingeschränkt. Für den Vergleich von Bewertungen ist es notwendig, dass die Experten dieselben Qualitätskriterien anwenden und diese einheitlich verstehen.

Wenn die Erfahrung von Personen sich auf die Effektivität der Befragung auswirkt, dann sind Maßnahmen zu überlegen, die insbesondere Personen mit weniger Bewertungserfahrung unterstützen. Dem Problem der mangelnden Bewerterübereinstimmung versucht man häufig durch Schulung der Evaluatoren und standardisierte Antwortmöglichkeiten zu begegnen. Beide Maßnahmen führen nach den Studien von Prichard *et al.* (1993, S.16ff.) nicht ausschlaggebend zu einer Verbesserung der Bewerterübereinstimmung. Maßgeblich für die geringe Bewerterübereinstimmung sind nach Prichard *et al.* unterschiedliche Auffassungen der Experten, wann ein Qualitätskriterium als erfüllt gelten kann.

Unter Metainformationen werden alle diejenigen Informationen des Fragebogens verstanden, die den Experten bei seiner Bewertungsaufgabe unterstützen. Dies können im Kontext von Qualitätskriterien stehende Leitfragen, Beschreibungen, Anleitungen, Erläuterungen und Literaturquellen sein. Ein Ziel dieser Metainformationen ist, den unerfahrenen Bewerter so zu unterstützen, dass der qualitative Unterschied seiner Hinweise verglichen mit einem Experten, der mehr Bewertungserfahrung hat, geringer wird.

These 2: Nach dem Studium der Metainformationen wird das Verständnis von Hauptfragestellungen gefördert, und der Bewerter kann fundiertere und reichhaltigere Hinweise zur Lernsystem-Qualität geben.

Über Metainformationen soll der Experte in den jeweiligen Qualitätsaspekt eingeführt und inspiriert werden, mehrdimensional zu denken. Metainformationen sind nicht nur für die Bewerter wichtig, sondern auch zur Auswahl von Qualitätskriterien für die Messung der Anforderungserfüllung. Die in einer Bewertung zu berücksichtigenden Qualitätskriterien werden daher ausführlich beschrieben. Zu den im Text verwendeten Fachbegriffen sind Definitionen abrufbar. Die Hauptfragestellung des Qualitätskriteriums wird durch eine Reihe von zusätzlichen detaillierteren Leitfragen ergänzt. Auch zu jeder Leitfrage kann der Bewerter Erläuterungen erhalten.

Die Leitfragen haben analog zum sogenannten sokratischen Gespräch<sup>25</sup> den Sinn, nur durch Fragen – nicht durch Antworten – Anhaltspunkte und Denkanstöße zur Bewertung eines Qualitätskriteriums zu liefern. Je komplexer ein Sachverhalt ist, desto weniger lässt er sich mit einer einzigen Frage abdecken, daher gibt es das Leitfragen-Konzept. Die Leitfragen müssen vom Experten nicht explizit schriftlich beantwortet werden. Sie fungieren bei der Meinungsbildung des Experten als Checkliste, damit kein wichtiger Aspekt außer Acht gelassen wird. Die Leitfragen sollen zudem helfen, die von den Experten wahrgenommene Qualität zu kanalisieren und zu verbalisieren.

Daneben stehen Ressourcen in Form von Literaturquellen bereit. Diese dienen zum weiterführenden Wissenserwerb und zur Fundierung der Qualitätskriterien.

Bei der in dieser Arbeit zu findenden Studie wurden Personen befragt, die nicht ausgewiesene Lernsystem-Experten waren. Daraus ergab sich eine besondere Anforderung an die Gestaltung und Durchführung der Expertenbefragung. Die Teilnehmer hatten die Gelegenheit, sich über die Dauer der Befragung zu schulen, d.h. ihre Wahrnehmung für Qualitätsbereiche zu schärfen und ihre Fähigkeit zur Artikulierung ihrer Qualitätshinweise zu erweitern. Durch die Zuhilfenahme der Metainformationen sollen die Experten die Qualitätskriterien besser einordnen, abgrenzen und verstehen können. Letztlich kann dies zu einer schnelleren Bearbeitung führen. Dieser positive Einfluss wurde aus der wie folgt beschriebenen Untersuchung geschlossen.

---

<sup>25</sup>Griechischer Philosoph Sokrates 470–399 v.u.Z: Fragen sind wichtiger als die Antworten.

In der Untersuchung von Fuhrmann Verdaguer (2005, S.116f.) wurden mit jeweils 5 unterschiedlichen Teilnehmern zwei Fragebögen mit freien Antwortmöglichkeiten zu den Fragestellungen verglichen. Die Fragebögen unterschieden sich nicht durch die Fragestellungen, sondern durch das Vorhandensein von Metainformationen. Der erste Fragebogen ohne Metainformation enthielt lediglich die Fragen, während der zweite Fragebogen mit Metainformationen zusätzlich zu den Fragen Erläuterungs- und Beschreibungstexte beinhaltete. Anhand der Fragebögen sollte von den Teilnehmern ein Softwareprototyp bewertet werden.

Fragebogen	Zeit in Minuten			Wörteranzahl		
	$(Z_{Min})$	$(Z_{Max})$	$(\bar{Z})$	$(W_{Min})$	$(W_{Max})$	$(\bar{W})$
ohne Metainformation	3	10	6	26	77	51
mit Metainformation	8	12	10	87	129	101

Tab. 6: Metainformationen nach Fuhrmann Verdaguer (2005)

Z := Gesamtzeit für das Lesen der Hauptfrage und der Metainformationen, sowie das eventuelle Nachfragen bei Problemen, das Nachdenken und die Beantwortung;

W := Wortanzahl;  $_{Min}$  := Minimalwert;  $_{Max}$  := Maximalwert;  $\bar{\phantom{x}}$  := Arithmetischer Mittelwert.

Tab. 6 zeigt, dass obwohl die Experten für den Fragebogen mit Metainformationen mehr Zeit benötigten, sie dafür aber im Vergleich zu der anderen Gruppe überproportional mehr schrieben. Es ergaben sich beim Fragebogen mit Metainformation das Verhältnis 5,9 Sekunden pro geschriebenem Wort und beim Fragebogen ohne Metainformation das Verhältnis 7 Sekunden pro Wort. Zu beachten ist dabei, dass bei umfangreichen Befragungen und einer damit verbundenen größeren Wortanzahl dieser Unterschied der Verhältnisse zu einer deutlichen Differenz der Gesamtzeit zwischen den zwei Gruppen führen kann. Der t-Test (Signifikanzniveau  $\alpha = 0,05$ ) ergab, dass die Mittelwerte der benötigten Zeit signifikant verschieden ( $p = 0,03$ ) sind, ebenso sind es die Mittelwerte der geschriebenen Wörter ( $p = 0,002$ ). Die Beobachtung der Experten und die Anzahl der Rückfragen ließen zudem erkennen, dass weniger Verständnisschwierigkeiten auftraten.

### 8.2.3 Fragefolge

Bei der grundsätzlichen Gestaltung eines Fragebogens bieten sich zwei alternative Lösungen an. Zum einen können sämtliche Fragen auf einer Bildschirmseite platziert werden, d.h. bei vielen Fragen und Informationen muss der Benutzer den sichtbaren Bildschirmausschnitt verschieben (scrollen). Zum anderen kann für jede Frage eine Bildschirmseite vorgesehen werden. Die nachfolgenden Seiten können bei dieser Aufteilung über Navigationsobjekte aufgerufen werden. Die Vorteile dieser beiden Alternativen fasst Theobald (2000, S.83) in der Tab. 7 auf der nächsten Seite zusammen.

Fragebogen auf einer Seite	Eine Frage pro Seite
<ul style="list-style-type: none"> <li>● Fragebogen im Gesamtüberblick; höhere Transparenz; überschaubarer Fragenkontext;</li> <li>● Zeitbedarf für den Probanden anhand der Länge abschätzbar; kein Gefühl der „Endlosbefragung“;</li> <li>● Datenübertragung nur zu Beginn und am Ende der Befragung; dadurch insgesamt meist geringere Wartezeiten;</li> <li>● Weniger anspruchsvolle Programmierung;</li> <li>● Ähnlichkeit mit konventionellem Papierfragebogen.</li> </ul>	<ul style="list-style-type: none"> <li>● Kein Scrollen notwendig; geringerer Eingabeaufwand;</li> <li>● Kürzere Ladezeit für die erste Frage;</li> <li>● Eingabeabhängige Verzweigungen und Datenübergabe innerhalb des Fragebogens handhabbar;</li> <li>● Isolierte Darstellung jeder einzelnen Frage;</li> <li>● Nachträgliche Änderungen durch Rücksprünge kontrollierbar;</li> <li>● Abbrecherverhalten erkennbar.</li> </ul>

Tab. 7: Vorteile verschiedener Fragefolgen nach Theobald (2000)

Für den in dieser Arbeit vorgestellten Fragebogen bot sich die Gestaltung „Eine Seite pro Frage“ an, denn die Platzierung von Metainformationen erfordert den gesamten verfügbaren Bildschirmplatz. Die abgegebene Antwort zum angezeigten Qualitätskriterium wird beim Verlassen der Seiten gespeichert. Somit sind bei einem Sitzungsabbruch des Experten oder vorübergehendem Internet-, Strom- und Computerausfall dessen Antworten gesichert. Der Teilnehmer kann somit zu einem späteren Zeitpunkt die Bewertung fortsetzen. Die bisherigen Antworten können durch Vor- und Zurückblättern zwischen den Qualitätskriterien jederzeit vom Bewerter verändert werden.

Der Fortschritt der Beantwortung des Fragebogens wird prozentual angezeigt (siehe Abb. 9 auf Seite 39). Da der Experte nicht den gesamten Fragebogen vor Augen hat, sondern nur immer ein Qualitätskriterium pro Bildschirmseite, stellt die Fortschrittsleiste einen Anhaltspunkt für seine Zeitplanung. Der Experte kann somit abschätzen, wieviel Zeit er ungefähr noch für die Bewertung der nachfolgenden Qualitätskriterien benötigt.

#### 8.2.4 Vortest des LSA-Fragebogens

Der LernSystemAnalyse-Fragebogen wurde in vier Probeläufen getestet. 6 Experten hatten die Aufgabe eine Software zu bewerten. Drei Experten traten als Bewerterteam an, d.h. sie bewerteten gemeinsam die Software. Die drei übrigen Experten wurden jeweils einzeln befragt. Bei der Software handelte es sich um ein Lernsystem namens „Wohnraumheizung“. Diese wurde für das Mensch-Maschine-Systeme-Praktikum des Fachgebiets Systemtechnik und Mensch-Maschine-Systeme der Universität Kassel entwickelt. Die Teilnehmer des Praktikums haben die Aufgabe, die Benutzerschnittstelle zur einer Wohnraumheizungssimulation durch die Entwicklung von aufgabenangemessenen grafischen Anzeigen zu vervollständigen.

Für die Befragung wurde ein Notebook benutzt, an dessen Grafikkarte zwei Monitore angeschlossen werden konnten. Der eine Bildschirm stand für die Anzeige des Fragebogens und die Eingaben des/der Experten zur Verfügung. Durch die synchrone Bildübertragung auf den zweiten Monitor konnte der Versuchsleiter die Interaktionen des Experten versteckt beobachten.

Auf einem zweiten PC wurde das Lernsystem Wohnraumheizung für die Inspektion des/der Experten gestartet.

Ein primäres Ziel der Vortest war, die Benutzerschnittstelle des entworfenen webbasierten Fragebogens zu testen. Dieses Ziel wurde den Experten gegenüber geheim gehalten. Die Experten wurden gebeten, die Qualitätskriterien zu gewichten, deren Erfüllung durch das Lernsystem zu bewerten und die Bewertung zu begründen.

Wichtige Nutzungsinformationen konnten über das Log-File-Protokoll bestimmter angeklickter Bildschirmobjekte gewonnen werden. Die Verweildauer auf einer Seite gab beispielsweise Aufschluss über die Bearbeitungszeit für einzelne Qualitätskriterien. Zusätzlich wurde jedes Anklicken auf den Schaltflächen des Fragebogens protokolliert. Die Auswertung der Logfiles zeigte entgegen der Erwartung, dass während einer Befragung ein Experte in einer Einzelbefragung im arithmetischen Mittel bei insgesamt vier Qualitätskriterien zweimal über eine Schaltfläche Metainformationen abrufen (siehe Tab. 8 auf Seite 47). Die Nutzung der Metainformationen wurde im UKap. 15.9 auf Seite 125 im Rahmen der Studie (siehe Teil IV) mit mehreren Personen noch einmal untersucht.

Eine Konsequenz aus der Durchführung der Vortests war, dass anstelle einer 10-Punkt-Skala zur Gewichtung und Bewertung im Fragebogen eine 5-Punkt-Skala implementiert wurde. Aus der Beobachtung der Mausbewegung der Teilnehmer wurde befunden, dass der Entscheidungsprozess der Experten für einen Wert innerhalb der 10-Punkt-Skala mehr Zeit in Anspruch nimmt. Zudem wünschten sich in der Nachbefragung zwei der drei Experten eine einstellbare mittlere Punktzahl auf der Skala.

Die Experten waren sehr gut an die Maussteuerung gewöhnt, während die Tastatureingaben von drei Experten als mühsam empfunden wurden. Die Gefahr, dass eine vorgegebene Skala zur Bewertung eines Qualitätskriteriums von verschiedenen Bewertern unterschiedlich interpretiert wird, kann mit Hilfe von Schiebereglern mit zwei Extremwerten verringert werden (Götte und Kümmerlein 1996, S.37). Daher wurde die Tastatureingabe von Punktzahlen durch entsprechende Schieberegler ersetzt, die mit der Maus zu steuern sind.

Nachdem die 10-Punkt-Skala bei einer späteren Befragung durch eine 5-Punkt-Skala ersetzt wurde, gab es bei der Nachbefragung zur Studie (siehe Teil IV) keine negativen Rückmeldungen zu den Schiebereglern.

Der Fragebogenentwurf sah vor, dass mehrere Fragen zu einem Qualitätskriterium zusammen durch eine Punktzahl gewichtet und bewertet werden sollten. Die Bildung von Mittelwerten zur Gewichtung und Bewertung über verschiedene Fragestellungen zu einem

Qualitätskriterium fiel drei Bewertern schwer. Dies zeigte sich in der Nachbefragung der Experten. Daher wurde zu jedem Qualitätskriterium eine die Einzelaspekte eines Qualitätskriteriums umfassende Hauptfrage formuliert. Die detaillierteren Fragestellungen führten zum Leitfragen-Konzept (siehe UKap. 8.2.2 auf Seite 41).

Die Idee, dem Experten seine ähnlich bewerteten und gewichteten Qualitätskriterien zur Erleichterung der Bewertungsaufgabe anzuzeigen, wurde verworfen zugunsten einer Ergebnismeldung mit Mittelwert und Streuung der bereits bewerteten Qualitätskriterien. Auf dieser Informationsseite ist auch eine Übersicht über alle Qualitätskriterien zu finden. Die Online-Statistik wurde allerdings entgegen der eigenen Einschätzung von allen Experten in der Hauptbefragung nicht genutzt.

Die Bewerber sollten die Möglichkeit haben, ihre getätigten Eingaben zu überprüfen und diese gegebenenfalls zu korrigieren. Dies wurde durch Navigationsobjekte ermöglicht, über die der Benutzer im Fragebogen vor- und zurück springen kann. Die Beobachtung der Experten ergab, dass die freie Navigationsmöglichkeit von den Experten sehr häufig genutzt wird.

Weiterhin wurde untersucht, ob die geschlossene und die freie Antwortmöglichkeit genutzt werden. Die drei Teilnehmer wurden bei der Vorbereitung auf die Befragung darauf hingewiesen, dass es für die Auswertung wichtig sei, dass die Bewertung auch begründet wird. Im Vortest wurde die freie Antwortmöglichkeit entgegen der Erwartung zur Begründung der Bewertung bei allen vier Qualitätskriterien genutzt (siehe Tab. 8 auf der nächsten Seite). Ähnlich wie bei der Methode des lauten Denkens formulierten die Experten ihre Überlegungen, Probleme, Handlungsalternativen etc.. Diese gaben dem Qualitätsregelteam Aufschluss darüber, an welchen Stellen des Lernsystems Verbesserungspotenzial zu finden ist.

Die Begründungstexte der Experten wurden dahingehend analysiert, ob sie die wenig aufschlussreichen Antworten wie „Ja“, „Nein“ oder „... ist (nicht) vorhanden“ enthielten. Die diese Antworten provozierenden Fragestellungen wurden daraufhin überarbeitet.

Zuletzt wurden auch die Ergebnisse der Gruppen- und Einzelbefragung verglichen. Dabei fiel das Ergebnis der Gruppe von drei Experten (X Y Z) deutlich schwächer aus als das aus den Einzelbewertungen der Experten A, B und C. Die Auswertung der Begründungstexte der Expertengruppe führte zu signifikant ( $p = 0,0024$ ) weniger Hinweisklassen im Vergleich zu der Anzahl der Hinweisklassen, die durch die Einzelbefragung der drei Personen A, B und C erhoben wurden. In Tab. 8 auf der nächsten Seite sind die erhobenen Hinweisklassen durch ein Slash-Zeichen separiert für die vier Qualitätskriterien eingetragen. Der Begriff Hinweisklassen steht für die vom Qualitätsregelteam zusammengefassten Themen, die für die Qualitätssicherung von Bedeutung sind. Dieselben Qualitätskriterien Mensch, Lernumfeld, Lerntechnologie und Lerninhalte/Lehrmethodik kamen bei allen vier Befragungsszenarien zur Anwendung. Die benötigte Zeit für die Begutachtung eines Softwareprototypen und dessen Bewertung betrug bei jeder Befragung eine Stunde.

Anzahl	(X Y Z)	A	B	C	p (t-test)
Hinweisklassen pro Qualitätskriterium	1/1/1/1	1/1/3/3	2/3/4/1	2/1/3/2	0,0024
Abfragen von Metainformationen	2/1/1/1	1/1/0/0	1/0/0/0	1/1/0/1	0,0468

Tab. 8: Gruppenbefragung vs. Einzelbefragung

Die Beobachtung der Gruppe ließ erkennen, dass sich die Gruppenmitglieder eher bei der Beantwortung der Fragestellungen behinderten, anstatt sich zu ergänzen. Die Gruppe brauchte mehr Zeit, um ihre Meinung den anderen zu erklären und um sich auf eine Aussage oder Formulierung zu einigen. Vermutlich hätte eine Person aus der Gruppe die Moderatorrolle übernehmen sollen. Gruppendynamische Prozesse sind schwer im Vorfeld vorherzusehen, daher wurden für die Studie (Teil IV) nur Einzelpersonen zugelassen.

### 8.3 Lernsystem-Konfiguration

Qualitätssicherung von Lernsystemen ist kein zielgruppen- oder lernumfeld- oder lerntechnik- oder didaktikorientiertes Verfahren, sondern es ist die Summe aus all diesen Aspekten. Daher ist die Bewertung von Lernsystemen eine anspruchsvolle interdisziplinäre Aufgabe. Zudem bemerkt Thaller (2000, S.65): „Software ist in ihrer Natur nach immateriell, und das erschwert ihre Beurteilung“. Er fügt allerdings sogleich einen möglichen Lösungsansatz hinzu: „Aber dennoch kann man diesem Produkt Kennzahlen zuordnen, kann diese weiche Ware durch Metriken in den Griff bekommen. Wir können dabei etwa an Fehler und Restfehler denken, ganz klare Indikatoren für die Qualität von Software. Wir können allerdings auch fragen, ob wir ein Benutzerhandbuch durch eine Kennzahl in den Griff bekommen“.

Die Qualität eines Lernsystems ergibt sich erst im Zusammenspiel einer Vielzahl von Aspekten. Es stellt sich die Frage: welche genau?

Als Kennzahlen für den Lernerfolg an Universitäten dienen häufig die Notengebung, Studienzahlen, Absolventen, Studienzeiten, Abbrecherquote, Betreuung von Studierenden, die interne Organisation der Lehre und die Ressourcen (Webler *et al.* 1993, S.37). Auf dieser Betrachtungsebene fällt es schwer, den Einfluss von Lernsystemen auf diese Kennzahlen nachzuweisen, da weitere Einflussgrößen nicht als konstant angenommen werden können. Für die Qualitätssicherung von Lernsystemen sind diese Kennzahlen völlig ungeeignet.

Nicht zielführend ist es daher, die objektive Qualität durch einmal definierte empirische Qualitätskriterien kontextunabhängig bestimmen zu wollen. Angemessener ist es, durch qualitätskriterienbasierte und kontextabhängige Befragung von Experten die subjektive Qualität vielschichtig und multiperspektivisch zu erheben, um anschließend auf die objektive Qualität zu schließen.

Eine systematische Bewertungsgrundlage für eine möglichst umfassende Wahrnehmung und Bewertung der Qualität eines Lernsystems stellt die „Lernsystem-Konfiguration“ dar. Die Lernsystem-Konfiguration strukturiert und ordnet Qualitätskriterien nach multiplen Qualitätsperspektiven für formative Kontext- und Produktevaluationen. Sie schafft eine Rahmenbedingung durch differenzierte Fragestellungen, die die Bewertung der Qualität des Lernsystems aus unterschiedlichen Blickrichtungen fundiert und ausgewogen fördert.

In dieser Arbeit werden Qualitätskriterien nicht als pragmatische Checkliste für notwendige und normative didaktische oder mediale Eigenschaften eines Lernsystems realisiert. Sie werden als Fragenkatalog verstanden, durch den Experten interviewt werden können. Der Katalog enthält neben der mit bestimmten Qualitätsaspekten verbundenen Hauptfragestellung umfangreiches Wissen, welches für die Beantwortung und Auswertung der Fragen nützlich ist.

Die Lernsystem-Konfiguration ist ein Ebenensystem, welches Qualitätskriterien hierarchisch ordnet. Die Ordnung in dieser Hierarchie erfolgt nach dem Detaillierungsgrad des repräsentierten Qualitätsaspekts und dem Verwandtschaftsgrad zu anderen Qualitätsaspekten. Nicht unerwähnt bleiben sollte, dass die Lernsystem-Konfiguration eine Ordnungsmöglichkeit der Qualitätskriterien darstellt. Begründet wird diese durch den Grad der inhaltlichen Gemeinsamkeiten der Qualitätskriterien.

Qualitätskriterien, die anderen Qualitätskriterien untergeordnet sind, werden Subkriterien genannt. Die oberste Ebene der Lernsystem-Konfiguration spannt vier Qualitätskriterienbereiche auf. Die Bewertung der Qualität erfolgt im Rahmen von durchgeführten Expertenbefragungen

- mit Sicht auf die am pädagogischen Prozess beteiligten **Menschen**,
- gemäß der Passung des Lernsystems zu den Bedingungen der **Lernumgebung**,
- unter Beachtung der Benutzbarkeit der implementierten **Lerntechnologie**,
- anhand der **Lerninhalte** und der **Lehrmethode**, die Lernprozesse initiieren und unterstützen sollen.

Die Qualitätskriterien entstanden nach der Sichtung von Einzelkriterien aus mehreren verfügbaren Kriterienkatalogen (siehe UKap. 6.4 auf Seite 17) und dem Abgleich mit dem Wissen aus den bislang durchgeführten eigenen Software-Projekten. Es erfolgte die Aussonderung von doppelten Kriterien, die Zusammenführung von Synonymen, die teilweise erforderliche Übersetzung, die Aufbereitung der Daten, die Erstellung einer systematischen, hierarchischen Liste mit Vernetzung der Kriterien durch Verweise und das Hinzufügen von Metainformationen (siehe UKap. 8.2.2 auf Seite 41). Auf den Qualitätskriterienkatalog wird in UKap. 9.4 auf Seite 52 eingegangen.

## 9 Die Software LernSystemAnalyse

Zur Unterstützung der Handhabung, Verankerung und Umsetzung des Qualitätsverbesserungsprozesses (siehe Kap. 7 auf Seite 24) wurde die LernSystemAnalyse-Software (LSA) implementiert. Beispielsweise wird der Benutzer bei der Pflege der Qualitätskriterien über ein Content-Management-System unterstützt. Weiterhin stehen Funktionen für die automatische Erstellung von webbasierten Fragebögen für formative Evaluationen bis hin zur computergestützten Auswertung zu Verfügung.

Die LernSystemAnalyse-Software ist ein sinnvolles Werkzeug zur Bewältigung von Aufgaben zur Qualitätssicherung eines Lernsystems. Die unterstützten Aufgaben betreffen insbesondere die Planung, Durchführung und Auswertung von Expertenbefragungen. Durch den Einsatz einer Software für diese Qualitätsaktivitäten werden unerwünschte Medienbrüche vermieden.

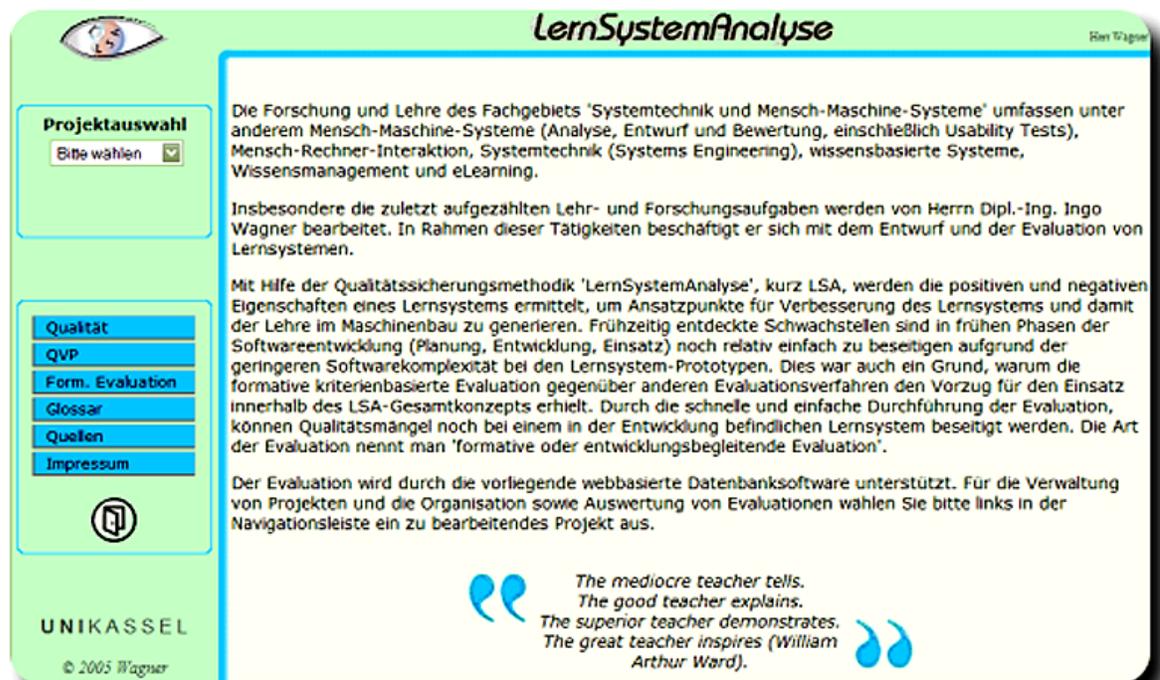


Abb. 11: GUI der LernSystemAnalyse

Bei der Implementierung wurden die Programmiersprachen PHP, Java (JSP, Servlets), JavaScript, HTML, XML, SQL, R und Ada verwendet. Die Applikation benötigt einen Apache und einen Tomcat Webserver, sowie eine MySQL-Datenbank. Die Implementierung erfolgte unter der Programmierumgebung Eclipse.

In den folgenden Unterkapiteln werden die wesentlichen Programmbereiche beschrieben, die innerhalb der Software verfügbar sind.

## 9.1 Projektverwaltung

In der Projektverwaltung der LernSystemAnalyse-Software können verschiedene Lernsystem-Projekte angelegt werden. Jedes Projekt kann durch Informationen wie beispielsweise Titel, Kurzbezeichnung, Kurzbeschreibung, Ziele, Status des Projekts, Interaktionsgrad (siehe UKap. 15.3.2 auf Seite 99), Lernsystem-Klasse (siehe Kap. 11 auf Seite 65) uvm. charakterisiert werden.

In diesem Programmbereich wird auch die Anforderungserfüllung in Prozent angezeigt, sobald ausgewertete Daten aus einer Evaluation zur Verfügung stehen. Der Anforderungserfüllungsgrad gilt nach Gleichung (1) auf Seite 7 als subjektive Qualität des Lernsystems. Diese Kennzahl ermöglicht den Vergleich von Lernsystemen und die gezielte Verbesserung von Lernsystemen.

Die Projektverwaltung ist skalierbar, sodass bei Bedarf von weiteren Funktionalitäten beispielsweise zur Projektzeit- und Aufgabenplanung frei verfügbare Projektmanagementsoftware wie z.B. PHProject<sup>26</sup> integriert werden kann.

Jedem Projekt können verschiedene Anforderungen, Personen, Evaluationen zugeordnet werden. Für diesen Vorgang gibt es eigene Programmbereiche, die in den folgenden Unterkapiteln beschrieben werden.

## 9.2 Anforderungen

Bevor ein Lernsystem implementiert wird, sind dessen Anforderungen zu spezifizieren. Für diesen Vorgang stellt die LernSystemAnalyse-Software Funktionen bereit.

Mit der LernSystemAnalyse-Software können Anforderungen jederzeit erstellt, aktualisiert und auch für andere Projekte wiederverwendet werden.

Bei der Formulierung einer Anforderung hilft das Konzept der Anforderungsschablone nach Rupp (2002, S.229). Durch die Orientierung an diese Anforderungsschablone wird eine einheitliche, verständliche und konsistente Gestalt von Anforderungen gewährleistet. Eine Anforderungsschablone ist ein Bauplan für die syntaktische Struktur eines Satzes. Eine Beispiel-Anforderung könnte nach diesem Schema wie folgt lauten: Das Lernsystem muss dem Lernenden die Möglichkeit bieten, personalisierte Lern- und Arbeitsergebnisse sichern und wiederverwenden zu können.

---

<sup>26</sup><http://www.phprojekt.com>

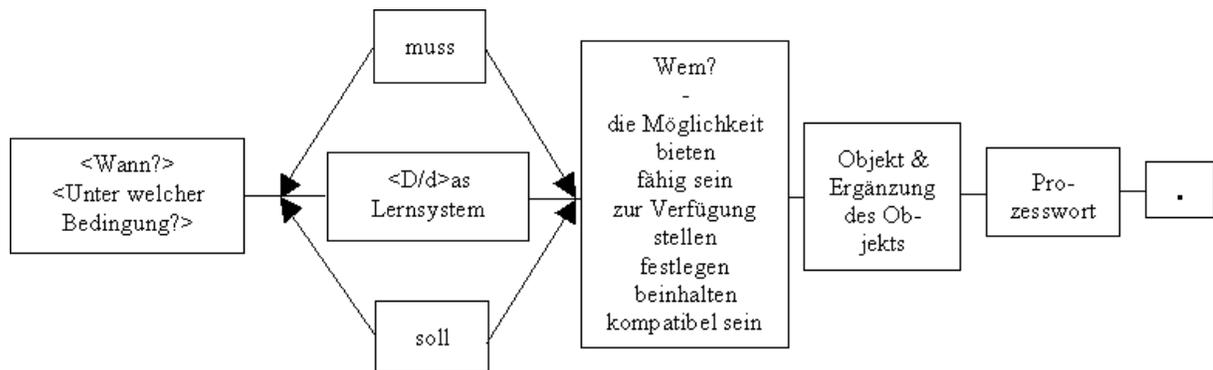


Abb. 12: Anforderungsschablone nach Rupp (2002, S.229)

Die Aktualisierbarkeit von Anforderungen ist nötig, da es in der Praxis schwierig ist, Lernsysteme von morgen durch Anforderungen von heute zu beschreiben. Hinzu kommt, dass Anforderungen oft dynamisch während der Entwicklung entstehen oder es sich herausstellt, dass bestehende Anforderungen zu ungenau formuliert sind.

Die Wiederverwendung von Anforderungen macht Sinn, da die Formulierung einer operationalen und wohl gestalteten Anforderung Zeit kostet.

Jeder Anforderung können Qualitätskriterien zugewiesen werden. Die Beziehungen der Anforderungen zu den Qualitätskriterien sind maßgebend für die Bildung der Qualitätsindikatoren (siehe UKap. 15.1 auf Seite 91). Qualitätsindikatoren helfen bei der Beurteilung des Projektfortschritts und beim Vergleich von Prototypen.

Mit Hilfe der Anforderungserfüllung als zentrales Element der Qualitätsmessung kann letztlich auf die Qualität des Lernsystem-Prototypen geschlossen werden. Entscheidungen für die (Nicht-) Verbesserung eines Merkmals oder einer Eigenschaft des Lernsystems können jederzeit nachvollzogen werden. Zu jeder Anforderung kann das Qualitätsregelteam Handlungsableitungen zur Verbesserung des Prototypen formulieren.

### 9.3 Personenverwaltung

Die Personenverwaltung hilft, die Projektmitglieder zu organisieren. Projektmitglieder können Personen des Managements, Qualitätsregel- und Entwicklerteams sein, sowie Experten, die an den Befragungen teilnehmen.

Wichtige Einstellmöglichkeiten in diesem Programmbereich betreffen die Zugangsdaten. Diese ermöglichen es, Rechte zur Benutzung der LernSystemAnalyse-Software zu verteilen und adaptive Benutzerschnittstellen bereit zu stellen.

Für die Fragestellungen in Unterkapitel 13.2 auf Seite 82 wurden demographische Angaben der Experten wie Alter, Geschlecht etc. festgehalten.

## 9.4 Qualitätskriterienkatalog

Die Qualitätskriterien werden in einer Datenbank gespeichert. Durch die Datenbankbasierung können die Qualitätskriterien leicht auch in anderen Softwarearchitekturen wiederverwendet werden.

Ein weiterer Vorteil stellt die Webbasierung dar. Hiermit kann von jedem zugelassenen internetfähigen PC auf die Datenbank zugegriffen werden. Die Datenbank ist auf Mehrbenutzerbetrieb ausgelegt. Daher können zu jedem Qualitätskriterium der Autor, das Erstellungsdatum, sowie eventuelle Überarbeiter und deren Änderungen gespeichert werden.

Die Qualitätskriterien können unterschiedlichen Lernsystem-Klassen zugeordnet werden (siehe Kap. 11 auf Seite 65). Auf diese Weise können Qualitätskriterien für bestimmte Lernsystem-Klassen heraus gefiltert werden. Dies erleichtert die Bearbeitung und Auswahl der Qualitätskriterien für eine bestimmte Lernsystem-Evaluation. Derzeitig enthält die Datenbank ausschließlich Qualitätskriterien, die für alle Lernsystem-Klassen gültig sind.

Qualitätskriterien können neben der Filterung nach Lernsystem-Klassen auch über die Schlüsselwörter, Kurzbezeichnungen oder über die Hierarchie gesucht werden.

Die Qualitätskriterien werden durch Hierarchie-Ebenen geordnet. Die erste Ebene wird als Lernsystem-Konfiguration (siehe UKap. 8.3 auf Seite 47) bezeichnet. Darunter können weitere Qualitätssubkriterien angeordnet werden. Die zweite Ebene entspricht den Qualitätskriterien, die antwortabhängig bei einer Expertenbefragung eingefügt werden.

Jedes Qualitätskriterium kann für eine Evaluation ausgewählt werden. Dazu wird das entsprechende Qualitätskriterium interaktiv markiert.

Die Qualitätskriterien selbst werden durch beschreibende Texte erklärt. Für die Bewertungsaufgabe ist zu jedem Qualitätskriterium eine Hauptfrage hinterlegt. Diese soll der Experte in Bezug auf das Lernsystem gewichten und bewerten. Für den Entscheidungsprozess stehen Leitfragen zur Verfügung. Zu jeder Leitfrage kann der Benutzer Informationstexte zur Erläuterung der Fragestellung abrufen.

Letztlich sind zu jedem Qualitätskriterium externe Quellen (Literatur, Websites, Personen etc.) zitiert. Die Quellen unterstreichen die Diskussion, Bedeutung und Herkunft der Qualitätskriterien.

Durch das Angebot dieser Metainformationen kann der Qualitätskriterienkatalog selbst auch als Lernsystem angesehen werden. Alle Qualitätskriterien einschließlich der Metainformationen können über ein Autorensystem bearbeitet werden.

Nachfolgend werden die Qualitätskriterien der ersten beiden Hierarchie-Ebenen kurz erläutert. Bei Qualitätskriterien der dritten Hierarchie-Ebene werden nachfolgend nur die Hauptfragen genannt. Der kursive Text entspricht der Hauptfrage eines Qualitätskriteriums.

## Qualitätskriterien der Lernsystem-Konfiguration

**Mensch** *Inwieweit halten Sie persönlich das Lernsystem für aufgabenangemessen?*

Ein Lernsystem ist dann aufgabenangemessen, wenn es den Benutzer durch seine Funktionalitäten und Möglichkeiten so unterstützt, dass er seine Aufgabe schneller, einfacher und besser bewältigen kann (in Anlehnung an DIN EN ISO 9241-10). Dies beinhaltet auch die Eigenschaft des Lernsystems, sich an dem mentalen Modell des Bewerter bezüglichen seiner persönlichen Arbeits- bzw. Lerngewohnheit zu orientieren und ihn bei seinen Aufgaben kompetent zu unterstützen.

**Lernumfeld** *Inwieweit liegt eine Passung des Lernsystems zum gegebenen Lernumfeld vor?*

Bei dieser Frage wird ein besonderes Augenmerk auf die äußeren Umfeldbedingungen des Lernens mit dem Lernsystem gelegt, insbesondere die Integrations- und Einbindungsmöglichkeiten des Lernsystems im jeweiligen Lehrkontext.

**Lerntechnologie** *Wie beurteilen Sie die technische Umsetzung und die Benutzbarkeit (Usability) des Lernsystems?*

Die Lerntechnologie ist das Medium über das die Vermittlung von Lerninhalten erfolgt. Es soll bewertet werden, welche technischen Eigenschaften des Lernsystems für die Benutzer (nicht) geeignet sind.

**Lerninhalte/Lehrmethode** *Wie bewerten Sie die Lerninhalte innerhalb des Lernsystems und die mit dem Lernsystem verbundene Lehrmethode?*

Der Begriff Lerninhalte steht für das über das Lernsystem zu erwerbende Wissen bzw. die anzueignenden Kompetenzen. Lerninhalte sind das Produkt von Lernprozessen, während die Lehrinhalte Ziele der Lehrprozesse sind. Lehrmethoden als didaktische Lehrform oder Lehrverfahren unterstützen die kognitive Konstruktion der Lerninhalte vom Lernenden. Die Fragestellung bezieht sich auf die Güte der Lerninhaltsvermittlung, d.h. die Qualität der Lerninhalte und der Lehrmethoden, die das Lernsystem repräsentiert.

## Qualitätssubkriterien zu „Mensch“

Wirksame Lernsysteme können nicht unabhängig davon konzipiert und bewertet werden, wer sie wie wofür benutzt und in welcher Umgebung sie eingesetzt werden. In der Praxis werden dennoch viele Lernsysteme an der Zielgruppe vorbei konstruiert. Dies ist kein böser Wille, sondern viel mehr eine logische Konsequenz, denn Lernsysteme werden zu weiten Teilen auf der Code-Ebene entworfen. Dabei besteht die Gefahr, dass der Mensch aus dem Blickfeld gerät.

Die Lernerorientierung wird oft als maßgeblich für den Erfolg eines Lernsystems ausgewiesen. Dabei wird häufig vergessen, dass es neben Lernenden noch die Anwendergruppe

der Lehrenden und der Administratoren/Entwickler gibt. Die rollenabhängige Aufgabenbewältigung in den drei Nutzerklassen sollte vom Lernsystem ermöglicht und angemessen unterstützt werden. Angemessen heißt, dass Werkzeuge, Dialoge und Bedienabläufe so auf die Aufgabe zugeschnitten sein sollten, dass das Arbeitsziel möglichst effizient erreicht werden kann. Hierbei sollten der Zeitaufwand, die mentale Anstrengung und die Anzahl der benötigten Arbeitsschritte möglichst gering sein.

**Administratoren und Entwickler** *Inwieweit ist das Lernsystem für Administratoren und Entwickler aufgabenangemessen?*

Administratoren sind für die Installation des Lernsystems und die anfallenden Wartungsarbeiten zuständig. Entwickler sind hingegen zuständig für das Design, die Implementierung, die Dokumentation und die Qualitätssicherung. Aufgabenangemessen bedeutet für diese Benutzer, dass das Lernsystem installierbar, anpassbar, portabel, wartbar und wiederverwertbar, gesetzes- und standardkonform und effizient (weiter) zu entwickeln ist.

**Lehrende** *Inwieweit halten Sie das Lernsystem für Lehrende aufgabenangemessen?*

Lehrende sind für die Lehre mit dem Lernsystem in den verschiedenen Bildungsinstitutionen verantwortlich. Das schließt Lehrende an Schulen, Universitäten, in der unternehmensinternen Weiterbildung oder der Erwachsenenbildung ein. Für den Lehrenden ist ein Lernsystem dann aufgabenangemessen, wenn es dessen Lehre effizient unterstützt.

**Lernende** *Inwiefern halten Sie das Lernsystem für Lernende aufgabenangemessen?*

Lernende erzielen bei gleichen Lernbedingungen oft unterschiedliche Lernerfolge. Dies ist vor allem auf Differenzen im Bereich der Vorkenntnisse, der Motivation und der grundsätzlichen intellektuellen Fähigkeiten der Lernenden zurückzuführen. Das Lernsystem soll trotz unterschiedlicher Persönlichkeitsmerkmale und individueller Präferenzen einen aufgabenangemessenen Rahmen bieten, in dem beim Lernenden Lernprozesse initiiert werden.

**Qualitätssubkriterien zum „Lernumfeld“**

Lernsystemen wird häufig vorgeworfen, sie würden zu einer Vereinsamung und zum Verlust der Kommunikationsfähigkeit der Lernenden beitragen. Dieser Vorbehalt ist sicherlich nicht bei allen Lernsystemen begründet. Das Lernumfeld von Lernsystemen sollte Kommunikations- und Kooperationsmöglichkeiten sowie eine angemessene persönliche Betreuung der Lernenden bieten.

**Kommunikation und Kooperation** *Wie bewerten Sie die Nutzung der Kommunikations- und Kooperationsmöglichkeiten im Lernsystem?*

Bei dieser Fragestellung geht es darum zu bewerten, ob das Lernsystem neben individuellem Lernen auch kollaborative Formen des Lernens ermöglicht und den Erwerb bzw. die Anwendung kommunikativer und kooperativer Kompetenz fördert.

**Betreuung** *Ist über das Lernsystem eine angemessene persönliche Betreuung des Lernenden durch den Lehrenden möglich?*

Eine angemessene persönliche Betreuung ist dann gewährleistet, wenn während des gesamten Lernprozesses eine Lernende-Lehrende Interaktion persönlich aufgebaut werden kann, bei der der Lernende beraten und angeleitet wird.

### **Qualitätssubkriterien zur „Lerntechnologie“**

Bei erfolgreichen Lernsystemen wird die Lerntechnologie nicht nur dazu verwendet, Informationen elektronisch zu speichern und zu verbreiten. Vor allem Interaktivität, Individualität und Adaptivität verwandeln passive Lernsysteme in aktive (intelligente) Lernsysteme. Viele Anwendungen nutzen diese Möglichkeit nur bedingt.

**Screen Design** *Wie beurteilen Sie das Screen Design des Lernsystems?*

Wenige Sekunden entscheiden über den Erfolg eines Lernsystems. Findet sich der Lernende zurecht, interessiert ihn die Informationsdarbietung, gefällt ihm das Screen Design, so bleibt er und kommt gerne wieder, um mit dem Lernsystem zu lernen.

### **Qualitätssubsubkriterien zum „Screen Design“**

**Bildschirmaufbau** *Wie beurteilen Sie den Bildschirmaufbau des Lernsystems?*

**Farben** *Sind Farben und Farbzusammenstellungen sinnvoll und verständlich eingesetzt?*

**Bilder/Metapher** *Sind die Bilder/Metapher nach Form, Inhalt und Häufigkeit zur Erklärung und Verdeutlichung von Lerninhalten, zur Orientierung, zur Navigation, zur Motivation der Lernenden geeignet?*

**Animation/Video** *Wie beurteilen Sie die Animationen und Videos im Lernsystem?*

**Audio** *Werden die auditiven Elemente als angenehm, verständlich, sinnvoll, motivierend und hilfreich empfunden?*

**Text** *Wie beurteilen Sie die Texte im Lernsystem?*

**Listen/Tabellen** *Sind die Listen/Tabellen gut lesbar und logisch strukturiert?*

**Hilfe** *Wie beurteilen Sie die Anwendbarkeit und den Umfang der angebotenen Hilfe?*

Über eine angemessene Hilfe erhält der Benutzer Antworten auf Fragen, die auf die Bedienung, die Struktur und den Einsatz des Lernsystems abzielen. Es gilt zu bewerten, ob Umfang und Inhalt der Hilfe vollständig, ausreichend, korrekt, aktuell und zielführend sind.

### **Qualitätssubsubkriterien zu „Hilfe“**

**Online-Hilfe** *Ist die geleistete Online-Hilfe adäquat, inhaltlich richtig, verständlich, anschaulich und nützlich?*

**Tutorial** *Wie beurteilen Sie die Hilfe durch das angebotene Tutorial?*

**Begleitmaterialien** *Ist das Begleitmaterial für Lernende und Lehrende differenziert, verständlich, übersichtlich, motivierend und hilfreich gestaltet?*

**Technische Handbücher** *Ist es mit dem Handbuch leicht eine Lösung für Ihre technischen Aufgaben/Probleme zu finden?*

**Benutzeranleitung** *Ist die Benutzeranleitung für die Bedienung und Benutzung des Lernsystems angemessen, d.h. ausführlich, vollständig, detailliert, genau, korrekt, anschaulich, einheitlich, eindeutig und verständlich?*

**Demo** *Werden die wichtigsten Eigenschaften und Funktionalitäten des Lernsystems ausreichend im Demo vorgestellt?*

Unter einem Demo (Demonstrationsprogramm) versteht man die selbstablaufende und selbsterklärende Vorführung eines Programms.

**Guided Tour** *Wie bewerten Sie die Verfügbarkeit einer Guided Tour?*

Eine Guided Tour führt den Lernenden auf einem didaktisch geeignetem Lernpfad durch die Lerninhalte des Lernsystems.

**Impressum** *Finden Sie im Impressum des Lernsystems, die von Ihnen erwarteten Kontaktinformationen zu den Anbietern des Lernsystems?*

**Verfügbarkeit** *Wie beurteilen Sie die Verfügbarkeit des Lernsystems?*

Die Verfügbarkeit des Lernsystems ist angemessen, wenn das Lernsystem ubiquitär einsetzbar/auffindbar ist, Lerninhalte exportiert und gedruckt werden können und das Lernsystem insgesamt schnell, zuverlässig, robust, stabil und kontrolliert arbeitet.

**Autorensysteme** *Wie beurteilen Sie die Verwendung von Autorensystemen im Lernsystem?*

Unter einem Autorensystem wird hier Software verstanden, die eine einfach zu handhabende Erstellung, Anpassung und Änderung von Lernsystemen ermöglicht.

**Sicherheit** *Wie schätzen Sie die Sicherheit des Lernsystems ein?*

Ein Lernsystem ist geschützt, wenn ein Benutzer das Lernsystem nicht unerwünscht manipulieren kann, seine Benutzerdaten nicht ungewollt für Dritte zugänglich sind und wenn Daten, insbesondere Beiträge, Lern- und Arbeitsergebnisse, gesichert werden.

**Eingabe-/Kommunikationsgeräte** *Wie gut sind die zur Verfügung stehenden Eingabe- und Kommunikationsgeräte zur Bedienung des Lernsystems geeignet?*

Die verwendeten Eingabe- und Kommunikationsgeräte unterstützen den Anwender bei der Interaktion mit dem Lernsystem. Es gilt zu bewerten, ob die angebotenen Eingabe- und Kommunikationsgeräte selbsterklärend, leicht installierbar und effizient benutzbar sind oder andere Geräte bevorzugt werden.

**Navigation und Steuerungsinteraktion** *Ist die Navigation zur effektiven und effizienten Steuerung des Lernsystems geeignet?*

Eine aufgabenangemessene Navigation dient dazu, den Benutzer schnell und einfach durch das Lernsystem zu führen und ihn nicht im Unklaren zu lassen über den aktuellen „Standort“ und den Zustand des Lernsystems. Eine schlechte Navigation führt hingegen zur Desorientierung. Ein zu kontrollierendes und zu steuerndes Lernsystem benötigt geeignete Bedienungs- und Orientierungselemente.

**Dialog** *Sind die Dialoge zwischen Mensch und Computer einfach, übersichtlich, logisch, sinnvoll, komfortabel und effizient zu führen?*

Ein Dialog zwischen dem Benutzer und dem Lernsystem entsteht dann, wenn sich der Benutzer in eine Kommunikationssituation versetzt fühlt.

**Individualisierung** *Wie bewerten Sie die Individualisierbarkeit (Customizing) des Lernsystems durch den Anwender?*

Als individualisierbar wird ein Lernsystem dann bezeichnet, wenn es an die jeweiligen Bedürfnisse, Vorstellungen und spezifischen Vorlieben des Anwenders angepasst werden kann. Diese Anpassung kann vom Anwender selbst vorgenommen werden, beispielsweise durch die Wahlmöglichkeit verschiedener Darstellungsformen, Zeichengrößen, Zoomfaktoren etc..

**Adaptivität** *Wie bewerten Sie die im Lernsystem befindlichen adaptiven Elemente in Bezug auf ihre Förderung des Lehr- und Lernprozesses?*

Ein Lernsystem ist adaptiv, wenn es den Unterstützungsbedarf der Benutzer diagnostiziert und das Ergebnis der Diagnose in geeignete (individuell abgestimmte) Lehraktivitäten und passende Informationsdarbietung umsetzt.

### **Qualitätssubkriterien zu „Lerninhalte/Lehrmethode“**

Die Qualität der Lerninhalte und der Lehrmethodik wird ähnlich wie bei Buchverlagen durch die Güte der Zusammenarbeit und der Kommunikation, zwischen den Trägern dieses Wissens

und den an der Implementierung und Veröffentlichung Beteiligten bestimmt. Die modernen Informations- und Kommunikationstechnologien vereinfachen die Inhaltserstellung und deren Veröffentlichung. Dies ist jedoch ein Vor- und Nachteil zugleich, denn jeder kann (ungefragt) zum Autor werden und seine Inhalte publizieren. Es findet häufig keine Vorauswahl oder kein Reviewprozess von Dritten statt. Als Konsequenz existieren Lernsysteme, die zwar zumeist technisch überzeugen, aber deren Lerninhalte bzw. die Lehrmethode oft als unüberdacht und unausgereift auffallen.

**Lernsystemklasse** *Wie bewerten Sie die Wahl der Lernsystem-Klasse für die Lerninhalte und die Lehrmethodik?*

Unter Lernsystem-Klasse (Bildungsdomäne) wird eine Typisierung von Lernsystemen verstanden. Die Zuordnung des Lernsystems zu einer bestimmten Lernsystem-Klasse ermöglicht die unterschiedliche Ansprache von Zielgruppen, die Einbettung bzw. Integration des Lernsystems in verschiedene Kontexte, die Verwendung bestimmter Informations- und Kommunikationstechnologien und die Umsetzung von speziellen Lerninhalten und Lehrmethoden.

**Informationsgehalt** *Wie bewerten Sie den Informationsgehalt des Lernsystems?*

Unter Informationsgehalt wird die Verständlichkeit, die Anschaulichkeit, die Lebendigkeit, der angemessene Umfang, die Aktualität und die Gültigkeit der Lerninhalte verstanden.

**Lehr-/Lernziele** *Wie bewerten Sie den Nutzen, die Benennung, die Erreichbarkeit der Lehr- und Lernziele des Lernsystems?*

Lehrziele sind diejenigen Ziele des Lehrenden, die er durch seine mediendidaktisch-organisatorische Gestaltung seiner Lehre zu erreichen versucht. Lernziele resultieren aus den individuell unterschiedlichen Interessen und Bedürfnissen der Lernenden. Lehr- und Lernziele sollten unter realistischen Annahmen für die Zielgruppe mit dem Lernsystem erreichbar, überprüfbar und von Nutzen sein.

**Lehrmethode** *Wie bewerten Sie die Lehrmethode des Lernsystems?*

Die Aufnahmefähigkeit der Menschen pro Zeiteinheit ist begrenzt. Eine Hauptaufgabe der Lehrmethode ist es, effiziente Lernprozesse zu initiieren und so zu unterstützen, dass der Lernende Lerninhalte kognitiv effektiv konstruieren und behalten kann.

**Vorwissen/Medienkompetenz** *Wie beurteilen Sie das Lernen und Lehren mit dem Lernsystem im Hinblick auf die Anknüpfung an das Vorwissen und an die Medienkompetenz der Anwender?*

Besonders hohe Anforderungen stellt das Informationszeitalter an die Fach-, Sozial- und Medienkompetenz der Menschen. Es gilt zu bewerten, ob das benötigte Vorwissen der Anwender zum effizienten Umgang mit dem Lernsystem und zum Verständnis der Lerninhalte ausreicht.

**Motivation/Aktivität** *Wie bewerten Sie den Beitrag des Lernsystems in Bezug auf die lernförderliche Motivation und Aktivität der Lernenden?*

Die Aktivierung und Motivation der Lernenden erfolgt in vielen Lernsystemen durch interaktive und multimediale Elemente. Durch diese Elemente, zeigen die Lernenden erfahrungsgemäß mehr Interesse an den Lerninhalten und sie beteiligen sich angeregter.

**Emotion** *Wird der Lernende durch das Lernsystem emotional angeregt, z.B. durch Dramaturgie, Humor etc.?*

Neurologen haben nachgewiesen, dass Emotionen wesentlich an der Aufnahme, Speicherung und Wiedergabe von Informationen beteiligt sind. Emotionen energetisieren, initiieren, verbinden und bewerten Gedanken und Ideen. Emotionen sind daher gezielt für die Vermittlung von Lerninhalten zu nutzen.

**Lernkontrollen** *Erhält der Lernende die Möglichkeit, seinen Lernfortschritt zu überprüfen?*

Gerade beim selbstgesteuerten Lernen, wo der Lehrende nicht immer anwesend ist, sollte für den Lernenden eine Möglichkeit bestehen, seinen Lernerfolg zu überprüfen. Über implementierte Lernkontrollen im Lernsystem kann ein Lernender Rückmeldung über seinen Lernfortschritt erhalten. Lernkontrollen sind häufig durch Tests zu Lerninhalten realisiert.

Die vorgestellte Qualitätssicherungsmethodik steht und fällt mit der Gültigkeit und Aktualität der verwendeten Qualitätskriterien. Die Qualitätskriterien sollten laufend mit den Vor-Ort-Erfahrungen, insbesondere den Lessons Learned und der Best Practise der Experten unter Berücksichtigung von ändernden Rahmenbedingungen aktualisiert werden. Fachwissen kann veralten, Personal kann wechseln, die Projekte können immer größer werden, die am Projekt beteiligten Personen werden immer mehr und die Produkte können verzweigter werden. Es sei ausdrücklich darauf hingewiesen, dass Qualitätskriterien in einer Welt von Änderungs- und Neuerungsprozessen nicht ohne ständige Überarbeitung überdauern können.

Neue Qualitätskriterien sollten so verfasst werden, dass sie möglichst allgemein in späteren Lernsystem-Projekten ebenfalls anwendbar sind. Für die Wartung, Anpassung und Erweiterung des Qualitätskriterienkatalogs wurden besondere Autorensysteme implementiert.

Der Qualitätskriterienkatalog umfasst derzeit 42 Qualitätskriterien. Ein Teil der Qualitätskriterien ist sicherlich auch in anderen Wissensgebieten anwendbar, jedoch sollte der Qualitätskriterienpool von Experten der neuen Wissensdomäne gegebenenfalls überarbeitet werden. In diesem Zusammenhang kann man den Zugang zum Qualitätskriterienpool für Autoren öffnen und zu einer Wiki-Plattform ähnlich wie die bekannte Wikipedia-Plattform<sup>27</sup> umbauen. Der Name Wiki leitet sich vom hawaiianischen „wiki wiki“ ab, was so viel wie

---

<sup>27</sup><http://de.wikipedia.org/wiki/Hauptseite>

„schnell“ bedeutet. Schnelligkeit der Dokumentenerstellung in Netzwerken war auch das Ziel des Namensgebers und Erfinders Cunningham (Cunningham und Leuf 2001). Auf diese Weise kann eine Qualitätsdiskussion entstehen, die zur Formulierung von fachübergreifenden Qualitätskriterien führt.

## 9.5 Evaluation

Die Funktionen aus dem Programmbereich Evaluation helfen, unterschiedliche Evaluationen zu verwalten, webbasierte Fragebögen für die Bewertung zu generieren, mit Hilfe von integrierten Statistikprogrammen auszuwerten und die Ergebnisse grafisch aufzubereiten. Der generierte interaktive Fragebogen enthält die zuvor ausgewählten Qualitätskriterien, vereinbarte Zugangsvoraussetzungen, eine Datenbankanbindung und eine besondere Interviewtechnik.

Da es bei der Durchführung der Evaluation nicht sinnvoll ist, den Bewerter mit zu vielen (überflüssigen) Fragen zu belasten, wurde bei der Durchführung ein adaptiver Fragenkatalog implementiert. Bürli (1975, S.115ff.) hatte erstmals im Jahr 1975 das „Branched testing“ vorgestellt. In seinem Test war die Bereitstellung der nachfolgenden Fragen von den Antworten der vorangegangenen Fragen abhängig. Es ist erstaunlich, dass sich derartig interaktive Fragebögen kaum durchgesetzt haben. Der Grund mag an den fehlenden Autorensystemen liegen, die es Nicht-Informatikern ermöglichen, Interaktivität in eigenen Fragebögen zu installieren.

Mit dem Programmbereich Evaluation der LernSystemAnalyse-Software steht nun ein solches Werkzeug zur Verfügung. Es kann mit wenigen Mausclicks ein Fragebogen mit selbstbestimmtem Inhalt zusammengestellt und automatisch für das Web generiert werden. Dieser enthält dann die vom Initiator der Befragung markierten und in eine gewollte Reihenfolge gebrachten Qualitätskriterien.

Die Interviewtechnik erlaubt ein „Nachfragen“ zur Laufzeit der Befragung durch sofortige Analyse der Überschreitung von Schwellenwerten zur Bewertung und Gewichtung eines Qualitätskriteriums aus der Lernsystem-Konfiguration. Der adaptive Fragebogen soll idealerweise das Gefühl vermitteln, dass es der Befragte mit einem interessierten und geschulten Interviewer zu tun hat, der die Benutzerantworten zur Laufzeit der Kommunikation korrekt auswertet, sinnvoll interpretiert und kontextbezogen reagiert. Wird ein Schwellenwert überschritten, werden dem Experten noch während der Befragung zusätzlich zu evaluierende kontextbezogene Qualitätskriterien angezeigt. Wenn eine verwandte Frage zuvor als unbedeutend oder mit hoher Punktzahl bewertet wurde, schließt die Interviewtechnik, dass für diesen Qualitätsaspekt die zu erwartende Quantität und Qualität der Antworten gering sein würde. Daher wird auf die Stellung von detaillierteren Fragen bei diesen Konstellationen im weiteren Verlauf zu Gunsten der Befragungsdauer verzichtet.

Die Befragungsoption mit den zusätzlichen Qualitätskriterien wirkt sich auf die Befragungsdauer und die zu veranschlagende Auswertungszeit aus (siehe UKap. 15.2 auf Seite 95). Daher ist die Interviewtechnik je nach benötigtem Umfang der Befragung individuell für jede Evaluation einstellbar. Damit sich die Befragung nicht endlos gestaltet, ist die Auswertung der Antworten einstufig, d.h. zu den antwortabhängigen Qualitätskriterien werden keine weiteren Fragen generiert.

Die Auswertung der mit dem Fragebogen erhobenen Daten ist ebenso leicht wie die Erstellung eines Fragebogens. Das integrierte Statistikprogramm R<sup>28</sup> und einige spezielle Eigenentwicklungen zur statistischen Analyse erlauben eine sofortige Auswertung der Daten und Präsentation der Ergebnisse. Für die Inhaltsanalyse, bei der keine automatische Auswertung erfolgen kann, werden spezielle aufgabenbezogene Datenanzeigen und Werkzeuge angeboten. Die Datenbank kann auch anderen Programmen außer der LernSystemAnalyse-Software den Zugriff und die Bearbeitung der Daten gewähren.

---

<sup>28</sup><http://www.r-project.org>

---

## Teil III

# Lernsysteme

Lernsysteme eröffnen die Möglichkeit, Wissen auf verschiedene Weise neben dem Frontalunterricht zu vermitteln. Im Lernalltag gibt es allerdings viele schlechte Lernsystem-Beispiele, so dass in vielen Fällen die „Alten“ Medien mehr als die „Neuen“ Medien überzeugen. Lömker (1998, S.20) empfiehlt: „Multimedia soll nicht um seiner selbst willen betrieben werden. Ausschlaggebend für den letztendlichen Erfolg sind die Fragen der pädagogischen und wirtschaftlichen Eignung der gewählten Strategien“.

Der für viele Bildungsaufgaben bedeutenste Vorzug eines Lernsystems ist, wenn es orts- und zeitunabhängiges Lehren und Lernen ermöglicht. Mit diesem Vorzug ist jedoch eine Schwäche verbunden. Die durch Nutzung der heutigen Informations- und Kommunikationstechnologie gewonnene Flexibilität erstreckt sich zumeist auf die Anwender und weniger auf die Entwickler. Fakt ist, dass die Erstellung von Lernsystemen, sowie das Lehren und Lernen weitaus mehr als ein paar Mausclicks erfordern. Die Entwickler müssen zumeist sehr viel Programmieraufwand investieren, um die hochgesteckten Erwartungen der Zielgruppen zu erfüllen. Die Erstellung eines einstündigen Online-Lernsystems erfordert einen durchschnittlichen Programmieraufwand von 300 Stunden.

Für die Praxis werden einfach und kostengünstig anzuwendende Techniken benötigt, mit denen die produktive Erstellung und Anwendung komplexer Lernsysteme in den Griff zu bekommen sind. Kerres (2001, S.366) formuliert in diesem Zusammenhang: „Die in vielen Fällen bemängelte didaktische Qualität von CBT und WBT<sup>29</sup> geht auch heute einher mit der weiterhin (zu) geringen Produktivität der didaktischen Medienentwicklung: Der Aufwand ist weiterhin so hoch, dass Qualitätskriterien oft zurückgestellt werden und die Entwicklung konzeptuell anspruchsvollerer Bildungsmedien verhindert wird“.

In dieser Arbeit werden Lernsysteme vorgestellt, die dem Anspruch an eine möglichst schnelle, intelligente und kostengünstige Implementierung gerecht werden. Da die Entwicklung und Qualitätssicherung von Lernsystemen nicht in einem Ein-Mann-Projekt zu bewerkstelligen sind, wurde auf die Beteiligung von Studenten an der Softwareentwicklung zurückgegriffen. Um den Aufwand einer Entwicklung gegenüber dem möglichen Einsatzbereich in der Lehre zu rechtfertigen, wurde darauf geachtet, dass die Lernsysteme auch in anderen Lehrkontexten außerhalb des Fachgebiets angewendet werden können. Besonders hervorzuheben sind diejenigen Projekte, bei denen ein Autorensystem zur flexiblen Anpassung der Lerninhalte implementiert wurde. Schenkel (2000a, S.29) bemängelt in diesem Zusammenhang: „Noch immer bestehen Lernprogramme sehr häufig aus einzelnen Lernmodulen, die, in ihrer proprietären Autorensprache programmiert, fest in das Lernprogramm eingebunden sind. ...

---

<sup>29</sup>CBT: Computerbasiertes Trainingsprogramm; WBT: Webbasiertes Trainingsprogramm

Einmal programmiert, kann die Lernsequenz nur schwer aufgelöst, in eine andere Programmiersprache übersetzt oder in einem anderen Lernprogramm wieder verwendet werden“.

Autorensysteme, bei denen Inhalt und Architektur des Lernsystems getrennt von Lehrenden oder Lernenden – ohne Programmierkenntnisse zu besitzen – bearbeitet werden können, sind ein Lösungsansatz auf den im nächsten Kapitel stärker eingegangen werden soll.

## 10 Autorensysteme

Von dem Ziel, dass der Lehrende oder Lernende die Software selbst individuell anpassen und erweitern kann, sind die meisten Lernsysteme noch weit entfernt. Ein vielversprechender Ansatz ist das Angebot sogenannter Autorensysteme. Durch sie können die Entwickler gewährleisten, dass die Benutzer Einfluss auf das Lernsystem nehmen können. Nur durch interessante, aktuelle und aufgabenangemessene Lerninhalte und Lehrmethoden ist der Entwicklungsaufwand gegenüber traditionellen Lehrformen über die lange Nutzungszeit zu rechtfertigen.

Leider werden häufig relativ statische und unflexible Lernsysteme entwickelt. Auf diese Weise sind viele Lernsystem-Prototypen trotz anfänglicher Euphorie genauso so schnell wieder verschwunden, wie sie entstanden. „In der euphorischen Phase standen nicht Fragen der Qualität im Vordergrund, sondern vielmehr quantitative Dimensionen wie Geschwindigkeit oder die Zahl der potenziellen oder tatsächlichen Nutzer bestimmter Angebote“ (Dahinden *et al.* 2004, S.104). „Derzeit werden die vielfältigen Erwartungen an Lernsysteme wie flexibles Lernen, Zeitersparnis, individuelles Lernen, Kostenersparnis, effektiveres Lernen, Verbesserung der Lernqualität nur zum Teil erfüllt“ (Haben 2002, S.78). Das Gute an dieser Entwicklung ist, dass nun vermehrt über die Qualitätssicherung und die bildliche Trennung von Spreu und Weizen nachgedacht wird.

Erfolgreiche Softwaresysteme blicken zumeist auf eine lange Geschichte und aufeinander aufbauende Versionen zurück. Die Wiederverwendbarkeit ist ein wichtiger Aspekt bei der Entwicklung von Lernsystemen. Seit Beginn des Software-Engineering wird die Wiederverwendung von Programmteilen oder -modulen aus früheren Lernsystem-Projekten als eine wirtschaftliche Möglichkeit gesehen, hochwertige zuverlässige Lernsysteme zu entwickeln. Stärker als bisher sollte die Wiederverwendbarkeit von Lernsystemen und als Nebenbedingung die Zusammenarbeit und die Kommunikation zwischen den Universitäten intensiviert werden. Lömker (1998, S.21) empfiehlt: „Da das Gesamtbudget für Hochschulbildung *ceteris paribus* kaum wesentlich zu steigern ist, kann dieser Wettbewerb nur effizient sein, wenn ein gesamtsystemischer Effizienzgewinn erwirtschaftet werden kann. Dieser kann durch einen erhöhten Austausch von vorbereiteten Materialien bei gleichzeitiger Modularisierung der Lehre geschehen“. Der Grad der Wiederverwendung bestimmt wesentlich die Kosten und den Zeitbedarf der Entwicklung, aber auch die Qualität des Lernsystems. Wiederverwendete

Komponenten sind durch ihren häufigen Einsatz in unterschiedlichen Anwendungen besser getestet und damit zuverlässiger.

Leider ist es, wie in den Anfängen der Software-Entwicklung an Universitäten oft traurige Praxis, lieber ein Lernsystem von Grund auf neu zu konzipieren, anstatt vorhandene Lösungen und Konzepte zu berücksichtigen und weiter zu entwickeln. Dies ist vor allem darauf zurück zu führen, dass die Qualitätssicherung und insbesondere die Dokumentation der Projekte zu kurz kommen, sodass eine Neuentwicklung lohnender erscheint als eine Revision. Die Neuentwicklungsstrategie bringt zwar kurzfristig sehr viele Lernsysteme hervor, aber erfahrungsgemäß wenig wirklich gute, da immer wieder von vorne angefangen wird und sprichwörtlich das Rad neu erfunden wird.

Die universitäre Bildungslandschaft braucht nicht nur viele, sondern auch gute Lernsysteme. Der Open-Source Markt bietet für Programmierer ein facettenreiches Spektrum. Da der Programmcode der aufwendigste Bestandteil jedes Lernsystems ist, ist seine Wiederverwendung sehr naheliegend. Die in der Praxis geringe Wiederverwendung könnte aber auch gerade mit dem Fokus auf Programmcode zu tun haben. Dieser entsteht am Schluss im Prozess der Softwareerstellung und ist zumeist spezialisiert für genau eine Aufgabe. Eine unmittelbare Wiederverwendung ist oft nicht möglich, wenn die Entwickler mögliche weitere Einsatzgebiete außerhalb des Projekts nicht vorhergesehen und entsprechende Schnittstellen ausgelegt haben, die eine umfassende Adaption des Codes vermeiden. Der goldene Weg ist, von Anfang an bei der Entwicklung des Lernsystems die Wartbarkeit, Ausbaufähigkeit und Wiederverwendung der Software zu erleichtern.

Im Unterkapitel 12.5 auf Seite 75 wird ein Autorensystem für webbasierte Planspiele vorgestellt. Durch dieses Beispiel wird deutlich, welches Potenzial in Autorensystemen steckt. Autorensysteme fördern die Entwicklung und Pflege von Lernsystemen in effizienter Weise. Insbesondere die Entwicklungsdauer versucht man, durch den Einsatz von Autorensystemen zu kürzen. In der Regel wird dabei ein Programm über bereitgestellte Funktionen nach einem „Baukasten-“ oder WYSIWYG-Prinzip<sup>30</sup> zusammengestellt. Dieses Vorgehen zur Anwendungsentwicklung wird auch als Rapid Application Development (RAD) bezeichnet.

Auf dem Markt existieren eine Reihe von Programmen, die sich Autorensysteme nennen. Diese bieten im Grunde oft nur eine alternative Programmier- bzw. Entwicklungsoberfläche gegenüber herkömmlichen Programmierumgebungen an. Komplexe Anwendungen müssen mit ihnen ebenso aufwendig entwickelt werden, wie das mit Programmierumgebungen zu gängigen Hochsprachen der Fall ist. Im Hinblick auf den Zeitaufwand für das Erlernen der Entwicklungsumgebungen, die eingeschränkten Programmiermöglichkeiten und die Hersteller- bzw. Plattformabhängigkeit, sind solche Werkzeuge keine nutzbringenden Autorensysteme.

---

<sup>30</sup>What you see is what you get - Bei WYSIWYG wird ein Dokument während der Bearbeitung am Bildschirm genauso angezeigt, wie es bei der Ausgabe aussieht.

Mit Autorensystemen sollte eine davon abweichende Strategie verfolgt werden, die man auch in Anlehnung an Kerres (2001, S.381f.) als Verwirklichung einer „weak automation“ bezeichnen kann: „Damit sind Systeme gemeint, die nicht beanspruchen, didaktische Medien selbsttätig zu generieren, sondern sich darauf beschränken, Autoren bei der Entwicklung solcher Medien zu unterstützen“.

Autorensysteme werden oft als der goldene Weg für die Realisierung von folgenden Qualitätssicherungsaspekten wie

- Wiederverwendbarkeit,
- Anpassbarkeit, Erweiterbarkeit, Aktualisierbarkeit, Wartbarkeit,
- Nutzerorientierung und
- Effizienz genannt.

Doch nicht nur ökonomische Aspekte sprechen für die Verwendung von Autorensystemen. Eine den Lernprozess unterstützende Lehrmethode behandelt den Lernenden nicht bloß als einen Rezipienten von Informationen. Eine Untersuchung von Petermandl belegt, dass Themen, die selbständig erarbeitet wurden, „stärker internalisiert werden, als solche, die vorgegeben sind und nur reproduziert werden müssen“ (Petermandl 1991, S.269). Durch die einfache Verwendung von Autorensystemen können Lehrkonzepte verwirklicht werden, bei denen der Lernende aktiv sein Wissen anwenden kann, um eigene Lernsysteme zu bauen. Lernende können durch entsprechende Autorensysteme zugleich Konsumenten und Produzenten eines Lernsystems sein.

Lernsystem-Produzenten sollten eine gute Übersicht über unterschiedliche Klassen von Lernsystemen haben, um je nach Situation ein geeignetes Lernsystem zu entwickeln. Das nachfolgende Kapitel enthält eine Beschreibung verschiedener Lernsystem-Klassen.

## 11 Lernsystem-Klassifikation

Im Bildungsbereich wird eine Vielzahl von Computerprogrammen eingesetzt. Es stellt sich die Frage, welche Software als Lernsystem zu bezeichnen ist und durch welche Merkmale sich Lernsysteme voneinander unterscheiden lassen. Ist eine Website wie die Suchmaschine Google etwa ein Lernsystem? Ein Lernsystem ist eine Software, welche Lehr- und/oder Lernprozesse durch Informations- und Kommunikationstechnik unterstützt. Die Suchmaschine Google ist sicherlich nicht als Lernsystem zu bezeichnen, sehr wohl aber einige Webseiten, die Google bei Eingabe eines Schlagwortes auflistet.

Ein gutes Kriterium zur Abgrenzung von Lernsystemen zu anderer Software sind die Zielgruppen. Lernsysteme richten sich an Lehrende und/oder Lernende. Die Abb. 13 auf der nächsten Seite zeigt unterschiedliche Lernsystem-Klassen (Wagner 2005, S.2).

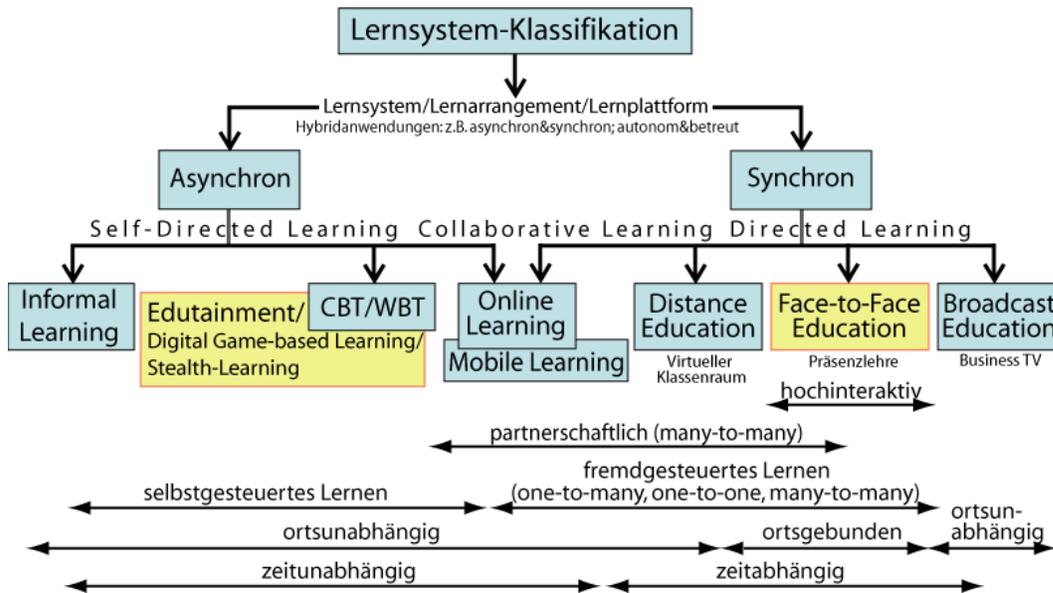


Abb. 13: Klassifikation von Lernsystemen

Die Lernsystem-Klassen sind häufig auch kombiniert anzutreffen, was durchaus aufgrund der Vorzüge und Schwächen jeder Lernsystem-Klasse für einen bestimmten Anwendungsbereich sinnvoll sein kann.

Viele der Begriffe aus Abb. 13 sind englischen Ursprungs, nicht zuletzt deswegen, weil Englisch die Informatikersprache ist. Zum Teil lassen sich in der Literatur auch deutsche Synonyme finden, diese sind allerdings zumeist weniger geläufig oder umstritten definiert.

**Informal Learning** bezeichnet das Lernen in nicht organisierten, unstrukturierten Lernprozessen. Bei dieser Art von Lernsystemen sind keine institutionell geprägten Lehrziele vordefiniert. Es sind auch keine Lernkontrollen und keine anerkannten Abschlüsse oder Zertifikate vorgesehen. Beispiele für solche Lernsysteme sind Informations- und Präsentationssysteme z.B. elektronische Bücher, Nachschlagewerke, Datenbanken, Websites, Dialog-, Assistenz-, Experten-, Diagnosesysteme und allgemein intellektuelle Werke. Faure *et al.* (1973, S.53) haben festgehalten, dass informelles Lernen etwa 70% allen menschlichen Lernens umfasst. Dohmen (1996, S.29) grenzt informelles Lernen von inzidentellem Lernen ab, das er als unbewußtes Gelegenheitslernen – oft ein Nebenprodukt anderer Tätigkeiten – kennzeichnet.

**CBT/WBT** Diese Abkürzungen stehen für Computer-Based- bzw. Web-Based-Training. Hierunter fallen Lernsysteme für programmierten Unterricht (Drill&Practise, Tutorials, hypermediale Lernsysteme) und Simulationen (Spiele und Mikrowelten). Bei der ersten Gruppe von Lernsystemen eignen sich die Lernenden mit Hilfe von Lehrtexten und Übungen selbstständig Kenntnisse und Fähigkeiten an. Bei der Simulation wird ein

reales (dynamisches) System nachgeahmt, um spielerisch und interaktiv Erkenntnisse über oder durch das modellierte System zu gewinnen.

**Edutainment** ist ein Kunstwort, welches sich aus den Begriffen „Education“ und „Entertainment“ zusammensetzt. Dieser Subklasse von CBT/WBT sind Lernsysteme zugeordnet, mit denen die Lernenden ihr Wissen spielerisch anwenden können. Beispiele für solche Lernsysteme sind Plan-, Rollen-, Strategiespiele und Quiz, die oft als sogenannte eGames (Lernspiele) bezeichnet werden. Edutainment orientiert sich an der Spiele- und Unterhaltungsbranche und schlägt dabei – oft versteckt (Stealth Learning) – eine Brücke zu den Lerninhalten bzw. zum Erwerb von Kompetenzen. Häufig umgesetzte Techniken sind interaktive Geschichtenerzählung, multimodale Benutzerschnittstellen und Simulation.

**Mobile Learning** Die jüngste Form des Lernens mit Neuen Medien stellen Mobilfunkdienste dar. Mittels Techniken wie WLAN, Handy, PDA, SMS, MMS<sup>31</sup> usw. kann ein Informationsaustausch zu Lehr- und Lernzwecken schnell, orts- und zeitunabhängig, aber immer noch mit hohen Kosten erfolgen. Der noch geringe Marktanteil von Mobile Learning wird voraussichtlich exponentiell steigen, wenn man bedenkt, dass im Durchschnitt fast jeder Mensch in Europa mindestens ein mobiles Gerät besitzt und benutzt.

**Online Learning** oder auch Open Distance Learning meinen das Lernen in Lern- und Informationsgemeinschaften (Communities of Practise, Learning/Online Communities) über speziell für Lernzwecke eingerichtete Foren, Wikis, Weblogs, Newsgroups, Chats, Instant-Messaging, Messageboards, E-Mail Workgroups etc.. Hier entscheidet nicht ein Lehrender über Lernzeit und -ort, sondern der Lernende. Die Lerninhalte entstehen größtenteils durch die aktiven Teilnehmer selbst.

**Distance Education** ist aus dem klassischen Fernunterricht entstanden. Statt Lehrmaterialien mit der Post zu verschicken, findet die Lehre hier online in einer Art virtuellem Klassenzimmer statt. Lehrende und Lernende sind dabei zumeist physisch voneinander getrennt. Über die zur Verfügung stehende Lerntechnologie haben die Teilnehmer die Möglichkeit, in einem angebotenen Kurs aktiv teilzunehmen. Beispielsweise können sie Fragen an Dozenten und Mitlernende stellen oder ihr Wissen synchron mit anderen teilen über Whiteboard, Shared Text, Application Sharing, Video- und Audio-Konferenz uvm.. Anerkannte Prüfungen, im Fall des Fernstudiums, finden in einer Präsenzveranstaltung (Face-to-Face) statt.

**Face-to-Face Education** ist eine traditionelle Präsenzlehrform, bei der verstärkt Neue Medien eingesetzt werden (Computer-enriched oder Computer-assisted Instruction). Die

---

<sup>31</sup>Wireless Local Area Network, Mobile phone, Personal Digital Assistant, Short Message Service, Multimedia Messaging Service.

Lernenden werden durch einen Dozenten, Lehrer oder Tutor betreut. Die Lehrorganisation kann dabei als Einzelunterricht (one-to-one) oder als Gruppenunterricht frontal (one-to-many) bzw. seminaristisch (many-to-many) ausgerichtet sein. Alle Teilnehmer befinden sich zu derselben Zeit am gleichen Ort (synchron). Lernen in einer Face-to-Face Lehrsituation kann hochinteraktiv sein, während bei allen anderen Lernsystem-Klassen Interaktivität technisch aufwendig unterstützt werden muss.

**Broadcast Education** Lernprozesse können über Medien wie Fernsehen (Teleteaching) und Radio angestoßen werden. Bei dieser Lehrform gibt es derzeit zumeist eindirektionale Beziehungen vom Lehrmedium zu den Lernenden, d.h. eine Möglichkeit zur bidirektionalen Kommunikation und Mitgestaltung des Lehrprozesses seitens der Lernenden besteht nicht. In Zukunft wird dieser Rückkanal mit fortschreitender Technik – beispielsweise durch interaktives digitales Fernsehen – sicherlich bald selbstverständlich.

## 12 Entwickelte Lernsysteme zur Systemtechnik

Im Fachgebiet Systemtechnik und Mensch-Maschine-Systeme<sup>32</sup> der Universität Kassel sind eine Reihe von Lernsystemen entstanden, die bis heute – dort wo es sinnvoll ist – erfolgreich in der Systemtechnik-Lehre eingesetzt werden.

Die in den nachfolgenden Unterkapiteln beschriebenen Lernsysteme sind mit aktiver Beteiligung von Studenten umgesetzt worden. Solche Entwicklungsprojekte bieten den Studenten die Chance zu einer intensiven Auseinandersetzung mit den Lerninhalten und Lehrmethoden. Die Vermittlung von Wissen über ein Lernsystem ist ein kognitiv bedeutender Schritt für das lückenlose Verstehen und den Transfer auf verschiedene praktische Lernanwendungen. Nebenbei wird der zukunftsweisende Erwerb von ausgefeilten Design- und Programmierkompetenzen gefördert.

Lernsystem	Kurzname	Lernsystem-Klasse	Autor(en)
Fuzzy-Logik CBT	FLCBT	CBT (Tutorial)	Mankowski
Fuzzy-Logik WBT	FLWBT	WBT (Tutorial)	Holbein
Fuzzy-Logik Kransteuerung	FLKran	WBT/Edutainment (Simulation/Autorensystem)	Bretz
Magnat-Schwebbahn	PS	WBT/Edutainment (Simulation/Planspiel)	Bretz/Wagner
Planspiel-Autorensystem	PAS	WBT/Edutainment (Simulation/Autorensystem)	Bretz/Wagner

Tab. 9: Lernsystem-Entwicklungen

<sup>32</sup><http://www.imat.maschinenbau.uni-kassel.de>

Mit den ersten drei Lernsystemen in der Tab. 9 auf der vorherigen Seite soll ein selbstgesteuertes Erlernen der Fuzzy-Logik-Theorie ermöglicht werden. Daher soll an dieser Stelle kurz auf die Methode eingegangen werden.

Die von ihrem Begründer Zadeh (1965, S.338ff.) entwickelte Theorie Fuzzy-Logik bietet in Abgrenzung zur traditionellen Logik nicht nur die binären Wahrheitswerte Wahr (1) und Falsch (0), sondern sie ermöglicht einen fließenden Übergang zwischen den Wahrheitswerten. Die Übersetzung des Wortes „Fuzzy“ mit „verschwommen, fusselig, kraus“ meint keine ungenaue Logik, sondern den Versuch, ungenaue und unscharfe Informationen des Menschen zu formalisieren und so diese durch eine Computertechnik auswertbar zu machen.

Fuzzy-Logik-Systeme kann man im weitesten Sinne als Expertensysteme betrachten, da sie Expertenwissen abbilden und berücksichtigen. Damit fallen Fuzzy-Logik-Systeme in den Bereich der Künstlichen Intelligenz (KI). Die Fuzzy-Logik ist eine Hilfe für Bewertungs-, Klassifikations- und Steuerungsaufgaben. Vester (1999, S.151) berichtet über seine Erfolge bei der Verwendung der Fuzzy-Logik-Methodik: „Mit der Erfassung weicher Daten auf der einen und der Beschränkung auf wenige Systemparameter auf der anderen Seite waren jedenfalls wichtige Grundvoraussetzungen geschaffen, um hochkomplexe Systeme mit wenigen Schlüsselvariablen repräsentativ abzubilden“. Weiterhin „hat man die Garantie, dass man zwar mehr oder weniger exakte, dafür aber niemals falsche Konzepte der Realität anbietet“ (Vester 1999, S.227).

Es werden Zugehörigkeitsfunktionen und unscharfe Regeln für eine Schlussfolgerung (Fuzzy-Inferenz) verwendet, die dem Erfahrungswissen eines oder mehrerer Experten entsprechen. Durch einen Wert zwischen 0 und 1 wird der Grad der Zugehörigkeit von Eingangsgrößen zu einer unscharfen Mengen definiert. Die Zustandsgrößen (linguistische Variablen) werden durch linguistische Terme, wie „sehr klein“ oder „ziemlich breit“ beschrieben. Die Regeln enthalten alle einen Bedingungsteil und einen Schlussfolgerungsteil (z.B. Wenn Temperatur hoch, dann Zulauf gering). Die unscharfen Schlussfolgerungen müssen zur Erledigung einer Aufgabe (z.B. Einstellung einer Stellgröße) in eindeutige scharfe Ergebnisse defuzzifiziert werden.

Für das Systemtechnik-Praktikum wurden zwei weitere Lernsysteme entwickelt. Während die Entwicklung der Fuzzy-Logik-Lernsysteme jeweils etwa drei Monate dauerte, benötigte die Entwicklung der Planspiel-Lernsysteme etwa ein Jahr. Zum Planspiel „Magnat-Schwebbahn“ und zum Planspiel-Autorensystem existiert eine Veröffentlichung (Wagner 2005). Daher werden in den folgenden Unterkapiteln die beiden Lernsysteme nur kurz beschrieben.

Neben diesen Lernsystemen wurden kollaborative Lernportale für die Studenten eingerichtet. Diese Online-Learning-, Groupware- bzw. Content-Management-Plattformen wurden nicht mit der Qualitätssicherungsmethodik evaluiert, da diese bis auf die eingebrachten dynamischen Lerninhalte keine Eigenentwicklungen darstellen. Über mehrere Semester wurde das kommerzielle Lotus Quickplace und die Open Source ATutor-Software angeboten. Die Nut-

zung der Lernportale seitens der bisherigen Maschinenbau-Studenten war sehr gering, d.h. von den Studenten wurden keine eigenen Beiträge verfasst. Diese Erfahrung macht unter Umständen deutlich, dass sich für das Lehrgebiet derzeit weniger kollaborative Foren zum Wissensaustausch anbieten, sondern eher die in den folgenden Unterkapiteln vorgestellten Lernsysteme.

## 12.1 Fuzzy-Logik CBT

Das FLCBT ist ein computerbasiertes Fuzzy-Logik-Lernsystem, welches aus einer studentischen Arbeit am Fachgebiet Systemtechnik und Mensch-Maschine-Systeme zur Unterstützung der Lehre hervorgegangen ist.

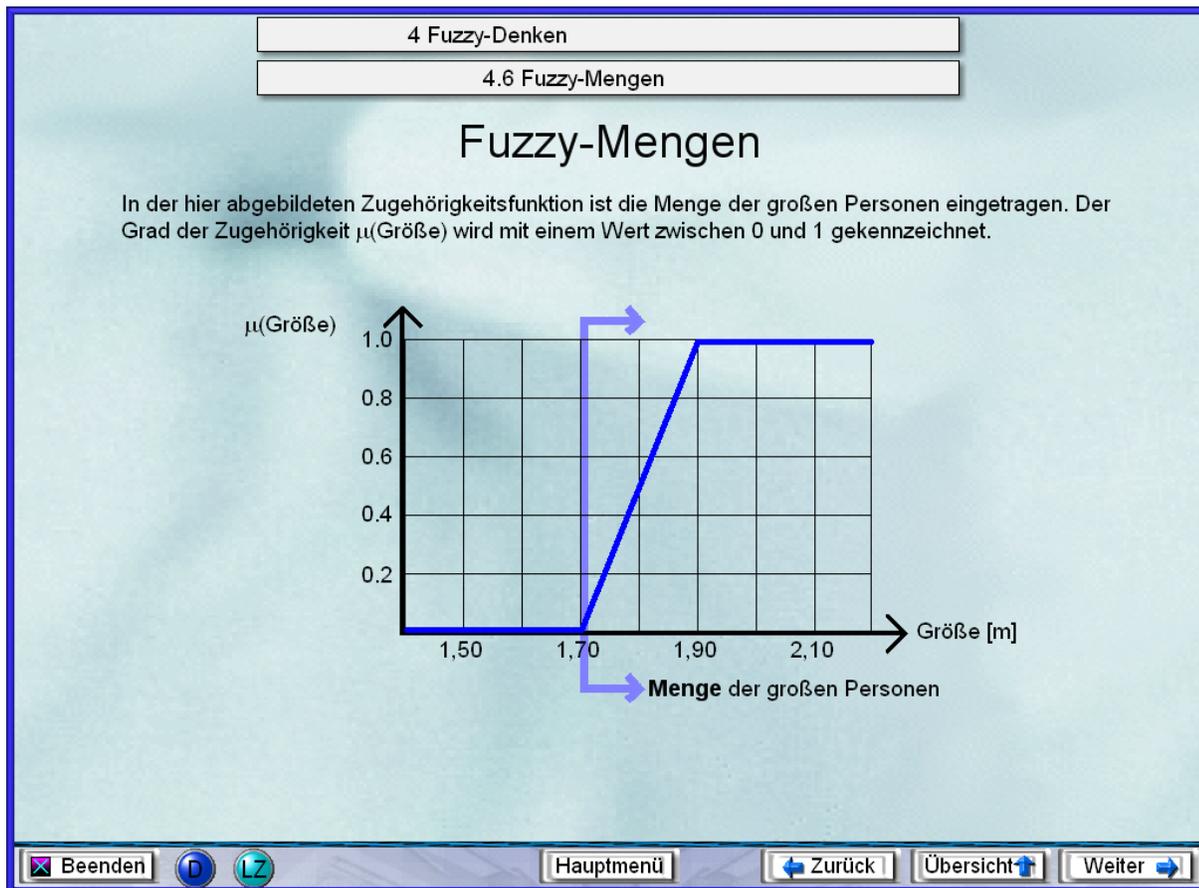


Abb. 14: Zugehörigkeitsfunktion im Fuzzy-Logik CBT

Da aus zeitlich-organisatorischen Gründen weder in der Vorlesung noch im Praktikum alle Aspekte der Fuzzy-Logik vertieft werden können, wurde das Angebot eines Lernsystems zum ergänzenden Selbststudium für die Studenten realisiert. Durch den Umfang der Lerninhalte

wird das Hinzuziehen weiterer Quellen – ausgenommen dem Besuch der Lehrveranstaltung – unnötig. Weiterführende Literatur zum Thema wird jedoch auch benannt.

Das FLCBT kann über ein Lernportal des Fachgebiets heruntergeladen werden. Nach dem Download sind keine weiteren Installationsarbeiten notwendig. Die Programmdatei ist durch Anklicken direkt einsatzbereit. Entwickelt wurde der Lernsystem-Prototyp mit dem kommerziellen Autorensystem Mediator des Herstellers MatchWare.

Das CBT-Lernsystem ist in verschiedene aufeinander aufbauende Lehreinheiten zur Fuzzy-Logik unterteilt. Die einzelnen Kapitel können je nach Interesse über ein Hauptmenü aufgerufen oder über die GuidedTour kennengelernt werden. Desweiteren helfen Lesezeichen, ein alphabetischer Index und eine Suchfunktion bei der Erschließung der Lerninhalte. Das Lernsystem wird damit verschiedenen Nutzergruppen mit unterschiedlichen Lerngewohnheiten und Vorwissen gerecht.

Das Lernsystem bietet eine Einführung in die Fuzzy-Logik und interaktive Lernkontrollen zu jedem Lehrkapitel und ist daher gemäß der behavioristischen Lerntheorie (Drill & Practice) umgesetzt.

### **12.2 Fuzzy-Logik WBT**

Das FLWBT ist ein webbasiertes Fuzzy-Logik-Lernsystem mit einem kognitivistischen Lehransatz. Das Lernsystem ist ebenfalls im Rahmen einer studentischen Arbeit am Fachgebiet Systemtechnik und Mensch-Maschine-Systeme verwirklicht worden.

In diesem Projekt wurden die Gestaltungsempfehlungen aus der Evaluation zum FLCBT-Lernsystem umgesetzt. Es ermöglicht eine orts- und zeitunabhängige Nutzung mit Hilfe der Webtechnologie. Das Lernsystem wird dadurch auf allen gängigen Computersystemen ohne Installationsarbeiten verfügbar. Das komplette Herunterladen des FLWBT-Lernsystems ist ebenfalls möglich, um auch ohne Internetverbindung damit arbeiten zu können.

Das FLWBT-Lernsystem ist im Gegensatz zum CBT-Prototyp ohne kommerzielle Entwicklungsumgebungen erstellt worden und deshalb herstellerunabhängig.

Die Lerninhalte sind in verschiedene aufeinander aufbauende Kapitel zur Fuzzy-Logik angeordnet. Durch die Einstreuung von interaktiv zu beeinflussenden Lernbeispielen in Gestalt von Java-Applets werden komplexe Sachverhalte in den Lehrtexten veranschaulicht (siehe Abb. 15 auf der nächsten Seite).

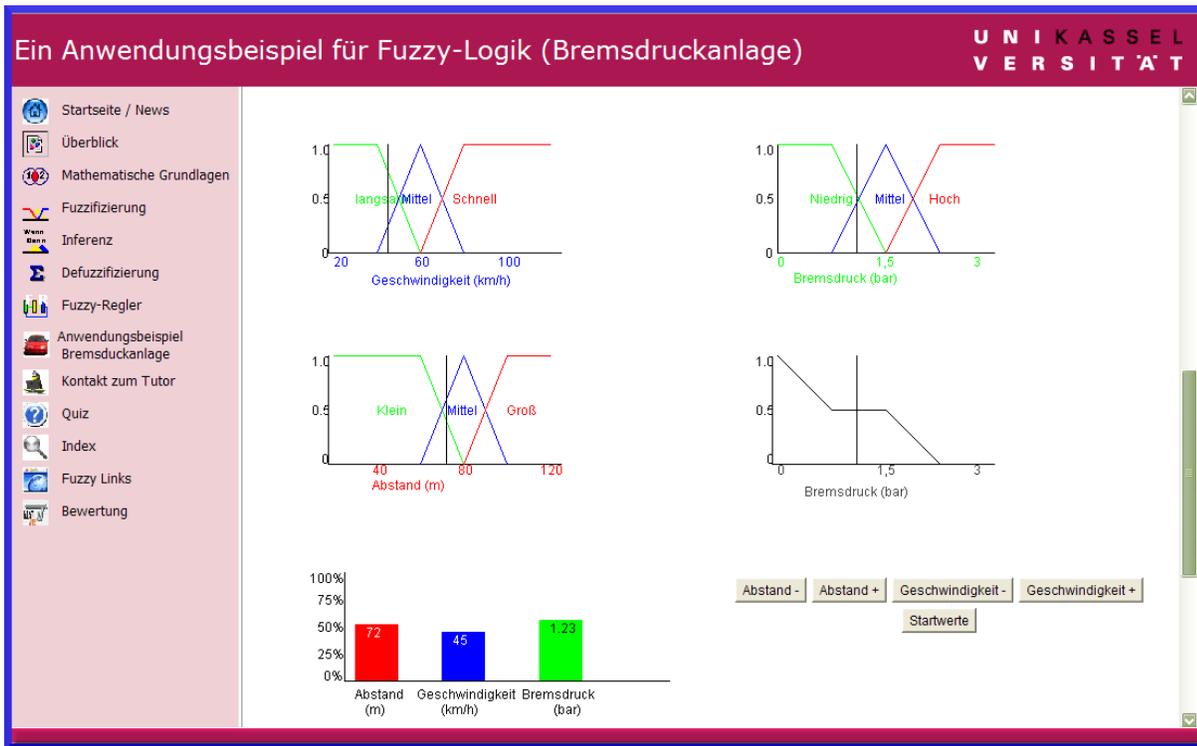


Abb. 15: Java-Applet im Fuzzy-Logik WBT

### 12.3 Fuzzy-Logik Kransteuerung

Beim dritten Lernsystem zur Fuzzy-Logik handelt es sich um eine Diplomarbeit ((Bretz 2005)). Aufgabe war es, ein Lernsystem für das im Fachgebiet Systemtechnik und Mensch-Maschine-Systeme angebotene Fuzzy-Logik-Praktikum zu entwickeln.

Im Unterschied zu den vorausgegangen Fuzzy-Logik-Lernsystemen finden die Lernenden eine ausschließlich experimentelle Simulationsumgebung vor, in der sie ihr Fuzzy-Logik-Wissen anwenden können. Damit entspricht das Lernsystem einer konstruktivistischen Lerntheorie.

Als Ausgangspunkt und Anregung für das Projekt diente das LoadSway-Applet von Sazonov.<sup>33</sup> Die in diesem Applet bereitgestellte Fuzzy-Logik-Programmibibliothek wurde in dem neu erstellten Lernsystem FLKran wiederverwendet. Zugleich fand eine Erweiterung der Funktionalität und eine Verbesserung der Authentizität der Simulation statt. Das Lernsystem verfügt über eine aufgabenangemessene Benutzerschnittstelle, die sich formatfüllend projizieren lässt.

<sup>33</sup>[http://people.clarkson.edu/~esazonov/neural\\_fuzzy/loadsway/LoadSway.htm](http://people.clarkson.edu/~esazonov/neural_fuzzy/loadsway/LoadSway.htm)

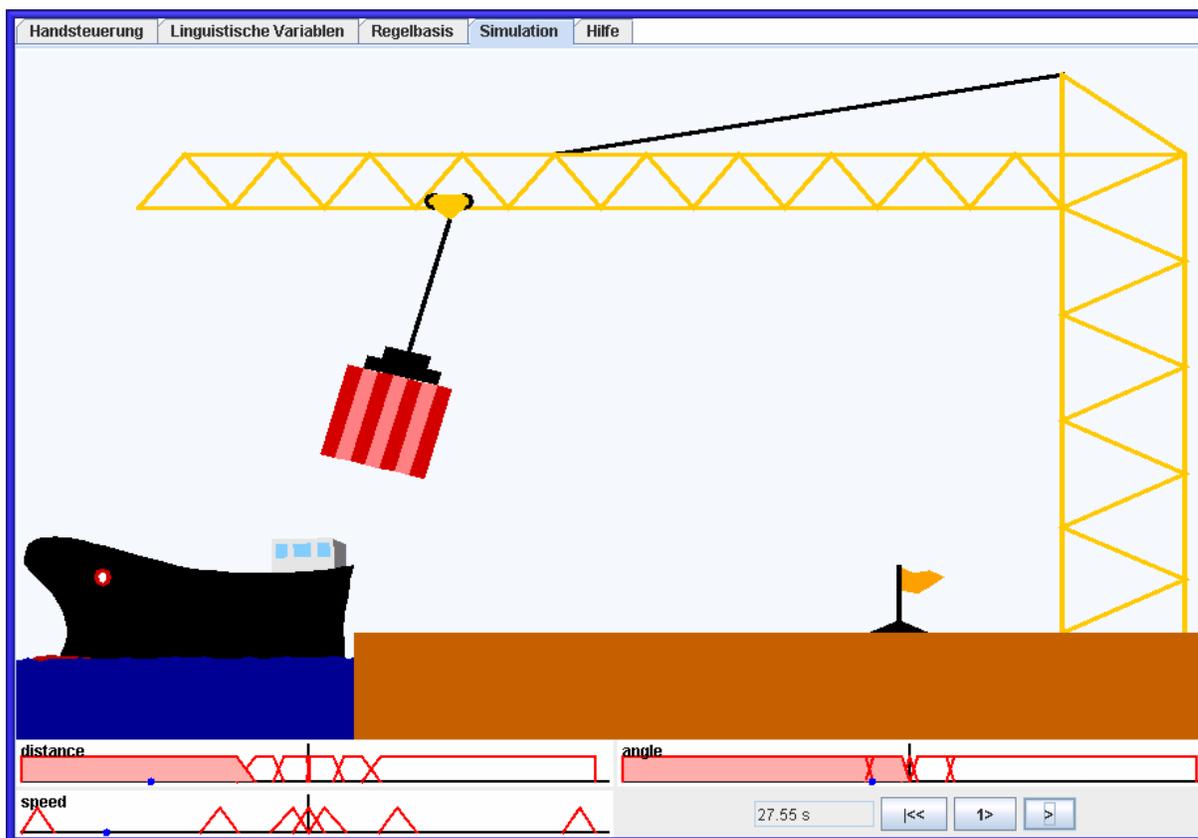


Abb. 16: Fuzzy-Logik Kransteuerung

Ähnlich wie bei der LoadSway-Applikation hat der Anwender die Aufgabe, ein Containerschiff zu entladen. Dazu kann er sich zunächst mit der Kransteuerung vertraut machen, indem er versucht über die manuelle Steuerung des Krans ein Szenario zu erfüllen. Mit den gesammelten Erfahrungen gilt es anschließend, über ein Autorensystem die Regelbasis und die Zugehörigkeitsfunktionen zur automatischen Kransteuerung zu entwerfen. Die Güte seines selbst entworfenen Fuzzy-Logik-Reglers kann er in der Simulation sofort begutachten und gegebenenfalls iterativ verbessern. Hilfreich dabei sind die Zeitmessung, der Einzelschrittmodus, sowie die Plots der Zugehörigkeitsfunktionen mit den aktuellen Werten der Eingangs- und Ausgangsgrößen des simulierten Systems.

## 12.4 Magnat Schwebbahn

Planspiele können wirkungsvoll die Theorie mit beruflicher Praxis verbinden. Sie ermöglichen eine Lehrmethode, die ohne das Lesen von Lehrtexten rein durch die interaktive Aufgabengestaltung Lernprozesse initiiert. Damit fördern sie ein aktives und selbstorganisiertes Lernen.

Das Planspiel „Magnat Schwebbahn“, kurz PS, ist das Vorläuferprojekt des in UKap. 12.5 auf der nächsten Seite vorgestellten Planspiel-Autorensystems. Viele bewährte Funktionalitäten und Konzepte des Planspiels sind vom Planspiel-Autorensystem übernommen worden. An anderen Stellen wurden im Planspiel-Autorensystem neue Eigenschaften bereit gestellt. Bevor auf das weiterentwickelte und aktuell verfügbare Planspiel-Autorensystem eingegangen wird, soll in diesem Unterkapitel der Vorgängerprototyp erläutert werden.

Die Entwicklung des Planspiels wurde im Rahmen eines E-Learning-Projektwettbewerbs von der Universität Kassel gefördert. Das Planspiel wird aktuell in der Systemtechnik-Lehre des Fachgebiets Systemtechnik und Mensch-Maschine-Systeme der Universität Kassel eingesetzt. Das Fallbeispiel Magnetschwebbahn wird seit Jahren in der Vorlesung und Übung behandelt.

The screenshot shows the game interface for 'Magnat Schwebbahn'. At the top, there are navigation links: [Abmelden](#), [Steuerung](#), [Funktionsmatrix](#), [Auswertung](#), [Glossar](#), and [Anleitung](#). Below these is a progress bar with 10 numbered circles, where the first circle is highlighted in red. The player's name is 'Spieler: nobody'. A 'nächste Runde' button with a circular arrow and a red 'X' for 'neues Spiel' is visible.

The interface is divided into three main colored boxes:

- Infrastruktur (Pink box):**
  - Fahrkomfort: -7..7, value 1
  - Anzahl MSB: 1..00, value 3
  - Anzahl Sektionen: 2..6, value 2
  - Zustand Infrastruktur: -7..7, value 0
- Verkehrnetz (Yellow box):**
  - Geschwindigkeit/Fahrzeit: -7..7, value 0
  - Platzangebot: 0..00, value 4000000
  - Auslastung: 0..100, value 0
  - Taktfrequenz: 1..7, value 1
  - Anzahl Haltepunkte: 2..00, value 2
  - Umweltverträglichkeit: -7..7, value -4
- Marktwert (Blue box):**
  - Fahrpreisniveau: 0..100, value 20
  - Kundenzufriedenheit: -7..7, value 1
  - Marktanteil: 0..100, value 15
  - Kundenzahl: 0..00, value 200000
  - Finanzmittel: 0..00, value 5700000

At the bottom left, there is a 'Kosten durch Eingriffe' table:

Kosten durch Eingriffe	Finanzmittel
Fahrkomfort	-400000
Anzahl MSB	-1000000
Fahrpreisniveau	-1000
<b>Summe</b>	<b>-1401000</b>

An 'aktualisieren' button is located below the table.

Abb. 17: Übersicht über Zustandsgrößen des Planspielsystems

Der Einsatz des Planspiels im betreuten Praktikum unterstützt die dozentenorientierte Lehre dadurch, dass die theoretisch erworbenen Kenntnisse aus der Vorlesung hier nun auf eine Aufgabenstellung transferiert und Lerninhalte praktisch angewendet werden. Die Studenten werden in die Rolle von Planern versetzt, die über den wirtschaftlichen Einsatz und Betrieb einer Magnetschwebbahn entscheiden. Der Name des Planspiels deutet dieses Ziel bereits an: Ein Magnat ist eine führende Persönlichkeit, die nach Reichtum strebt. Die Lernenden fungieren als Manager bzw. Regler des dynamischen Systems.

Das zugrunde liegende Modell zur Simulation im Planspiel wurde didaktisch abstrahiert und vereinfacht, wobei wirtschaftliche und ferner technische Aspekte im Vordergrund stehen. Somit wird ein Ausschnitt der Realität betrachtet. Alle für das Systemverständnis notwendigen Informationen sind im Planspiel enthalten. Der Fortschritt bei der Aufgabe der Planspielteams, ihren virtuellen Magnetschwebbahn-Betrieb wirtschaftlich zu halten, gilt als Hinweis auf den Lernerfolg.

### **12.5 Planspiel-Autorensystem**

Im Fachgebiet Systemtechnik und Mensch-Maschine-Systeme an der Universität Kassel wurde ein flexibel einsetzbares Planspiel-Autorensystem (PAS) entwickelt. Dieses Projekt wurde im Rahmen der E-Learning Projektförderung durch Hessen-Media unterstützt.

Das PAS verhilft Lehrenden oder Lernenden dazu, schnell und einfach webbasierte Planspiele zu konstruieren. Die Erstellung eines eigenen Planspiels verlangt dabei nicht, dass die modellbildenden Autoren Programmierkenntnisse besitzen müssen.

Betont sei hierbei, dass die Software nicht nur für Planspiel-Produzenten interessant ist, sondern auch für Dozenten, Lehrer und Ausbilder, die neben dem vorteilhaften Einsatz der Planspielmethode auch die Modellbildung von Systemen praktisch und effizient vermitteln bzw. trainieren wollen. Bei herkömmlichen webbasierten Planspielen entscheiden lediglich die Produzenten des Planspiels über das simulierte Modell, nicht die Zielgruppe, die das Planspiel in der Lehre nutzt. Beim PAS haben die Benutzer hingegen alle Fäden in der Hand, um beliebige Problemsituationen aus der Realität in einem Modell abzubilden, es anschließend zu verifizieren und zu validieren.

Das Planspiel-Autorensystem gestattet eine Gestaltung der Lehre, die nicht nur durch reines Nachvollziehen eines simulierten fremden Modells geprägt ist, sondern den Erwerb von weiteren Kompetenzen zur Systembeobachtung, -beschreibung, -modellierung und zur Systemkritik begünstigt. Die Zielgruppe entscheidet darüber, was wesentlich für die Abbildung eines beobachteten Ausschnitts der Realität ist und kann durch die Simulation prüfen, ob die Entscheidung richtig war.

Die integrierte Simulationsumgebung ermöglicht, bezogen auf die vorausgegangene Modellbildung, eine spielerische Erschließung des thematisierten Systems. Im Unterschied zu vielen

anderen Planspielen sind die wirkenden Zustandsgrößen und Beziehungen von den Lernenden abrufbar. Durch diese Transparenz können die Eigendynamik von komplexen Systemen besser erforscht und eigene Strategien zur Erfüllung eines Planspielziels entwickelt und angewendet werden.

Bei der Gestaltung der grafischen Benutzeroberfläche wurde insbesondere auf eine intuitive und aufgabenangemessene Bedienbarkeit der Software geachtet. Schlüssel für eine gute Bedienbarkeit ist die Übersichtlichkeit der Benutzerschnittstelle. Das PAS wurde in dieser Hinsicht bewusst mit einem schlichten aufgeräumten Screen Design versehen, das die aufgabenangemessene Visualisierung und Bearbeitung eines komplexen dynamischen Systems ermöglicht. Die Informations- und Eingabeseiten sind erwartungskonform durch kontextlogische Hyperlinks vernetzt. Die Benutzerschnittstelle stellt nur die real benötigten Funktionalitäten für die Zielgruppen bereit.

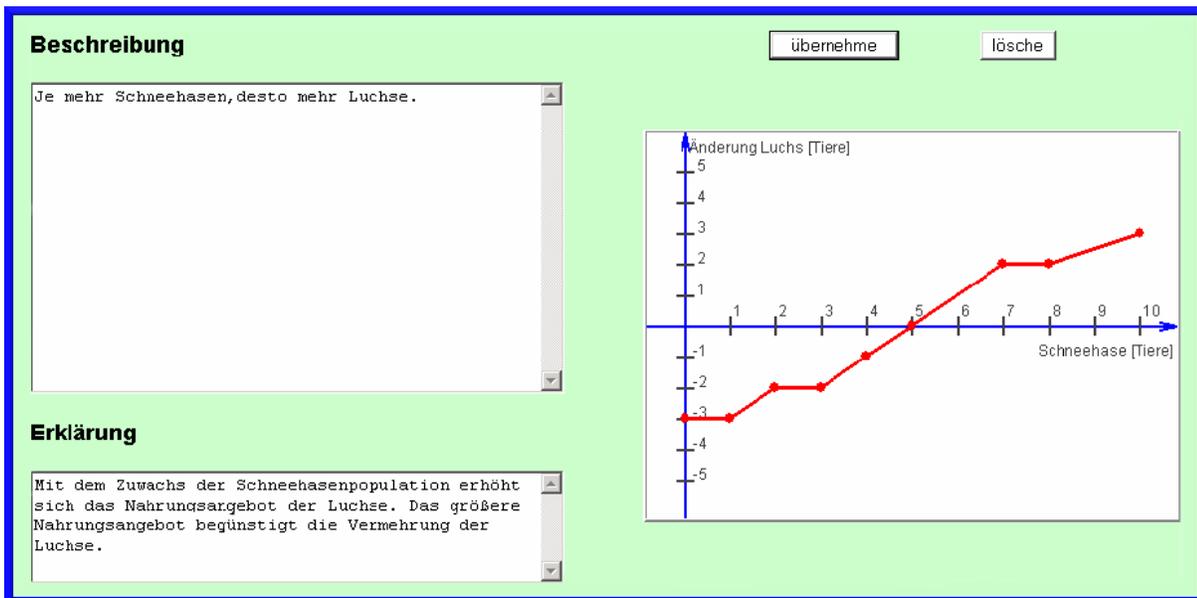


Abb. 18: Bearbeitung einer Beziehung zwischen zwei Zustandsgrößen

## Teil IV

# Studie zur Expertenbefragung

Der Bedarf einer formativen Evaluation wurde in Teil II der vorliegenden Arbeit diskutiert. Dort wurde die Expertenbefragung als formative Evaluationsmethode vorgestellt. Das Potenzial dieser Methode soll in dieser Arbeit nicht durch den Vergleich mit anderen Evaluationsmethoden zu einem bestimmten Lernsystem abgegrenzt werden, sondern durch die mehrfache Erprobung in unterschiedlichen Lernsystem-Projekten. Der Schwerpunkt der Betrachtung bildet dabei die zu erzielende Quantität und Qualität der Bewertungshinweise von unterschiedlichen Experten. Für die Potenzialabschätzung der Expertenbefragung ist es wichtig, die maßgeblichen Einflussgrößen für reichhaltige Bewertungshinweise zu kennen und diese für die Planung von Expertenbefragungen zu berücksichtigen. Im Rahmen der in diesem Teil beschriebenen Studie werden nicht nur mögliche Einflussgrößen zu den durchgeführten Expertenbefragungen genannt, sondern darüber hinaus werden Planungshilfen für zukünftige Expertenbefragungen abgeleitet.

Die Analyse der durchgeführten Expertenbefragungen dient somit primär dazu, Informationen über die Richtung und das Gewicht der möglichen Einflussgrößen zu erhalten und sekundär, um entwickelte Lernsystem-Prototypen qualitativ zu vergleichen und Verbesserungen dieser einzuleiten. Für die zuletzt genannte Motivation wurden Qualitätsindikatoren definiert und bewertet.

In diesem Teil der Arbeit wird zunächst die Versuchsplanung und die Durchführung der Expertenbefragungen erläutert. Anschließend werden die Studienergebnisse vorgestellt und interpretiert.

## 13 Versuchsplanung

Für den Nachvollzug, die Interpretation und die Beurteilung der Wiederholbarkeit der Studienergebnisse ist eine Versuchsplanung (Experimental Design) notwendig. Bei wissenschaftlichen Studien erfolgt eine sorgfältige Planung mit einer ausführlichen Schilderung des grundsätzlichen Vorgehens. Die Versuchsplanung hilft auch den Initiatoren, selbst gezielt zu besseren Ergebnissen zu gelangen.

Zur Versuchsplanung gehören nach Johannsen (1993, S.162) die

1. Darlegung unabhängiger und abhängiger Variablen;
2. Hypothesenbildung über erwartete Versuchsergebnisse;
3. Beschreibung der Versuchsdurchführung, -bedingungen und -wiederholungen;
4. statistische Auswertung.

Ziele der Studie waren, die Qualität der Lernsystem-Prototypen und das Potenzial der Expertenbefragung unter möglichst authentischen Bedingungen eines formativen Entwicklungsprozesses abzuschätzen. Während das Ziel, über Expertenbeurteilungen auf die Qualität des Lernsystems zu schließen, den befragten Personen bekannt war, wurde das Ziel zur Abgrenzung des Potenzials der Expertenbefragung geheim gehalten. Da sich sowohl die Qualitäts- als auch die Potenzialabschätzung auf dieselben Urdaten beziehen, konnte die Versuchsplanung für beide Zielsetzungen angewendet werden.

Neben der Inhaltsanalyse von Begründungstexten zur Lernsystem-Bewertung wurden folgende statistische Verfahren angewendet: Deskriptive Statistik, Permutation, t-Test, Regressions- und Korrelationsanalyse.

### **13.1 Expertenanzahl**

Für die Untersuchung des Potenzials der Expertenbefragung in einem formativen Entwicklungsprozess ist es bedeutsam, dass die üblicherweise zur Verfügung stehenden Personengruppen befragt werden. Die Zusammensetzung der Expertenpopulation entsprach den Zielgruppen der untersuchten Lernsysteme. Damit wurde die Forderung nach realen Endbenutzern erfüllt (Rubin 1994, S.29, Nielsen 1993, S.196). Studenten können in Bezug auf die Lernsysteme, die in der universitären Lehre zum Einsatz kommen, grundsätzlich als repräsentative Endbenutzer angesehen werden. Neben Studenten wurden Administratoren, Entwickler, Lehrende bzw. Wissenschaftler befragt. Während Studierende noch relativ einfach und kostengünstig rekrutiert werden können, ist die Verfügbarkeit der zuletzt genannten Expertengruppe mit hohem Erfahrungsgrad im Bereich der Softwarebewertung schon eher begrenzt.

Eine Faustregel besagt, dass rund 1000 Befragte repräsentative Ergebnisse für die Gesamtbevölkerung bringen, die ordnungsgemäß nach Zufallsstichprobe oder Quote ausgewählt wurden. Die Zufallsstichprobe wird auch Random-Stichprobe genannt. Die Befragten werden zufällig beispielsweise aus dem Telefonbuch ausgewählt. Bei der Quote-Auswahltechnik wird die Grundgesamtheit – meist der Gesamtbevölkerung – in kleinerem Maßstab nachgebaut.

Mit zunehmender Expertenanzahl können Verteilungen besser bestimmt, Ausreißer identifiziert und Schlussfolgerungen zur Qualität des Lernsystems abgesichert werden.

Eine große Expertenanzahl ist bei formativen Evaluationen allerdings nicht ökonomisch. Bevor ein Nutzen durch eine Expertenbefragung entstehen kann, kosten Experten in der Regel Zeit und Geld in Abhängigkeit von deren Anzahl. Die Ergebnisse der Evaluation sollen möglichst schnell zur Verfügung stehen, noch bevor der aktuelle Prototyp wesentlich weiterentwickelt wurde. Je mehr Zeit zwischen dem Bewertungszeitraum und der Fortentwicklung des Prototyps vergeht, desto weniger wertvoll sind die Hinweise aus den Expertenbefragungen.

Aus ökonomischen Gründen ist das Qualitätsregelteam daher bestrebt, die Anzahl der Experten so gering wie nötig zu halten.

Nach Rubin (1994, S.93) sollte die Anzahl etwa bei 10 bis 12 Testpersonen liegen. Für einen weniger formalen Test würden nach Rubin vier bis fünf Testpersonen ausreichen, um 80% der vorhandenen Probleme zu finden.

Nielsen (2000) empfiehlt lieber mehrmals Tests mit wenigen Personen und jeweils anschließenden Verbesserungszyklen durchzuführen, als mit erheblich mehr Experten die restlichen 20% der vorhandenen Probleme bei einer Software in einem Durchgang zu finden.

Nielsen (1993, S.173) nennt folgende Gleichung zur Planung der Testpersonenanzahl:

$$\text{Gefundene Benutzbarkeitsprobleme}(i) = N \cdot (1 - (1 - \lambda)^i) \quad (2)$$

$i$  Personenanzahl,  $N$  Gesamtanzahl der Probleme,  $0 \leq \lambda \leq 1$  Problemanzahl, die eine Person im Durchschnitt findet (mit 0:=0 Prozent; 1:=100 Prozent).

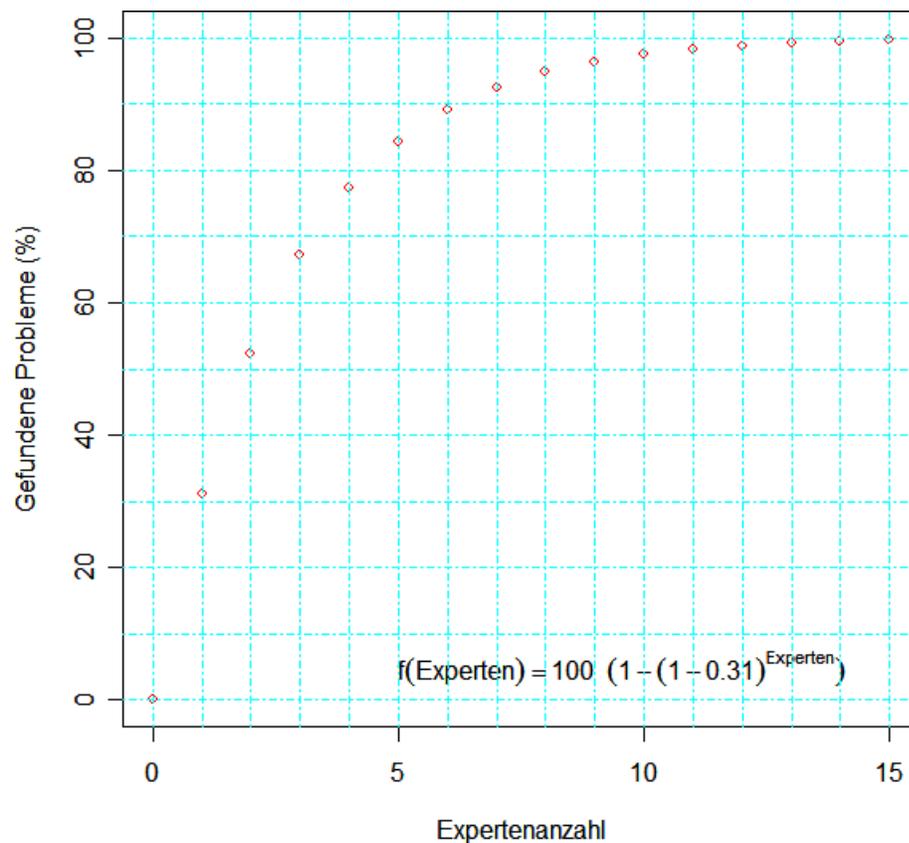


Abb. 19: Expertenanzahl nach Nielsen (1993)

$\lambda$  ist eine unbekannte Variable, die schwierig zu schätzen ist. Nielsen empfiehlt für  $\lambda$  den Erfahrungswert 0,31. Problematisch an diesem Erfahrungswert ist, dass dieser in der Realität sicherlich nicht bei jedem Experten konstant ist.

Der Wert  $N$  erhält notwendigerweise den normierten Wert 100%, da man in der Praxis die Gesamtanzahl der Probleme nicht kennt und diese erst durch die Expertenbefragung herausfinden will.

Die Funktion in Abb. 19 auf der vorherigen Seite ist eine Sättigungskurve. Mit zunehmender Expertenanzahl werden immer weniger neue Probleme aufgedeckt. Vielmehr werden die schon bekannten Probleme häufiger genannt. Mit 5 Personen sollen beinahe 85% der Probleme gefunden werden.

Ein Nachteil der Normierung ist, dass beispielsweise die Prozentzahl 100% bei einem Lernsystem 1000 oder 10 gleich schwer wiegende Probleme bedeuten kann. Die Qualität eines Lernsystems ist jedoch in den beiden Fällen unterschiedlich zu bewerten. Für die Qualitätssicherung von Lernsystem-Prototypen bedarf es genauere Abschätzungen zum Qualitätsvergleich von Prototypen und zur effizienten Einplanung von Experten.

Eine weitere Berechnung des Stichproben- bzw. Versuchsumfangs findet sich bei Kleppmann (2003, S.83): „Welcher Versuchsumfang ist erforderlich, wenn ein bestimmter Unterschied mit hoher Wahrscheinlichkeit gefunden werden soll, falls er existiert... Man kann zeigen, dass bei einem Stichprobenumfang  $n$  bzw. einem Versuchsumfang  $N = 2 \cdot n$  (Gesamtanzahl der Messwerte) von

$$\text{Versuchsumfang } N = 2 \cdot n = 60 \cdot \left( \frac{\sigma}{\Delta\mu} \right)^2 \quad (3)$$

ein wahrer Unterschied zwischen den Gruppen (Effekt)  $\Delta\mu$  mit hoher Wahrscheinlichkeit entdeckt wird, wenn die Standardabweichung der Einzelwerte  $\sigma$  beträgt“.

Deutlich wird in dieser Gleichung (3), dass übertragen auf die Expertenbefragung und den Vergleich zweier Prototypen die Standardabweichung mit der Expertenanzahl korreliert. Die wahre Standardabweichung lässt sich im Vorfeld einer Befragung allerdings sehr schwer abschätzen, da sie durch zufällig abweichende Expertenmeinungen zustande kommt. Die Standardabweichung  $\sigma$  muss für die Gleichung bekannt sein, ebenso muss der Unterschied zwischen zwei Mittelwerten von Stichproben festgelegt werden.

Somit tragen die Gleichungen (2) auf der vorherigen Seite und (3) nur sehr vage dazu bei, die Frage nach einer effizienten Expertenanzahl in der Planungsphase einer formativen Evaluation zu beantworten. Die Frage, mit welcher Expertenpopulation sich wieviele Qualitätshinweise erheben lassen, bleibt unbeantwortet.

Die Anzahl ist maßgeblich vom Expertengrad abhängig, wie aus Untersuchungen von Bailey (2001) und Hertzum *et al.* (2002, S.662) zu schließen ist. Mit dem Begriff Expertengrad sind hier die Vorerfahrung, das Wissen, die Kenntnisse, die Ansichten, die Fähigkeiten etc. des Experten gemeint.

Bei der Untersuchung von Bailey wurden verschiedene Usability-Labore zur Analyse derselben Software beauftragt. Es zeigen sich deutliche Unterschiede bezüglich der Effizienz der Usability-Labore:

Usability-Labore	A	B	C	D	E	F	G	H	I
Usability-Experten	2	7	1	1	3	1	1	3	7
Anzahl der durchgeführten Tests	7	6	6	50	9	5	11	4	6
Gefundene Probleme	26	150	17	10	68	75	30	18	20

Tab. 10: Vergleich von Usability-Laboren nach Bailey (2001)

Ebenso treten Unterschiede beim Vergleich von einzelnen Experten auf. Dies bestätigt die folgende Untersuchung, bei der die Experten die gleiche Software untersuchen sollten:

Experte	1	2	3	4	5	6	7	Total
Alle gefundenen Probleme	135	39	25	9	7	4	1	220
Gefundene schwerwiegende Probleme	14	3	7	2	5	1	1	33

Tab. 11: Usability-Inspektionen (Hertzum *et al.* 2002, S.662f.)

Die nachfolgend aufgeführte Untersuchung von Molich *et al.* (1998, S.189ff.) ist ähnlich aufgebaut wie die zuerst genannte. Zusätzlich wurde jedoch der Zeitaufwand (ohne die Zeit der Testpersonen) ausgewertet. Dies ist eine wichtige Größe für die Bestimmung der Effizienz einer Analyse.

Usability-Labore	A	B	C	D
Zeitaufwand Usability-Experten	26	70	24	84
Anzahl der von den Experten durchgeführten Tests	18	5	4	5
Gefundene Problemanzahl	4	98	25	35

Tab. 12: Usability-Tests nach Molich *et al.* (1998)

Für die Planung der Expertenanzahl für die eigene Studie wird aufgrund der genannten Untersuchungen und Empfehlungen angenommen, dass 15 Experten mehr als ausreichend sind.

Im Hinblick auf eine ökonomische Durchführung der Expertenbefragung wird im UKap. 15.4 auf Seite 109 eine gegenüber den genannten Abschätzungen verbesserte Gleichung diskutiert. Zur Parametrisierung der Gleichung wurden die Lernsystem-Prototypen aus Kap. 12 auf Seite 68 untersucht.

## 13.2 Fragestellungen

Bevor man sich für die qualitätskriterienbasierte Expertenbefragung entscheidet, sollte das Potenzial dieser Methode zur formativen Evaluation von Lernsystemen bekannt sein. Aufgrund der multiplen vieldimensionalen Einflussgrößen und den häufig ungenannten Annahmen zu einer Studie sind methodenvergleichende Untersuchungen von nur geringem Wert. Nutzbringender sind Untersuchungen, die die Einflussgrößen aus mehreren realen Projekten systematisch bestimmen und nennen. Nur so kann das Potenzial der qualitätskriterienbasierten Expertenbefragung angemessen eingeschätzt und auf andere Kontexte übertragen werden.

Folgende detailliertere Fragestellungen sollen helfen, diese Hauptfrage durch eigene Studien belegt zu beantworten:

- Eignen sich die **Qualitätsindikatoren** *PN*, *HA* und *GB* als Kennzahlen zur Anforderungserfüllung von Lernsystemen? (siehe Kap. 15.1 auf Seite 91)
  - PN:** Der Qualitätsindikator *PN* bezieht sich auf die in den Begründungstexten enthaltene Kritik. Kritik ist die prüfende Beurteilung nach einem begründeten Maßstab, die mit der Abwägung von Wert (Positiv) und Unwert (Negativ) eines Sachverhalts einhergeht. Das Verhältnis von positiver und negativer Kritik führt zur Bildung des Qualitätsindikators *PN*.
  - HA:** Der Qualitätsindikator *HA* ist ein mit dem Qualitätsindikator *PN* ähnliches Verhältnis mit dem Unterschied, dass das Qualitätsregelteam positive und negative Handlungsableitungen zur Qualität des Lernsystems vermerkt. Aus diesem Verhältnis wird der Qualitätsindikator *HA* ermittelt.
  - GB:** Die Experten hatten die Aufgabe, die Qualitätskriterien zu gewichten und deren Umsetzung im Lernsystem zu bewerten. Aus den durch den Experten vergebenen Punktzahlen zur Gewichtung und Bewertung wird der Qualitätsindikator *GB* berechnet.
- Wie hoch ist der benötigte **Zeitaufwand** für eine Expertenbefragung?
- Welche Einflüsse wirken sich auf die Anzahl der **Hinweisklassen** aus? (siehe UKap. 15.3 auf Seite 97)
  - Hinweisklassen:** Die freien Antworten der Experten zur Begründung ihrer Bewertung werden klassifiziert und durch einen Hinweisklassentext inhaltlich zusammengefasst.
- Welche Korrelation besteht zwischen der Expertenanzahl und der kumulierten Anzahl der **Hinweisklassen**? Lohnt sich ab einer bestimmten Expertenanzahl die Befragung von weiteren Personen im Hinblick auf das Aufwand-Nutzenverhältnis nicht mehr?

- Kann anhand der kumulierten Anzahl von **Monopol-Hinweisklassen** zu einer Expertenbefragung auf den Nutzen der Befragung einer weiteren Person geschlossen werden? (siehe UKap. 15.5 auf Seite 112)

**Monopol-Hinweisklassen:** Führt ein Begründungstext eines Experten zur Bildung einer Hinweisklasse, d.h. die Begründungstexte der anderen Experten werden nicht dieser speziellen Hinweisklasse inhaltlich zugeordnet, dann wird diese als Monopol-Hinweisklasse definiert.

- Wie verhält sich die **Standardabweichung** zur Bewertung und Gewichtung von Qualitätskriterien?
- Wird die freie **Antwortmöglichkeit** konstruktiv von den Experten genutzt?
- Zu welchen **Qualitätskriterien** werden die meisten Hinweisklassen generiert?
- Wie häufig werden die für den Bewertungsprozess verfügbaren **Metainformationen** (Leitfragen, Beschreibung, Erläuterung, Anleitung, Fundierung etc.) zum Qualitätskriterium genutzt?

Diese Fragestellungen werden im Kap. 15 auf Seite 90 wieder aufgegriffen werden, um sie durch die Studienergebnisse gestützt zu beantworten.

### 13.3 Versuchsvariablen

Bei der Versuchsplanung werden verschiedene Einflussgrößen auf Expertenbefragungen angenommen. Diese werden nachfolgend als nicht kontrollierte, kontrollierte, abhängige und unabhängige Variablen klassifiziert.

**Nicht kontrollierte Variablen** finden in der Versuchsplanung keine Berücksichtigung. Dies sind normalerweise Variablen von denen angenommen wird, dass sie keinen oder nur einen sehr geringen Einfluss auf die Untersuchungsergebnisse bewirken. Zufällige Einflüsse der nicht kontrollierten Variablen führen dazu, dass die Versuchsergebnisse streuen. Hierzu werden der Zeitpunkt und Zeitraum der Befragung sowie die physiologische Leistungskurve der Experten gerechnet. Denkbare Einflüsse wie Geschlecht, Reife/Alter, Motivation (Interesse, Leistungsbereitschaft) und unterschiedliches Sprachvermögen der Experten werden als konstant oder gering angenommen. Für den Bereich universitäre Lehrbewertung durch Studierende fand Hage (1996) anhand einer Studie für einen repräsentativen Querschnitt von deutschen Hochschulen heraus, dass unterschiedliche Sachkenntnis, Reife der Studierenden, sowie das Geschlecht der Studierenden und der Lehrenden die Lehrbewertung nicht verzerrt.

**Kontrollierte Variablen** sind mögliche Einflussgrößen, die im Versuch konstant gehalten werden:

1. Gestaltung des Fragebogens (Layout, Webbasierung etc.);
2. Qualitätskriterien (Fragestellungen, Metainformationen etc.);
3. Untersuchungsumgebung (normaler Arbeitsplatz), Stressbedingungen (keine), eingesetzte Computertechnik (CPU, Arbeitsspeicher, etc.).

**Unabhängige Variablen** sind diejenigen Einflussgrößen, die kontrolliert variiert werden:

1. Expertenanzahl;
2. Erfahrungsgrad der Experten;
3. Lernsystem.

**Abhängige Variablen** sind diejenigen Messgrößen, die durch die zuvor genannten Variablen beeinflusst werden:

1. Anzahl der Begründungstexte und (Monopol-) Hinweisklassen;
2. Bepunktung gewichteter und bewerteter Qualitätskriterien;
3. Befragungsdauer und Anzahl der Fragen.

Die Experten nahmen an einer Vor-, Haupt- und Nachbefragung teil. Die Vor- und Nachbefragungen dienten hauptsächlich dazu, den Expertengrad der Teilnehmer festzulegen. Bei der Expertenbefragung wird auf Experten mit unterschiedlicher Vorerfahrung zurückgegriffen. Die Hauptbefragung führte zu den im Kap. 15 auf Seite 90 diskutierten Ergebnissen. Das folgende Kapitel erläutert, wie die Hauptbefragung der Experten organisiert war.

## 14 Durchführung

Die Hauptbefragung erfolgte mit einem webbasierten LernSystemAnalyse-Fragebogen (siehe Abb. 9 auf Seite 39). Im Gegensatz zu klassischen Online-Befragungen wurde jedoch sicher gestellt, dass tatsächlich auch die eingeladene Person und keine andere die Fragen beantwortet hat. Allen Experten wurden dieselben Informationen zum Ablauf der Befragung, zur Bedienung des Fragebogens und zum untersuchten Prototyp zur Verfügung gestellt.

Bei internen Evaluationen kann es zu einer ungewollten Beeinflussung der Bewertungsergebnisse kommen. In der Praxis besteht die Gefahr, dass die Befragten dem Initiator gegenüber ungewollt Antworten geben, die in die vermeintlich gewünschte Richtung gehen (Sponsorship-Effekt, soziale Erwünschtheit). Dieser Mechanismus läuft völlig im Unterbewusstsein ab. Bei der Durchführung wurde den Befragten daher erklärt, dass jede Bewertung wertvoll ist, egal ob mit positivem oder negativem Inhalt. Die Bewerter wurden aufgefordert, keine Sympathiepunkte zu vergeben und das Lernsystem möglichst objektiv anhand der Qualitätskriterien zu beurteilen.

Ein weiterer für die Experten zu beachtender Aspekt betraf den Umgang mit den Qualitätskriterien. Die Aufgabe der Expertenbefragung war es, möglichst viele Hinweise zur Qualität mit Hilfe einer vom Detaillierungsgrad der Untersuchung abhängigen Anzahl von Qualitätskriterien zu erlangen. Die Experten wurden also gebeten, sich gründlich mit dem Lernsystem auseinander zu setzen. Nach eigenem Ermessen sollten sie danach oder parallel dazu die Qualitätskriterien studieren und die darin enthaltenen Fragen gewichten und bewerten. Neben Punkt-Skalen zur Einstellung der Gewichtung und der Bewertung standen zur Begründung der Punktvergabe Texteingabefelder zur Verfügung. Diese Begründungstexte konnten beispielsweise Hinweise enthalten, ob die Anwendung des Lernsystems für den Experten einfach, bequem, schnell, lehrreich etc. war.

Während die von den Experten abgegebene Bepunktung einzelner Qualitätskriterien infolge der standardisierten Antwortmöglichkeit direkt ausgewertet werden konnte, mussten die freien Antworten zur Begründung der Bepunktung aufwändiger analysiert werden. Der Aufwand bestand darin, dass die Begründungstexte der Experten zur Bewertung größtenteils unpräzise Formulierungen zum Lernsystem-Prototypen enthielten. Expertenhinweise wie „Die Benutzeroberfläche ist ziemlich originell“ oder „Ich bin mit der Navigation des Lernsystems zufrieden“ erforderten eine händische Inhaltsanalyse durch einen Menschen, der die Inhalte direkt oder aus dem situativen Kontext heraus angemessen zu interpretieren vermochte. Die in den Begründungstexten enthaltenen Hinweise zur Qualität des Lernsystems konnten erst nach diesem Zwischenschritt zu einer Ableitung von Hinweisklassen und Priorisierung von Verbesserungsmaßnahmen beitragen.

Mit zunehmender Anzahl von Expertenbefragungen kam es vor, dass die Experten sich in ihren Aussagen wiederholten. Im Umfeld von Expertenbefragungen ist die Redundanz von Hinweisen nicht überflüssig, sondern erwünscht. Redundante Hinweise sind ein wichtiger

Schlüssel zur Interpretation sprachlich unpräziser Informationen. Durch redundante Informationen können beispielsweise unvollständige, fehlerhafte Sätze, Stichpunkte etc. besser verstanden werden.

Die Begründungstexte wurden nach der Befragung vom Qualitätsregelteam gelesen mit dem Ziel, inhaltlich zusammenhängende Abschnitte zu bilden. Weiterhin wurden diese Abschnitte nach ihrem in Bezug zum Lernsystem positiv, neutral oder negativ klingenden Inhalt klassifiziert. Das Verhältnis von Lob und Tadel ging in die Bildung eines Indikators zur Messung der Qualität des Lernsystems ein. Die Qualitätsindikatoren in Form von Kennzahlen halfen, Verbesserungsmaßnahmen zu entscheiden und diese nach Dringlichkeit umzusetzen. Hauptsächlich dienten sie jedoch zur Überprüfung der Anforderungserfüllung der Lernsystem-Prototypen (siehe UKap. 15.1 auf Seite 91).

Es kam vor, dass Experten Hinweise gaben, die über das Thema der Fragestellung hinausgingen oder sehr detailliert ausfielen, sodass sie besser zu anderen Qualitätskriterien passten. In diesem Fall wurden die Antworten neu zugeordnet. Die Zuordnungen mussten sich nicht unbedingt auf die abgefragten Qualitätskriterien beziehen, sondern hier konnten auch andere Qualitätskriterien aus der Datenbank hinzu gezogen werden. Eine Zuordnung wurde durch eine spezielle Kennzeichnung des Begründungstextes markiert. Durch die Kennzeichnung konnten die Abschnitte der Begründungstexte von der LernSystemAnalyse-Software gezählt werden.

Nachfolgend findet sich eine Auflistung von Begründungstexten, die bereits in Informationsabschnitte gegliedert und die dem Qualitätskriterium Lernumfeld mit der Fragestellung „Inwieweit liegt eine Passung des Lernsystems zum gegebenen Lernumfeld vor?“ zugeordnet worden sind. Die Begründungstexte wurden von Experten nach Begutachtung des Planspiels Magnat-Schwebebahn abgegeben. Anschließend wurden die Begründungstexte nach inhaltlichen Gesichtspunkten gegliedert, d.h. in sinnvolle Informationsabschnitte aufgeteilt. Die in Klammern angegebenen Nummern hinter den Begründungstexten entsprechen den Identifikationsnummern der für den Inhalt verantwortlichen Autoren.

1. Das System lässt sich sicher gut in Seminaren verwenden und sollte auch relativ problemlos einsetzbar sein (6).
2. Gezielte Einsetzung als zusätzliche Übung zum Erlernten wird abgedeckt (6).
3. Gute Integration des Lernsystems in das Lernumfeld (5).
4. Würde zur Lehr- und Lernkultur passen (5).
5. Computer und Internet stehen zumeist zur Verfügung (5).
6. Ein Lernsystem in dieser Form habe ich im Maschinenbaustudium bisher noch nicht kennen gelernt (13).
7. Die Gewichtung habe ich sehr hoch gesetzt, da die Frage den Zugang der Studenten zum Lernsystem impliziert (12).

8. Das Planspiel harmonisiert wenig mit den vorhandenen universitären Strukturen (15).
9. Im Bereich der Gewichtung, ist die Passung des Lernsystems zum gegebenen Umfeld (hier der FB 15) nach meiner Ansicht nach durchschnittlich. Es ist nicht zwingend erforderlich das eine Passung vorliegt, da meiner Ansicht nach ein jeder das Lernen und Verstehen will, was ihm angeboten wird, solange dieses Themengebiet nicht gleich zu umfangreich und fächerübergreifend (transdisziplinär) dargestellt wird (20).
10. Bei der Einbettung in das Lernumfeld ist darauf zu achten, dass interessierte Lernende außerhalb des Fachbereichs nicht ausgeschlossen werden (20).
11. Das Planspiel hat fachübergreifendes Potenzial (20).
12. Die Einbindung in den gesamten Bereich Maschinenbau ist allein schon durch den Fakt gegeben, dass es um eine Schwebbahn geht. Was eine transdisziplinäre Einbindung betrifft, kann man noch mehr erreichen, wenn man ggf. die Erfahrungen und das Wissen von anderen Fachbereichen mit einbezieht (20).
13. Das Lernsystems eines Planspiels passt in erster Ansicht gar nicht in den Fachbereich Maschinenbau, da hier eher theoretisches Wissen vermittelt wird, als das praktisch auf Probleme eingegangen wird. Daher ist hier eine Umstellung nötig, da es ja nicht nur an den Studenten liegt, denen ein höherer Praxisbezug ihres Studiums recht wäre, ob das Planspiel Erfolg hätte. Wichtig ist ja auch die Bereitschaft von Professoren und Dozenten, die neu zur Verfügung stehenden Möglichkeiten entsprechend zu bearbeiten und anzubieten. Allerdings gibt es im Maschinenbau meiner Meinung nach diverse Ansatzpunkte, an denen die Anwendung eines Planspiels sinnvoll wäre, um einfach den gewünschten Praxisbezug zumindest in Ansätzen herzustellen (22).
14. Das Lernsystem halte ich für den Maschinenbaustudenten für sinnvoll und lehrreich, da es Einblicke und einen Einstieg in die Systemtechnik gewährt und den Studenten in der Informationstechnik schult (19).
15. Als Simulation einer realen Problematik in der man mehrere voneinander abhängigen Variablen optimieren muss, um ein Ziel zu erreichen, würde ich sagen, dass es auf jedem Fall zur Ausbildung von Ingenieuren sehr gut passt (14).
16. Das Lernsystem passt eigentlich gut zum Lernumfeld des Fachbereichs Maschinenbau (21).
17. Es werden die Information für das Planspiel weit verteilt. So ist es notwendig sich etwas „umzuschauen“ und nach den verschiedenen Lösungen zu schauen. Was ja den Lernmethoden einer Universität angepasst ist. Wobei dort (Uni) eher die Texte gegeben werden als die Diagramme (Funktionsmatrix) zum Verstehen der Eingangs- zu den Ausgangswerten. Doch ist es auch durch die Diagramme noch nicht für jüngere Schulklassen geeignet, da dort dann eine genauere Verknüpfung der einzelnen Werte gegeben sein müsste (21).

18. Das Lernsystem passt zum Lernumfeld, soweit Rechner und Browser vorhanden sind (3).

Diese Fülle an Informationen gilt es, durch die Einführung von Hinweisklassen auf die wesentlichen Aussagen zu reduzieren. Daher wurden alle Begründungstexte nach gleichen oder ähnlichen Inhalten analysiert. Gab es mehrere Begründungstexte, die zu einem bestimmten qualitätsrelevanten Thema Stellung nahmen, so wurde vom Qualitätsregelteam eine Hinweisklasse gebildet, die die Begründungstexte im wesentlichen zusammenfasst mit der Angabe einer Referenz auf die Bewerter. Hinweisklassen dienen dazu, die Informationen aus verschiedenen Quellen in Bezug auf einen angesprochenen Qualitätsaspekt zu bündeln. Dadurch wird die Informationsmenge besser für die Auswertung handhabbar. Bei der Ableitung von Hinweisklassen musste das Qualitätsregelteam sensibel vorgehen, damit durch die Zusammenfassung keine für die Qualitätsmessung relevanten Informationen verloren oder umgedeutet werden.

Die abgeleiteten Hinweisklassen zu den obigen Begründungstexten lauten:

1. Das Lernsystem lässt sich gut in das Lernumfeld integrieren (3, 5, 6, 14, 19, 20, 21).
2. Die Verwendung eines Planspiels im Maschinenbaustudium ist sicherlich ungewöhnlich und bedarf einer erkennbaren Anknüpfung an die Lerninhalte der Vorlesung (6, 13, 15, 20, 22).
3. Das Planspiel hat auch fächerübergreifendes Potenzial (20, 22).
4. Für den Schulbereich sollten die Texte und die Funktionsmatrix zum einfacheren Verständnis aufbereitet sein (21).

Die Zahlenwert-Referenzen am Ende einer Hinweisklasse geben die Identifikationsnummern der Experten an, deren Begründungstexte zur Bildung der Hinweisklasse geführt haben. Bei der zuletzt aufgeführten Hinweisklasse handelt es sich, um eine Monopol-Hinweisklasse, da nur ein Experte zu diesem Hinweis beigetragen hat. Dies ist an der Zuordnung von nur einer Identifikationsnummer eines Experten erkennbar. Über die zugeordneten Referenzen ist es möglich, (Monopol-) Hinweisklassen automatisch zählen zu lassen für das Evaluationscontrolling (siehe beispielweise UKap. 15.3 auf Seite 97ff.);

Zusätzlich zu der Expertenbewertung findet eine Bewertung des Lernsystems durch das Qualitätsregelteam statt. Dieses formuliert und priorisiert Handlungsableitungen für eine bessere Erfüllung von Anforderungen. Die Handlungsableitungen entsprechen einer Bewertung auf der Grundlage der über die Qualitätskriterien zugeordneten Hinweisklassen und der Erfahrung des Qualitätsregelteams.

Folgende Handlungsableitungen wurden zu den obigen Hinweisklassen entwickelt:

- Es wurde ein Planspiel entworfen, welches das systemtechnische Anwendungsbeispiel Magnetschwebbahn aus der Vorlesung Systemtechnik 1 aufgreift. Im Planspiel können die in der Vorlesung erworbenen Kenntnisse zur Systemtechnik praktisch angewendet werden. Der betreuende Tutor muss gegebenenfalls bestimmte Techniken für bestimmte Aufgaben und Fragestellungen empfehlen.
- Das Planspiel ist für die Realisierung einer kooperativen Ausbildung konzipiert. Die Selbststeuerung würde viele Lernende überfordern und weniger der Plan als das Spiel würden im Vordergrund stehen. In einem sogenannten Blended Learning-Ansatz (Kombination von Neuen Medien und Präsenzlehre) muss der Tutor die Studenten zur Verinnerlichung der Lerninhalte motivieren und coachen. Auf diese Weise können die Lerninhalte anschaulich über die Planspielmethodik vermittelt werden.
- Es ist nicht damit zu rechnen, dass nach dem Finden einer Lösungsstrategie sich die Lernenden weiterhin mit dem Planspiel beschäftigen. Dies ist für die Erreichung der Lernziele, die man mit der Planspielmethode angeht auch nicht notwendig. Die Komplexität des Modells ist für zwei Lehrstunden ausreichend. Um eine flexible Anpassung an andere Unterrichtsszenarien zu gewährleisten, sollten bei einer Revision Funktionalitäten angeboten werden, die eine Abwandlung der Modelldaten ermöglichen.

Die Expertenbewertung (Bewertungspunkte und Begründungstexte) und die Qualitätsregeltteam-Bewertung (Handlungsableitungen) führen zu der Bestimmung der Anforderungserfüllung (siehe hierzu Qualitätsindikatoren in Kap. 15 auf der nächsten Seite). Diese Kennzahl zur Qualität des Lernsystems dokumentiert den Fortschritt der Entwicklung, und sie ermöglicht den Vergleich von Prototypen.

Eine Anforderung zum Planspiel Magnet-Schwebbahn wird hier als Beispiel angegeben:

Das Lernsystem soll inhaltlich zu der Vorlesung Systemtechnik 1 passen und ergänzend den Erwerb von impliziten Wissen fördern, das durch die Präsenzlehre nur zum Teil kommunizierbar ist.

In der Vorlesung lernen die Studenten theoretisch Techniken kennen, komplexe Systeme zu betrachten, zu verstehen, zu beschreiben, zu bewerten und zu beherrschen. Das Lernsystem soll an dieses in der Vorlesung vermittelte Vorwissen anknüpfen und zum Aufbau praktischer Problemlösungskompetenz anhand des Fallbeispiels Magnetschwebbahn führen. Ziel dieses Lernsystems ist, dem Lehrenden ein didaktisches Medium zu bieten, mit dem die Studenten unter Anleitung die erworbenen Kenntnisse aus der Vorlesung vertiefen, anwenden und übertragen können. Ein Tutor soll dabei die Lehrziele verdeutlichen und den Studenten bei der Anwendung von systemtechnischen Methoden hilfreich zur Seite stehen.

In der Abb. 20 sind die genannten Arbeitsschritte in zeitlicher Reihenfolge dargestellt.

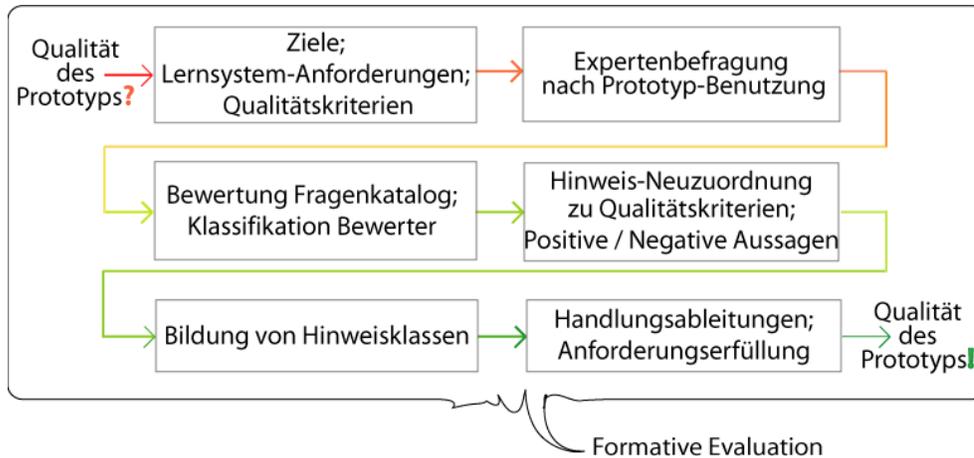


Abb. 20: Qualitätskriterienbasierte Analyse

## 15 Auswertung

Bevor die Fragestellungen aus UKap. 13.2 auf Seite 82 gestützt durch die Studienergebnisse beantwortet werden, sind zunächst einige Begriffe und deren Abkürzungen einzuführen. Diese werden für das Verständnis der nachfolgenden Unterkapitel vorausgesetzt. Weitere Begriffe sind im Kontext der Unterkapitel erläutert.

**Id, E<sub>Anz</sub>** Als Experten werden alle Personen bezeichnet, die sich intensiv mit dem Lernsystem auseinandergesetzt haben und mit Hilfe des LSA-Fragebogens das Lernsystem bewertet haben. Den Experten wird eine Identifikationsnummer (Id) zugeteilt, die eine anonymisierte Auswertung ermöglicht. Für die Expertenanzahl wird die Abkürzung E<sub>Anz</sub> vereinbart.

**Lernsystem LS** Evaluiert wurden drei Lernsysteme zur Fuzzy-Logik (**FLCBT**, **FLWBT**, **FLKran**), ein Planspiel (**PS**) und ein Planspiel-Autorensystem (**PAS**) zur Systemtechnik (siehe Kap. 12 auf Seite 68).

**Korrelationskoeffizient Kor** Der Korrelationskoeffizient gibt an, wie stark eine Beziehung zwischen zwei oder mehr quantitativen statistischen Variablen ist.

**Signifikanz p** Der p-Wert steht für die Wahrscheinlichkeit für das Zutreffen der Gegenhypothese. Den p-Wert erhält man durch einen Signifikanztest. Für die Annahme einer Hypothese  $H_0$  wird ein Wert  $\alpha$  als Signifikanzniveau festgelegt. Überschreitet der p-Wert dieses Signifikanzniveau, muss die Hypothese  $H_0$  verworfen werden.

## 15.1 Qualitätsindikatoren

**Frage:** Eignen sich die Qualitätsindikatoren PN, HA und GB als Kennzahlen zur Anforderungserfüllung von Lernsystemen?

**Begrifflichkeit:**

**Qualitätsindikatoren Qi** Um verschiedene Lernsystem-Prototypen besser miteinander vergleichen und die Qualität dieser abschätzen zu können, werden Qualitätsindikatoren benötigt. Für diese Aufgabe wurden drei Qualitätsindikatoren auf der Grundlage der Bewertungen zu den untersuchten Lernsystemen berechnet.

**Begründungstexte Bg** Freie Antworten der Experten zur Begründung der Bewertung (siehe UKap. 15.3.1 auf Seite 98).

**PN** Der Qualitätsindikator  $PN$  bezieht sich auf die in den Begründungstexten enthaltene Kritik. Diese wird für jedes Qualitätskriterium ( $Q_k$ ) differenziert nach positiv ( $P$ ) und negativ ( $N$ ) klingendem Inhalt gezählt, um das prozentuale Verhältnis  $PN_{Q_k} = \frac{P \cdot 100\%}{P+N}$  zu bilden. Ein  $PN_{Q_k}$ -Verhältnis von 50% bedeutet demzufolge, dass genauso viel positive wie negative Kritik in den Begründungstexten der Experten zu einem Qualitätskriterium gezählt wurde.

Zu jeder Anforderung eines Lernsystems wurden Qualitätskriterien vereinbart, durch die die Anforderungserfüllung gemessen werden soll. Zur Berechnung des  $PN_{\text{Anforderung}}$ -Verhältnisses werden gemäß dieser Vereinbarung die entsprechenden  $PN_{Q_k}$ -Verhältnisse arithmetisch gemittelt:  $PN_{\text{Anforderung}} = \overline{PN_{Q_k}}$ . Sind dieselben Qualitätskriterien mehrfach den Anforderungen zugewiesen, entspricht dies einer Gewichtung durch mehrfach eingehende  $PN_{Q_k}$ -Verhältnisse. Der  $PN$ -Qualitätsindikator ist der arithmetische Mittelwert der  $PN_{\text{Anforderung}}$ -Verhältnisse über alle Anforderungen:  $PN = \overline{PN_{\text{Anforderung}}}$ .

**HA** Das Qualitätsregelteam verfügt über detailliertes Projektwissen und ist in der Lage, auf mehrere Prototypen und Projekte zurückzublicken. Daher wird das Qualitätsregelteam zu den Lernsystemen mit Berücksichtigung der Anforderungen befragt. Das Qualitätsregelteam vermerkt dabei positive ( $P_{HA}$ ) und negative ( $N_{HA}$ ) Handlungsableitungen zum Lernsystem. Eine positive Handlungsableitung bedeutet, dass sich der Prototyp bezogen auf einen Qualitätsaspekt bewährt hat und hier keine Verbesserungsmaßnahme empfohlen wird. Eine negative Handlungsableitung signalisiert hingegen, dass der Prototyp bezüglich eines Qualitätsaspekts verbessert werden muss. Die gezählten positiven und negativen Handlungsableitungen führen zur Bildung des folgenden Verhältnisses:  $PN_{HA} = \frac{P_{HA} \cdot 100}{P_{HA} + N_{HA}}$ . Der Qualitätsindikator  $HA$  entspricht dem arithmetischen Mittelwert der  $PN_{HA}$ -Verhältnisse über alle Anforderungen:  $HA = \overline{PN_{HA}}$ .

**GB** Die Experten haben die Aufgabe, mehrere durch Qualitätskriterien beschriebene Qualitätsaspekte im Hinblick auf das Lernsystem zu gewichten und die Umsetzung im Lernsystem zu bewerten. Die Bewertung und Gewichtung erfolgt anhand einer 5-Punkt-Skala. Die Gewichtungs- ( $G$ ) und Bewertungspunkte ( $B$ ) der  $n$  Experten zu einem Qualitätskriterium werden arithmetisch gemittelt. Die gewichteten Bewertungen eines Qualitätskriterium  $GB_{Q_k} = \frac{\sum_{i=1}^n G_i \cdot B_i}{n}$  werden den zugeordneten Anforderungen zugewiesen und arithmetisch gemittelt:  $GB_{Anforderung} = \overline{GB_{Q_k}}$ . Der Qualitätsindikator  $GB$  entspricht dem Mittelwert der gewichteten Bewertungen über alle Anforderungen:  $GB = \overline{GB_{Anforderung}}$ .

**Statistische Methode:** Korrelationsanalyse Pearsons Produkt-Moment mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme 1:** Wenn alle Qualitätsindikatoren die Anforderungserfüllung desselben Lernsystem-Prototyps anzeigen, so sollten sie untereinander korrelieren.

**Annahme 2:** Die Fuzzy-Logik-Lernsysteme (FLCBT, FLWBT, FLKran) sind in der angegebenen Reihenfolge nacheinander entwickelt und evaluiert worden. Die Evaluationsergebnisse der Vorgänger standen den zeitlich nachfolgenden Prototypen zur Verfügung. Mit jedem neuen Prototyp wurden die Anforderungen überarbeitet, und es wurden falls erforderlich neue Anforderungen definiert. Mit jeder Neuentwicklung wurde das Ziel verfolgt, die neuen Anforderungen besser zu erfüllen als es der Vorgängerprototyp in Bezug auf die alten Anforderungen tat.

Es wird daher erwartet, dass die Qualitätsindikatoren mit den Lernsystemen FLCBT, FLWBT und FLKran ansteigen, um den Projekterfolg zu dokumentieren. Der Projekterfolg wird dadurch begründet, dass von Prototyp zu Prototyp folgende wesentliche Verbesserungen umgesetzt wurden:

**FLCBT zu FLWBT:** Herstellerunabhängigkeit und Portabilität durch Webtechnologie; Interaktive Java-Applets zur Veranschaulichung von Lerninhalten.

**FLWBT zu FLKran:** Verzicht auf Textlektionen; Interaktive Simulation.

**Annahme 3:** Die beiden Lernsystem-Prototypen PS und PAS sind sehr ähnlich. Die in beiden Lernsystemen implementierte Planspiel-Anwendung unterscheidet sich im PAS von der des PS lediglich durch die Änderung des Seitenkopfes und die Verfügbarkeit einer Auswertungsgrafik. Mit dem PAS ist es möglich, eigene Planspielmodelle für die Simulation zu entwerfen, während das PS bereits ein vollständiges Modell beinhaltet. Ansonsten ist die Gestaltung des PAS mit der des PS vergleichbar.

Es könnte daher erwartet werden, dass das PAS durch seine größere Flexibilität für unterschiedliche Anwendungsbereiche besser bewertet wird. Das Ergebnis des PAS wird jedoch vermutlich dadurch belastet, dass die Schwächen des Planspiels auch gleichzeitig im PAS auftreten. Daher wird erwartet, dass die Anforderungserfüllung

des PS im Vergleich zum PAS und FLKran in etwa gleich oder geringer bewertet wird. Dies ist dadurch begründet, dass das PS, das PAS und das FLKran hochinteraktive Lernsysteme mit einem konstruktivistischen Lehransatz sind. Durch diesen Lehransatz können die Lerngruppen einen spielerischen Ehrgeiz entwickeln, was sich positiv auf die Lernmotivation auswirkt. Dies wurde beim Einsatz der Lernsysteme in Lehrveranstaltungen beobachtet. Die Qualitätsindikatoren sollten dies durch eine höhere Anforderungserfüllung gegenüber den behavioristischen bzw. kognitivistischen Lehransätzen des FLCBT und FLWBT bestätigen.

LS	PN	HA	GB	Qi
FLCBT	31%	48%	59%	46%
FLWBT	41%	60%	69%	57%
FLKran	56%	84%	78%	73%
PS	67%	83%	69%	73%
PAS	61%	82%	69%	71%

Qi	Kor	p
PN und HA	0,96	0,01
GB und HA	0,79	0,12
PN und GB	0,61	0,27

Tab. 13: Qualitätsindikatoren

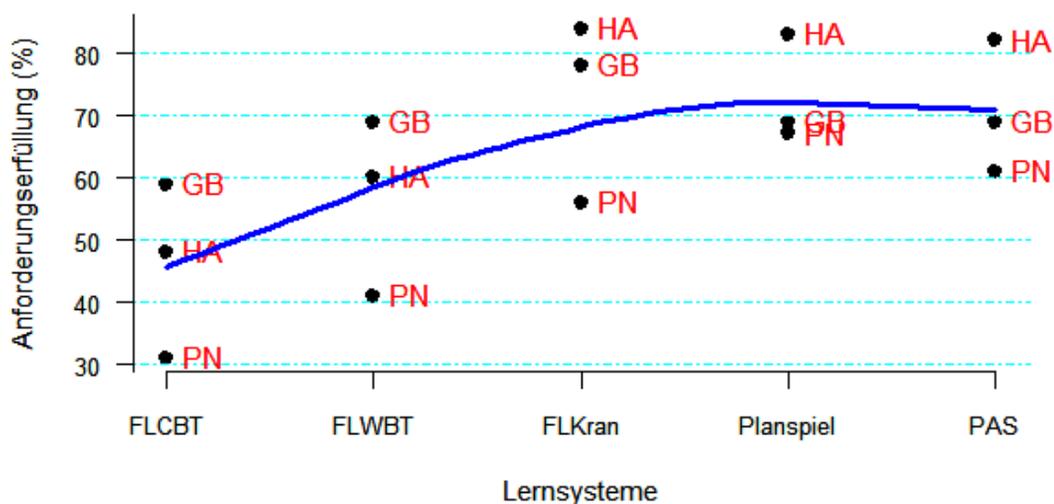


Abb. 21: Qualitätsindikatoren im Vergleich

## Schlussfolgerungen

**Zur Annahme 1:** Die Korrelationskoeffizienten weisen auf eine vorhandene ausgeprägte Korrelation zwischen den drei Qualitätsindikatoren hin. Ein stark signifikanter Zusammenhang ( $p = 0,01$ ) konnte durch die Korrelationsanalyse lediglich zwischen den Qualitätsindikatoren PN und HA nachgewiesen werden (siehe Tab. 13 auf der vorherigen Seite).

**Zur Annahme 2:** Die steigenden Werte der Anforderungserfüllung aller Qualitätsindikatoren aus Tab. 13 für die Lernsysteme FLCBT, FLWBT und FLKran bestätigen die Annahme, dass mit jeder Revision der FL-Lernsystem-Prototypen eine Verbesserung stattgefunden hat.

**Zur Annahme 3:** Die PS- und PAS-Anforderungserfüllung fallen recht ähnlich aus. Im Grunde ist das Ergebnis des PAS durchaus positiv zu bewerten, denn schließlich wird es mit einem für die Systemtechnik-Lehre zugeschnittenen Planspiel verglichen. Die Modellbildung des PS benötigte einen Arbeitstag. Für das Kennenlernen des PAS und die eigene Modellbildung nahmen sich die Bewerter durchschnittlich etwa eine Stunde Zeit. Das Planspielmodell des PS wurde mit einem achtfach höheren Zeitaufwand entwickelt. Die Experten hatten vermutlich nicht genügend Zeit, um sich ausgehend von der gestellten Modellbildungsaufgabe weitere lernwirksame Modelle und Einsatzfelder zu überlegen. Außerdem wurden überwiegend Studierende ohne selbsttätige Lehrerfahrung befragt. Diese Gründe sind vermutlich die Ursache, warum das PAS keine höhere Anforderungserfüllung als das PS erzielen konnte.

**Fazit** Die Qualitätsindikatoren GB, PN und HA eignen sich für die Messung der Anforderungserfüllung, da sie alle drei den aufgestellten Annahmen entsprechen. Bei der GB konnte allerdings keine signifikante Korrelation zu den anderen Qualitätsindikatoren nachgewiesen werden. Jedoch ist der Verlauf der GB im Vergleich zu den Verläufen der anderen zwei Qualitätsindikatoren quantitativ nicht völlig anders gelagert (siehe Abb. 21 auf der vorherigen Seite).

Weder die Experten noch das Qualitätsregelteam neigen im Mittel dazu, ausschließlich positive oder negative Kritik zum Lernsystem abzugeben. Das Verhältnis ist vielmehr von der Qualität des Lernsystems abhängig.

Mit den drei Qualitätsindikatoren als Kennzahlen ist ein Vergleich der Qualität von Lernsystem-Prototypen in einem formativen Evaluationskontext möglich. Die Berechnung der Qualitätsindikatoren hat den Vorteil, dass sie schnell durchgeführt werden kann. Qualitätsindikatoren können auch als Kennzahlen für die Konzentration von Qualitätssicherungsaktivitäten auf einzelne nicht erfüllte Anforderungen dienen.

## 15.2 Zeitaufwand

**Frage:** Wie hoch ist der benötigte Zeitaufwand für eine Expertenbefragung?

**Begrifflichkeit:**

**Zeitaufwand** Der Zeitaufwand betrifft die Zeit für die Vorbereitung, Bewertung und Auswertung von Expertenbefragungen.

**Vorbereitungszeit** Für die Vorbereitung und Administration einer Expertenbefragung wird ca. eine Stunde veranschlagt.

**Bewertungszeit** Der Zeitaufwand für den Experten setzt sich aus der Dauer der Auseinandersetzung mit dem Lernsystem-Prototypen und der Zeit für die anschließende Beurteilung anhand von Qualitätskriterien zusammen.

**Auswertungszeit** Es werden ca. 3 Stunden für die Auswertung der Expertenbefragung benötigt.

**Spannweite, Max, Min, Median, Mittelwert** Die Spannweite (engl. Range) ist definiert als Differenz zwischen dem Maximalwert Max und dem Minimalwert Min der von den Experten benötigten Bewertungszeit. Der Median, auch Lagemaß genannt, ist der mittelste Wert nach einer Rangordnung von Werten. Der Wert des Mittelwerts entspricht der Summe aller Zeitaufwände geteilt durch die Anzahl der Experten (arithmetischer Mittelwert). Die in Tab. 14 auf der nächsten Seite zur Bewertungszeit eingetragenen Zahlenwerte haben die Einheit Stunden pro Experte.

**F** Mittlere Fragen- bzw. Qualitätskriterienanzahl pro Experte. Bei den Lernsystemen FLKran und PAS wurden keine antwortabhängigen Zusatzfragen freigeschaltet. Daher entspricht die Fragenanzahl den vier Qualitätskriterien der Lernsystem-Konfiguration.

**Statistische Methoden:** Deskriptive Statistik, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme 1:** Die verwendete LernSystemAnalyse-Software ermöglicht eine automatische Generierung von webbasierten Fragebögen. Die Webtechnologie ermöglicht eine gleichzeitige Befragung von mehreren Experten.

**Annahme 2:** Es muss damit gerechnet werden, dass die freie Beantwortung der Fragen für die Experten erheblich länger dauert als die Auswahl von standardisierten Antwortmöglichkeiten.

LS	Fragen ( $\bar{F}$ )	Bewertungszeit in Stunden				
		(Spannweite)	(Min)	(Max)	(Median)	(Mittelwert)
FLCBT	15,8	8	1	9	4	3,8
FLWBT	20,1	3	1	4	2	2,7
FLKran	4	2	1	3	1	1,3
PS	15,7	3	1	4	2	2,3
PAS	4	2	1	3	1	1,5

Tab. 14: Zeitaufwand

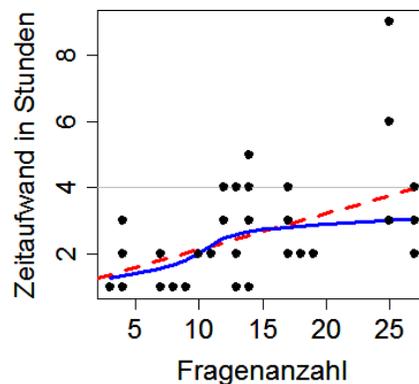


Abb. 22: Korrelation Fragenanzahl – Bewertungszeit

## Schlussfolgerungen

**Zur Annahme 1:** Die Auswertungszeit beträgt in der Regel eine halbe Stunde pro Experte. Diese Zeit ist vorwiegend für die Inhaltsanalyse der freien Antworten nötig. Eine Evaluation mit 6 Experten ist somit innerhalb eines Tages (8 Stunden) durchzuführen. Daher können von einem Tag auf den nächsten Tag bereits die von der Evaluation abgeleiteten Gestaltungsmaßnahmen zur Verbesserung des Prototypen umgesetzt werden. Die Evaluationsmethodik ist somit geeignet für die formative Qualitätssicherung.

**Zur Annahme 2:** Die Beantwortung der Fragen dauert zwar lange (etwa 6 Fragen/Stunde), aber sie ist gerade durch die freie Antwortmöglichkeit der Experten von besonderem Wert für das Qualitätsregelteam. Bei einer Empfehlung einer angemessenen Fragenanzahl ist nicht nur der Zeitaufwand zu berücksichtigen, sondern auch die durch die Fragen erhobenen Hinweise zur Qualität des Lernsystems. Mit diesem Aspekt beschäftigt sich das Unterkapitel 15.3.3 auf Seite 101.

**Fazit:** Die Untersuchung zeigt eine hochsignifikante ( $p=1,094e-06$ ) Korrelation zwischen Fragenanzahl und Zeitaufwand eines Experten (siehe Abb. 22 auf der vorherigen Seite). Aus der Lage der Messwerte aus Abb. 22 auf der vorherigen Seite wird geschlossen, dass diese Korrelation nicht unbedingt linear verläuft, sondern vermutlich eher s-förmig wie die eingezeichnete Anpassungskurve. Bei dieser Anpassungskurve ist deutlich zu erkennen, dass ab einer Fragenanzahl von 12 die Steigung deutlich flacher ausfällt. Die überwiegende Anzahl von Experten teilt sich mit zunehmender Fragenanzahl offenbar ihre Bewertungszeit so ein, dass sie den Grenzwert von vier Stunden nicht überschreitet. Abb. 22 zeigt, dass nur in extremen Ausnahmefällen eine Befragung länger als vier Stunden dauern kann. Daher sind endlose Bewertungen, verursacht durch eine zu hohe Fragenanzahl nicht zu erwarten. Für den Bereich bis 12 Fragen gilt uneingeschränkt die Annahme, dass im Mittel über alle Experten jede Frage zu einer äquidistanten Erhöhung der Bewertungszeit führt.

### 15.3 Hinweisklassen

**Frage:** Welche Einflüsse wirken sich auf die Anzahl der Hinweisklassen aus?

**Begrifflichkeit:**

**Hinweisklassen Hk** Offene Antwortmöglichkeiten sind nach Hahne (2003, S.118) schwierig auszuwerten und schlecht zu vergleichen. Daher wurden vom Qualitätsregelteam thematisch ähnliche Begründungstexte von verschiedenen Experten zu Hinweisklassen zusammengefasst.

**Annahme 1:** Als wesentlicher Einfluss auf die Anzahl der Hinweisklassen wird die Anzahl der Begründungstexte zu den gewichteten und bewerteten Qualitätskriterien angenommen. Je mehr Begründungstexte von den Experten verfasst werden, desto mehr werden darauf bezogene Hinweisklassen vom Qualitätsregelteam formuliert. Anders ausgedrückt, ohne Begründungstexte existieren auch keine abgeleiteten Hinweisklassen.

**Annahme 2:** Die Anzahl der **Begründungstexte** könnte vom **Interaktionsgrad** des Lernsystems, von der **Fragenanzahl**, vom **Expertengrad** der befragten Personen und von der **Anforderungserfüllung** des Lernsystems abhängen.

**Fazit:** In den folgenden Unterkapiteln werden die genannten Einflüsse aus den Annahmen 1 und 2 erläutert und untersucht.

### 15.3.1 Begründungstexte

**Frage:** Korreliert die Anzahl der Begründungstexte mit der Anzahl der Hinweisklassen?

**Begrifflichkeit:**

**Anzahl der Begründungstexte Bg** Die Experten wurden gebeten, ihre Bewertung und Gewichtung zu begründen, um die Bewertungsergebnisse reichhaltiger und nachvollziehbar zu halten. Die Texte, die die Experten in die freien Antworttextfelder des Fragebogens notiert haben, wurden vom Qualitätsregelteam gelesen und in thematische Abschnitte gegliedert. Die Begründungstexte wurden nach der Gliederung automatisch gezählt.

**Statistische Methoden:** Regressionsanalyse, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme:** Um möglichst den Einfluss von unterschiedlichen Personencharakteristika in engen Grenzen zu halten, wurden zwei Experten mit gleichem Expertengrad 4 ausgewählt, die alle 5 Lernsysteme (FLCBT, FLWBT, FLKran, PS, PAS) evaluiert hatten.

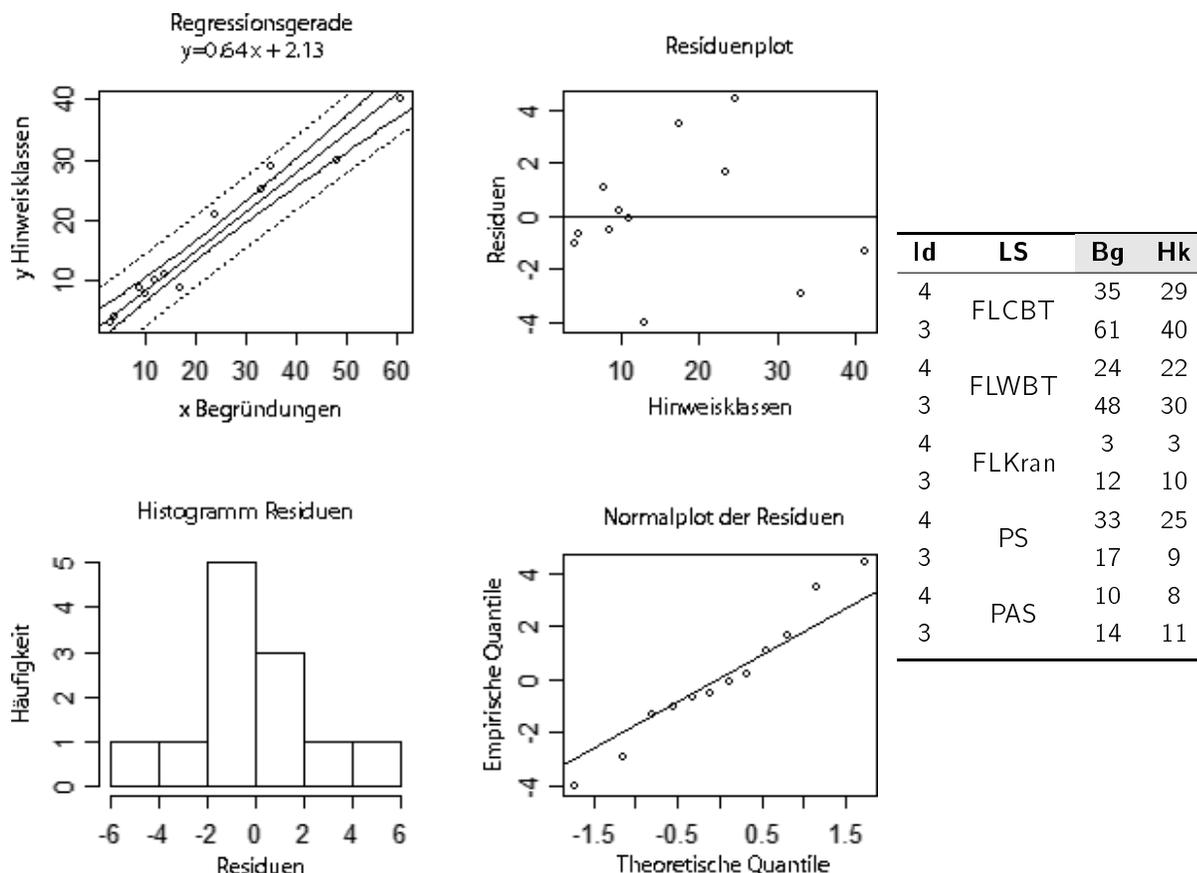


Abb. 23: Korrelation Begründungstexte – Hinweisklassen

## Schlussfolgerungen

**Zur Annahme:** Die Regressionsgerade in der Abb. 23 auf der vorherigen Seite gibt nur den Trend an. Die reale Modellgleichung muss hingegen durch den Nullpunkt gehen. Der Residuenplot zeigt die Abweichungen von der Regressionsgerade. Aus den Untersuchungsergebnissen lässt sich exemplarisch für die zwei Experten der hochsignifikante Zusammenhang ( $p = 4,745e-8$ ) ermitteln, dass je mehr Begründungstexte von den Experten geschrieben werden, desto mehr Hinweisklassen gebildet werden. Ähnliche Ergebnisse erhält man, wenn man die weiteren 34 befragten Experten auswertet, allerdings ist hier der Einfluss weiterer Variablen (z.B. der Erfahrungsgrad der Experten) nicht von dem Einfluss der Anzahl der Begründungstexte zu trennen.

**Fazit:** Aufgrund der Korrelation zwischen Begründungstexten und Hinweisklassen kann bei Kenntnis der Anzahl der Begründungstexte über die Regressionsgerade aus Abb. 23 auf der vorherigen Seite auf die Anzahl der Hinweisklassen geschlossen werden. Für die Untersuchung von weiteren Einflussgrößen auf die Anzahl der Hinweisklassen werden in den nachfolgenden Unterkapiteln diejenigen Rahmenbedingungen betrachtet, die zu einer Erhöhung der Anzahl von Begründungstexten führen.

### 15.3.2 Interaktionsgrad

**Frage:** Korreliert der Interaktionsgrad eines Lernsystems mit der Anzahl der Begründungstexte?

#### Begrifflichkeit:

**Interaktionsgrad  $I_g$**  Lernsysteme unterscheiden sich stark hinsichtlich der integrierten interaktiven Lernobjekte. Das Gewicht der interaktiven Lernobjekte gegenüber anderen Lernobjekten wurde bei den untersuchten Lernsystemen vom Qualitätsregelteam anhand einer 4-Punkt-Skala bewertet. Punkte werden beispielsweise nicht für interaktive Navigationsobjekte vergeben, sondern für Lernobjekte, die die aktive Auseinandersetzung des Lernenden mit den Lerninhalten fördern. Eine geringe Punktzahl des Interaktionsgrades deutet darauf hin, dass eher eine passive Auseinandersetzung des Lernenden stattfindet. Dies ist beispielsweise der Fall, wenn der Lernende hauptsächlich Lerninhalte durch das Lesen von Lehrtexten internalisieren kann.

**B<sub>g</sub>** Anzahl der Begründungstexte (siehe UKap. 15.3.1 auf der vorherigen Seite)

**A<sub>e</sub>** Anforderungserfüllung eines Lernsystems (siehe UKap. 15.3.5 auf Seite 107)

**Statistische Methoden:** Regressionsanalyse, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme:** Um möglichst den Einfluss von unterschiedlichen Personencharakteristika in engen Grenzen zu halten, wurden zwei Experten mit gleichem Expertengrad 4 ausgewählt, die alle 5 Lernsysteme evaluiert hatten.

LS	lg	Ae
FLCBT	1	46%
FLWBT	2	55%
FLKran	4	65,5%
PS	3	71,6%
PAS	4	69,5%

Tab. 15: Korrelation Interaktionsgrad – Anforderungserfüllung

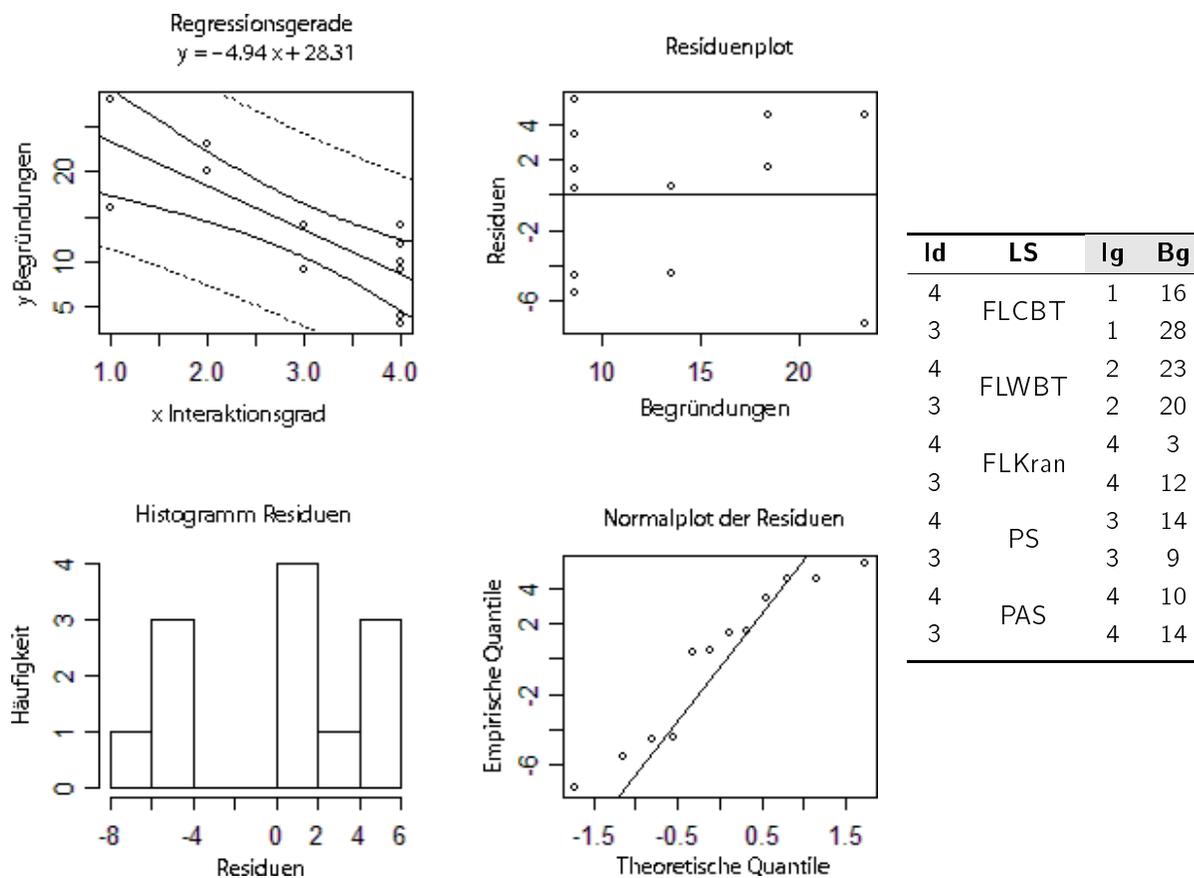


Abb. 24: Korrelation Interaktionsgrad – Begründungstexte

## Schlussfolgerungen

**Zur Annahme:** Bei dem Einfluss des Interaktionsgrades auf die Anzahl der Begründungstexte handelt es sich um eine hochsignifikante Einflussgröße ( $p = 0,001663$ ).

**Fazit:** Die Korrelationsanalyse von Interaktionsgrad und der Anzahl der Begründungstexte zeigt, dass bei interaktiven Lernsystemen weniger Begründungstexte geschrieben werden. Unter Umständen liegt dies daran, dass die Benutzer eines hochinteraktiven Lernsystems zu weiten Teilen selbst am Lehrerfolg beteiligt sind. Sie müssten damit sich selbst bzw. ihre Lernstrategien bewerten. Gefragt wird im Rahmen des Qualitätskriterienkatalogs jedoch vorwiegend nach Aspekten, die das Lernsystem direkt betreffen. Die befragten Experten haben keine Selbstreflexion ihres Beitrags zum Erfolg des Lernsystems in den abgegebenen Begründungstexten erkennen lassen. Bei interaktiven Lernsystemen sollte daher für eine ganzheitliche Evaluation des Lernsystems, die Fragebogen- bzw. Interviewmethode durch Beobachtungsmethoden ergänzt werden.

Neben diesem Einfluss ist signifikant ( $p = 0,048$ ;  $Kor = 0,88$ ), dass mit zunehmendem Interaktionsgrad die Lernsysteme die Anforderungen besser erfüllen (siehe Tab. 15 auf der vorherigen Seite). Damit sind interaktive Lernsystem-Entwürfe generell zu befürworten.

„Providing more interactivity in user interfaces appears to have a substantial positive effect on learning. Interactive interfaces allow

- learners to control, manipulate and explore material, or
- the computer to periodically ask learners to answer questions that help them to integrate the material,
- or both.

Interaction may be useful because it encourages learners to elaborate more. For this to occur, the interaction must be cognitively engaging (i.e., not merely selecting hyperlinks)“ (Human Factors International 2001).

### 15.3.3 Fragenanzahl

**Frage:** Korreliert die Fragenanzahl mit der Anzahl der Begründungstexte?

**Begrifflichkeit:**

**Fragenanzahl F** Die Expertenbefragungen werden mit einer unterschiedlichen Qualitätskriterienanzahl durchgeführt. Beim PAS und FLKran Lernsystem waren vier Qualitätskriterien der Lernsystem-Konfiguration vorgesehen. Bei den übrigen

Lernsystemen wurden hingegen je nach Bewertung und Gewichtung der Qualitätskriterien der Lernsystem-Konfiguration weitere Qualitätskriterien eingereiht. Jedes Qualitätskriterium enthielt eine Hauptfrage. Je nach Qualitätskriterienanzahl hatten die Experten eine entsprechende Fragenanzahl zu beantworten.

**Bg** Anzahl der Begründungstexte (siehe UKap. 15.3.1 auf Seite 98).

**Statistische Methoden:** Regressionsanalyse, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme:** Um möglichst den Einfluss von unterschiedlichen Personencharakteristika in engen Grenzen zu halten, wurden zwei Experten mit gleichem Expertengrad 4 ausgewählt, die alle 5 Lernsysteme evaluiert hatten.

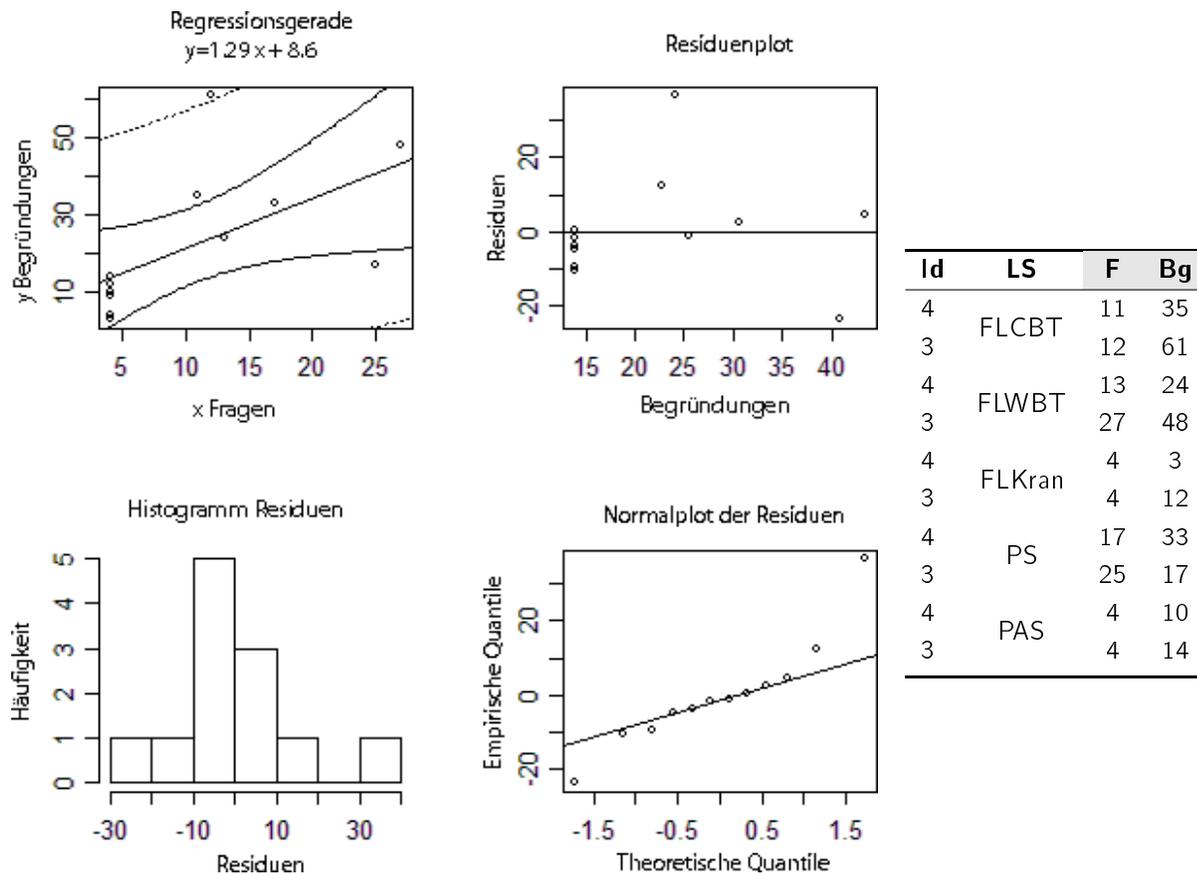


Abb. 25: Korrelation Fragenanzahl – Begründungstexte

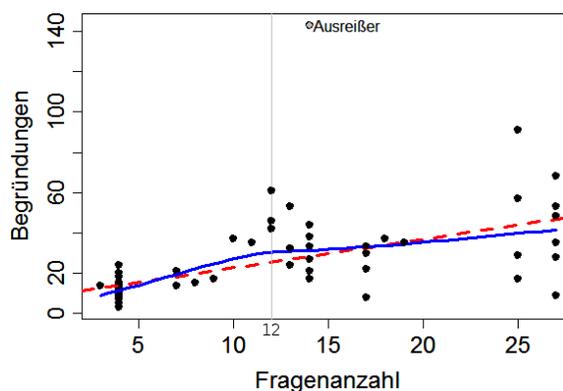


Abb. 26: Korrelation Fragenanzahl – Begründungstexte

### Schlussfolgerungen

**Zur Annahme:** Die These, dass mehr Fragen zu mehr Begründungstexten führen, ist signifikant ( $p = 0,04$ ).

**Fazit:** Interessant ist, dass die Anpassungskurve in Abb. 26 ähnlich wie die in Abb. 22 auf Seite 96 verläuft. Die Steigung der Kurve fällt in beiden Abbildungen bei einer Fragenanzahl von 12 geringer aus. Dieser Effekt ist darauf zurückzuführen, dass wie bereits angedeutet die Experten sich ihre Bewertungszeit für ein Qualitätskriterium einteilen, sodass sie nicht über den Grenzwert von vier Stunden kommen. Um die Gefahr einer oberflächigen Bewertung in Grenzen zu halten, ist eine Fragenanzahl mit freier Antwortmöglichkeit von über 12 nicht zu empfehlen.

#### 15.3.4 Expertengrad

**Frage:** Korreliert der Expertengrad mit der Anzahl der Begründungstexte?

**Begrifflichkeit:**

**Expertengrad Eg** In UKap. 13.1 auf Seite 78 wurde bereits angemerkt, dass sich Experten sehr wohl hinsichtlich ihrer Erfahrung unterscheiden können und daher für eine differenziertere Auswertung klassifiziert werden sollten. Je nach Vorwissen werden verschiedene Expertengrade unterschieden: 1:=Anfänger, 2:=Laie, 3:=Fortgeschrittener, 4:=Professional. Bei den Untersuchungen werden insbesondere der Expertengrad  $<3$  und  $>2$  unterschieden. Ein Wert  $>2$  bedeutet, dass der Experte über ausreichend Vorwissen im Bereich der Bewertung von Lern- und Softwaresystemen besitzt. In der Literatur wird häufig bei dieser Unterscheidung von Nicht-Experten und Experten gesprochen. Diese Begriffe sind sehr unglücklich gewählt, denn jeder Mensch ist zu einem gewissen Grad Experte in einer

Disziplin, sobald dieser Wissen über die Disziplin verfügt. Der Skalenwert, also der Expertengrad, wird auf der Grundlage der Vor- und Nachbefragung des Experten vom Qualitätsregelteam festgelegt.

**Zeichen** Mittlere Anzahl eingegebener Zeichen der Experten für Begründungstexte.

**Zeitaufwand** Mittlerer ganzzahliger Zeitaufwand der Experten für die Bewertung eines Lernsystems.

**Statistische Methoden:** Deskriptive Statistik, t-Test, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme 1:** Zu vermuten ist, dass Experten mit höherem Expertengrad bei der Bepunktung von Qualitätskriterien weniger voneinander abweichen als Experten mit weniger Erfahrung.

**Annahme 2:** Erwartet werden kann, dass die Experten mit höherem Expertengrad ein differierendes Bewertungspunktniveau haben als Experten mit weniger Erfahrung.

**Annahme 3:** Experten mit höherem Erfahrungsgrad werden vermutlich mehr Text schreiben und sich länger mit der Bewertungsaufgabe befassen als Experten mit weniger Erfahrung.

**Annahme 4:** Es wird angenommen, dass die Begründungstexte von Experten mit einem höheren Erfahrungsgrad zu mehr Hinweisklassen führen als die der Experten mit weniger Erfahrung.

LS	Eg	Zeichen	Zeitaufwand
FLCBT	> 2	4229	4
	< 3	3756	4
FLWBT	> 2	5313	3
	< 3	3543	3
FLKran	> 2	541	1
	< 3	1948	1
PS	> 2	2878	2
	< 3	4336	2
PAS	> 2	1589	1
	< 3	1623	2

Tab. 16: Bewertungsunterschiede bezüglich des Expertengrades

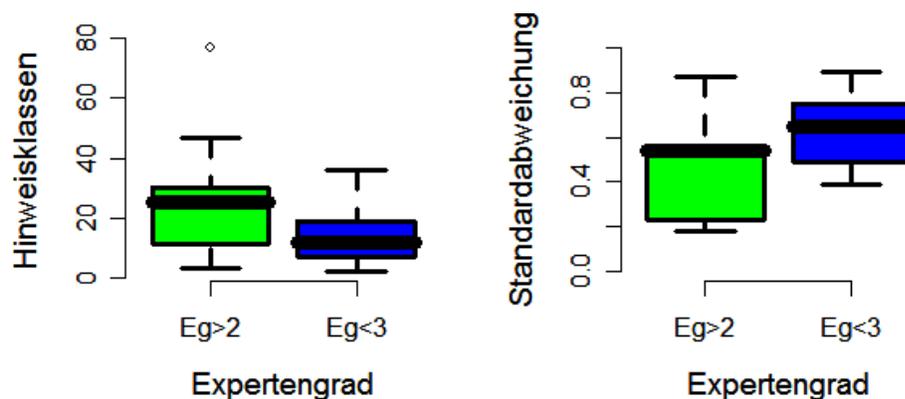


Abb. 27: Korrelation Expertengrad – Hinweisklassen und Standardabweichung der Bewertungspunktzahlen

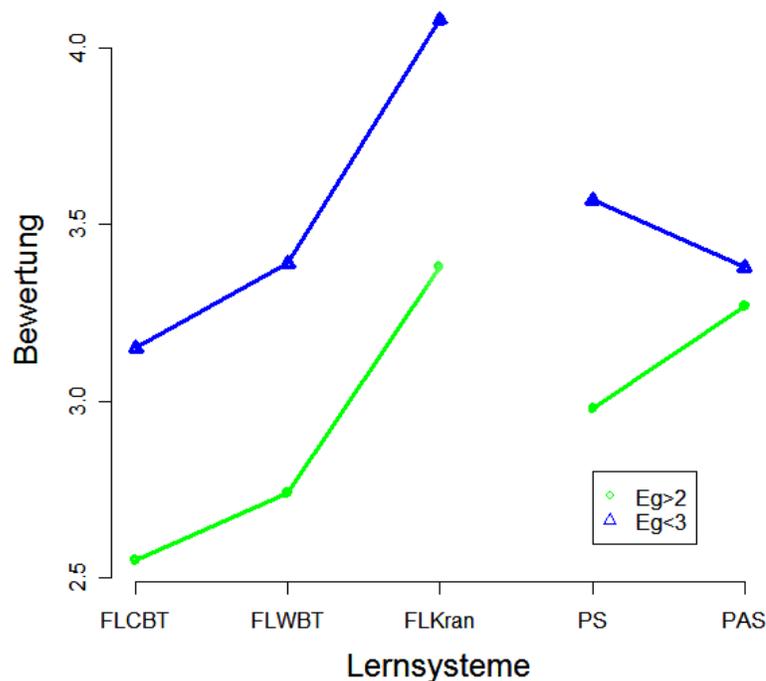


Abb. 28: Ungewichtete Bewertung

### Schlussfolgerungen

**Zur Annahme 1:** Die Annahme, dass Experten mit höherem Expertengrad weniger bei ihrer Bepunktung des Lernsystems voneinander abweichen, wird verworfen. Die Signifikanz beträgt  $p = 0,3314$ .

**Zur Annahme 2:** In Abb. 28 wurde das Bewertungsniveau je nach Expertengrad analysiert. Zu erkennen ist deutlich, dass das Niveau zur Bepunktung bei den Expertengruppen

etwa um einen Bewertungspunkt auseinander liegt (außer beim PAS). Der t-Test ergab einen p-Wert von 0,04322. Beim PAS haben die meisten Experten mit einem höheren Erfahrungsgrad offenbar die Nutzungsmöglichkeiten des PAS gegenüber dem PS erkannt. Die Experten mit weniger Erfahrung haben das PAS sogar schlechter bewertet als das PS.

Die Experten mit Erfahrungsgrad  $>2$  sind zurückhaltender mit der Vergabe von höheren Punktzahlen. Wenn die arithmetisch gemittelten Bewertungspunktzahlen normallisiert werden, liegen die beiden Kurven beinahe übereinander (außer beim PAS). Dies zeigt, dass die Experten unabhängig von ihrem Erfahrungsgrad sehr wohl in der Lage sind, Lernsysteme vergleichbar einzustufen. Dieser Sachverhalt erhöht die Validität der Bewertungsdaten für die Berechnung des Qualitätsindikators GB (siehe UKap. 15.1 auf Seite 91).

**Zur Annahme 3:** Es ist keine Regel aus den Versuchsergebnissen ableitbar, dass Experten mit höherer Erfahrung mehr Zeichen schreiben oder sich länger mit der Bewertungsaufgabe befassen ( $p = 0,9008$ ).

**Zur Annahme 4:** Aus den Begründungstexten lassen sich bei Experten mit einem höherem Erfahrungsgrad hochsignifikant ( $p = 0,001851$ ) mehr Hinweisklassen formulieren. Dies spricht sehr dafür, Experten mit möglichst hohem Expertengrad auszuwählen.

**Fazit:** Wie auch andere Studien (siehe UKap. 13.1 auf Seite 78) belegen, hat der Erfahrungsgrad der Experten durchaus einen Einfluss auf den Nutzen der Expertenbefragung. Letzlich muss das Qualitätsregelteam entscheiden unter Abwägung der genannten Vorzüge und der ökonomischen Zwänge, ob sich der Einsatz von Experten mit höherem Erfahrungsgrad rechnet. Die Ergebnisse der durchgeführten Evaluationen machen deutlich, dass Experten mit unterschiedlichem Erfahrungsgrad durchaus wertvolle Hinweise auf die Qualität eines Lernsystems geben können. Die befragten Studenten konnten teilweise durchaus mit den Leistungen höher qualifizierter Experten mithalten. Dies ist insofern sehr erfreulich, da im universitären Kontext sehr viele Studenten zur Verfügung stehen, die sich für Expertenbefragungen bereit erklären könnten. Studenten haben zudem den Vorzug, dass sie der Zielgruppe von Lernsystemen entsprechen.

Im Rahmen von Lehrveranstaltungsevaluierung gelangt Bagel (2001, S.9) belegt durch eine eigene Studie zu dem Schluss, dass Zweifel an der Gültigkeit studentischer Urteile unberechtigt seien: „Insgesamt sind die Studierenden sehr wohl in der Lage, differenzierte und konsistente Urteile in der Lehrbewertung zu treffen. Sie werfen keineswegs alles in einen Topf, sie können Stärken und Schwächen erkennen und mitteilen. Der Verdacht, studentische Stellungnahmen kommen eher einem bloßen Beliebtheitswettbewerb gleich, konnte nicht bestätigt werden. Ebenso wenig wurde bestätigt, dass Anforderungen und Strenge von Nachteil für die Lehrbeurteilung seien. Schließlich lie-

ßen sich keine Belege finden, die eine unzureichende Sachkenntnis der Studierenden vermuten lassen“.

### 15.3.5 Anforderungserfüllung

**Frage:** Korreliert die Anforderungserfüllung eines Lernsystems mit der Anzahl der Begründungstexte?

**Begrifflichkeit:**

**Qualitätsindikatoren  $Q_i$**  Die drei Qualitätsindikatoren PN, HA und GB resultieren aus den Bewertungen des Lernsystems (siehe UKap. 15.1 auf Seite 91).

**Anforderungserfüllung  $A_e = \bar{Q}_i$**  Aus den Qualitätsindikatoren wird der arithmetische Mittelwert berechnet. Das Ergebnis entspricht der Anforderungserfüllung:

$$A_e = \bar{Q}_i = \frac{\sum_{j=0}^2 Q_{ij}}{\sum_{j=0}^2 \text{sgn}(Q_{ij})} \quad (4)$$

Je größer der Wert eines Qualitätsindikators ist, desto höher wird die Anforderungserfüllung ausfallen. Die Signum-Funktion (sgn) ermöglicht es, einen Qualitätsindikator unberücksichtigt zu lassen und trotzdem die Anforderungserfüllung mit den übrigen Qualitätsindikatoren zu ermitteln. Dieser Fall kann eintreten, wenn in der Praxis die Auswertung zu einem Qualitätsindikator noch nicht verfügbar ist.

**Statistische Methoden:** Regressionsanalyse, Pearson Produkt-Moment Korrelation mit Signifikanzniveau  $\alpha = 0,05$ .

**Annahme 1:** Um möglichst den Einfluss von unterschiedlichen Personencharakteristika in engen Grenzen zu halten, wurden zwei Experten mit gleichem Expertengrad 4 ausgewählt, die alle 5 Lernsysteme evaluiert hatten.

**Annahme 2:** Ein geläufiger Spruch von Entwicklern ist, dass gutes Design nicht auffallen würde, schlechtes hingegen schon. Dies würde bedeuten, dass Experten eher dazu neigen überwiegend negative Kritik am Lernsystem zu üben.

**Annahme 3:** Es liegt die Vermutung nahe, dass je schlechter das Lernsystem die Anforderungen erfüllt, mehr Verbesserungshinweise gegeben werden. Dies würde zu einer Erhöhung der Anzahl der Begründungstexte führen.

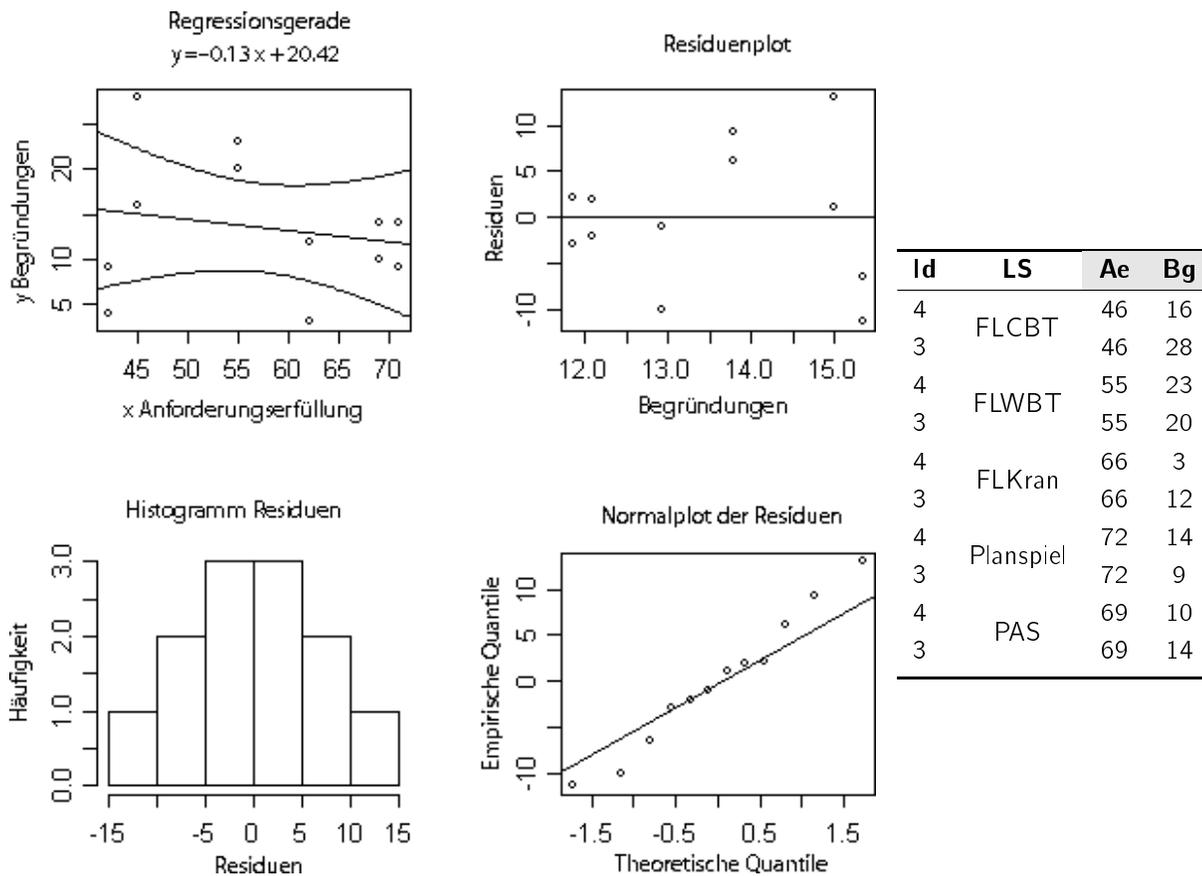


Abb. 29: Korrelation Anforderungserfüllung – Begründungstexte

## Schlussfolgerungen

**Zur Annahme 2:** Wie in Abb. 21 auf Seite 93 zu erkennen ist, treten beim Qualitätsindikator PN sehr wohl Verhältnisse über und unter 50% auf. Das bedeutet, dass die Experten nicht die guten Gestaltungselemente eines Lernsystems übersehen und überwiegend die schlechten Gestaltungselemente in den Begründungstexten ihrer Bewertung nennen.

**Zur Annahme 1 und 3:** Der Einfluss der Anforderungserfüllung auf die Anzahl der Begründungstexte wird nicht signifikant ( $p = 0,3463$ ) durch die Versuchsergebnisse bestätigt.

**Fazit:** Es wird daher geschlossen, dass die Anforderungserfüllung (:=Qualität eines Lernsystems) keinen bemerkenswerten Einfluss auf die Anzahl der Begründungstexte hat.

## 15.4 Hinweisklassen – Expertenanzahl

**Fragen:** Welche Korrelation besteht zwischen der Expertenanzahl und der Anzahl der Hinweisklassen? Lohnt sich ab einer bestimmten Expertenanzahl die Befragung von weiteren Personen im Hinblick auf das Aufwand-Nutzenverhältnis nicht mehr?

**Begrifflichkeit:**

**Hinweisklassen** Thematisch ähnliche Begründungstexte von verschiedenen Experten werden vom Qualitätsregelteam zu Hinweisklassen zusammengefasst.

**Monopol-Hinweisklasse** Liegt bei der Formulierung einer Hinweisklasse nur eine Expertenbegründung zugrunde, wird die Hinweisklasse als Monopol-Hinweisklasse spezifiziert.

**F** Fragen- bzw. Qualitätskriterienanzahl (siehe UKap. 15.3.3 auf Seite 101)

**Ig** Interaktionsgrad des Lernsystems (siehe UKap. 15.3.2 auf Seite 99)

**Eg** Expertengrad der befragten Personen (siehe UKap. 15.3.4 auf Seite 103)

**E<sub>Anz</sub>** Expertenanzahl (siehe UKap. 13.1 auf Seite 78)

**Statistische Methoden:** Permutation, Regressionsanalyse.

**Annahme 1:** Die Vorteile der webbasierten schriftlichen Befragung liegen in einer wenig aufwändigen Durchführung, die in allen Lebensphasen eines Lernsystems stattfinden kann. Der größte Kostenfaktor ist die Rekrutierung der Experten. Die Einplanung von vielen Experten verteuert das Verfahren. Daher ist die Expertenanzahl möglichst im Hinblick auf den Aufwand und den Nutzen zu minimieren. Zu wenig Experten führen jedoch dazu, dass die Ergebnisse statistisch sehr vage sind und/oder zu wenig Hinweisen zur Qualität des Lernsystems führen.

**Annahme 2:** Die Anzahl von Hinweisklassen ist nicht von der Qualität des Lernsystems abhängig (siehe UKap. 15.3.5 auf Seite 107), sondern von der Evaluationsmethode, dem Interaktionsgrad des Lernsystems, der Fragenanzahl des Qualitätskriterienkatalogs und dem mittleren Expertengrad. Dies bedeutet, dass bei Kenntnis dieser Einflussgrößen die Anzahl der zu erwartenden Hinweisklassen im Vorfeld einer Befragung abgeschätzt werden. Diese Information hilft insbesondere bei der Festlegung der benötigten Expertenanzahl.

**Annahme 3:** In Abb. 30 auf der nächsten Seite sind die kumulierten Hinweisklassen in Bezug zur Expertenanzahl pro untersuchtem Lernsystem aufgetragen. Um Reihenfolgeeffekte von Experten zu vermeiden, wurden die Hinweisklassen mit allen möglichen Personen-Reihenfolgen berechnet. Die arithmetischen Mittelwerte aller kumulierten Hinweisklassen zu einer bestimmten Expertenanzahl und einem Lernsystem ergaben die Funktionsverläufe in der Abbildung. Diese beschreiben den wahrscheinlichsten Verlauf einer Befragung.

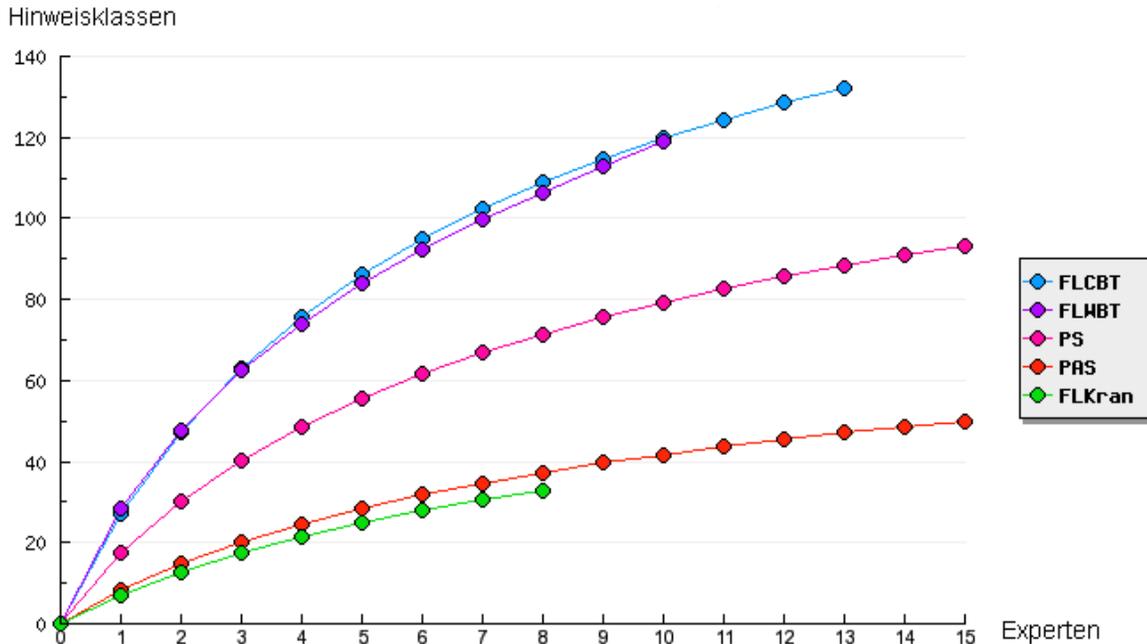


Abb. 30: Kumulierte Hinweisklassen im Prototypvergleich

**Zur Annahme 1:** Aus Abb. 30 lässt sich entnehmen, dass bereits mit 5 Experten ein Drittel der durch die Expertenbefragung möglichen Hinweisklassen abgeleitet werden. Für die restlichen zwei Drittel sind circa 30 Experten notwendig. Daher ist die Expertenbefragung nur mit wenigen Experten effizient.

**Zur Annahme 2 und 3:** Der exponentielle Verlauf der kumulierten Hinweisklassen in Abhängigkeit von der Anzahl der Experten  $E_{Anz}$  ist bei jedem Lernsystem-Projekt ähnlich. Unterschiedlich ist hauptsächlich die Amplitude  $A$ . Diese ist von den bereits diskutierten Einflussgrößen auf die Anzahl der Hinweisklassen abhängig. Unter Berücksichtigung von Annahme 2 wird folgendes Berechnungsverfahren der Amplitude  $A_{Plan}$  empfohlen:

$$A_{Plan}(E_{Anz}) = \begin{cases} (52 + F - 6 \cdot lg) \cdot E_g \cdot (1 - \exp(-B_{Plan} \cdot E_{Anz})) & \text{für } F > 4 \\ (10 + F \cdot 16/lg) \cdot E_g \cdot (1 - \exp(-B_{Plan} \cdot E_{Anz})) & \text{für } F < 5 \end{cases} \quad (5)$$

Die Konstanten Fragenanzahl  $F$ , Interaktionsgrad des Lernsystems  $lg$  und Erfahrungsgrad der Experten  $E_g$  müssen in der Planungsphase zu einer Expertenbefragung festgelegt werden. Die weiteren verwendeten Konstanten in der Modellgleichung wurden durch Regressionsanalyse von den durchgeführten Expertenbefragungen hergeleitet.

Je nach Fragenanzahl  $F$  unterscheidet die Gleichung (5) zwei Funktionen. Diese Unterscheidung ist nötig, da zwei Interviewtechniken angewendet werden können. Im

Fall der Fragenanzahl  $F < 5$  werden für nur Fragen aus der Lernsystem-Konfiguration – ohne Zusatzfragen – gestellt. Bei umfangreicheren Befragungen  $F > 4$  werden detailliertere Fragestellungen zugelassen (siehe UKap. 9.5 auf Seite 60).

$$B_{\text{Plan}} = 0,17 \quad (6)$$

Der Parameter  $B_{\text{Plan}}$  wird aus der Tab. 17 vereinfachend als Konstante festgelegt. Mit Kenntnis der Parameter  $A_{\text{Plan}}$  und  $B_{\text{Plan}}$  kann der Verlauf der Hinweisklassen in Abhängigkeit von der Anzahl der Experten prognostiziert werden.

$$\text{Hinweisklassen}_{\text{Plan}}(E_{\text{Anz}}) = A_{\text{Plan}} \cdot (1 - \exp(-B_{\text{Plan}} \cdot E_{\text{Anz}})) \quad (7)$$

Anhand der Modellgleichung (7) kann die erforderliche Anzahl von Versuchspersonen je nach gewünschter Anzahl von Hinweisklassen abgelesen werden.

Die Parameterberechnung kann in der Tab. 17 nachvollzogen werden. Die Parameter  $A$  und  $B$ , die als Referenzwerte dienen, wurden durch die Regressionsanalyse der Evaluationsdaten ermittelt. Die Parameter  $A_{\text{Plan}}$  und  $B_{\text{Plan}}$  wurden über die Gleichungen (5) und (6) berechnet.

LS	A	B	$A_{\text{Plan}}$	$B_{\text{Plan}}$	$\Delta A_{\text{Plan}}$	$\Delta B_{\text{Plan}}$
FLCBT	141,9	0,19	136	0,17	5,9	0,019
FLWBT	132,3	0,2	136	0,17	-3,7	-0,03
FLKran	44,4	0,17	50	0,17	-1,6	0,003
PS	100	0,17	50	0,17	8	-0,007
PAS	55,6	0,14	51	0,17	1,6	-0,027

Tab. 17: Schätzung der Hinweisklassen

## Schlussfolgerungen

**Fazit:** Mit Gleichung (7) ist es im Vorfeld einer Expertenbefragung möglich, eine ökonomische Expertenanzahl zu ermitteln. Die Entscheidung für eine bestimmte Expertenanzahl ist davon abhängig, wieviele Hinweisklassen insgesamt benötigt werden. Sind die Einflussgrößen auf die Anzahl der Hinweisklassen, also die Fragenanzahl des Qualitätskriterienkatalogs, der Interaktionsgrad des Lernsystems und der Experten-grad ( $E_g$ ), festgelegt, kann anhand der Gleichung (7) die voraussichtlich benötigte Expertenanzahl berechnet werden.

## 15.5 Monopol-Hinweisklassen – Expertenanzahl

**Frage:** Kann anhand der kumulierten Anzahl von Monopol-Hinweisklassen zu einer Expertenbefragung auf den Nutzen der Befragung einer weiteren Person geschlossen werden?

**Begrifflichkeit:**

**Begründungstexte** Schriftliche Begründungen der Experten zu ihrer Bewertung des Lernsystems anhand von Qualitätskriterien. In diesem Unterkapitel sind die Begründungstexte gemeint, die zur Formulierung von Hinweisklassen heran gezogen worden sind.

**B<sub>g</sub>** Anzahl der Begründungstexte

**Hinweisklasse** Eine Hinweisklasse wird auf der Grundlage von zusammengefassten Begründungstexten der Experten vom Qualitätsregelteam formuliert.

**Monopol-Hinweisklasse** Liegt bei der Formulierung einer Hinweisklasse nur ein Begründungstext eines Experten zugrunde, wird eine Hinweisklasse als Monopol-Hinweisklasse spezifiziert. Die Untersuchung in diesem Unterkapitel soll zeigen, dass Monopol-Hinweisklassen ein Indikator für den Grenznutzen der Expertenbefragung darstellen. Durch die kumulierte Zählung der Monopol-Hinweisklassen in Abhängigkeit von der Expertenanzahl soll prognostiziert werden, ob es sich gemessen am Aufwand-Nutzen-Verhältnis lohnt, einen weiteren Experten zu befragen.

**E<sub>Anz</sub>** Expertenanzahl

**Statistische Methoden:** Permutation, Regressionsanalyse.

**Annahme:** Während einer Befragung, bei der die Experten zu verschiedenen Terminen nacheinander eingeladen werden, kann durch Auswertung der kumulierten Monopol-Hinweisklassen der Befragungsnutzen prognostiziert werden. Je größer die positive Steigung der Kurve der Monopol-Hinweisklassen ausfällt, desto mehr lohnt sich die Befragung eines weiteren Experten (siehe Abb. 31 auf der nächsten Seite).

Monopol-Hinweisklassen

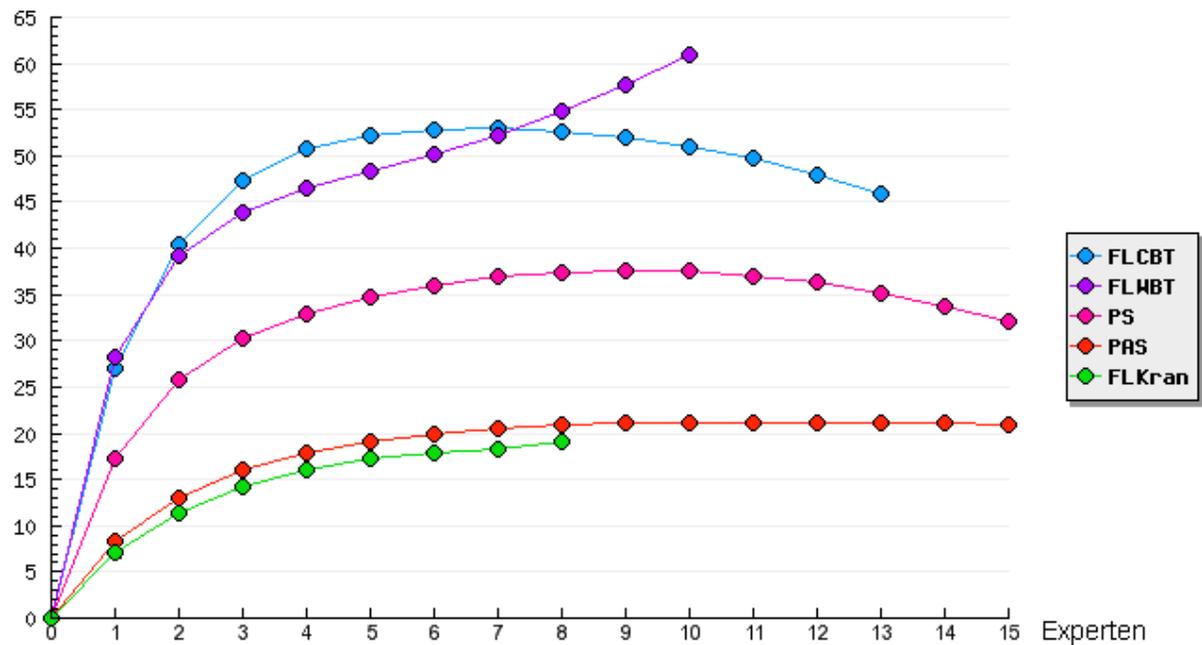


Abb. 31: Kumulierte Monopol-Hinweisklassen im Projektvergleich

## Schlussfolgerungen

**Zur Annahme:** Im Vergleich zu den übrigen Prototypen hat der Verlauf der Kurve des FLWBT-Lernsystems eine abweichende charakteristische Form. Dieser Verlauf ist damit begründet, dass es einen Experten gab, der extrem die Leistung (Hinweise pro Zeit) der übrigen Experten übertraf. Diese Leistung ist dafür verantwortlich, dass die Anzahl der Monopol-Hinweisklassen bis zu der Expertenanzahl 10 ansteigt und sich nicht reduziert, wie es im Vergleich zu den anderen Lernsystem-Hinweisklassen-Funktionen zu vermuten wäre. Der Experte stellt somit einen Ausreißer dar.

Aus den Kurvenverläufen aus Abb. 31 lässt sich folgende Modellgleichung aufstellen:

$$\text{Monopol-Hinweisklassen} = 14 \cdot E_{\text{Anz}} \cdot \exp(-0,08 \cdot E_{\text{Anz}}) \quad (8)$$

Die Gleichung ist in Abb. 32 auf der nächsten Seite dargestellt.

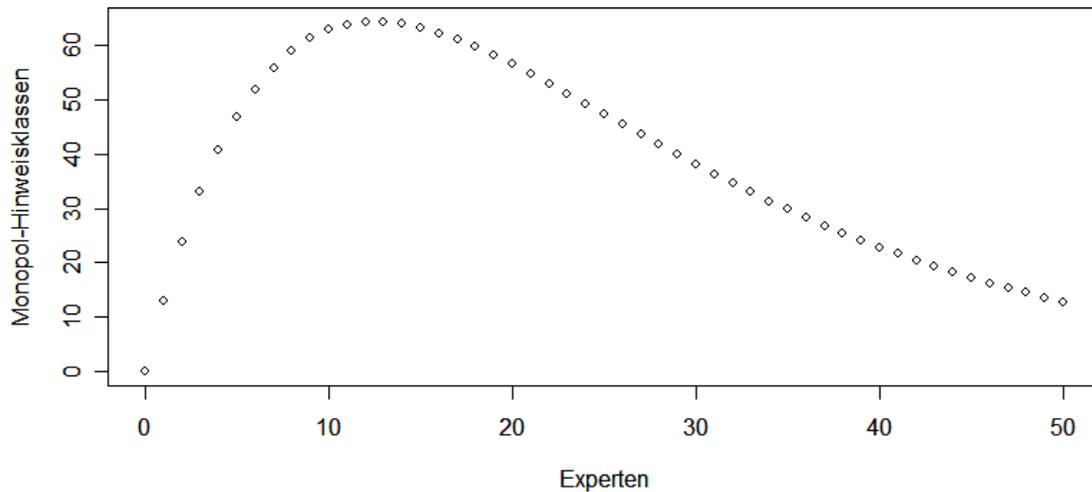


Abb. 32: Kumulierte Monopol-Hinweisklassen

**Fazit:** Durch die Abb. 32 lässt sich eine für Expertenbefragungen typische Empfehlung nachvollziehen, die besagt, dass eine Bewerteranzahl von etwa 5 Personen ausreichend ist. Ab dieser Anzahl wird die Steigerungsrate sichtlich geringer bis sie ins negative umkehrt. Bei 12 Experten ist ein Maximum erreicht, darüber hinaus nimmt die Zahl der Monopol-Hinweisklassen durch Zunahme inhaltlich ähnlicher Begründungstexte zu den Hinweisklassen ab.

Neben den Monopol-Hinweisklassen sind die Hinweisklassen von Interesse, die mehr als nur einmal bestätigt worden sind. Als Entscheidungskriterium für die effektive Expertenanzahl, soll nicht das Maximum der Monopol-Hinweisklassen-Funktion dienen, sondern das Maximum der Kurve mit beispielsweise 3 Begründungstexten zu den Hinweisklassen. Als Planungshilfe für eine solche Fragestellung kann die Verteilung der Begründungstexte zu den Hinweisklassen beim PS-Lernsystem in Abb. 33 auf der nächsten Seite dienen.

In Abb. 33 auf der nächsten Seite sind gewisse Gesetzmäßigkeiten zu erkennen, die vermutlich auch für eine höhere Expertenanzahl gelten. Die Anzahl der Begründungstexte zu den Hinweisklassen kann nicht größer als die Anzahl der Experten sein. Sie ist immer kleiner, da der Fall, dass eine Hinweisklasse auf der Grundlage von Begründungstexten aller Experten mit zunehmender Expertenanzahl unwahrscheinlich wird. Mit zunehmender Expertenanzahl nehmen die Monopol-Hinweise nach dem Erreichen eines Maximums ab, während die Anzahl der Hinweisklassen mit mehr als einem zugrunde liegenden Begründungstext zunimmt. Zudem erhöht sich die kumulierte Anzahl der Begründungstexte zu den Hinweisklassen. Dieser Zusammenhang ist in Abb. 33 und 34 auf der nächsten Seite exemplarisch für die Expertenbefragung zum Planspiel-Lernsystem (PS) abgebildet.

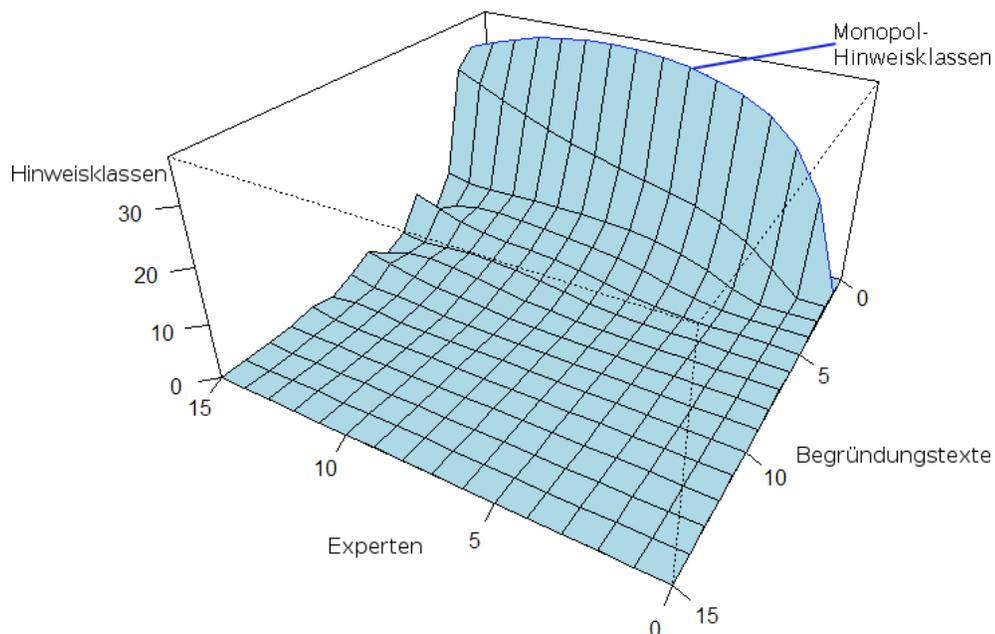


Abb. 33: PS-Perspektive der Begründungstexte

Die Abb. 33 lässt sich auch als 2D-Darstellung darstellen. In Abb. 34 können die Konturen der 3D-Perspektive besser wahrgenommen werden. Die verwendeten Farben stehen für eine bestimmte Anzahl der Hinweis-klassen.

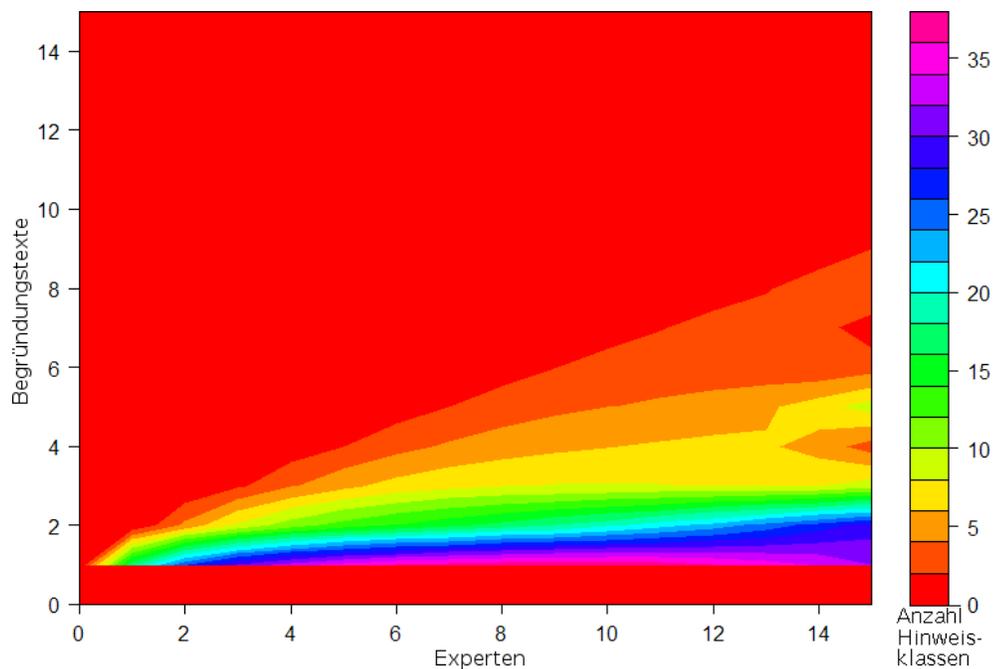


Abb. 34: PS-Konturen der Begründungstexte

Die Charakteristik der Abbildung 33 bzw. 34 auf der vorherigen Seite lässt sich auch für die anderen untersuchten vier Lernsysteme nachweisen. Diese Tatsache ermöglicht es, ein Modell zu entwickeln, welches die Ergebnisse der durchgeführten Befragungen nachbildet und auch bei einer Expertenanzahl, größer als 15, anwendbar ist.

Die folgende Gleichung kann durch Kenntnis der Einflussgrößen auf die Anzahl der Hinweis-klassen angepasst werden:

$$\text{Hinweisklassen} = \frac{100}{E_{\text{Anz}}} \cdot (E_{\text{Anz}} - 1,25 \cdot Bg) \cdot \exp(-0,05 \cdot (E_{\text{Anz}} - 1,25 \cdot Bg)) \quad (9)$$

Die Modellgleichung wird für das Planspiel-Lernsystem in den folgenden Abbildungen 35 und 36 grafisch angezeigt. Die Abbildungen stellen die möglichen Verteilungen der Anzahl der Expertenbegründungen in Abhängigkeit von den Hinweisklassen und den bis zu 50 Experten dar.

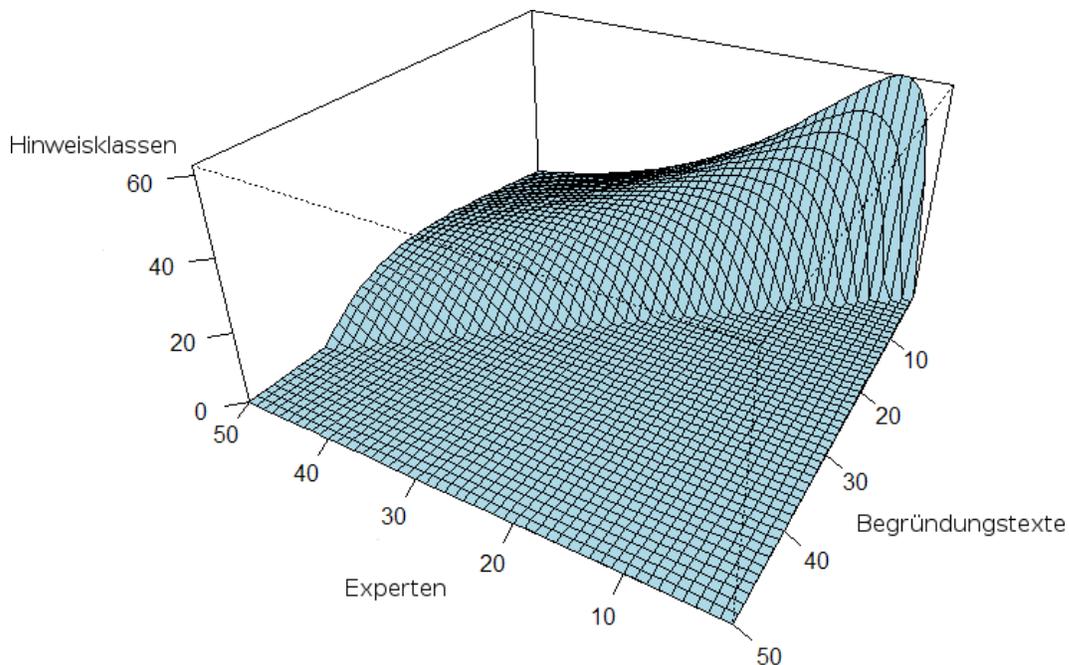


Abb. 35: Modell-Perspektive der Begründungstexte

Ein Vergleich mit der Abb. 33 auf der vorherigen Seite zeigt, dass die Modellgleichung zwar idealisiert, aber ein brauchbares Planungsinstrument für Expertenbefragungen ist.

Das 3D-Modell lässt sich ebenfalls wieder auf eine 2D-Sicht reduzieren:

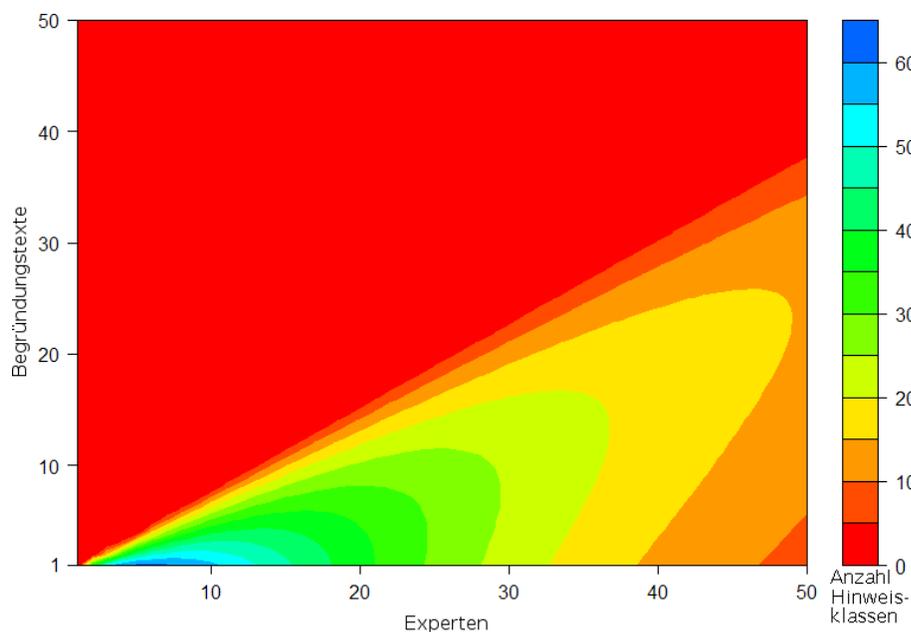


Abb. 36: Modell-Konturen der Begründungstexte

Eingesetzt werden kann das Modell bei Fragestellungen zu einer ökonomischen Expertenanzahl, bei der als Nebenbedingung z.B. ein Maximum einer bestimmten Anzahl von Begründungstexten erreicht wird. Die Nebenbedingung ist sinnvoll, da aufgrund der Anzahl von Bestätigungen Qualitätsverbesserungsmaßnahmen priorisiert werden können. Sollen beispielsweise möglichst viele Hinweisklassen mit mindestens drei Begründungstexten vorliegen, dann bedarf es mehr als 10 Experten.

## 15.6 Standardabweichung der Bepunktung

**Fragen:** Wie verhält sich die **Standardabweichung** zur Bewertung und Gewichtung von Qualitätskriterien?

**Begrifflichkeit:**

$E_{Anz}$  Expertenanzahl

**Sa** Die Standardabweichung ist ein Maß der Streuung.

**G** Die Experten sollten die Qualitätskriterien in Bezug auf das Lernsystem anhand einer 5-Punkt-Skala gewichten.

**M, LU, LT, LI/LM** Qualitätskriterien der Lernsystem-Konfiguration: Mensch M, Lernumfeld LU, Lerntechnologie LT, Lerninhalte/Lehrmethode LI/LM (siehe UKap. 9.4 auf Seite 52ff.).

**Statistische Methoden:** Deskriptive Statistik, Permutation, Regressionsanalyse.

**Annahme 1:** Die Bepunktung zur Bewertung und Gewichtung entspricht nicht einer Zufallsverteilung. Bei dieser würde eine über die gesamte 5-Punkt-Skala gleichverteilte Vergabe der Punkte erfolgen.

**Annahme 2:** Mehrere Faktoren können eine hohe Standardabweichung bei Gewichtungen und Bewertungen hervorrufen:

- die Anzahl von Experten,
- stark differierende Experteneigenschaften (Erfahrung, Motivation, Beruf etc.) oder
- die Verständlichkeit des Qualitätskriteriums.

Durch diese Faktoren können mehr oder weniger starke Ausreißer in der Bepunktung entstehen. Mit zunehmender Expertenanzahl sinkt allerdings der Einfluss von Ausreißern und die Reliabilität der Bepunktung steigt. Standardnormalverteilungen bilden sich erst ab einer gewissen Anzahl >100 heraus.

Angestrebt wird also eine möglichst kleine Standardabweichung der Bepunktung. Doch ab welchem Wert ist die Standardabweichung konkret als klein zu bezeichnen? Eine angemessene Einschätzung der Größe der Standardabweichung kann anhand der Beispieldaten-sätze aus der nebenstehenden Tabelle erfolgen. Spalte A enthält die Anzahl der Antworten mit dem Wert 1 und Spalte B die Anzahl der Antworten mit dem Wert 4. Zeilenweise wird zu der Menge A und B die Standardabweichung  $S_a$  berechnet. Die Werte der Standardabweichungen schaffen eine Grundlage, die Standardabweichungen zu den Expertenbefragungen angemessen einzuschätzen.

A	B	$S_a$
1	4	1.34
1	8	1
1	12	0.83
1	16	0.73
1	20	0.65

Für die Abb. 37 auf der nächsten Seite wurde die minimale, maximale und arithmetisch mittlere Standardabweichung über alle möglichen Reihenfolgen von Experten berechnet. Aus der Abbildung lässt sich ablesen, welche Spanne von Standardabweichungen bei welcher Expertenanzahl zu erwarten ist. Die mittlere Standardabweichung entspricht der wahrscheinlichsten Standardabweichung.

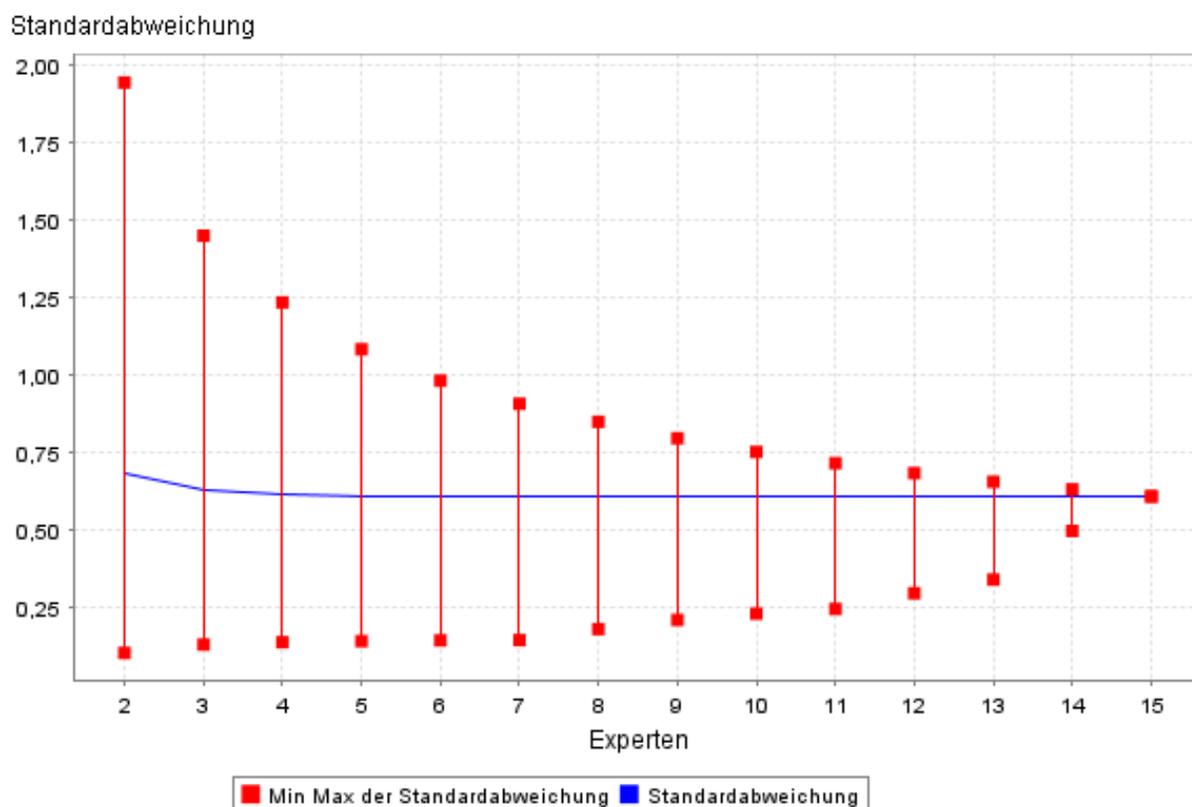


Abb. 37: Standardabweichungen zur Bewertung des PAS

Tab. 18 zeigt, wie die Standardabweichung zu den vier Qualitätskriterien Mensch M, Lernumfeld LU, Lerntechnologie LT, Lerninhalte/Lehrmethode LI/LM je nach Lernsystem-Projekt LS ausfällt.

LS	M		LU		LT		LI/LM	
	( $\bar{G}$ )	(Sa)						
FLCBT	4.4	0.74	3.6	0.92	4.1	0.83	4.3	0.72
FLWBT	3.9	1.20	3.1	1.27	3.9	0.74	3.9	0.99
FLKran	4.0	0.71	3.3	1.25	4.1	0.93	3.4	0.90
PS	3.6	0.95	3.2	1.12	3.7	1.25	3.3	1.16
PAS	3.9	0.81	3.4	0.98	4.0	0.63	3.5	1.23

Tab. 18: Standardabweichung der Gewichtung

In der Tab. 19 auf der nächsten Seite finden sich die zu den Lernsystemen berechneten Regressionsgleichungen zur maximalen Standardabweichung der Bewertung.

LS	$Sa_{\max}(E_{Anz}) =$
FLCBT	$(1.40 - 1.3e^{(-0.18E_{Anz})})^{-1}$
FLWBT	$(1.59 - 1.3e^{(-0.16E_{Anz})})^{-1}$
FLKran	$(1.44 - 1.3e^{(-0.25E_{Anz})})^{-1}$
PS	$(1.71 - 1.3e^{(-0.15E_{Anz})})^{-1}$
PAS	$(1.46 - 1.3e^{(-0.18E_{Anz})})^{-1}$

Tab. 19: Regressionsanalyse Standardabweichung der Bewertung

In Abb. 38 sind die aus der Expertenbefragung zum PAS ermittelte maximale Standardabweichung zur Bewertung und die entsprechende Regressionsgleichung eingetragen (vgl. mit Tab. 19).

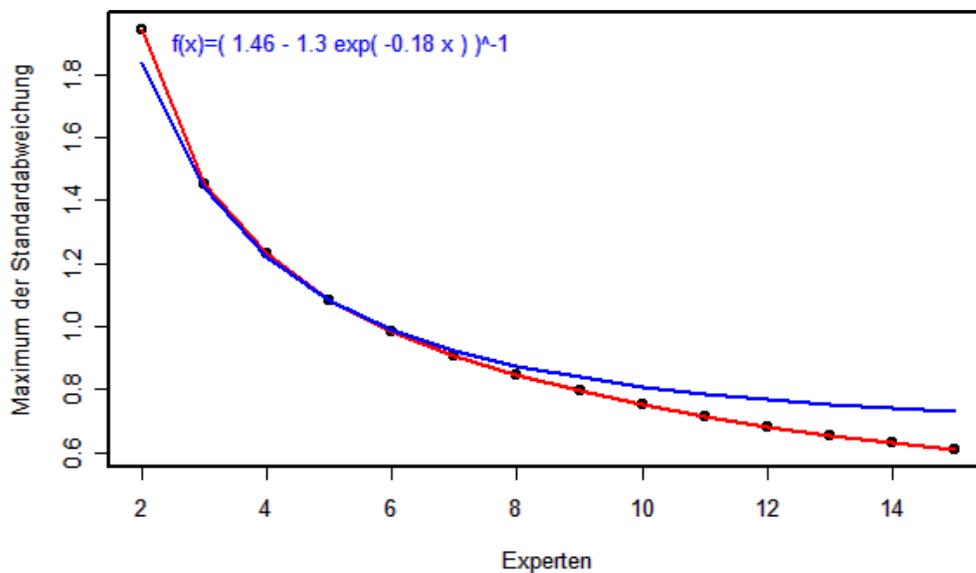


Abb. 38: Maximale Standardabweichungen zur Bewertung des PAS

## Schlussfolgerungen

**Zur Annahme 1:** Die Standardabweichungen in Tab. 18 auf der vorherigen Seite liegen in einem Wertebereich, der einen Vergleich von erreichten Punktzahlen zweier Prototypen erlaubt. Die Mittelwerte der Gewichtung liegen nicht bei dem Punktwert 3 (Erwartungswert bei zufälliger Gewichtung). Daher handelt es sich nicht um eine Gleichverteilung, sondern eine Verteilung mit einem ausgeprägten Maximum. Die Standardabweichung liegt um den Punktwert 1. Diese fällt bei einer 5-Punkt-Skala und bei einer kleinen Expertenanzahl nicht ungewöhnlich hoch aus.

Bei den Qualitätskriterien Mensch, Lernumfeld und Lerntechnologie gibt es vergleichbare Gewichtungen über alle Lernsystem-Prototypen. Beim Qualitätskriterium Lerninhalte und Lehrmethode ist ein unterschiedliches Niveau zu beobachten. Die Lernsysteme lassen sich hier je nach verwirklichter Interaktivität der Programme in zwei Gruppen einteilen. Die Lerninhalte und die Lehrmethode werden bei den weniger interaktiven Lernsystemen signifikant höher gewichtet.

Dies liegt vermutlich daran, dass die Lerninhalte in interaktiven Lernsystemen nicht so greifbar für die Experten sind. Bei den interaktiven Lernsystemen geht es vorwiegend um den methodischen Kompetenzerwerb, der nicht durch einen expliziten Lehrtext vermittelt wird. Aus der Nachbefragung der teilnehmenden Experten folgt, dass es offenbar zwei mögliche Denkrichtungen zur Festlegung der Gewichtungspunktzahl gab. Es gab Personen, die argumentierten, dass die Lerninhalte erst durch das Lernsystem entstünden und daher das Qualitätskriterium weniger wichtig für die Bewertung hielten. Es stellte sich die Frage, soll ich nun die von mir erzeugten, die denkbaren oder die vom Lernsystem bereitgestellten Lerninhalte (vor/ohne Interaktion) gewichten? Zum anderen wurden Lerninhalte in interaktiven Lernsystemen überhaupt nicht wahrgenommen, sondern nur der persönlich wahrgenommene anschließende Lernerfolg. Die realisierten Lernsystemkomponenten im Sinne des Konstruktivismus haben offenbar mit Erfolg dazu geführt, dass die Anwender gar nicht merkten, dass sie gerade lernen. Vermutlich aufgrund dieser Unsicherheit der Experten fällt die Gewichtung geringer aus und nicht etwa dadurch begründet, dass die Lerninhalte oder die Lehrmethode generell bei interaktiven Lernsystemen im Hintergrund stehen. Dies zeigt beispielsweise die Tab. 21 auf Seite 125, wo der Informationsgehalt als Subkriterium von „Lerninhalte und Lehrmethode“ den vordersten Platz in der Qualitätsdiskussion der Experten einnimmt.

**Zur Annahme 2:** Die Untersuchungen belegen, dass die Spanne zwischen minimaler und maximaler Standardabweichung mit zunehmender Expertenanzahl geringer wird und sich der Lagebereich für die wahre Standardabweichung der Grundgesamtheit einschränkt (siehe Abb. 37 auf Seite 119). Die Spanne sollte möglichst klein sein, um verlässliche Aussagen über die erzielte Bewertungspunktzahl machen und um Prototypen angemessen miteinander vergleichen zu können. Je größer die Expertenanzahl einer Stichprobe, desto schmaler wird die Spanne der Standardabweichungen. Mit dieser Tendenz steigt die Wahrscheinlichkeit, die wahre Bewertungspunktzahl der Grundgesamtheit der Experten zu erkennen.

Daraus könnte man schließen, dass möglichst viele Experten zu befragen sind, wenn da nicht die steigenden Kosten und der wachsende Zeitbedarf für die Evaluation wäre. Unter 5 Experten muss das Ergebnis noch starken zufälligen Schwankungen unterworfen sein. Ab 5 Experten verläuft die Standardabweichung annähernd linear mit flacher negativer Steigung.

**Fazit:** Der Funktionsverlauf der maximalen Standardabweichung in Abb. 38 auf Seite 120 ist für die Bestimmung der ökonomischen Anzahl von Experten geeignet. Ähnliche Verläufe wie in dieser Abbildung ergaben sich auch für die anderen Lernsystem-Evaluationen. Die Mittelwerte der Parameter der Gleichungen aus Tab. 19 auf Seite 120 führen zu folgender Näherung, anhand der die maximale Standardabweichung bei gegebener Expertenanzahl geschätzt werden kann.

$$S_{\max}(E_{\text{Anz}}) = (1.5 - 1.3 \cdot e^{(-0.2E_{\text{Anz}})})^{-1} \quad (10)$$

Die Gleichung hilft zusätzlich zu den in UKap. 15.4 auf Seite 109 erläuterten Abbildungen und Gleichungen bei der Planung einer effizienten Expertenanzahl und bei der Einschätzung der Unsicherheit beim Vergleich von Lernsystem-Prototypen. Unterscheiden sich beispielsweise zwei Lernsystem-Prototypen nur sehr gering, muss mit einer höheren Anzahl von Experten (>5) operiert werden, als wenn der Unterschied groß ist.

## 15.7 Freie Antwortmöglichkeit

**Frage:** Wird die freie Antwortmöglichkeit konstruktiv von den Experten genutzt?

**Begrifflichkeit:**

**Bg** Begründungstexte (siehe UKap. 15.3.1 auf Seite 98)

**Hk** Hinweisklassen (siehe UKap. 15.4 auf Seite 109)

LS	Bg	Hk	Bg/Hk
FLCBT	38.1	15.5	2.46
FLWBT	49.8	27.8	1.79
FLKran	11.3	7.1	1.59
PS	26.7	17.3	1.54
PAS	11.5	8.3	1.39

Tab. 20: Begründungstexte und Hinweisklassen

## Schlussfolgerungen

**Fazit:** In der Tab. 16 auf Seite 104 wird die Anzahl der geschriebenen Zeichen aus den Begründungstexten aufgeschlüsselt nach Lernsystem und Expertengrad. Wenn man

berücksichtigt, dass auf eine DIN A4-Seite 4335 Zeichen (Times New Roman Schriftgrad 12; Seitenränder oben/links/rechts 2.5cm, unten 2cm) passen, dann ist die geschriebene Zeichenanzahl stattlich. Aber wie steht es mit der Qualität der Begründungstexte, d.h. können Hinweisklassen zur Qualitätssicherung vom Qualitätsregeltteam formuliert werden?

Tab. 20 auf der vorherigen Seite zeigt auf, wieviele Qualitätsaspekte vor und nach der Inhaltsanalyse der Begründungstexte zur Bewertung der Experten erhoben wurden. Nach der Inhaltsanalyse führen die Hälfte der Hinweise zur Formulierung von Hinweisklassen, aufgrund derer Verbesserungsmaßnahmen zur Qualitätssicherung des Lernsystems empfohlen werden können.

Im Unterschied zu standardisierten Antwortmöglichkeiten sind also nicht alle Themenbereiche, die ein Experte niederschreibt von Interesse für die Qualitätssicherung. Dies ist objektiv ein Nachteil. Hinzu kommt die aufwändige Inhaltsanalyse, die nicht zu automatisieren ist (siehe Kap. 14 auf Seite 85).

Bei standardisierter Antwortmöglichkeit können zwar alle Antworten automatisch und schnell ausgewertet werden, aber die freie Antwortmöglichkeit hat den Vorzug der erhöhten Variabilität. Ohne Vorgabe von Antworten können unverfälschte, detaillierte und facettenreiche Qualitätsthemen erhoben werden, die den Gedankengängen der Experten direkt entspringen. Die Eindrücke, Erfahrungen, Wertungen etc. der Experten werden nicht auf vorgegebene und vorgedachte Antwortmöglichkeiten konvertiert.

Die Bedenken, dass bei freier Antwortmöglichkeit zu wenig Hinweise zusammen kommen, hat sich nicht bestätigt. Die freie Antwortmöglichkeit wird konstruktiv genutzt.

## 15.8 Qualitätskriterien

**Frage:** Zu welchen Qualitätskriterien werden die meisten Hinweisklassen generiert?

**Begrifflichkeit:**

**Qualitätskriterium** Die Qualitätskriterien werden in UKap. 9.4 auf Seite 52 kurz beschrieben. In Tab. 21 auf Seite 125 sind die Qualitätskriterien der Lernsystem-Konfiguration fett hervor gehoben. Insgesamt richtet sich die Reihenfolge der aufgeführten Qualitätskriterien nach der Anzahl der zugeordneten Hinweisklassen.

**Aspekt** Unterschieden wird, ob ein Qualitätskriterium eher didaktische und/oder technische Qualitätsaspekte beinhaltet.

**Hk** Hinweisklassenanzahl (siehe UKap. 15.3 auf Seite 97)

**Annahme:** Zu den Qualitätskriterien werden unterschiedlich viele Begründungstexte über alle Lernsystem-Evaluationen formuliert. Daraus resultiert eine unterschiedliche Anzahl von Hinweisklassen zu den Qualitätskriterien. Angenommen wird, dass dies größtenteils an der Art der Hauptfragestellung liegt. Manche Fragen lassen sich kurz und knapp beantworten. Andere Fragen erfordern längere Stellungnahmen. Die Experten werden zu bestimmten abgefragten Qualitätsaspekten mehr schreiben, wenn diese ihnen bezogen auf ihr Vorwissen vertrauter erscheinen und / oder wenn der jeweilige Qualitätsaspekt für sehr bedeutsam gehalten wird.

Qualitätskriterium	Aspekt	Hk
Informationsgehalt	didaktisch	131
Screen Design	technisch	98
Verfügbarkeit	technisch	88
Lehrmethode	didaktisch	76
Hilfe	didaktisch	67
<b>Lernumfeld</b>	didaktisch	63
Navigation/Steuerungsinteraktion	technisch	62
Motivation/Aktivität	didaktisch	40
Vorwissen/Medienkompetenz	didaktisch	36
Lernsystemklasse	techn./didakt.	34
Text	didaktisch	31
Lernkontrollen	didaktisch	31
Bilder/Metapher	techn./didakt.	29
Lehr-/Lernziele	didakt.	28
<b>Mensch</b>	techn./didakt.	27
Administratoren/Entwickler	technisch	26
Farben	technisch	23
<b>Lerninhalte/Lehrmethode</b>	didaktisch	21
Betreuung	didaktisch	19
Animation/Video	technisch	18
Dialog	technisch	18
Lernende	didaktisch	17
<b>Lerntechnologie</b>	technisch	16
Listen/Tabellen	technisch	13
Lehrende	didaktisch	12
Individualisierung	didaktisch	11

*Tabelle wird auf der nächsten Seite fortgesetzt.*

Qualitätskriterium	Aspekt	Hk
<i>Tabelle wird von der vorherigen Seite fortgesetzt.</i>		
Eingabe-/Kommunikationsgeräte	technisch	10
Autorensysteme	technisch	10
Emotionen	didaktisch	9
Kommunikation und Kooperation	didaktisch	8
Sicherheit	technisch	7
Adaptivität	technisch	6
Audio	technisch	3
Impressum	technisch	1

Tab. 21: Qualitätskriterien – Hinweisklassen über alle Evaluationen

## Schlussfolgerungen

**Zur Annahme:** Nach Tab. 21 werden bei einigen Qualitätskriterien deutlich mehr Hinweisklassen aus den Begründungstexten der Experten abgeleitet. Somit wird die Annahme gestützt.

**Fazit:** Aus den Ergebnissen in Tab. 21 lässt sich folgern, welche Qualitätskriterien für die Experten bedeutsam sind. Die Aufstellung ergibt eine Rangordnung, bei der die Bedeutung des Informationsgehalts und Screen Designs eines Lernsystems einen vordersten Rang einnimmt. Die Tab. 21 legt aber auch offen, welche Qualitätskriterien sich weniger für eine Expertenbefragung eignen. In jedem Fall können Prioritäten für die Qualitätssicherung der untersuchten Lernsysteme festgelegt werden.

## 15.9 Metainformationen

**Frage:** Wie häufig werden die für den Bewertungsprozess verfügbaren Metainformationen (Leitfragen, Beschreibungen, Erläuterungen, Fundierung) zum Qualitätskriterium genutzt?

**Begrifflichkeit: Metainformationen** Zum besseren Verständnis der Hauptfragestellung eines Qualitätskriteriums wurden im Fragenkatalog Metainformationen hinterlegt, die der Experte bei Bedarf abrufen kann (siehe UKap. 8.2.2 auf Seite 41)

Bei allen Evaluationen wurden insgesamt 750 Bewertungen zu Qualitätskriterien abgegeben. Dabei riefen die Experten 203 mal Metainformationen zum Qualitätskriterium auf.

## Schlussfolgerungen

**Fazit:** Die Abruftrate ist zwar unerwartet gering, sodass die Frage berechtigt ist, ob der Aufwand-Nutzen für die angebotene Wissensbasis zur Unterstützung der Experten gerechtfertigt ist. Man muss sich allerdings fragen, welcher Schaden entsteht, wenn etwa bei einem Viertel der Qualitätskriterien Bewertungen unter Unsicherheit aufgrund von mangelnden Informationsmöglichkeiten abgegeben werden. Dies wäre dann der Fall, wenn die 203 Informationsabfragen wegfielen.

Das Informationsangebot im Rahmen des Qualitätskriterienkatalogs wird nicht immer, aber auch nicht selten genutzt. Gründe für das Verhalten der Bewerter könnte mangelndes Interesse und das Ziel möglichst schnell die Evaluation abschließen zu wollen sein. Oder positiv gedacht, dass das primär sichtbare Informationsangebot der Qualitätskriterien weitgehend keine Fragen offen lässt.

Zusammen mit den Untersuchungen aus UKap. 8.2.2 auf Seite 41 lässt sich folgern, dass Metainformationen sehr wohl den Bewertungsprozess in sinnvoller Weise unterstützen.

## 16 Übertragbarkeit der Ergebnisse

Die Expertenbefragung lässt sich mit geeigneten Werkzeugen und Methoden schnell, einfach und kostengünstig durchführen. Dadurch zeichnet sie sich für den Kontext von kürzeren Entwicklungszeiten und schmalen Projektbudgets aus. Sie ist nicht auf eine bestimmte Phase im Entwicklungsprozess beschränkt. Ist das Verbesserungspotenzial aufgedeckt, können – falls nötig – mit dem Einsatz weiterer Evaluationsverfahren oder spezieller Qualitätskriterien Nachuntersuchungen stattfinden. In den meisten Fällen können direkt, ohne weitere Untersuchungen des Lernsystems, Verbesserungsmaßnahmen abgeleitet werden.

Es stellt sich allerdings die Frage, inwieweit die Studienergebnisse auf andere Lernsysteme oder Qualitätskriterienkataloge übertragbar sind. Die Signifikanz der Ergebnisse wurde in der Auswertung zu den verschiedenen Fragestellungen angegeben. Der p-Wert gibt an, wie wahrscheinlich es ist, dass der untersuchte Zusammenhang sich anders darstellt als angenommen. Der p-Wert liefert keine Antwort auf die Frage, ob bei einer zweiten Untersuchung ebenfalls ein signifikantes Ergebnis erzielt wird. Er wird aber nicht selten so interpretiert, als wenn er dies täte (Tversky und Kahneman 1971, S.105ff., Oakes 1986).

Es wurden daher mehrere reale Lernsystem-Prototypen untersucht. Die Signifikanztests halfen, Einflussgrößen auf das Potenzial von Expertenbefragungen zu identifizieren. Mit dem Wissen über diese Einflussgrößen können je nach Umfeld die Ergebnisse der Studie auf gleichartige Qualitätssicherungsprojekte und Expertenpopulationen übertragen werden. Werden die Versuchsplanung und -durchführung, die Qualitätskriterien, das Befragungsmedium etc. erheblich geändert, so ist damit zu rechnen, dass die aus der Studie gewonnenen Parameter zur Schätzung der Hinweisklassen neu kalibriert werden müssen (siehe UKap. 15.4 auf Seite 109). Die Modellgleichungen selbst können sicherlich unverändert übernommen werden.

## 17 Ausblick

Die Integration von Lernsystemen in den Lehralltag wird von vielen Lehrenden als ein sinnvoller Weg zur Professionalisierung und Qualitätssicherung der universitären Lehre gesehen. Folgende E-Learning-Trends sind dabei zu beobachten:

- Pull-Prinzip: Gemeint ist die Eigenverantwortung für das Lernen ohne starke Anleitung und Betreuung.
- WYSWYE (What you see is what you can edit): Ein zukunftsweisender Trend ist die Bereitstellung von Autorensystemen. Durch die vereinfachte Erstellung von Lernsystemen über entsprechende Benutzerschnittstellen können auch Lernende zur Lerninhaltsproduktion beitragen. Lernende werden Konsumenten und zugleich Produzenten bzw. Qualitätssicherer von Lerninhalten. Sie treffen autonome Entscheidungen über Lerninhalte. Die zwei genannten Prinzipien führen zu einer Neubestimmung der Rolle der Studierenden.
- Flexibilisierung/Verfügbarkeit/Globalisierung: Orts- und Zeitungebundenheit, internationales Lehren und Lernen.
- Just-in-time/Learning-by-doing: Arbeiten = Lernen oder Integration des Lernens in die Arbeitsprozesse,
- E-Kollaboration/E-Kooperation/E-Kommunikation/Wissensmanagement: Das Lernen ist fest eingebunden in kollaborative und kooperative Lernprozesse (Grune 2000, S.76). Unser Leben ist geprägt von Lernerfahrungen und so liegt es nahe, dieses Wissen anderen zugänglich zu machen und im Team zu lernen (Soziales Teamwork). Zwischen der Notwendigkeit und diesem Schritt der Wissensbereitstellung liegt jedoch ein tiefer Graben. Durch Wissensmanagement wird erarbeitetes Wissen allgemein zugänglich und somit erst wertvoll. Wissensakquirierung, -speicherung, -(ver-)teilung sollten Studenten schon in der Universität aktiv erleben und selber praktizieren. Das Wissen ist letztlich im Lernnetzwerk gespeichert: nicht Wissensbesitz, sondern die Weitergabe und Weiterentwicklung in der Learning Community werden honoriert (Flehsig 1996, S.189). Die Fähigkeit zur Gruppenarbeit unter Nutzung moderner Informationstechnologie, ist eine unabdingbare und erfolgsentscheidende Schlüsselqualifikation von Ingenieuren.

Befürworter bezeichnen häufig den Einsatz von Lernsystemen als didaktische Innovation, organisatorische Revolution oder finanzpolitische Sparmaßnahme mit effizienzsteigernder Wirkung (Hauff 1998, S.11).

Kritiker der Neuen Medien bemängeln hingegen zumeist den indirekten persönlichen Kontakt der Studierenden zum Dozenten über ein technisches Medium. In diesem Zusammenhang

stehen die wissenschaftlich nicht haltbaren Argumente, dass Lernsysteme zur Verkümmern von Sozial- und Kommunikationsverhalten, Phantasie, intuitivem und assoziativem Denken beitragen. Mit dem nutzbringenden Gebrauch des Internets für Lehr- und Lernzwecke verhält es sich nicht anders wie bei der geschichtlichen Einführung des Telefons. Zu jener Zeit wurde behauptet, Menschen würden in Zukunft nicht mehr persönlich miteinander reden. Heute verwenden wir das Telefon selbstverständlich neben vielen Kommunikationsarten als zusätzliche Kulturtechnik.

Eigene Erfahrungen belegen, dass diejenigen, die den Umgang mit den Neuen Medien beherrschen, weniger Ängste und Befürchtungen bezüglich einer technologie- und mediengestützten Lern- und Lehrgesellschaft äußern. Immer mehr Jugendliche, Schüler und Studenten werden einen PC und den Zugang zum Internet nutzen und somit ganz selbstverständlich mit elektronischen Mitteln lernen und kommunizieren.

Neue Medien können als ein dauerhafter Hype gesehen werden, die die Möglichkeit beinhalten, Wissen zielgerichteter, kostengünstiger und unabhängig von Zeit und Ort zu organisieren.

Für den Bereich Maschinenbau gibt es derzeit viel zu wenig Lernsysteme, um mehr als nur punktuell die Lehre zu unterstützen. In der Maschinenbau-Lehre werden multimediale Lernsysteme als wirkungsvolle Werkzeuge zur Unterstützung der Lehre zwar wahrgenommen, aber nach wie vor nur verhalten eingesetzt. Es ist allerdings zu erwarten, dass der Einsatz von Lernsystemen auch in der ingenieurwissenschaftlichen Ausbildung zunehmend an Bedeutung gewinnen wird, sei es als Lehr-Lernmedium, als Kommunikationsmedium oder als Informationsmedium. Hier sind vor allem Lernsysteme interessant, die komplexe Systeme veranschaulichen. Das Internet bietet Techniken für den Maschinenbau, die wie kaum eine andere Technik dazu geeignet ist, Daten für die Vor- und Nachbereitung von Lehrveranstaltungen zu verarbeiten und zu verbreiten, sowie Visualisierungs-, Simulations- und Kommunikationsprozesse zu unterstützen.

Zwei Einsatzstrategien von Lernsystemen werden nach Hauff (1998, S.11) in der Lehre hauptsächlich realisiert:

- „Universitäre Lehre, wie sie traditionell in der personalen Präsentation im Seminarraum und im Hörsaal stattfindet, soll durch multimediale Elemente unterstützt und damit anschaulicher und effizienter gestaltet werden.
- Grundlagenwissen kann in Netzen oder auf CD-ROMs bereitgestellt werden, um von den Studierenden im Selbststudium parallel zu Lehrveranstaltungen bearbeitet zu werden. Der Anteil der im Netz zu studierenden Inhalte kann dabei so stark gesteigert werden, dass ganze Studienabschnitte online studierbar sind“.

Der erste Ansatz kann als „evolutionär“ bezeichnet werden, da der Einsatz von Lernsystemen in der Lehre sich in vielen kleinen Schritten vollzieht und kleinere „Mutationen“ in bestimmten Lehrgegenständen dafür sorgen, dass die Lernsysteme sich in die gegebenen Lehrstrukturen einfügen und diese verbessern.

Der zweite Ansatz geht etwas weiter, die Lernsysteme nehmen eine fest verankerte und teilweise dominierende Rolle im Lehrbetrieb ein. Die Idee der virtuellen Hochschule, die am ehesten der Lehrform des Fernstudiums entspricht, ist diesem "revolutionären" Ansatz zuzuordnen.

Überschätzt wird häufig die Leistungsfähigkeit einiger Lernsysteme. Viele Lernsysteme haben nicht den Anspruch eines „Nürnberger Trichters“, sondern sie entfalten ihr Potenzial in Verbindung mit einer Präsenzlehrveranstaltung. Bei den durchgeführten Expertenbefragungen wird die Notwendigkeit dieser Verbindung sehr deutlich. Einzelne Experten merkten zu den verschiedenen FLCBT- und FLWBT-Lernsystemen an: Das Lernsystem ist zwar gut, aber es kann den Dozenten nicht ersetzen.

Mit der in Zukunft zu erwartenden zunehmenden Anzahl von Autoren- und Lernsystemen steigt auch der Bedarf für eine effiziente und fest in den Entwicklungsprozess integrierte Qualitätssicherung. Der erfolgreiche Einsatz von Lernsystemen wird für bestimmte Lehr- und Lernsituationen nur gelingen, wenn die Entwickler diese in einer angemessenen Zeit, zu annehmbaren Kosten und in einer überzeugenden Qualität zur Verfügung stellen können. Bei der Suche nach gegenwärtigen Lösungsansätzen zur Qualitätssicherung von Lernsystemen begibt man sich schnell auf „Neuland“ aufgrund der Dynamik von Lernsystem- und Internetentwicklungen. „Auf diesem Neuland sind bis jetzt noch keine ausgewachsenen Bäume von etablierten Forschungsparadigmen zu erkennen, sondern allenfalls zarte Pflänzchen erster Forschungsbemühungen“ (Dahinden *et al.* 2004, S.103). Wenige verfügbare wissenschaftliche Studien sind übertragbar bzw. verallgemeinerungsfähig. Für die Evaluationsforschung gilt zur Zeit, dass es zwar hier und da Studien, Gutachten, Publikationen und Lehrbücher zum Thema „Evaluation von Lernsystemen“ gibt, doch fachübergreifende Sammelbände, in denen evaluatorisches Wissen gebündelt und integriert wird, fehlen weitgehend (Stockmann 2000, S.31).

Im Rahmen dieser Arbeit wurde eine ansehnliche Anzahl von Lernsystem-Projekten evaluiert. Die eingesetzte Qualitätssicherungsmethodik wurde detailliert beschrieben. Im Rahmen der Qualitätssicherungsmethodik wurden neben der Expertenbefragung noch weitere Evaluationsmethoden genannt. Das Potenzial dieser Methoden muss in Zukunft ähnlich wie die durchgeführte Studie zur Expertenbefragung noch untersucht werden. Insbesondere der Bedarf eines Qualitätskriterienkatalogs für Evaluationsmethoden ist vorhanden. Ein solcher Katalog würde dann Entwicklern, Planern etc. helfen, für ihre Qualitätssicherung und ihren Evaluationskontext die geeignete Methode auszuwählen.

Derzeitig haben Lernsystem-Entwickler und Anbieter nun die Chance, die beschriebene Qualitätssicherungsmethodik und die Planungshilfen für Expertenbefragungen in eigenen Lernsystem-Projekten anzuwenden. Durch das in der Studie aufgezeigte Wissen können Expertenbefragungen ökonomischer geplant und mittels der LernSystemAnalyse-Software schnell umgesetzt und ausgewertet werden.

## Teil V

# Verzeichnisse

## Abbildungsverzeichnis

1	Qualitätsbeurteilung nach Wallmüller . . . . .	7
2	Subjektive und objektive Qualität . . . . .	9
3	Teufelsquadrat nach Sneed . . . . .	11
4	Evaluationsschritte des DIAMANTA Projektlebenszyklus nach Tiemann et al.	13
5	Evaluationsebenen Evaluationsnetz . . . . .	15
6	Wirkungskreise des Qualitätsverbesserungsprozesses . . . . .	26
7	Entwicklungsprozess . . . . .	27
8	Formativer Wirkungskreis . . . . .	31
9	Fragebogen: Qualitätskriterium Mensch . . . . .	39
10	Beschreibung des Qualitätskriteriums Mensch . . . . .	40
11	GUI LernSystemAnalyse . . . . .	49
12	Anforderungsschablone nach Rupp . . . . .	51
13	Klassifikation von Lernsystemen . . . . .	66
14	Fuzzy-Logik CBT . . . . .	70
15	Fuzzy-Logik WBT . . . . .	72
16	Fuzzy-Logik Kransteuerung . . . . .	73
17	Übersicht über Zustandsgrößen des Planspielsystems . . . . .	74
18	Beziehung zwischen zwei Zustandsgrößen . . . . .	76
19	Expertenanzahl nach Nielsen . . . . .	79
20	Qualitätskriterienbasierte Analyse . . . . .	90
21	Qualitätsindikatoren im Vergleich . . . . .	93
22	Fragenanzahl – Zeitaufwand . . . . .	96
23	Begründungstexte – Hinweisklassen . . . . .	98

24	Interaktionsgrad – Begründungstexte . . . . .	100
25	Fragenanzahl – Begründungstexte . . . . .	102
26	Fragenanzahl – Begründungstexte . . . . .	103
27	Expertengrad – Hinweisklassen und Standardabweichung der Bewertungspunktzahlen . . . . .	105
28	Ungewichtete Bewertung . . . . .	105
29	Anforderungserfüllung – Begründungstexte . . . . .	108
30	Hinweisklassen im Projektvergleich . . . . .	110
31	Monopol-Hinweisklassen im Projektvergleich . . . . .	113
32	Monopol-Hinweisklassen . . . . .	114
33	PS-Perspektive der Begründungstexte . . . . .	115
34	PS-Konturen der Begründungstexte . . . . .	115
35	Modell-Perspektive der Begründungstexte . . . . .	116
36	Modell-Konturen der Begründungstexte . . . . .	117
37	Standardabweichungen PAS . . . . .	119
38	Max. Standardabweichungen PAS . . . . .	120

## Tabellenverzeichnis

1	Lehrveranstaltungsevaluierung im WS2004/2005: FB15 . . . . .	5
2	Evaluationsmethoden . . . . .	15
3	WWW-, mündliche, telefonische und schriftliche Befragung . . . . .	36
4	Vor- und Nachteile der Computerbefragung nach Fuchslocher und Fitz . . . . .	37
5	Vorhersagerate für Endbenutzer-Probleme nach Desurvivre . . . . .	41
6	Metainformationen nach Fuhrmann Verdaguer . . . . .	43
7	Vorteile verschiedener Fragefolgen nach Theobald . . . . .	44
8	Gruppenbefragung vs. Einzelbefragung . . . . .	47
9	Lernsystem-Entwicklungen . . . . .	68
10	Usability-Labore nach Bailey . . . . .	81
11	Usability-Inspektionen nach Hertzum <i>et al.</i> . . . . .	81
12	Usability-Tests nach Molich <i>et al.</i> . . . . .	81
13	Qualitätsindikatoren . . . . .	93
14	Zeitaufwand . . . . .	96
15	Interaktionsgrad – Anforderungserfüllung . . . . .	100
16	Bewertungsunterschiede Expertengrad . . . . .	104
17	Schätzung der Hinweisklassen . . . . .	111
18	Standardabweichung Gewichtung . . . . .	119
19	Regressionsanalyse Standardabweichung . . . . .	120
20	Begründungstexte und Hinweisklassen . . . . .	122
21	Qualitätskriterien – Hinweisklassen . . . . .	125

## 18 Glossar

**Anforderungen** an ein Lernsystem entstehen, wenn Menschen bestimmte oder potenzielle Bedürfnisse und Präferenzen an das Lernsystem stellen (werden). Muss-Anforderungen müssen mit dem Lernsystem-Produkt auf jeden Fall erreicht werden. Hingegen sind Soll-Anforderungen weniger verbindlich. Eine Anforderungsanalyse sorgt für die Richtigkeit und Vollständigkeit der Anforderungen für die Implementierung.

**Applet** bzw. **Java-Applet** ist ein zumeist interaktives Programm, welches im Internetbrowser gestartet und angezeigt wird. Für den Ablauf ist das Java-Runtime-Environment erforderlich.

**Autorensysteme** Durch sie wird die Entwicklung und Pflege von Lernsystemen in effizienter Weise gefördert, da die Programmierarbeit und die Qualitätssicherung durch Standardisierung und Wiederverwendung von bereits optimiertem Programmcode beschleunigt werden kann.

**Blended Learning** Kombination von E-Learning und Präsenzlehre.

**CBT** Computer Based Training: Offline-Lernsystem.

**Datenbank** In einer Datenbank können unterschiedliche Daten organisiert und abgerufen werden. Dies geschieht über SQL-Befehle (Structured Query Language). Häufig eingesetzte Datenbanksysteme sind MySQL, MS ACCESS, Oracle, PostgreSQL u.a.

**Didaktik** ist die Theorie des Lehrens und Lernens. Ballstaedt (1997, S.12) definiert didaktisches Design als die planmäßige und lernwirksame Entwicklung von Lernumgebungen (von der Bedarfsanalyse bis zur Evaluation) auf wissenschaftlicher Grundlage.

**E-Learning** Lehren und Lernen mit Hilfe elektronischer Medien.

**Evaluation** ist eine sachgerechte Beurteilung bzw. Bewertung. Formative Evaluation findet interaktiv während der Entwicklung von Prototypen statt. Summative Evaluation erfolgt am vollständig entwickelten Lernsystem.

**Evaluationsmethoden** In dieser Arbeit wird die Expertenbefragung mittels Qualitätskriterienkatalog als eine für den formativen Entwicklungsprozess passende Evaluationsmethode identifiziert, die schwerpunktmäßig in dieser Arbeit beschrieben wird. Weitere Evaluationsmethoden sind Interview, lernzielorientierte Tests, Gruppendiskussion, Beobachtung, Fallstudie/Erfahrungsberichte, Dokumentenanalyse, Usability Labor etc. (Meister *et al.* 2004, S.11).

**Expertenbefragung** ist eine Evaluationsmethode, über die Hinweise von Experten als Wissensträger zu Qualitätsaspekten gewonnen werden.

**GUI** Graphical User Interface, zu deutsch grafisches Benutzerinterface.

**Hinweisklassen** Der Begriff steht für die Anzahl der Kategorien, die thematisch zusammengefasste Aussagen der Experten zur Qualität des Lernsystems enthalten. Die Bedingung für die Zuordnung und Benennung einer Hinweisklasse ist, dass mindestens eine Aussage eines Experten existiert, die als Grundlage zur Ableitung einer Verbesserungsmaßnahme des Lernsystems dienen kann.

**HTML** HyperText Markup Language ist eine Textauszeichnungssprache. Durch sie wird die Darstellung / Formatierung eines Textes in einem Browser vereinbart.

**Java** ist eine sehr vielseitig einsetzbare, plattformunabhängige Programmiersprache. Sie wird beispielsweise für die Webprogrammierung von Java-Applets, Java-Applikationen und JSP-Webseiten (Java Server Pages) benötigt. JSP-Webseiten benötigen eine Serverumgebung wie z.B. den Jakarta Tomcat Server.

**Javascript** ist nicht zu verwechseln mit Java. Es ist eine clientseitig auszuführende in HTML-Text integrierte Programmiersprache.

**Kap.** Abkürzung für Kapitel

**Lebenslanges Lernen** wird beschrieben als notwendige Bereitschaft sich aus- und weiterzubilden über die Schulzeit bzw. das Studium hinaus ein Leben lang, um mit den ständigen Veränderungen durch immer wiederkehrende Umlernprozesse Schritt zu halten. Das Studium kann als eine Basis- und Initialqualifikation gesehen werden.

**Lebensphasen** Der Entwicklungszyklus eines Lernsystems kann in unterschiedliche technisch-organisatorische Abschnitte untergliedert werden. In der Systemtechnik werden in typischen Lebensphasenmodellen beispielsweise die Phasen Vorentwicklung, Hauptentwicklung, Detailentwicklung, Systembau, Systemeinführung und Systemnutzung unterschieden.

**Lerntheorien** Nach der Definition von Gage und Berliner (1996, S.230) ist Lernen der Prozess, durch den ein Organismus sein Verhalten als Resultat von Erfahrung dauerhaft ändert. Lerntheoretiker beschäftigen sich seit Anfang des 20. Jahrhunderts mit den Gesetzmäßigkeiten dieser Verhaltensänderungen. Mit den entwickelten Lerntheorien oder Lernparadigmen sind verschiedene Auffassungen verbunden, die die Entwicklung von Lernsystemen in der Vergangenheit geprägt haben. Die wissenschaftstheoretischen Positionen lassen sich vor allem den Richtungen des Behaviorismus, Kognitivismus und Konstruktivismus zuordnen. Die bekanntesten behavioristischen Theorien stammen von Pawlow, I. (1849–1936) und Skinner, B. (1904–1990). Lernen ist nach deren Theorie ein konditionierter Reflex: Ein bestimmter Stimulus ruft ein Verhalten hervor. Im Unterschied zum Behaviorismus versucht der Kognitivismus aufbauend auf den Untersuchungen von Köhler, W. (1887–1967), die im Gehirn stattfindenden Verarbeitungsprozesse, die den Input in Output umsetzen, durch Modelle zu beschreiben. Mit der konstruktivistischen Auffassung von Piaget, J. (1896–1980) ist die Ansicht verbunden, dass Wissen an das Individuum gebunden ist und in der handelnden Auseinandersetzung mit der Umwelt konstruiert wird.

**Lernsystem** Ein Lehr- und Lernsystem ist eine mediendidaktische (Bildungs-) Umgebung, Software bzw. Programm mit der Zielsetzung Lehr- und Lernprozesse zu ermöglichen. In dieser Arbeit wird aufgrund der schnelleren Lesbarkeit stellvertretend der Begriff „Lernsystem“ verwendet.

**Log-Files** Die Nutzung von Software kann durch sogenannte Log-Files aufgezeichnet werden. Auf diese Weise können Administratoren die Interaktionen der Benutzer mit der Software auswerten und nachvollziehen.

**LSA** Abkürzung für die LernSystemAnalyse-Software, die ein Werkzeug zur Qualitätssicherung von Lernsystemen darstellt.

**Medienkompetenz** bezeichnet den verantwortungsvollen und kompetenten Umgang mit Medien verschiedenster Natur.

**Metainformationen** sind zusätzlich verfügbare Informationen für den Bewertungsprozess eines Experten.

**Monopol-Hinweisklassen** Liegt bei der Formulierung einer Hinweisklasse nur ein Begründungstext eines Experten zugrunde, wird eine Hinweisklasse als Monopol-Hinweisklasse spezifiziert.

**Multimedia** steht für den Einsatz von vielen Medien zur Vermittlung und Übertragung von Informationen und kennzeichnet die Bedienung von mindestens zwei der menschlichen Sinne. „Multimedia begann, als das erste Klavier ins Stummfilmkino geschoben wurde“ (Schulmeister 1992, S.19). Die Multimedia-Koordinationsstelle der Universität Kassel ist der zentrale Ansprechpartner für Lehrende, die über die traditionellen Lehrformen hinaus, die Neuen Medien in ihre Veranstaltungen einbinden möchten (<http://www.uni-kassel.de/hrz/e-learning>).

**Navigation** ist hier die Steuerung/Bedienung einer Software.

**PHP** Personal HomePagetools ist eine serverseitige Programmiersprache zur Entwicklung dynamischer Webseiten. Der PHP-Code wird zumeist in den HTML-Text einer Datei eingebettet. Der PHP-Code wird vom Server ausgeführt, beispielsweise um Daten einer Datenbank in den HTML-Text zu integrieren. Die vervollständigte HTML-Seite wird dann im Browser des anfordernden Client-Rechners angezeigt.

**Prototyping** hilft bei der Spezifikation der Entwicklung von Lernsystemen durch die schnelle Entwicklung von Prototypen. Der Prototyp selbst sollte erweiterbar, aber nicht notwendigerweise vollständig sein (siehe UKap. 7.5 auf Seite 30).

**Prozess- und Projektmanagement** heißt nichts anderes als das Planen, Organisieren, 'Staffing' (richtige Person an der richtigen Stelle), Entscheiden, Anordnen, Durchführen, Überwachen, Kontrollieren, Koordinieren, Inanghalten eines Projekts.

**Qualität** In der Kategorienlehre des griechischen Philosophen Aristoteles (geboren 348 vuz) stellt die Qualität eine der Kategorien dar, die zu den Grundbestimmungen aller Dinge gehört. Qualität, lat. Beschaffenheit/Eigenschaft, ist in erster Linie die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen (Deutsches Institut für Normung 1992). Qualität ist in zweiter Linie die Güte oder der Wert eines Objektes. Qualität ergibt sich immer aus der Übereinstimmung zwischen den Erwartungen hinsichtlich einer Leistung und der dann tatsächlich erbrachten Leistungen (Merfeld und Ogrzall 1998, S.18). Auf E-Learning angewandt, kann Qualität definiert werden als die Beschaffenheit eines Lernsystems und der damit verbundenen Eignung, die zwischen E-Learning-Anbietern und den Zielgruppen vereinbarten Eigenschaften zu erfüllen (Astleitner 2002).

**Qualitätsindikatoren** Auf der Grundlage von verschiedenen Qualitätsindikatoren können Aussagen zur Qualität gefällt werden. Die Qualitätsindikatoren in Form von Kennzahlen helfen, Lernsystem-Prototypen zu vergleichen, Maßnahmen zur Anforderungserfüllung zu entscheiden und diese nach Dringlichkeit umzusetzen.

**Qualitätskriterien** sind 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“ (Fricke 2000, S.75). Für eine Evaluation von Lernsystemen sind Qualitätskriterien unverzichtbar. Qualitätskriterien fokussieren und strukturieren Qualitätsaspekte, die ein befragter Experte bei der Evaluation eines Lernsystems berücksichtigen soll (siehe UKap. 6.3 auf Seite 16). Die Qualitätskriterien werden durch die Lernsystem-Konfiguration hierarchisiert nach Qualitätsaspekten Mensch, Lernumfeld, Lerntechnologie, Lerninhalte und Lehrmethodik.

**Qualitätsmanagement** ist eine auf der Mitwirkung aller Projektmitglieder basierende Managementmethode einer Organisation, die Qualität in den Mittelpunkt stellt und die durch Zufriedenstellung der Kunden auf langfristigen Geschäftserfolg sowie auf Nutzen für die Mitglieder der Organisation und für die Gesellschaft zielt (Deutsches Institut für Normung 1992). Qualitätsmanagement von E-Learning betrifft demnach alle planenden, steuernden und organisatorischen Aktivitäten, die die Eigenschaften von Lernprodukten und -prozessen im Rahmen webbasierter Instruktion beeinflussen (Astleitner 2002).

**Qualitätsregelteam** Das Qualitätsregelteam sorgt für die konsequente Umsetzung der systematischen und ganzheitlichen Qualitätssicherung von Lernsystemen.

**Qualitätssicherung** Unter diesem Begriff sind alle qualitätsbezogenen Maßnahmen zu verstehen, die dauerhaft die geforderte Qualität gewährleisten und die Zahl der notwendigen Nacharbeit, Beanstandungen, Änderungen verringert.

**QVP** Der QualitätsVerbesserungsProzess stellt eine Organisation zur kontinuierlichen Verbesserung des Lernsystems dar, die durch während der Entwicklung regelmäßig stattfindende Evaluationen erwirkt wird.

**RAD** Ein Rapid Application Development ist ein Programmiersystem, welches die schnelle Implementierung eines laufenden Programms unterstützt.

**Selbststeuerung** Die Lernenden bestimmen Breite und Tiefe ihres Lernprozesses ohne regulierende Eingriffe (oder Zwang) von außen.

**Systemtechnik** Die Systemtechnik (Systems Engineering) lehrt Aspekte und Konzepte zum Verständnis komplexer Systeme und bietet Techniken zur Gestaltung (Modellbildung), Beschreibung, Bewertung und Beherrschung von Systemen.

**Telematik** verknüpft die Technologiebereiche Telekommunikation und die Informatik.

**UKap.** Abkürzung für Unterkapitel

**Vorentwicklung** In der Vorentwicklungsphase erfolgt die Planung der Lernsystem-Entwicklung z.B. die Aufstellung von Anforderungen, Qualitätskriterien und Evaluationsmethoden.

**WBT** Web Based Training: internetbasiertes Lernen.

## Literaturverzeichnis

- Akao, Y. (1990). An Introduction of Management Engineering. In: *Quality function deployment integrating customer requirements into product design* (Akao, Y., Hrsg.). Productivity Press. Portland.
- Astleitner, H. (2002). Qualitätsmanagement von E-Learning. [http://www.sbg.ac.at/erz/as/AS\\_QUAL2.DOC](http://www.sbg.ac.at/erz/as/AS_QUAL2.DOC).
- Bagel, T. (2001). Studentische Urteile zur Lehr- und Studienqualität. Erwartungen, Erfahrungen und Probleme. <http://www.uni-konstanz.de/FuF/SozWiss/fg-soz/ag-hoc/Vortaege/StudentLehreStudquali.pdf>.
- Bailey, B. (2001). How reliable is usability performance testing? <http://www.human-factors.com/downloads/sep012.htm>.
- Ballstaedt, S.-P. (1997). *Wissensvermittlung. Die Gestaltung von Lernmaterial*. Psychologie Verlags Union. Weinheim.
- Blom, H. (2000). *Der Dozent als Coach*. Kriftel. Neuwied.
- BmBF (2002). *Bericht der Bundesregierung zur Bildung für eine nachhaltige Entwicklung*. Bundesministerium für Bildung und Forschung BmBF. Referat Öffentlichkeitsarbeit. Bonn. [http://www.bmbf.de/pub/bb\\_bildung\\_nachhaltige\\_entwicklung.pdf](http://www.bmbf.de/pub/bb_bildung_nachhaltige_entwicklung.pdf).
- BmBF (2005). *DLR – Projektträger Neue Medien in der Bildung*. Bundesministerium für Bildung und Forschung BmBF. Bonn. <http://www.medien-bildung.net>.
- Bretz, S. (2005). *Ein Lernsystem zur Fuzzy-Logik am Beispiel einer Kransteuerung*. Universität Kassel (Diplomarbeit). Kassel.
- Bürli, A. (1975). Computer-unterstützte und verzweigte Testdurchführung. *Schweizerische Zeitschrift für Psychologie* (2), 7–10.
- Cunningham, W. und B. Leuf (2001). *The Wiki Way: Quick collaboration on the Web*. Addison-Wesley Longmann. Boston.
- Dahinden, U., P. Kaminski und Niederreuther R. (2004). Content is King. Gemeinsamkeiten und Unterschiede bei der Qualitätsbeurteilung aus Angebots- vs. Rezipientenperspektive. In: *Gute Seiten — schlechte Seiten. Qualität in der Onlinekommunikation* (Beck, K., Schweiger, W., Wirth, W., Hrsg.). Verlag Reinhard Fischer. München.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life by Charles Darwin*. Murray. London.

- DeGEval (2003). *Kompendium der Standards zur Selbstevaluation*. Gesellschaft für Evaluation e.V.. Mainz.  
Zur Vorlage bei der Mitgliederversammlung der DeGEval am 8.10.03 in Hamburg.  
<http://www.degeval.de/dokumente/Kompendium%20der%20Standards%20zur%20Selbstevaluation.pdf>.
- Dekanat Maschinenbau (2005). *Mitteilungen aus dem Dekanat Maschinenbau. Nr. 8 (31.1.2005)*. FB15 Maschinenbau. Kassel.
- Deming, W. E. (1982). *Quality, Productivity, and Competitive Position*. Massachusetts Institute of Technology. Massachusetts.
- Deming, W. E. (1986). *Out of the Crisis*. MIT Press. Cambridge, Mass.
- Desurvivre, H. W. (1994). Faster, cheaper! Are Usability Inspection Methods as effective as empirical testing?. In: *Usability Inspection Methods* (Nielsen, J., Mack, R., L., Hrsg.). John Wiley. New York.
- Deutsches Institut für Normung (1992). *DIN EN ISO 8402: Qualitätsmanagement und Qualitätssicherung, Begriffe*. Beuth Verlag. Berlin.
- Deutsches Institut für Normung (2000). *DIN EN ISO 9000:2000: Qualitätsmanagement, Statistik, Umweltmanagement: Anwendungshilfen und Normensammlungen*. Beuth Verlag. Berlin.
- Deutsches Institut für Normung (2004). *Aus- und Weiterbildung unter besonderer Berücksichtigung von E-Learning. Referenzmodell für Qualitätsmanagement und Qualitätssicherung. Planung, Entwicklung, Durchführung und Evaluation von Bildungsprozessen und Bildungsangeboten*. Beuth Verlag. Berlin.
- Dick, E. (2000). *Multimediale Lernprogramme und telematische Lernarrangements. Einführung in die didaktische Gestaltung*. BW Bildung und Wissen. Nürnberg.
- Dohmen, G. (1996). *Das lebenslange Lernen. Leitlinien einer modernen Bildungspolitik*. BMBF. Bonn.
- Ehlers, U.-D. (2004). *Qualität im E-Learning aus Lernaltersicht. Grundlagen, Empirie und Modellkonzeption subjektiver Qualität*. VS Verlag für Sozialwissenschaften. Wiesbaden.
- Evaluationsnetz (2005). Kriterienkataloge.  
<http://www.evaluationsnetz.de/content.php?cat=3&id=320>.
- Ewen, V. und H. Gelszus, R. (1990). Der Computer als Befragungsinstrument. *Marktforschung & Management*.

- Faure, E., F. Herrera, A.-R. Kaddura, H. Lopes, A. V. Petrowski, M. Rahnema und F. Champion Ward (1973). *Wie wir leben lernen. Der UNESCO-Bericht über Ziele und Zukunft unserer Erziehungsprogramme..* Rowohlt. Reinbek.
- Flehsig, K.-H. (1996). *Kleines Handbuch didaktischer Modelle.* Neuland. Eichenzell.
- Fricke, R. (1995). Evaluation von Multimedia. In: *Information und Lernen mit Multimedia* (Issing, L. J., Klimsa, P., Hrsg.). Psychologie Verlags Union. Weinheim.
- Fricke, R. (2000). Qualitätsbeurteilung durch Kriterienkataloge. Auf der Suche validen Vorhersagemodellen. In: *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme auf dem Prüfstand* (Schenkel, P., Tergan, S.-O., Lottmann, A., Hrsg.). Bildung und Wissen. Nürnberg.
- Fuchslocher, H. und M. Fitz (1992). Computergestützte Marktforschung. Erfahrungen aus dem Bereich der Messemarktforschung. In: *Handbuch des Electronic Marketing* (Hermanns, A., Flegel, V., Hrsg.). Beck. München.
- Fuhrmann Verdaguer, E. (2005). *Design, Implementation and Testing of an Expert Self Diagnosis System, S.D.S..* Universität Kassel (Diplomarbeit). Kassel.
- Gage, N. L. und C. Berliner, D. (1996). *Pädagogische Psychologie.* Psychologie Verlags Union. Weinheim.
- Gigerenzer, G. (1997). Intelligent denkt, wer sich nicht auf Logik verlässt. *GEO Denken. Die alltägliche Erleuchtung.*
- Gräber, W. (1990). *Das Instrument MEDA. Ein Verfahren zur Beschreibung, Analyse und Bewertung von Lernprogrammen.* Institut für die Pädagogik der Naturwissenschaften. Kiel.
- Gräber, W. (1996). *Kriterien und Verfahren zur Sicherung der Qualität von Lernsoftware in der beruflichen Weiterbildung.* Institut für die Pädagogik der Naturwissenschaften. Kiel.
- Grune, C. (2000). *Lernen in Computernetzen. Analyse didaktischer Konzepte für vernetzte Lernumgebungen.* KoPäd-Verlag. München.
- Götte, A. und K. Kümmerlein (1996). Der Einsatz von Multimedia in der Marktforschung. *Planung & Analyse.*
- Haben, M. (2002). Beim E-Learning dazugelernt. *IT Management.*
- Hage, N. (1996). *Lehrevaluation und studentische Veranstaltungskritik. Projekte, Instrumente und Grundlagen.* bmbf. Bonn.

- Hahne, F. (2003). *Interaktive Websites*. FranzisVerlag GmbH. Poing.
- Hamborg, K.-C. (2002). Gestaltungsunterstützende Evaluation von Software. Zur Effektivität und Effizienz des IsoMetrics<sup>L</sup> Verfahrens. In: *Mensch & Computer 2002. Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*. (Herczeg, M., Prinz, W., Oberquelle, H., Hrsg.). Teubner. Stuttgart, Leipzig, Wiesbaden.
- Hampe-Neteler, W. (1994). *Software-ergonomische Bewertung zwischen Arbeitsgestaltung und Softwareentwicklung*. Lang. Frankfurt a. M.
- Hauff, M. (1998). *media@uni-multi.media? Entwicklung – Gestaltung – Evaluation neuer Medien*. Waxmann. Münster, New York, München, Berlin.
- Haupt, K. (1996). Perspektiven der Marktforschung. *Planung & Analyse*.
- Hertzum, M., N. E. Jacobsen und R. Molich (2002). Usability Inspections by Groups of Specialists: Perceived Agreement in Spite of Disparate Observations. In: *Extended Abstracts of the ACM CHI 2002 Conference (Minneapolis, MN)*. ACM Press. New York.
- Hippler, H.-J., F. Meier und N. Schwarz (1988). Erste Erfahrungen mit der Erprobung eines interaktiven Befragungs- und instruktionssystems (IBIS). *ZUMA-Nachrichten*.
- Hippler, H.-J. und A. Beckenbach (1992). Das persönlich-mündliche Interview am Scheideweg?. *Planung & Analyse*.
- Hoepner, G. (1994). *Computereinsatz bei Befragungen*. Gabler. Wiesbaden.
- Human Factors International (2001). Multimedia and Learning. <http://www.human-factors.com/downloads/oct01.asp>.
- ISO 13407 (1999). *Human-centred design processes for interactive systems*. ISO. Genf.
- Jeffries, R., J. R. Miller, C. Wharton und K. M. Uyeda (1991). User interface evaluation in the real world. A comparison of four techniques. In: *Proceedings ACM CHI'91 Conference*. New Orleans.
- Jelitto, M. (2005). Kriterienkataloge.  
<http://www.evaluiieren.de/evaluati.on/methoden/0001.htm>.
- Johannsen, G. (1993). *Mensch-Maschine-Systeme*. Springer-Verlag. Berlin, Heidelberg.
- Johannsen, G., J. Heuer und M. Tiemann (1997). Participative User Interface Design for Industrial Processes. In: *ErgoCon'97 Conference*. Palo Alto.
- Judd, C. M., E. R. Smith und L. H. Kidder (1991). *Research methods in Social Relations*. Harcourt Brace Jovanovich College Publishers. Fort Worth.

- Karat, C., R. L. Campbell und T. Fiegel (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In: *Proceedings ACM CHI'92 Conference*. Monterey.
- Kaufmanns, R. (2002). Schulbank im Gepäck. Erfolgskriterien für mobiles E-Learning. *IT Management*.
- Kerres, M. (2001). *Multimediale und telemediale Lernumgebungen. Konzeption und Entwicklung*. Oldenbourg. München, Wien.
- Klauer, K. J. (1985). Framework for a theory of teaching. *Teaching and teacher education*.
- Kleppmann, W. (2003). Versuchsplanung – Produkte und Prozesse optimieren. In: *Praxisreihe Qualitätswissen* (Brunner, F. J., Hrsg.). Hanser Verlag. München, Wien.
- Kubicek, H., H.-J. Braczyk, D. Klumpp, G. Müller, W. Neu, E. Raubold und A. Roßnagel (1998). *Lernort Multimedia – Jahrbuch Telekommunikation und Gesellschaft*. Deckers Verlag. Heidelberg.
- Laube, J. (1994). Internationale Erfahrungen mit computerunterstützten telefonischen Befragungen in der amtlichen Statistik. *Wirtschaft & Statistik*.
- Liggesmeyer, P. (2002). *Software-Qualität. Testen, Analysieren und Verifizieren von Software*. Spektrum, Akademischer Verlag. Heidelberg, Berlin.
- Lömker, K. (1998). Ausgewählte Problembereiche zur Integration multimedialer Lehre in die Hochschulen. In: *media@uni-multi.media? Entwicklung - Gestaltung - Evaluation neuer Medien* (Hauff, M., Hrsg.). Waxmann. Münster, New York, München, Berlin.
- Lüthje, J. (1997). Verfahren und Elemente systemischer Qualitätsentwicklung. Beispiele aus der Universität Hamburg. In: *Qualitätsentwicklung in einem differenzierten Hochschulsystem* (Stifterverband für die Deutsche Wissenschaft, Hrsg.). Symposium. Bonn.
- Mandl, T. (2001). *Die automatische Bewertung der Qualität von Wissensprodukten. Woran erkennt der Computer eine „gute“ Internet-Seite?*. Informationswissenschaft Universität Hildesheim. Arbeitsbericht Nr. 3/2001 (Projekt AQUAINT). Hildesheim. [http://www.uni-hildesheim.de/%7Emandl/Publikationen/Ab\\_aquaint02.pdf](http://www.uni-hildesheim.de/%7Emandl/Publikationen/Ab_aquaint02.pdf).
- Meier, A. (1995). Qualitätsbeurteilung von Lernsoftware durch Kriterienkataloge. In: *Evaluation multimedialer Lernprogramme und Lernkonzepte* (Schenkel, P., Holz, H., Hrsg.). BW Bildung und Wissen. Nürnberg.
- Meister, D. M., S.-O. Tergan und P. Zentel (2004). Evaluation von E-Learning – Eine Einführung. In: *Evaluation von E-Learning – Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven* (Meister, D., M., Tergan, S.-O., Zentel, P., Hrsg.). Waxmann. Münster.

- Merfeld, M. und P. Ogrzall (1998). Qualitätssicherung im Arbeitskreis deutscher Bildungsstätten. In: *Reihe Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe* (Bundesministerium für Familie, Senioren, Frauen und Jugend, Hrsg.). Number 20. BMFSFJ. Berlin.
- Molich, R., N. Bevan, I. Curson, S. Butler, E. Kindlund, D. Miller und J. Kirakowski (1998). Comparative Evaluation of Usability Tests. In: *Proceedings of the Usability Professionals Association Conference (UPA98)*. Washington D.C.
- Naumann, A., J. Waniek und J. F. Kreams (1999). Wissenserwerb, Navigationsverhalten und Blickbewegungen bei Text und Hypertext. *Aktuelle Online-Forschung: Trends, Techniken, Ergebnisse*. [http://www.dgof.de/tband99/pdfs/i\\_p/naumann.pdf](http://www.dgof.de/tband99/pdfs/i_p/naumann.pdf).
- Niegemann, H. N., S. Hessel, D. Hochscheid-Mauel, D. Hochscheid-Mauel und K. Aslanski (2004). *Kompendium E-Learning*. Springer x.media.press. Berlin.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In: *Proceedings ACM CHI'92 Conference*. Monterey.
- Nielsen, J. (1993). *Usability Engineering*. Academic Press Professional. San Diego.
- Nielsen, J. (1994). Heuristic evaluation. In: *Usability inspection methods* (Nielsen, J., Mack, R., L., Hrsg.). John Wiley. New York.
- Nielsen, J. (2000). Why You Only Need to Test With 5 Users. <http://www.useit.com/alertbox/20000319.html>.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Wiley. New York.
- Petermandl, M. (1991). *Optimierung des Einsatzes von Medien in der beruflichen Weiterbildung*. Erich Schmidt. Berlin.
- Prichard, W. H. J., T. Micceri und A. J. Barret (1993). A review of computer-based training materials. Current state of the art (instruction and interaction). *Educational Technology Research & Development*.
- Rössler, P. (2004). Qualität aus transaktionaler Sicht. Zur gemeinsamen Modellierung von User Quality und Sender Quality. In: *Gute Seiten – schlechte Seiten. Qualität in der Onlinekommunikation* (Beck, K., Schweiger, W., Wirth, W., Hrsg.). Verlag Reinhard Fischer. München.
- Rubin, J. (1994). *Handbook of Usability Testing*. John Wiley&Sons. New York.
- Rupp, C. (2002). *Requirements-Engineering und -Management. Professionelle, iterative Anforderungsanalyse für die Praxis*. Carl Hanser Verlag. New York.

- Schenkel, P. (2000a). *Lerntechnologien in der beruflichen Bildung. Zehn Thesen*. Bundesinstitut für Berufsbildung. Bonn.
- Schenkel, P. (2000b). *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme*. BW Bildung und Wissen. Nürnberg.
- Schott, F. (2000). Evaluation aus theoriegeleiteter, ganzheitlicher Sicht. In: *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme* (Schenkel, P., Tergan, S.-O., Lottmann, A., Hrsg.). BW Bildung und Wissen. Nürnberg.
- Schulmeister, R. (1992). *Grundlagen hypermedialer Lernsysteme. Theorie-Didaktik-Design*. Oldenbourg. München.
- Schulmeister, R. (1998). Medien und Hochschuldidaktik. Welchen Beitrag können neue Medien zur Studienreform leisten?. In: *media@uni-multi.media? Entwicklung – Gestaltung – Evaluation neuer Medien* (Hauff, M., Hrsg.). Waxmann. Münster, New York, München, Berlin.
- Schulmeister, R. (2005). Kriterien didaktischer Qualität im E-Learning zur Sicherung der Akzeptanz und Nachhaltigkeit. In: *E-Learning in Hochschulen und Bildungszentren* (Euler, D., Seufert, S., Hrsg.). Oldenbourg. München.
- Scriven, M. (1967). The methodology of evaluation. In: *Perspectives of curriculum evaluation* (Scriven, M., Tyler, R. W., Gagné, R. M., Hrsg.). Rand McNally. Chicago.
- Shneiderman, B. (2002). *User Interface Design*. Mitp-Verlag. Bonn.
- Sneed, H. M. (1987). *Software-Management*. Verlagsgesellschaft R. Müller. Köln.
- Sproull, L. S. (1986). Using Electronic Mail for Data Collection in Organizational Research. *Academy of Management Journal* (29), 159–169.
- Stockmann, R. (2000). *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Leske und Budrich. Opladen.
- Stockmann, R. (2004). Konzepte und Methoden für die Evaluation von E-Learning. In: *Evaluation von E-Learning – Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven* (Meister, D., M., Tergan, S.O., Zentel, P., Hrsg.). Waxmann. Münster.
- Stufflebeam, D. L. (1972). Evaluation als Entscheidungshilfe. In: *Evaluation, Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (C. Wulf, Hrsg.). Piper. München.
- Tergan, S.-O. (1996). Evaluation of software for computer-based Learning. In: *Staff training in media use for learning and teaching* (Wedekind, J., Hrsg.). Deutsches Institut für Fernstudienforschung. Tübingen.

- Tergan, S.-O. (2000). 10+1 Fragen zur Evaluation von Bildungssoftware. In: *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand* (Schenkel, P., Tergan, S.-O., Lottmann, A., Hrsg.). BW Bildung und Wissen. Nürnberg.
- Tergan, S.-O. (2001). Qualitätsbeurteilung von Bildungssoftware mittels Kriterienkatalogen. Problemaufriss und Perspektiven. *Unterrichtswissenschaft*.
- Thaller, G. E. (2000). *Softwaremetriken einsetzen, bewerten, messen*. Verlag Technik. Berlin.
- Theobald, A. (2000). *Das World Wide Web als Befragungsinstrument*. Gabler. Wiesbaden.
- Thomé, D. (1989). *Kriterien zur Bewertung von Lernsoftware*. Dr. Alfred Hüthig Verlag. Heidelberg.
- Tiemann, M., E. A. Averbukh und G. Johannsen (1996). *Evaluation and Measurement Procedure. Interner Bericht D4 DIAMANTA. ESPRIT Projekt 20507*. Fachgebiet Systemtechnik und Mensch-Maschine-Systeme. Universität Kassel. Kassel.
- Tversky, A. und D. Kahneman (1971). Belief in the law of small numbers. *Psychological Bulletin* (76), 105–110.
- Vester, F. (1999). *Die Kunst vernetzt zu denken. Ideen und Werkzeuge für einen neuen Umgang mit der Komplexität*. DVA. Stuttgart.
- Virzi, R. A., J. F. Sorce und L. B. Herbert (1993). A Comparison of Three Usability Evaluation Methods: Heuristic, Think Aloud and Performance Testing. In: *Designing for diversity* (Human Factors and Ergonomics Society, Hrsg.). Human Factors and Ergonomics Society. Santa Monica.
- Wagner, I. (2005). *Einsatz und Evaluation eines nutzergerechten Autorensystems für webbasierte Planspiele*. Kassel University Press GmbH. Kassel.
- Wallmüller, E. (1995). *Ganzheitliches Qualitätsmanagement in der Informationsverarbeitung*. Hanser. München, Wien.
- Walraet, B. (1991). *A discipline of software engineering*. North-Holland. New York.
- Watt, J. H. (1997). Using the Internet for Quantitative Survey Research  
. <http://www.SwiftInteractive.com/white%20papers/Watt/InternetSurvey.html>.
- Webler, W.-D., V. Domeyer und B. Schieber (1993). *Lehrberichte. Empirische Grundlagen Indikatorauswahl und Empfehlungen zur Darstellung der Situation der Lehre in Lehrberichten*. Bock Verlag. Bad Honnef.

- Will, C. (1997). Neue Medien – Neue Marktforschung: Zum Einsatz von Neuen Medien in der Marktforschung. *Marktforschung & Management*.
- Will, H., A. Winteler und A. Krapp (1987). *Evaluation in der beruflichen Aus- und Weiterbildung. Konzepte und Strategien*. Sauer. Heidelberg.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control* (8), 338–353.
- Zimmer, G. und E. Psaralidis (2000). Der Lernerfolg bestimmt die Qualität einer Lernsoftware! Evaluation von Lernsoftware als logische Rekonstruktion von Handlungen. In: *Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand* (Schenkel P., Tergan, S.-O., Lottmann, A., Hrsg.). BW Bildung und Wissen. Nürnberg.