

# Online Lehrevaluationssysteme

Anforderungen und Implementation  
am Beispiel des Medizinischen Curriculums

Marc Gumpinger



München 2008



Aus  
dem Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie  
der Universität München  
Direktor: Prof. Dr. Ulrich Mansmann

**Online-Lehrevaluationssysteme:  
Anforderungen und Implementation  
am Beispiel des Medizinischen Curriculums**

Dissertation  
zum Erwerb des Doktorgrades der Humanbiologie  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

vorgelegt von  
Marc Gumpinger  
aus München  
im Jahr 2008

Mit Genehmigung der Medizinischen Fakultät  
der Universität München

Berichterstatter: Prof. Dr. Ulrich Mansmann

2. Berichterstatter: Prof. Dr. Michael Meyer

Mitberichterstatter: Prof. Dr. Helmut Bartels

Prof. Dr. Rolf Holle

Mitbetreuung durch: Dr. Martin Fischer

Dekan: Prof. Dr. Dietrich Reinhardt

Tag der mündlichen Prüfung: 7. April 2008





*für Karin und Mia*





## **Danksagung**

Neben der Annahme des Zusatzaufwands, den die Betreuung einer Doktorarbeit bedeutet, möchte ich mich bei meinem Doktorvater Herrn Prof. Dr. Ulrich Mansmann besonders für die Freiheit bei der Themenwahl und deren Umsetzung sowie für seine hilfreichen Hinweise und seinen Rückhalt sehr herzlich bedanken.

Mein großer Dank gilt dem Evaluationsbeauftragten der Medizinischen Fakultät der Ludwig-Maximilians-Universität München, Herrn Dr. Martin Fischer für die Mitbetreuung, die Einbringung seiner langjährigen Evaluationserfahrung und die Unterstützung beim Piloteinsatz.

Bei Herrn Prof. Dr. Dr. h.c. Reinhard Putz und Herrn Prof. Dr. Dr. Walter Neupert möchte ich mich für die Abnahme der Promotionsvorprüfung und ihr Interesse an meiner Arbeit bedanken. Zusätzlich möchte ich mich bei Herrn Prof. Putz für seine Anregung zur Durchführung der Studie und zur Funktion der Vertreter in der Systemumsetzung bedanken.

Weiterhin bedanke ich mich bei Herrn Kirsten Ingmar Heiss für seine initialen Hinweise zum Thema meiner Doktorarbeit und bei Frau Veronika Kopp, M.A. sowie Herrn Dipl.-Ing. Matthias Holzer für ihre inhaltlichen Anregungen und die Unterstützung bei der Durchführung des Piloteinsatzes.

Schließlich bedanke ich mich bei den Vertretern der Fakultäten, die an der Studie teilgenommen haben für ihre Zeit und das Engagement bei der Beantwortung der Fragen.



# Inhaltsübersicht

|  |     |
|--|-----|
| <b>Abbildungsverzeichnis</b>   | vii |
| <b>Tabellenverzeichnis</b>   | ix  |
| <b>Abkürzungsverzeichnis</b>   | xi  |
| <b>1 Einleitung</b>  | 1   |
| 1.1 Bedeutung der Lehrevaluation . . . . .                                     | 1   |
| 1.2 Begriffliche Abgrenzung . . . . .  | 2   |
| 1.3 Studentische Veranstaltungskritik . . . . .                                | 4   |
| 1.4 Forschungslücke . . . . .  | 5   |
| 1.5 Zielsetzung und Fragestellung . . . . .                                    | 6   |
| <b>2 Methoden</b>  | 9   |
| 2.1 Methodische Grundlagen . . . . .   | 9   |
| 2.2 Besondere Aspekte des medizinischen Curriculums . . . . .                  | 24  |
| 2.3 Evaluationspraxis an den medizinischen Fakultäten in Deutschland . . . . . | 27  |
| <b>3 Ergebnisse</b>  | 43  |
| 3.1 Technische Anforderungen . . . . .   | 43  |
| 3.2 Implementation . . . . .   | 58  |
| 3.3 Piloteinsatz . . . . .   | 80  |
| <b>4 Diskussion</b>  | 85  |
| 4.1 Auswirkungen . . . . .   | 85  |
| 4.2 Schwachstellen und Weiterentwicklungspotenziale . . . . .                  | 90  |
| 4.3 Interpretation . . . . .   | 92  |
| <b>5 Zusammenfassung</b>   | 95  |
| <b>A Fragen der Studie</b>   | 97  |
| <b>B Vollständiger Evaluationsbericht des Piloteinsatzes</b>                   | 99  |
| <b>C Technische Beschreibung der E-Learning-Schnittstelle</b>                  | 113 |
| <b>Literaturverzeichnis</b>  | 115 |



# Inhaltsverzeichnis

|  |     |
|--|-----|
| <b>Abbildungsverzeichnis</b>   | vii |
| <b>Tabellenverzeichnis</b>   | ix  |
| <b>Abkürzungsverzeichnis</b>   | xi  |
| <b>1 Einleitung</b>  | 1   |
| 1.1 Bedeutung der Lehrevaluation . . . . .                                     | 1   |
| 1.2 Begriffliche Abgrenzung . . . . .  | 2   |
| 1.3 Studentische Veranstaltungskritik . . . . .                                | 4   |
| 1.4 Forschungslücke . . . . .  | 5   |
| 1.5 Zielsetzung und Fragestellung . . . . .                                    | 6   |
| <b>2 Methoden</b>  | 9   |
| 2.1 Methodische Grundlagen . . . . .   | 9   |
| 2.1.1 Rechtsnormen . . . . .   | 9   |
| 2.1.2 Interessensgruppen . . . . .   | 10  |
| 2.1.3 Planung . . . . .  | 13  |
| 2.1.4 Untersuchung . . . . .   | 18  |
| 2.1.5 Auswertung . . . . .   | 20  |
| 2.1.6 Effizienz . . . . .  | 24  |
| 2.2 Besondere Aspekte des medizinischen Curriculums . . . . .                  | 24  |
| 2.2.1 Neue Lehrformen . . . . .  | 24  |
| 2.2.2 Modellstudiengänge . . . . .   | 25  |
| 2.2.3 Fragmentierte Veranstaltungen . . . . .                                  | 25  |
| 2.2.4 Komplexe Organisationsstruktur . . . . .                                 | 26  |
| 2.2.5 Homogene Prüfungsstruktur . . . . .                                      | 27  |
| 2.3 Evaluationspraxis an den medizinischen Fakultäten in Deutschland . . . . . | 27  |
| 2.3.1 Erhebungsform . . . . .  | 28  |
| 2.3.2 Externe Evaluation . . . . .   | 29  |
| 2.3.3 Leistungsindikatoren . . . . .   | 30  |
| 2.3.4 Berücksichtigung von Störvariablen . . . . .                             | 32  |
| 2.3.5 E-Learning . . . . .   | 33  |
| 2.3.6 Zufriedenheit . . . . .  | 33  |
| 2.3.7 Rücklaufquote . . . . .  | 36  |
| 2.3.8 Notwendige Verbesserungen . . . . .                                      | 37  |
| 2.3.9 Anforderungen an ein ideales Online-Evaluationssystem . . . . .          | 39  |

|   |     |
|---|-----|
| <b>3 Ergebnisse</b>   | 43  |
| 3.1 Technische Anforderungen . . . . .                        | 43  |
| 3.1.1 Rahmenbedingungen . . . . .                             | 43  |
| 3.1.2 Stammdatenbezug . . . . .                               | 47  |
| 3.1.3 Fragen . . . . .  | 48  |
| 3.1.4 Untersuchungseinheiten . . . . .                        | 49  |
| 3.1.5 Erhebung . . . . .                                      | 51  |
| 3.1.6 Auswertung . . . . .                                    | 54  |
| 3.1.7 Integration . . . . .                                   | 57  |
| 3.2 Implementation . . . . .                                  | 58  |
| 3.2.1 Fragebogen . . . . .                                    | 58  |
| 3.2.2 Fragen . . . . .  | 60  |
| 3.2.3 Berichte . . . . .                                      | 65  |
| 3.2.4 Rankings . . . . .                                      | 69  |
| 3.2.5 Unterstützende Funktionen . . . . .                     | 75  |
| 3.2.6 E-Learning . . . . .                                    | 78  |
| 3.3 Piloteinsatz . . . . .                                    | 80  |
| 3.3.1 Einrichtung . . . . .                                   | 81  |
| 3.3.2 Durchführung . . . . .                                  | 82  |
| 3.3.3 Bewertung . . . . .                                     | 83  |
| <b>4 Diskussion</b>   | 85  |
| 4.1 Auswirkungen . . . . .                                    | 85  |
| 4.1.1 Dozentenbezogene Evaluation . . . . .                   | 85  |
| 4.1.2 Integration der Dozenten . . . . .                      | 87  |
| 4.1.3 Integration der Studierenden . . . . .                  | 88  |
| 4.1.4 Effizienz . . . . .                                     | 89  |
| 4.1.5 Methodische Konformität . . . . .                       | 90  |
| 4.2 Schwachstellen und Weiterentwicklungspotenziale . . . . . | 90  |
| 4.2.1 Störvariablen . . . . .                                 | 91  |
| 4.2.2 Statistische Instrumente . . . . .                      | 91  |
| 4.2.3 Rankings . . . . .                                      | 92  |
| 4.3 Interpretation . . . . .                                  | 92  |
| <b>5 Zusammenfassung</b>                                      | 95  |
| <b>A Fragen der Studie</b>                                    | 97  |
| <b>B Vollständiger Evaluationsbericht des Piloteinsatzes</b>  | 99  |
| <b>C Technische Beschreibung der E-Learning-Schnittstelle</b> | 113 |
| <b>Literaturverzeichnis</b>                                   | 115 |







# Abbildungsverzeichnis

|      |  |    |
|------|--|----|
| 1.1  | Aufbau der Arbeit . . . . .  | 8  |
| 2.1  | Entwicklung der Erhebungsformen . . . . .                                | 29 |
| 2.2  | Externe Evaluation . . . . .   | 29 |
| 2.3  | Leistungsorientierte Mittelvergabe . . . . .                             | 31 |
| 2.4  | Lehrpreise . . . . .   | 31 |
| 2.5  | Dozenten-Ranking . . . . .   | 32 |
| 2.6  | Berücksichtigung von Störvariablen . . . . .                             | 32 |
| 2.7  | Angebot von E-Learning-Kursen . . . . .                                  | 33 |
| 2.8  | Gründe der Zufriedenheit (technische Aspekte) . . . . .                  | 34 |
| 2.9  | Gründe der Zufriedenheit (allgemeine Aspekte) . . . . .                  | 36 |
| 2.10 | Konsequenzen aus der Teilnahme bzw. Nicht-Teilnahme an der Evaluation .  | 36 |
| 2.11 | Notwendige Verbesserungen . . . . .                                      | 37 |
| 2.12 | Anforderungen an ein ideales Online-Evaluationssystem . . . . .          | 39 |
| 3.1  | Bezug der Erhebungen auf unterschiedliche Einheiten . . . . .            | 59 |
| 3.2  | Klassendiagramm der dozentenbezogenen Evaluation . . . . .               | 59 |
| 3.3  | Dozentenbezogener Fragebogen für einen Termin . . . . .                  | 60 |
| 3.4  | Zugang zu einem neuen Fragebogen per E-Mail . . . . .                    | 61 |
| 3.5  | Klassendiagramm des Bezugs der Erhebungen auf unterschiedliche Einheiten | 62 |
| 3.6  | Dynamischer Bezug der Fragebögen auf die Fragen . . . . .                | 63 |
| 3.7  | Klassendiagramm des dynamischen Fragebogens . . . . .                    | 63 |
| 3.8  | Anlegen einer neuen Frage . . . . .                                      | 64 |
| 3.9  | Zustandsautomat einer Frage im Review-Verfahren . . . . .                | 65 |
| 3.10 | Auszug eines automatisch generierten Berichts . . . . .                  | 68 |
| 3.11 | Klassendiagramm des Berichts . . . . .                                   | 69 |
| 3.12 | Automatisch erstelltes Ranking für den Evaluationsbeauftragten . . . . . | 72 |
| 3.13 | Individuelles Ranking für einen Dozenten . . . . .                       | 73 |
| 3.14 | Individuelles Ranking für einen Dozenten mit Stärken-/Schwächenanalyse . | 74 |
| 3.15 | Klassendiagramm des Rankings . . . . .                                   | 75 |
| 3.16 | Darstellung eines Termins im Lehrzettel . . . . .                        | 76 |
| 3.17 | Zustandsautomat eines Termins . . . . .                                  | 77 |
| 3.18 | Einrichtung einer Vertretung . . . . .                                   | 78 |
| 3.19 | Nutzung einer Vertretung . . . . .                                       | 79 |
| 3.20 | Klassendiagramm der Vertretung . . . . .                                 | 79 |
| 3.21 | Klassendiagramm der Evaluation von E-Learning-Kursen . . . . .           | 80 |



## Tabellenverzeichnis

|     |   |    |
|-----|---|----|
| 2.1 | Störvariablen . . . . .                                 | 17 |
| 3.1 | Instrumente der univariaten Datenbeschreibung . . . . . | 55 |



## Abkürzungsverzeichnis

|       |   |
|-------|---|
| ÄAppO | Approbationsordnung für Ärzte   |
| API   | Application Programming Interface   |
| BDSG  | Bundesdatenschutzgesetz   |
| BeMA  | Bericht der Sachverständigenkommission zur Bewertung der Medizinischen Ausbildung |
| BGBI  | Bundesgesetzblatt   |
| CHE   | Centrum für Hochschulentwicklung  |
| ebd.  | ebenda  |
| ERM   | Entity Relationship Model   |
| GB    | Gigabyte  |
| HILVE | Heidelberger Inventar zur Lehrveranstaltungs-Evaluation                           |
| HRG   | Hochschulrahmengesetz   |
| IBE   | Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie   |
| IT    | Informationstechnologie   |
| KVP   | Kontinuierlicher Verbesserungsprozess   |
| LDAP  | Lightweight Directory Access Protocol   |
| LMU   | Ludwig-Maximilians-Universität München  |
| LOM   | Leistungsorientierte Mittelvergabe  |
| MB    | Megabyte  |
| MHz   | Megahertz   |
| NPM   | New Public Management   |
| NSM   | Neues Steuerungsmodell  |

|      |   |
|------|---|
| PDF  | Portable Document Format                                  |
| POL  | Problemorientiertes Lernen                                |
| RFID | Radio Frequency Identification                            |
| TED  | Telefon-Dialog  |
| UML  | Unified Modeling Language                                 |
| URL  | Uniform Resource Locator                                  |
| vhb  | Virtuelle Hochschule Bayern                               |
| ZEvA | Zentrale Evaluations- und Akkreditierungsagentur Hannover |

# 1 Einleitung

## 1.1 Bedeutung der Lehrevaluation

Der Ursprung der Lehrevaluation an deutschen Hochschulen liegt in den studentischen Vorlesungsrezensionen Mitte bis Ende der 1960er Jahre. Die Ziele lagen in der Demokratisierung und Modernisierung der Hochschulen durch eine Reform der Organisationsstrukturen sowie in der Stärkung der Mitbestimmungsrechte des akademischen Mittelbaus und der Studierenden. Mit dem Abflauen der Studentenbewegung und wegen methodischer Schwächen der Pionierarbeiten von Hochschuldidaktikern<sup>1</sup> zum Thema der Lehrevaluation, ging auch das Interesse an der Lehrevaluation in den 1970er Jahren wieder zurück (vgl. Spiel & Gössler 2001, S. 11). Ende der 1980er Jahre thematisierten Medien die Lehrevaluation wieder stärker und brachten das Thema damit zurück in das Rampenlicht der Öffentlichkeit. 1989 veröffentlichte ‚Der Spiegel‘ erstmals ein Ranking deutscher Universitäten aus der Beschreibung der Studienbedingungen durch die Studierenden. Wenngleich die Methodik dieses Rankings kontrovers diskutiert wurde (vgl. Kriz 1994 sowie Hornbostel & Daniel 1994), so weckte es doch ein breites Interesse für die Missstände an den Universitäten. Während die Ursprünge der universitären Lehrevaluation in Deutschland also primär von Modernisierungs- und Mitbestimmungsgedanken geprägt waren, so sind die Faktoren, welche die Lehrevaluation heute beeinflussen diversifizierter.

Unterschiedlichste Interessensgruppen haben Erwartungen und Forderungen an die Qualität der Lehre, die im Rahmen der modernen Lehrevaluation gemessen und kontinuierlich verbessert werden soll.

Studierende sind daran interessiert, durch die Lehre eine Ausbildung zu erhalten, die ihnen möglichst gute Voraussetzungen für die berufliche Laufbahn schafft. Gerade bei der langen Ausbildung in der Medizin, stellt eine annehmbare Gestaltung des Wegs dorthin eine wichtige Nebenbedingung dar.

In einem rohstoffarmen Land wie Deutschland, liegt eine gute Ausbildung der Studenten auch im Interesse der breiten Öffentlichkeit, da ihr Beitrag zum Volkseinkommen allen Bürgern zugute kommt. Insbesondere der steuerzahlende Teil der Öffentlichkeit bzw. private Geldgeber als Financiers der Hochschulen haben die berechnete Erwartung und einen damit zusammenhängenden Auskunftsanspruch (vgl. Künzel 2006 sowie el Hage 1996, S. 22), dass die Investition ihrer Steuern in möglichst kurzer Zeit eine möglichst hohe Rendite in Form von steuerzahlendem Humankapital erbringt. Die Lehre stellt dabei den entsprechenden Transformationsprozess dar, der möglichst effektiv und effizient gestaltet werden muss, um

---

<sup>1</sup> Zur besseren Lesbarkeit wird in dieser Arbeit einheitlich bei gemischtgeschlechtlichen Personengruppen auf die Nennung der weiblichen Form verzichtet. Selbstverständlich sind damit jedoch auch Hochschuldidaktikerinnen, Studentinnen, Dozentinnen usw. gemeint.

diesem Anspruch gerecht zu werden. Mit zunehmender Finanzmittelknappheit wächst der Druck auf die Hochschulen, die Umsetzung dieses Ziels quantitativ zu belegen. Die Politik geht dazu über, die Messergebnisse der Lehrevaluation als Kennzahlen für ein neues Verständnis der Ressourcenallokation einzusetzen, das im Rahmen des Neues Steuerungsmodells<sup>2</sup> (NSM) einen Übergang von der Kameralistik hin zu output-orientierten Prinzipien darstellt (vgl. Küpper & Tropp 2004, Wottawa 2001, Nickel & Zechlin 2000).

Um die finanzielle Ausstattung zu sichern und möglicherweise sogar auszubauen, haben die Universitäten selbst ein sehr großes Interesse daran, eine qualitativ hochwertige Lehre anzubieten, die sich durch valide Evaluationsmethoden nachprüfbar quantifizieren lässt. Neben diesen pekuniären Interessen müssen die Universitäten aber auch ihre Reputation in der Scientific Community in einem zunehmend globaleren Bildungsmarkt verteidigen. Im Rahmen der Profilbildung stellt das Angebot hochwertiger Lehre ein wichtiges Kriterium in der Konkurrenz um gute Studierende und Lehrende dar. Mit der Einführung von Studiengebühren in Deutschland wird sich dieser Wettbewerb noch verstärken. So konstatiert die Generalsekretärin der Hochschulrektorenkonferenz Ebel-Gabriel (2006, S. 12): „Eine Hochschule, die ihre Exzellenz in diesen Bereichen nicht darlegen und wirksam dokumentieren kann, wird im Wettbewerb nicht bestehen.“

Die Dozenten, als direkte Erbringer der Leistung „Lehre“ stehen derzeit noch in dem Ruf, die Lehre zugunsten der Forschung zu vernachlässigen. Diesem Ungleichgewicht soll jedoch mit zunehmender Gewichtung der Lehrleistung bei der Besoldung im Rahmen der Leistungsorientierten Mittelvergabe (LOM) sowie bei Habilitationen und Rufen entgegengewirkt werden (Wissenschaftsrat, 1992, S. 59 f.). Dadurch ist ein Anstieg des Interesses an der Lehrevaluation auch bei den Dozenten zu erwarten.

### 1.2 Begriffliche Abgrenzung

Qualitativ hochwertige Lehre zu erbringen, ist von breiter und zentraler Bedeutung. Gleichzeitig wird die Bestimmung dessen, was qualitativ hochwertige Lehre konkret ist, als sehr schwierig beschrieben (vgl. Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) 2001, S. 99). In der Literatur wird eine Vielzahl von Versuchen unternommen, den abstrakten Begriff „Qualität“ im Sinne von „guter Lehre“ zu instrumentalisieren (für die medizinische Ausbildung vgl. Burger 2003 und Bargel 2003; allgemein vgl. Rindermann 2001b, S. 35–41, Kromrey 2001b, S. 29–35, Mellis 2000, S. 386 ff., Wottawa & Thierau 1998, S. 43–45, Kromrey 1994, S. 92 f. und Richter 1994, S. 7–14).

Ebenfalls häufig beschrieben wird der Begriff „Evaluation“ (vgl. Deutsche Gesellschaft für Evaluation 2002, S. 13–16, Rindermann 2001b, S. 10–13, Wottawa & Thierau 1998, S. 13–15). Im Rahmen dieser Arbeit soll unter dem Überbegriff „Evaluation“ die „systematische Analyse und empirische Untersuchung von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zum Zwecke ihrer Bewertung und Modifikation“ (Rossi et al.

---

<sup>2</sup> Das NSM ist in der Literatur auch „auch unter der Bezeichnung ‚New Public Management‘ (NPM) oder ‚Wirkungsorientierte Verwaltungsführung‘ bekannt“ (Nickel & Zechlin 2000, S. 154)



1999 zitiert nach Rindermann 2001b, S. 10) verstanden werden. Unter „Lehrevaluation“ soll die Evaluation der Lehre und unter „Lehrveranstaltungsevaluation“ die Evaluation von Lehrveranstaltungen verstanden werden.

Scriven (1980) führte die Unterscheidung in formative (begleitende) und summative (zusammenfassende) Evaluation ein (vgl. auch Kromrey 2001a, S. 24 f.) und brachte damit zusätzlich zu den Definitionen des Begriffs „Evaluation“ ein Verständnis zum Ausdruck, über das heute weitgehend Einigkeit herrscht: das Prädikat einer guten Lehre darf nicht als Zustand missverstanden werden. Vielmehr muss gute Lehre als Verbesserungsprozess gesehen werden, der in Anlehnung an das Managementkonzept des Kontinuierlichen Verbesserungsprozesses<sup>3</sup> (KVP) aus den Teilschritten des Planens, der Durchführung, der Überprüfung und des erneuten Verbesserns besteht (vgl. Wottawa 2001).

Da die Evaluation hohe methodische Ansprüche stellt, empfiehlt der Wissenschaftsrat (1996, S. 23–32) ein „zweistufiges Verfahren, das interne Selbstevaluation und externe Begutachtung durch fachnahe Sachverständige miteinander verbindet“ (ebd. S. 23). In der ersten Stufe („interne Evaluation“ oder auch „Selbstevaluation“ genannt) unterzieht sich der zu evaluierende Fachbereich selbst einer kritischen Überprüfung von Studium und Lehre und fasst die Ergebnisse in einem Bericht zusammen. In der zweiten Stufe („externe Evaluation“) führen Sachverständige („Peers“) auf der Basis der Selbstevaluation eine externe Begutachtung („Peer Review“<sup>4</sup>) durch und verfassen einen Bericht mit Empfehlungen (vgl. Mittag 2006, Mittag et al. 2003, Bülow-Schramm 2000, S. 174 ff., Hochschulrektorenkonferenz 2004 und aus europäischer Sicht: European Association for Quality Assurance in Higher Education 2005). Bei der externen Evaluation sind die Aktivitäten von Nordverbund und der Zentralen Evaluations- und Akkreditierungsagentur Hannover (ZEvA) sowie speziell in der Medizinischen Lehre der Bericht der Sachverständigenkommission zur Bewertung der Medizinischen Ausbildung (BeMA) (vgl. Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) 2001) hervorzuheben.

Ein zentrales Element der Evaluation ist die vom Wissenschaftsrat (1996, S. 25) geforderte und in § 6 Satz 2 Hochschulrahmengesetz (HRG) kodifizierte Beteiligung der Studierenden an der Bewertung der Qualität der Lehre. In der Praxis kommen die Hochschulen dieser Forderung mit der Durchführung von Umfragen unter den Studenten und der Verwertung der darauf aufbauenden Ergebnisse nach. Berücksichtigt man die Tatsache, dass gut Zweidrittel der medizinischen Fakultäten in Deutschland keine externe Evaluation durchführen (siehe Abschnitt „Externe Evaluation“ auf den Seiten 29–30), so kommt der studentischen Lehrveranstaltungskritik eine noch bedeutendere Rolle zu.

---

<sup>3</sup> Einen Überblick über den KVP liefern beispielsweise Kostka & Kostka 2002.

<sup>4</sup> Gelegentlich wird das gesamte zweistufige Verfahren als „Peer Review“-Verfahren bezeichnet, obwohl dieser Begriff streng genommen nur die Evaluation durch Sachverständige in der zweiten Stufe beschreibt.

### 1.3 Studentische Veranstaltungskritik

Gerade aber diese studentische Lehrveranstaltungskritik steht selbst häufig in der Kritik. Wie von Rindermann (2001b, S. 163–206) ausführlich dargestellt, ist die dahinter stehende Kontroverse mannigfaltig und kann hier nur in Auszügen behandelt werden, um ein grundlegendes Verständnis dafür zu schaffen. Im Zentrum der Kritik steht die Frage, ob die studentische Lehrveranstaltungskritik den Gütekriterien der Objektivität, Reliabilität und Validität genügen und als Maß zur Beurteilung der Lehrqualität herangezogen werden kann.

So führt Kromrey (1994, S. 92 ff.) ganz grundsätzlich aus, dass die Urteile der Studierenden für die individuelle Rückmeldung an einen Dozenten konzipiert wurden und dafür auch gut geeignet seien. Der in der Praxis häufig direkt aus den Urteilen abgeleitete Vergleich zwischen mehreren Veranstaltungen oder Dozenten sei jedoch nicht zulässig. Marsh & Roche (1997) dagegen zeigen, dass über Veranstaltungen gemittelte Beurteilungen von Studierenden messgenau und stabil sind. Rindermann (2003, S. 88 f.) gibt an, dass die Übereinstimmung der Beurteilungen zwischen den Studenten mit einer Interraterreliabilität von  $r = .3$  gleich hoch liegen wie die von Gutachtern im Peer-Review-Verfahren zur Beurteilung wissenschaftlicher Manuskripte. Zudem weist auch ein Vergleich der Urteile von Studierenden mit denen von Fremdbeurteilern (insbesondere Gasthörer, andere Dozenten, Hochschuldidaktiker oder frühere Studierende) eine hohe Übereinstimmung auf (vgl. Rindermann 2003, S. 91). Kromrey (1994, S. 97–107) entgegnet dem jedoch, dass die Einzelurteile so stark voneinander abweichen, dass die Bildung von Mittelwerten ihrer statistischen Aufgabe, einem zufällig gewählten Urteil möglichst nahe zu kommen, nicht gerecht wird.

Ein weiterer Kernpunkt in dieser Diskussion ist die Volatilität der Beurteilungen der Studierenden gegenüber unterschiedlichsten lehrangebotsunabhängigen Einflüssen (siehe hierzu auch Abschnitt „Störvariablen“ auf den Seiten 16–17). So fasst Radler (1997, S. 73) die Ergebnisse von Kromrey (1994, S. 109) beispielsweise auf den Einfluss der Motivation wie folgt zusammen: „Dozenten, die vor interessierten Studenten lehren, werden gelobt. Wer hingegen die weitaus schwierigere Arbeit hat, Pflichtstoff zu lehren, erntet Tadel.“ Während das Vorhandensein solcher Bias-Variablen nicht geleugnet werden kann, lässt sich daraus nicht gleichzeitig der Schluss ziehen, dass studentische Urteile generell nicht verwertbar sind (vgl. Spiel & Engel 2001, S. 76–79). In diesem Zusammenhang sei auf den von Naftulin et al. (1973) vorgestellten „Doctor Fox“-Effekt<sup>5</sup> verwiesen, nachdem weder Studierende noch Fachpublikum vor sogar gravierenden Fehlurteilen gefeit sind.

---

<sup>5</sup> In ihrem ursprünglichen Experiment ließen Naftulin et al. (1973) einen professionellen Schauspieler eine von ihnen konzipierte Vorlesung vor einem Fachpublikum halten, ohne dass dieser ein Fachwissen zu dem Thema hatte. Trotzdem erhielt der Schauspieler annähernd 80 % positive Bewertungen, die ihn u. a. als herausragenden Psychiater beschrieben. Bemerkenswert ist dabei vor allem, dass nicht nur der Schauspieler keine Fachkenntnisse zu dem vorgetragenen Thema hatte, sondern auch der Vortrag selbst inhaltlich keinerlei Substanz hatte und dem vermeintlichen Wissenschaftler nach dem Vortrag auch Fragen gestellt wurden. Dieses Experiment wurde mit ähnlichen Ergebnissen mehrfach vor Fachpublikum und Studenten wiederholt. Die Ergebnisse führten die Autoren auf die professionelle Darbietung zurück.

Kaum angezweifelt werden kann dagegen der Ansatz, dass Studierende ihre Zufriedenheit mit den Lehrveranstaltungen und den Dozenten zum Ausdruck bringen können (vgl. Herzwurm 2000). Dies ist durchaus auch als eine Form der Kundenzufriedenheit zu sehen, die in den USA mit den teilweise erheblichen Studiengebühren, bereits Usus ist und mit der Einführung von Studiengebühren auch in Deutschland zunehmend an Bedeutung gewinnen wird. Der in diesem Zusammenhang entstehende Druck auf die Universitäten, zu den von Meffert (1998, S. 1179) beschriebenen „marktorientierten Problemlösern mit hoher Dienstleistungscompetenz“ zu werden, wird in diesem Zusammenhang weiter steigen und eine verstärkte Orientierung der zu erbringenden Leistungen an den Studenten erfordern. Es darf bei dieser Diskussion auch nicht übersehen werden, dass es den Studenten in der Regel nicht um einen Angriff auf die Dozenten geht, sondern schlicht um das eigene Fortkommen (vgl. Wissenschaftsrat 1992, S. 63). Zu beachten ist dabei jedoch, dass eine Zufriedenheit bei allen Mitgliedern einer Zielgruppe schwer erreichbar ist (vgl. Kromrey 2001b, S. 41) und die Zufriedenheit der Studierenden alleine nicht ausreicht, um darauf Handlungsmaßnahmen aufzubauen, welche die Lehre insgesamt betreffen. Dies soll an folgendem Dilemma beispielhaft kurz verdeutlicht werden. Es wird sich kaum ein Studierender finden lassen, der mit einer schlechten Note zufrieden ist. Eine reine Ausrichtung der Maßnahmen an der Zufriedenheit der Studierenden würde in diesem Beispiel dazu führen, die Noten leistungsunabhängig anheben zu müssen. Dies würde die Benotung an sich jedoch ad absurdum führen und schließlich die Chancen der Studierenden auf dem Arbeitsmarkt zunichte machen, was wiederum ebenfalls die Zufriedenheit der Studierenden (bzw. Absolventen) erheblich beeinträchtigt.

Studentische Urteile sind also *ein* Mosaikstein in einem Gesamtbild, das den Zustand der Lehre beschreibt. Um dieses Gesamtbild möglichst solide zu gestalten, müssen die studentischen Urteile um weitere Informationen wie objektive Fakten<sup>6</sup> zu den Rahmenbedingungen (vgl. Wissenschaftsrat 1996, S. 29 f. sowie Mellis 2000, S. 389 f.) sowie um weitere Urteile von Absolventen, Dozenten und idealerweise auch von externen Gutachtern ergänzt werden. Darauf aufbauend kommt dann der Interpretation dieser Zustandsbeschreibung sowie den daraus abgeleiteten Handlungsmaßnahmen eine besondere Bedeutung zu.

## 1.4 Forschungslücke

Wie im vorangegangenen Abschnitt dargelegt, ist die Informationsquelle der studentischen Urteile zweifelsohne essentiell, gefordert, kodifiziert und wegen dem nicht unerheblichen Aufwand, der mit einem mehrstufigen Evaluationsverfahren verbunden ist (vgl. Reissert & Konnerth 2001, S. 190 f.), häufig auch die gewichtigste – mitunter sogar die einzige (vgl.

<sup>6</sup> Es sei an dieser Stelle aber darauf hingewiesen, dass auch die Verwendung sog. ‚hard facts‘ als Datengrundlage für die Evaluation nicht ohne Interpretationsprobleme und damit nicht automatisch vollständig valide sind. So führt Hornbostel (2001, S. 88 f.) als Beispiel die Qualität des Bibliotheksangebots an, die einen großen Interpretationsspielraum zulässt. So kann es z. B. sein, dass dezentrale Bestände nicht erfasst werden oder bestimmte Bestände nicht für alle Interessensgruppen zugänglich sind. Weiterhin können die Öffnungszeiten oder die Recherchemöglichkeiten die Qualität beeinflussen.

Hochschulrektorenkonferenz 2003, S. 23). Es muss deshalb auf breiter Basis versucht werden, diese Informationsquelle so gut es geht, an die bestehenden Anforderungen anzupassen, um eine umfassende Akzeptanz erreichen zu können, auf der die Verbesserungsmaßnahmen aufbauen können. Diese breite Basis stellt derzeit eine Vielzahl an Publikationen aus dem Bereich der empirischen Sozialforschung und der Statistik dar (vgl. allgemein: Bortz & Döring 2006, Diekmann 2005, el Hage 1996, S. 85–119, Kromrey 1994, Kromrey 2001a, Wottawa & Thierau 1998, S. 131 ff., Wissenschaftsrat 1996, Wolf et al. 2001, S. 97–107, Frank 1997, Kaplan & Norton 1992, Richter 1994, S. 52 ff., Hungerbühler 1997 und speziell zur Konstruktion der Fragebögen: Rindermann 2001b, el Hage 1996, S. 120–139, Stangl 2000, Peus et al. 2005b, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) 2001, S. 100 um nur eine Auswahl der bekannteren Arbeiten zu nennen).

Bisher nicht<sup>7</sup> dargestellt ist jedoch eine Übersetzung dieser Anforderungen in technische Anforderungen an die Systeme der Informationstechnologie (IT), die für die Erhebung und Auswertung der studentischen Urteile zunehmend zum Einsatz kommen und damit zunehmend an Bedeutung gewinnen. Wie sich zeigen wird, kann die fehlende Erfüllung solcher technischer Aspekte gravierende Auswirkungen auf die Durchführung (z. B. auf den Aufwand) und die Ergebnisse (z. B. auf die eindeutige Bewertbarkeit der Dozenten) haben. Wie sich aber auch zeigen wird, kann der Einsatz von Computersystemen bei der Implementation bestimmter Anforderungen dazu beitragen, die Umsetzung der Evaluation effektiver und effizienter zu gestalten und damit mittelbar zu einer Verbesserung der Lehre beitragen. Die vorliegende Arbeit postuliert deshalb die Einbeziehung technischer Disziplinen wie die Informatik in den interdisziplinären Kanon in dem die Evaluierung in der Literatur (vgl. Wottawa & Thierau 1998, S. 66 f., Koutala 1997, S. 54) gesehen wird.

### 1.5 Zielsetzung und Fragestellung

Reissert & Konnerth (2001, S. 178) stellen fest, dass kaum noch grundsätzlich über das Ob der Evaluation gestritten wird, sondern in erster Linie nur noch über das Wie und die daraus zu ziehenden Konsequenzen. Genau dieses Wie wird zunehmend durch Aspekte der IT beeinflusst. So bietet die IT viele Möglichkeiten, Probleme abzufangen. Gleichzeitig birgt eine nicht anforderungsgerechte Umsetzung jedoch auch die Gefahr, neue Probleme zu generieren.

Das Ziel der Arbeit besteht deshalb darin, die Anforderungen an ein Evaluationssystem aus unterschiedlichen Perspektiven herzuleiten und diese anhand eines konkreten Systems in der Praxis zu veranschaulichen. Dazu stellt die Arbeit relevante Anforderungen aus der Literatur und der Praxis dar, um daraus Verfahren für die Umsetzung ableiten zu können. Damit soll die Arbeit eine Grundlage dafür bieten, bestehende Evaluationssysteme zu verbessern und neue Evaluationssysteme anforderungsgerechter zu entwickeln.

---

<sup>7</sup> Abgesehen von Arbeiten wie beispielsweise der von Postert (2000), deren Fokus ein anderer ist und in welchen die hier beschriebenen Zusammenhänge qualitativ und quantitativ nur unzureichend dargestellt werden.

Der Fokus liegt dabei auf dem von Rindermann (2001b, S. 57 f.) als zentral beschriebenen Teil der Umfragen und deren Auswertungen unter besonderer Berücksichtigung des medizinischen Curriculums, das im Gegensatz zu anderen Curricula besondere Ansprüche stellt. So konkretisiert soll im weiteren Verlauf der Arbeit damit auch der Begriff des „Lehrevaluationssystems“ (bzw. kurz „Evaluationssystem“) verstanden werden, falls er nicht anders spezifiziert ist. Außerhalb des Spektrums dieser Arbeit liegen die Erhebung unterstützender Informationen für den Selbstbericht sowie weitere Ebenen des mehrstufigen Verfahrens.

Im Speziellen soll die Arbeit folgende Fragen behandeln:

1. Welchen methodischen Anforderungen muss ein Evaluationssystem genügen?
2. Welche Aspekte sind aus der Sicht der Evaluationspraxis relevant?
3. Auf welche technischen Anforderungen lassen sich diese Erkenntnisse verdichten?
4. Wie lassen sich die abstrakten Anforderungen konkret umsetzen?
5. Welche Bedeutung hat eine solche Umsetzung für die Evaluationspraxis?

Durch die Beantwortung dieser Fragen soll die Arbeit die im vorherigen Abschnitt aufgezeigte Forschungslücke füllen und erste Anknüpfungspunkte zwischen den Disziplinen der empirischen Sozialforschung sowie der Statistik auf der einen Seite und der für die Evaluation zunehmend an Bedeutung gewinnenden Disziplin der Informatik auf der anderen Seite herstellen, um damit insbesondere den speziellen Aspekten des medizinischen Curriculums gerecht zu werden.

Zur Beantwortung dieser Fragen setzt die Arbeit ihrem Titel entsprechend die folgenden drei Meilensteine um.

1. Herleitung der Anforderungen an ein Evaluationssystem aus Methoden und Praxis
2. Umsetzung der Anforderungen zu einem konkreten System
3. Darstellung des Einsatzes dieses Systems in der Praxis

Die einzelnen Kapitel behandeln dabei folgende Themen um diese Meilensteine zu erreichen.

Das zweite Kapitel „Methoden“ zeigt die methodischen Grundlagen der Lehrevaluation, erweitert diese um die besonderen Aspekte des medizinischen Curriculums und stellt das relevante Umfeld der Evaluationspraxis anhand einer im Rahmen dieser Arbeit durchgeführten empirischen Studie unter den medizinischen Fakultäten in Deutschland dar.

Das dritte Kapitel „Ergebnisse“ verdichtet die im zweiten Kapitel dargestellten Rahmenbedingungen zu technischen Anforderungen an ein Evaluationssystem, zeigt deren Umsetzung anhand eines im Rahmen dieser Arbeit entwickelten, voll funktionsfähigen Software-Systems und beschreibt dessen Piloteinsatz in der Praxis anhand ausgewählter Veranstaltungen im

Sommersemester 2006 an der Medizinischen Fakultät der Ludwig-Maximilians-Universität München (LMU).

Das vierte Kapitel „Diskussion“ überträgt die Ergebnisse auf andere Fakultäten, beschreibt die vorhandenen Schwachstellen und fügt die Erkenntnisse in den Gesamtkontext der Lehrevaluation ein.

Abbildung 1.1 visualisiert diesen Aufbau der Arbeit.

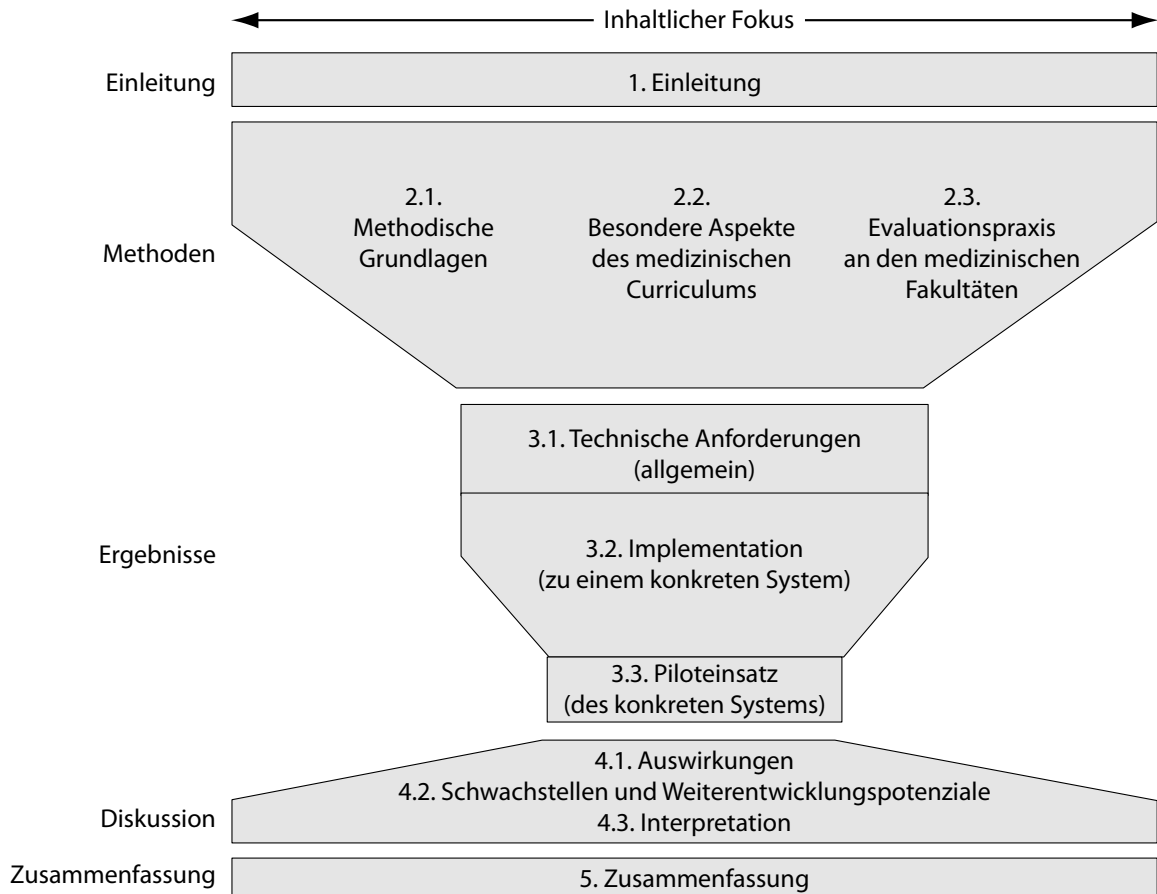


Abbildung 1.1: Aufbau der Arbeit

## 2 Methoden

Das vorliegende Kapitel zeigt die methodischen Grundlagen zur Evaluation aus der Literatur der empirischen Sozialforschung sowie der Statistik und erweitert diese um die speziellen Aspekte des medizinischen Curriculums. In dem darauf folgenden empirischen Teil wird die Evaluationspraxis an den medizinischen Fakultäten in Deutschland dargestellt, die auf einer im Rahmen der vorliegenden Arbeit durchgeführten Umfrage unter diesen Fakultäten analysiert wurde. Sowohl die theoretischen Grundlagen zur Evaluation wie auch die empirische Analyse der Praxis konzentrieren sich dabei dem Fokus dieser Arbeit entsprechend auf die Aspekte, aus denen sich Anforderungen an ein Evaluationssystem ableiten lassen.

### 2.1 Methodische Grundlagen

Bei der Darstellung der methodischen Grundlagen soll eine möglichst umfassende Übersicht über die für ein System relevanten Anforderungen vorgenommen werden, die sich aus der Literatur ableiten lassen. Die Darstellung erhebt dabei keinen Anspruch auf Vollständigkeit und soll umgekehrt auch nicht den Eindruck erwecken, dass ein System nur dann einsetzbar ist, wenn es alle Anforderungen erfüllt. Vielmehr sollen wichtige Aspekte herausgearbeitet werden, die das Potential haben, mittelbar zur Verbesserung der Lehrqualität beizutragen.

Berücksichtigt werden in diesem Kapitel auch europäische Empfehlungen der Bologna-Deklaration. Deren Anwendung in vollem Umfang auf das medizinische Curriculum ist zwar umstritten (vgl. Gerke et al. 2006). Da die Ausführungen zur Evaluation und Qualitätssicherung nicht Kern dieser Kritik sind (vgl. Troschke 2006, S. 3) und sich in der Literatur eine grundsätzliche Offenheit zur Anwendung abzeichnet (vgl. Hochschulrektorenkonferenz 2005), geht auch die vorliegende Arbeit auf diese Empfehlungen ein – wenn auch in einem der andauernden Debatte angemessen geringem Rahmen.

#### 2.1.1 Rechtsnormen

Die grundsätzliche Anforderung, Lehrveranstaltungen zu evaluieren ist in mehreren Normen des Bundesrechts<sup>8</sup> festgeschrieben. So regelt § 6 HRG<sup>9</sup>: „Die Arbeit der Hochschulen in Forschung und Lehre, bei der Förderung des wissenschaftlichen Nachwuchses sowie der Erfüllung des Gleichstellungsauftrags soll regelmäßig bewertet werden. Die Studierenden sind

---

<sup>8</sup> Für eine ausführlichere Darstellung der relevanten Rechtsnormen bei der Berücksichtigung der länderspezifischen Regelungen sei auf Hochschulrektorenkonferenz (2001) und el Hage (1996, S. 27–35) verwiesen.

<sup>9</sup> Die vorliegende Arbeit bezieht sich auf die Fassung der Bekanntmachung vom 19. Januar 1999 (BGBl. I S. 18), zuletzt geändert durch Artikel 1 des Gesetzes vom 27. Dezember 2004 (BGBl. I S. 3835) unter Berücksichtigung der Entscheidung des Bundesverfassungsgerichts vom 26. Januar 2005 (2 BvF 1/03).

bei der Bewertung der Qualität der Lehre zu beteiligen. Die Ergebnisse der Bewertungen sollen veröffentlicht werden.“ § 2 Abs. 9 ÄAppO<sup>10</sup> legt fest: „Lehrveranstaltungen sind regelmäßig auf ihren Erfolg zu evaluieren. Die Ergebnisse sind bekannt zu geben.“ § 41 Abs. 2 Nr. 4 ÄAppO setzt für die Zulassung eines Modellstudiengangs voraus, dass „[...] eine sachgerechte begleitende und abschließende Evaluation des Modellstudiengangs gewährleistet ist.“

Zusammengefasst legen diese Normen eine grundsätzliche Verpflichtung zur Evaluation fest. Sie fordern dabei, dass die Studierenden an der Evaluation zu beteiligen sind und die Ergebnisse der Evaluation veröffentlicht werden sollen. Konkretere Vorgaben für die Umsetzung und Ausgestaltung der Evaluation geben diese Normen jedoch nicht. Es ist konkludent davon auszugehen, dass die Evaluation im Rahmen der sonstigen Rechtsnormen und der Möglichkeiten der Evaluationsträger dem Stand der Wissenschaft entsprechen soll.

Eine besondere Beachtung gilt dem Datenschutz, der grundlegend im Bundesdatenschutzgesetz (BDSG) geregelt ist. Nach § 1 Abs. 1 BDSG<sup>11</sup> ist der „Zweck dieses Gesetzes [...], den Einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.“ Aus § 3a BDSG lässt sich insbesondere die Vorgabe ableiten, dass die Datenerhebung so weit wie möglich anonym erfolgen muss.<sup>12</sup> Bereits aus diesen Normen ergibt sich damit die Anforderung, dass ein Evaluationssystem Umfragen grundsätzlich anonymisiert durchführen muss. Dem Verbotsprinzip mit Erlaubnisvorbehalt des BDSG entsprechend, können persönliche Daten gem. § 4 Abs. 1 BDSG jedoch dann erhoben, verarbeitet und genutzt werden, wenn der Betroffene eingewilligt hat. Sollen – wie unter Abschnitt „Evaluationszeitpunkte“ auf Seite 18 beschrieben – Längsschnittanalysen durchgeführt werden, bei denen ein Bezug zwischen den Antworten und dem Beantwortenden hergestellt werden muss, ist dies nur dann zulässig, wenn der Betroffene dazu ausdrücklich zustimmt. Ein Evaluationssystem muss dieser Anforderung und der in § 4 Abs. 3 BDSG näher beschriebenen Ausgestaltung über die Informationen der Betroffenen entsprechen.

### 2.1.2 Interessensgruppen

#### Studierende, Absolventen und Studienabbrecher

Mit den Studierenden wurde bereits im vorigen Abschnitt eine zentrale Interessensgruppe<sup>13</sup> eingeführt. Die Aussagekraft der studentischen Urteile ist – wie in der Einleitung unter Abschnitt „Studentische Veranstaltungskritik“ auf den Seiten 4–5 dargestellt – umstritten.

---

<sup>10</sup> Die vorliegende Arbeit bezieht sich auf die Fassung vom 27. Juni 2002 (BGBl. I S. 2405), zuletzt geändert durch Artikel 71 des Gesetzes vom 21. Juni 2005 (BGBl. I S. 1818).

<sup>11</sup> Die vorliegende Arbeit bezieht sich auf die Fassung vom 20. Dezember 1990, zuletzt geändert durch Gesetz zur Änderung des Bundesdatenschutzgesetzes und anderer Gesetze vom 18. Mai 2001.

<sup>12</sup> Sofern dies gem. § 1 Abs. 2 Nr. 2 BDSG nicht durch spezielles Landesrecht anders geregelt ist. Da es im Rahmen dieser Arbeit um einen Überblick über die rechtlichen Grundlagen geht, wird ausschließlich Bezug auf das Bundesrecht genommen, ohne spezielle Regelungen auf der Ebene der Länder zu berücksichtigen.

<sup>13</sup> In der Literatur wird synonym häufig auch der Begriff des „Stakeholders“ verwendet, der aus dem betriebswirtschaftlichen Konzept des Stakeholder-Relationship-Managements (SRM) übernommen ist.



Da die Einbeziehung der Studierenden interessanterweise<sup>14</sup> jedoch explizit kodifiziert ist, kann davon ausgegangen werden, dass der Gesetzgeber dieser Diskussion klar entgegen möchte. Dieser Ansatz wird durch die europäischen Rahmenvorgaben weiter unterstrichen, in welchen die Rolle der Studenten bei der Qualitätssicherung herausgehoben wird (vgl. European Ministers of Education 2003, European Ministers of Education 2005, European Association for Quality Assurance in Higher Education 2005). Die Frage, ob Studenten an der Evaluation zu beteiligen sind, soll aus diesem Grund nicht weiter strapaziert werden. Für den Rahmen dieser Arbeit wird davon ausgegangen, dass die Studierenden eine wichtige Interessensgruppe sind, die bei der Evaluation mit einzubeziehen sind.<sup>15</sup>

Die Einbeziehung der Studenten in die Evaluation ist jedoch nicht auf die aktive Zeit des Studiums beschränkt. So besteht ein weiteres zentrales Ziel der Evaluation darin, nicht nur die Prozess-, sondern auch die Ergebnisqualität zu erfassen (vgl. Wissenschaftsrat 1996, S. 30, European Ministers of Education 1999, Kromrey 2001a, S. 45, Schmidt 2005, S. 44), zu der die unter dem Begriff der „employability“ verstandenen Berufschancen gehören. Um diese erfassen zu können, wird in der Literatur ebd. gefordert, auch die Absolventen an der Evaluation zu beteiligen, um diese Informationen aus erster Hand aufnehmen zu können. Neben den Berufschancen der Studenten, die das Studium erfolgreich abgeschlossen haben, empfiehlt Schmidt (2005, S. 43) darüber hinaus, auch die Studienabbrecher an der Evaluation zu beteiligen, um sie nach den Gründen des Scheiterns befragen zu können.

Neben den Studierenden existieren jedoch noch weitere Interessensgruppen, deren Beteiligung an der Evaluation inhärent ist bzw. die ein nicht kodifiziertes Anrecht auf eine Beteiligung haben. Wie in der Einleitung auf Seite 5 dargestellt, ist es wichtig, die Interessensgruppen als ‚Kunden‘ des Evaluationssystems umfassend zu berücksichtigen, um die Akzeptanz der Evaluation auf breiter Basis zu sichern und damit die Grundlage zu schaffen, aus den Evaluationsergebnissen auch tatsächlich Konsequenzen ableiten zu können. Diese weiteren Interessensgruppen sollen nachfolgend dargestellt werden.

## **Dozenten**

Eine für die Akzeptanz des Systems essentielle Interessensgruppe sind die Dozenten. Nachdem im vorangegangenen Abschnitt dargestellt wurde, dass aufgrund der gesetzlichen Vorgaben

---

<sup>14</sup> Interessant ist insbesondere, dass der Gesetzgeber die Teilnahme der Studierenden ausdrücklich fordert, dagegen mehrstufige Evaluationsverfahren mit keinem Wort erwähnt. Trotzdem beteiligen Evaluationsagenturen Studierende an mehrstufigen Evaluationsverfahren nur uneinheitlich (vgl. Hopbach 2005, S. 21) obwohl auch die European Association for Quality Assurance in Higher Education (2005) die Beteiligung der Studenten in dieser Evaluationsform explizit fordert.

<sup>15</sup> Zur Erfüllung der weitergehenden Forderung der European Association for Quality Assurance in Higher Education (2005, S. 18), insbesondere die Zufriedenheit der Studenten zu erfassen, sei auf die pädagogisch-psychologischen Aspekte der Formulierung entsprechender Fragen verwiesen, die nicht Bestandteil dieser Arbeit sind. Die Zufriedenheit ist in den USA bereits eine zentrale Messgröße, weil sich die amerikanischen Hochschulen über die von den Studenten gezahlten Studiengebühren finanzieren müssen. Nicht zuletzt mit dem Einzug der Studiengebühren hierzulande ist abzusehen, dass dieser Aspekt auch in Deutschland eine noch größere Rolle spielen wird (siehe Abschnitt „Studentische Veranstaltungskritik“ auf Seite 5).

die Studierenden an der Evaluation zu beteiligen sind und dies – wie sich im Lauf der Arbeit zeigen wird – auch in der Evaluationspraxis faktisch so umgesetzt wird, muss alles daran gesetzt werden, die Kontroverse um die Beteiligung der Studierenden zu entschärfen. Da die Evaluation nicht Selbstzweck ist, sondern Grundlage für darauf aufbauende Konsequenzen wie die leistungsorientierte Mittelvergabe (vgl. Wissenschaftsrat 1996, S. 13), die insbesondere die Dozenten betreffen, muss die Evaluation von den Dozenten mitgetragen werden. Andernfalls ließen sich die geforderten Konsequenzen zwar kurzfristig aufkotroyieren, langfristig aber wohl nur schwer umsetzen.

Ein Instrument zur Bindung der Dozenten an die Evaluation ist die kontinuierliche Rückmeldung von Evaluationsergebnissen (vgl. Balk 2000, S. 34 f.). Um den für die Durchsicht und Interpretation der Ergebnisse benötigten kognitiven Aufwand zu minimieren, sollen die Ergebnisse möglichst weit reichend aufbereitet und individualisiert sein. Den Dozenten soll es dadurch möglich sein, mit effizientem Einsatz knapper Ressourcen wie insbesondere der Zeit, einen Bezug zur Evaluation aufbauen zu können, der durch die Individualisierung über die allgemeinen Gesamtberichte hinausgeht. Wolf et al. (2001, S. 104) bieten den Dozenten in dem von ihnen dargestellten Instrument darüber hinaus die Möglichkeit, eigene Fragen in die Evaluation einbringen zu können. Mit dieser Funktionalität kann bereits die Erhebung individualisiert und der Nutzen der Auswertung zusätzlich erhöht werden.

### **Evaluationsverantwortliche**

Kromrey (2001a, S. 36) stellt die Bedeutung eindeutiger Kompetenzzuweisungen bei der Evaluation als wichtige Grundlage für eine Nutzen bringende Evaluation dar. Dementsprechend finden sich in den Fakultäten zunehmend eigens benannte Evaluationsverantwortliche, die eine organisatorisch festgelegte Verantwortung für die Durchführung der Evaluation innehaben. Im Fokus dieser Interessensgruppe stehen neben der Qualität und Effizienz der Evaluation, die in den folgenden Abschnitten aufgrund des größeren Umfangs ausführlich dargestellt werden, die Möglichkeit, umfassende Auswertung zu erhalten.

Neben den bisher dargestellten Interessensgruppen, die direkt mit einem Evaluationssystem interagieren, bestehen in der Hochschulleitung und der Öffentlichkeit zwei weitere Interessensgruppen, die mit einem Evaluationssystem nur indirekt in Kontakt treten.

### **Hochschulleitung**

Dem zunehmenden Wettbewerb zwischen den Universitäten (vgl. Ebel-Gabriel 2006, S. 12) und der Einbeziehung von Konzepten der privatwirtschaftlichen Managementlehre im Rahmen des Neuen Steuerungsmodells inhärent ist das Interesse der Hochschulleitung<sup>16</sup> an einer qualitativ hochwertigen und effizienten Evaluation, die einen wichtigen Faktor für eine zunehmende Marktorientierung und die Ergebnissteuerung nach Controlling-Aspekten darstellt. Über die systeminterne Effizienz hinaus hat die Hochschulleitung ein besonderes Interesse

---

<sup>16</sup> Unter „die Hochschulleitung“ wird hier der organisatorische Bereich verstanden, der nicht bereits durch die Gruppe der Evaluationsbeauftragten abgedeckt wird.

daran, dass sich das strategisch wichtige Evaluationssystem nahtlos in die IT-Landschaft der Universität einfügt.

### **Breite Öffentlichkeit**

Für den Wissenschaftsrat (1996, S. 12) besteht eines der vier primären Ziele der Evaluation in der Schaffung der Transparenz auch für die breite Öffentlichkeit. Dabei sollen die mit öffentlichen Mitteln geförderten Universitäten ihrer Pflicht nachkommen, Rechenschaft über die Verwendung der eingesetzten Mittel abzulegen (vgl. ebd. S. 39). In diesem Zusammenhang stellt Lange (2002, S. 9) fest: „Bereits Mitte der 90er Jahre war die Leistungsvermutung zugunsten der Hochschulen in der Öffentlichkeit nicht mehr gegeben. Es galt das Wort, das vom Anfang der 70er Jahre stammt, ‚Man kann in die Hochschulen oder in das Hochschulsystem in Deutschland eine Milliarde hineinwerfen und es macht noch nicht einmal plumps‘.“ Ein Evaluationssystem muss deshalb so konzipiert sein, dass die ermittelten Ergebnisse nicht nur für universitätsinterne Maßnahmen wie die leistungsorientierte Mittelvergabe, sondern auch als Beitrag zur Rechenschaft über große Finanzvolumina unmittelbar vor Haushaltspolitikern und damit mittelbar vor der (Steuer zahlenden) Öffentlichkeit dienen können. Ein zentraler Ansatz zur Erfüllung dieser Anforderung ist eine möglichst weit reichende Absicherung der grundlegenden Anforderungen empirischer Sozialforschung, die in den folgenden drei Abschnitten dargestellt werden.

#### **2.1.3 Planung**

Wie in den vorangegangenen Abschnitten hergeleitet, soll ein Evaluationssystem zur Datenerhebung und zur Erstellung von Auswertung eingesetzt werden können, wobei die Erhebung mit dem Instrument der Befragung durchgeführt wird (vgl. Kromrey 2005, S. 62). Damit muss ein auf Befragungen aufbauendes Evaluationssystem den methodischen Anforderungen der empirischen Sozialforschung genügen. Im Speziellen handelt es sich dabei um die Anforderungen aus den Phasen der Planung, Untersuchung und Auswertung (vgl. Bortz 2005, S. 2–11). Aus den Abhängigkeiten dieser Phasen untereinander lassen sich Funktionsanforderungen an ein Evaluationssystem ableiten, das in ein umfassendes Evaluationsverfahren (vgl. Nippert 2003, S. 47) eingebettet werden kann.

#### **Merkmalsträger**

Ziel der Befragung ist die Messung bestimmter Parameter wie beispielsweise die (abstrakte) Qualität der Lehre. Dabei müssen insbesondere die Erhebungseinheiten eindeutig definiert sein, um den formalen Voraussetzungen zu genügen (vgl. Kromrey 2005, S. 65). Zentrale Elemente sind dabei die Dozenten im Speziellen und die nachfolgend weiter beschriebenen Lehreinheiten im Allgemeinen.

**Personen** In der Praxis kommen dozentenbezogene Evaluationsinstrumente wie Lehrpreise, Rankings oder die leistungsorientierte Mittelvergabe zum Einsatz (siehe Abschnitt

„Leistungsindikatoren“ auf den Seiten 21–23), die in hohem Maße auf der Messung personenbezogener Variablen aufbauen. Auch der Wissenschaftsrat fordert eine dozentenspezifische Lehrevaluation (vgl. Wissenschaftsrat 2006a, S. 87, Wissenschaftsrat 2006c, S. 93). Wenn also an einer beliebigen Stelle im Evaluationsprozess Aussagen über einen Dozenten getroffen werden sollen, so müssen sich die Umfrageantworten eindeutig auf den zu bewertenden Dozenten beziehen. Dieser leicht verständliche Zusammenhang gestaltet sich jedoch als höchst komplex, da ein gängiges Prinzip der Evaluationspraxis – nämlich die Gleichsetzung einer Veranstaltung mit einem Dozenten – damit nur noch dann zulässig ist, wenn jeder Termin einer Veranstaltung von ein und demselben Dozenten gehalten wird.<sup>17</sup> Schon bei der Berücksichtigung von Vertretungen (wegen Krankheit oder anderer Abwesenheitsgründen) ist dieser Fall in der Praxis jedoch nur noch eine seltene Ausnahme. Insbesondere im stark fragmentierten medizinischen Curriculum (siehe Abschnitt „Fragmentierte Veranstaltungen“ auf den Seiten 25–26) ist diese Annahme schließlich gänzlich zu verwerfen, da hier einzelne Termine einer Veranstaltung regelmäßig von unterschiedlichen Dozenten gehalten werden.

**Lehreinheiten** Würde bei einer Evaluation jedoch nur die feingranuläre Erhebung auf der Ebene der Dozenten durchgeführt werden, entstünden mitunter erhebliche Redundanzen. So müssten die Studierenden veranstaltungsübergreifende Fragen wie beispielsweise nach der Zufriedenheit mit Abstimmung der Lehrinhalte zwischen den Veranstaltungen sowie personenbezogene Fragen mehrerer Termine ein und desselben Dozenten mehrfach beantworten. Zudem zeigen van den Bussche et al. (2006), dass Studenten einen ganzen Studienabschnitt anders bewerten, als die Summe dessen einzelner Veranstaltungen. Aus diesem Grund können diese Informationen nicht direkt aus den Antworten auf die Veranstaltungen abgeleitet werden. Eine Abhängigkeit der Ergebnisse von der Veranstaltungsform stellen auch el Hage (1996, S. 61–67) und Wottawa (2001, S. 158) dar. So zeigt beispielsweise el Hage ebd., dass Vorlesungen schlechter bewertet werden, als andere Veranstaltungsarten. Peus et al. (2005a, S. 1) sehen die Abdeckung aller Ebenen der Lehre als eine Grundlage, um die Qualität der Lehre effektiv verbessern zu können. Zur Berücksichtigung dieser Effekte müssen bei der Evaluation neben der dozentenspezifischen auch Erhebungen bezogen auf ganze Lehreinheiten wie Veranstaltungen, Module oder Semester durchgeführt werden können. Damit kann schließlich auch der Forderung von Wolf et al. (2001, S. 98) nach zielgruppenspezifischen Fragebögen entsprochen werden.

**Räumliche Ressourcen** Neben der Evaluation von Dozenten und Lehreinheiten wird in der Literatur gefordert, auch Räume zu evaluieren (vgl. Putz 2003, S. 107 ff.). Dass die räumliche Verteilung in der medizinischen Lehre mit den besonderen Anforderungen des medizinischen Curriculums noch nicht optimal ausgestaltet ist, zeigen auch die Empfehlungen

---

<sup>17</sup> Der abstrakte Begriff der „Veranstaltung“ beschreibt in dieser Arbeit einen organisatorischen Themenkomplex, der mehrere „Termine“ umfasst. Eine Instanz einer abstrakten Veranstaltung gehört immer einer konkreten Ausprägung (z. B. Vorlesung, Tutorial oder Seminar) an. So umfasst beispielsweise die Veranstaltung „Innere Medizin: Methodenseminar“ u. a. die Termine „Herzkatheter am 21.7.2006 14:00–15:30“, „Endoskopie am 28.7.2006 14:00–15:30“ und „Angiographie am 4.8.2006 14:00–15:30“.

des Wissenschaftsrats (vgl. z. B. Wissenschaftsrat 2003a, S. 74 f., Wissenschaftsrat 2003b, S. 89, Wissenschaftsrat 2002b, S. 80).

## Variablen

Neben der Festlegung der Erhebungseinheiten gilt es in der Planungsphase, die zu erhebenden Variablen zu erfassen. Wie unter Abschnitt „Methoden“ auf Seite 9 bereits festgelegt wurde, liegt der Fokus auf der Darstellung von Aspekten, die durch ein Evaluationssystem umgesetzt bzw. unterstützt werden können. Insbesondere ausgeklammert werden dementsprechend Anforderungen an die inhaltliche Ausgestaltung der Fragen sowie der damit verbundenen Operationalisierung der Gütekriterien. Für eine ausführliche Darstellung dieser Aspekte sei auf Rindermann (2001b, S. 55 ff.) verwiesen.

**Fragearten** Um ein umfassendes Bild bei der Befragung zu erhalten, müssen bei der Evaluation unterschiedliche Arten von Fragen gestellt werden können, die sich in offene und geschlossene Fragen unterteilen lassen (vgl. Rindermann 2003, S. 94, Mandl et al. 1999, S. 38). Offene Fragen erlauben eine Freitextantwort und bieten damit die Möglichkeit, schwer vorhersehbare Aspekte zu erfassen. Geschlossene Fragen dagegen erlauben eine Kategorisierung der Antworten. Die Antworten geschlossener Fragen ermöglichen somit in Abhängigkeit vom jeweiligen Skalenniveau Vergleiche und unterschiedliche Formen der Ergebnisverdichtung. Geschlossene Fragen lassen sich in solche mit einfacher (Single-Choice-Fragen) oder mehrfacher Antwortmöglichkeit (Multiple-Choice-Fragen) unterteilen. Eine spezielle Ausprägung der Single-Choice-Fragen sind bipolare Fragen, zu deren Beantwortung ein diskretes Kontinuum zwischen zwei Polen zur Verfügung steht (vgl. Wolf et al. 2001, S. 99). Bipolare Fragen kommen insbesondere für die Einstellungsmessung mithilfe der Likert-Technik zum Einsatz (vgl. Diekmann 2005, S. 209–215).<sup>18</sup> Im Gegensatz zu nicht weiter spezifizierten Single-Choice-Fragen implizieren bipolare Fragen ein höheres Skalenniveau und erlauben damit weitergehende Auswertungen. Bipolare Fragen sind ein zentrales Erhebungsinstrument. So baut das in der Literatur häufig zitierte Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE) überwiegend auf bipolaren Fragen auf (vgl. Rindermann 2001b).

Ein weitgehend akzeptierter und in der Praxis (vgl. Nippert 2003, S. 62) häufig umgesetzter Einsatzbereich der auf studentischen Urteilen basierenden Evaluation liegt in der Messung der Zufriedenheit (vgl. Herzwurm 2000). Kromrey weist jedoch mehrfach (vgl. Kromrey 2001a, S. 54 f., Kromrey 2005, S. 65 f.) darauf hin, dass verschiedene Studenten unterschiedliche Anforderungen an die Lehre haben und bei der statistischen Informationsverdichtung durch die Berechnung von Mittelwerten die Ergebnisse ‚entsubjektiviert‘ werden. Um diesem Problem entgegenzuwirken und die Verdichtung aussagekräftiger zu gestalten, schlagen Hornbostel (2001, S. 94–97) und Herzwurm (2000, S. 401–403) vor, es den Studenten bei der Erhebung zu ermöglichen, die Relevanz der Fragen zur Erreichung ihrer individuellen Zufriedenheit zu

---

<sup>18</sup> Ein typisches, mit der Likert-Technik aufgespanntes Kontinuum, reicht von „Ich stimme überhaupt nicht zu“ bis „Ich stimme voll zu“.

gewichten. Durch diese zusätzliche Information der Gewichtung lassen sich aber nicht nur die Mittelwerte aussagekräftiger gestalten, sondern auch eine explizite Gegenüberstellung von Zufriedenheit und Relevanz herleiten, wie Herzwurm (2000, S. 402) dies anhand eines „Studienzufriedenheitsportfolio[s] einer Lehrveranstaltung“ zeigt.

**Störvariablen** Die Validität studentischer Urteile wird in der Literatur kontrovers diskutiert (vgl. Spiel & Engel 2001, Rindermann 2001b, S. 200–206, Hornbostel 2001, S. 91–94).<sup>19</sup> Eine zentrale Frage lautet dabei, inwiefern Störvariablen<sup>20</sup> die Studierenden bei ihren Urteilen beeinflussen. Die in der Einleitung dargestellte Erkenntnis (siehe Seite 4), dass bereits Äußerlichkeiten der Dozenten die Studierenden in ihren Urteilen beeinflussen ist dabei nicht neu.<sup>21</sup> Nachdem die Evaluationsergebnisse jedoch in verdichteter Form zunehmend zu gravierenden Konsequenzen wie beispielsweise bei der leistungsorientierten Mittelvergabe führen, fordern Moosbrugger & Hartig (2001, S. 58), dass bei der Evaluation „relevante Studienbedingungsmerkmale der Studierenden wie auch der Universitäten mitzuerheben und statistisch zu kontrollieren [sind], um eine angemessene Interpretation der Untersuchungen zu ermöglichen.“<sup>22</sup> In der Literatur werden folgende, nicht unumstrittenen Merkmale aufgeführt.

---

<sup>19</sup> Diese Diskussion soll anhand der unterschiedlichen Auffassung zur Berechnung der für viele Bereiche relevanten Mittelwerte veranschaulicht werden. So stellen Rindermann (2003, S. 90) und Richter (1994, S. 59) fest, dass die Bewertungen von mindestens 10–15 Studierende gemittelt werden müssen, um aussagekräftige Daten zu erhalten. Rindermann (2001a, S. 69) gibt an, dass – in Stichprobengrößen von 10 bis 20 Studierenden und mehr – mittlere Interraterreliabilitäten von  $r = .80$  bis  $r = .90$  erzielbar sind. Die Veranstaltungsmittelwerte gäben somit repräsentativ das durchschnittliche Meinungsbild der Studierenden wieder. Kromrey (1994, S. 97 ff.) widerspricht dem aber, solange nicht bestimmte Störfaktoren explizit berücksichtigt werden, welche dazu führen, dass die Urteile extrem widersprüchlich sind und gemittelte Werte nur ein statistisches Artefakt eines Gesamteindrucks liefern, der so nicht existiert.

<sup>20</sup> Nach Spiel & Engel (2001, S. 62) werden unter Störvariablen (bzw. Biasvariablen) im Rahmen der Lehrevaluation „[...] Variablen bezeichnet, die nicht mit dem Lehr- und Lernprozess in Zusammenhang stehen, jedoch systematisch mit dem Antwortverhalten der Studierenden kovariieren.“

<sup>21</sup> Auf die Bedeutung der Äußerlichkeiten wies bereits Max Weber in seinem Vortrag „Wissenschaft als Beruf“ im Jahre 1922 hin als er feststellte: „Nun ist es aber eine Tatsache, dass der Umstand, dass die Studenten einem Lehrer zuströmen, in weitestgehendem Maße von reinen Äußerlichkeiten bestimmt ist: Temperament, sogar Stimmfall, – in einem Grade, wie man es nicht für möglich halten sollte.“ (Weber 1922 zitiert nach Soellner et al. 2005, S. 72).

<sup>22</sup> Wie gravierend diese Auswirkungen dabei sind, stellt Kromrey (1994, S. 109) speziell bezogen auf globale Zufriedenheitsfragen wie folgt dar: „Wer als Dozent oder Dozentin vor freiwilligen Hörern mit Wahlmöglichkeiten zwischen Angebotsalternativen lehren darf, wessen Hörer persönliches Interesse mitbringen, wessen Hörer regelmäßig vor- und/oder nacharbeiten und die Veranstaltung regelmäßig besuchen und wessen Hörer ‚studien erfahren‘ sind, erhält [auf einer 5-Punkt-Beurteilungsskala] 3,6 Punkte positivere Urteile als der Kollege oder die Kollegin mit entgegengesetzter Konstellation. Wird die Analyse fakultätsweise durchgeführt und können die Veranstaltungen noch nach generell ‚beliebtem‘ oder ‚unbeliebtem‘ Stoff eingeordnet werden, ist der so ‚vorhersagbare‘ lehrunabhängige Punkteanteil sogar noch höher.“

Tabelle 2.1: Störvariablen

| Merkmal  | Literatur  |
|--|--|
| Alter der Studierenden                             | Moosbrugger & Hartig 2001, S. 51–54  |
| Anwesenheit  | Spiel & Engel 2001, S. 67  |
| Anzahl der Studierenden in der Veranstaltung       | el Hage 1996, S. 61–67<br>Spiel & Engel 2001, S. 67  |
| Fachart  | van den Bussche et al. 2006, S. 3–5  |
| Intrinsische/extrinsische Motivation <sup>23</sup> | el Hage 1996, S. 61–67<br>Engel & Krekeler 2001, S. 126 f.<br>Kromrey 1994, S. 94 ff.                          |
| Geschlecht der Lehrenden                           | el Hage 1996, S. 74–77<br>Spiel & Engel 2001, S. 67  |
| Geschlecht der Studierenden <sup>24</sup>          | el Hage 1996, S. 74–77<br>Spiel & Engel 2001, S. 67  |
| Noten  | el Hage 1996, S. 77–81<br>Schmidt 2005, S. 43  |
| Raumverhältnisse                                   | Spiel & Engel 2001, S. 67  |
| Studienabschnitt/-erfahrung                        | el Hage 1996, S. 61–67<br>Kromrey 1994, S. 108<br>van den Bussche et al. 2006, S. 6<br>Wolf et al. 2001, S. 92 |
| Veranstaltungsart                                  | el Hage 1996, S. 61–67<br>van den Bussche et al. 2006, S. 5  |

Eine Erhebung solcher Störvariablen kann dabei sowohl explizit in eigenen Fragen wie auch implizit durch den Bezug der Informationen aus Fremdsystemen erfolgen. Diese Merkmale wurden konkret benannt, da bei ihrer Umsetzung mit impliziter Erfassung jeweils für sich genommen ganz spezifische Anforderungen an ein Evaluationssystem entstehen (siehe Abschnitt „Stammdatenbezug“ auf den Seiten 47–48).

<sup>23</sup> Laut Kromrey (1994, S. 94 ff.) bewerten Studierende, die eine Veranstaltung besuchen müssen schlechter, als solche, die eine Veranstaltung freiwillig besuchen.

<sup>24</sup> Wobei el Hage (1996, S. 77) darauf hinweist, dass „[...] Studentinnen nicht generell Hochschullehrer bzw. Hochschullehrerinnen anders bewerten als ihre männlichen Kommilitonen, sie aber einzelne Aspekte anders gewichten.“

## 2.1.4 Untersuchung

### Evaluationszeitpunkte

Im Rahmen der Untersuchung wird bei der Evaluation nach Scriven (1980) zwischen normativen und summativen Verfahren unterschieden (vgl. auch Kromrey 2001a, S. 24 f., Peus et al. 2005a, S. 2–5, Schratz 1992, S. 237). Beim normativen Verfahren wird die Evaluation Semester begleitend durch mehrere Erhebungen durchgeführt, während bei der summativen Evaluation eine Erhebung am Ende des Semesters durchgeführt wird. Hungerbühler (1997, S. 39) stellt dabei jedoch heraus, dass „sich diese beiden Formen meist nicht klar voneinander trennen [lassen] und [...] eher als extreme Pole eines Kontinuums aufgefasst [werden].“

Mandl et al. (1999, S. 35) fordern die Möglichkeit, Längsschnittanalysen durchführen zu können, bei denen sich die Entwicklung der Antworten personenbezogen über die Zeit hinweg nachvollziehen lassen. In diesem Spezialfall, bei dem der Bezug zur Person zumindest systemintern hinterlegt wird, muss jedoch sichergestellt werden, dass die Umsetzung den Bestimmungen des Datenschutzes entspricht (siehe Abschnitt „Rechtsnormen“ auf Seite 10).

Weiterhin zeigen van den Bussche et al. (2006, S. 5 f.), dass zeitnahe Erhebungen zu anderen Ergebnissen führen, als zeitversetzte. Da unterschiedliche, nicht konstante Einflüsse zu diesen Unterschieden führen, kann nicht verallgemeinert werden, ob eine möglichst frühe oder möglichst späte Evaluation den ‚wahren‘ Einstellungen näher kommen. Auch el Hage (1996, S. 59–60) relativiert den Einfluss des Messzeitpunkts. Um eine grundsätzliche Erinnerungsverzerrung (Recall-Bias) zu vermeiden, sollte die Evaluation jedoch möglichst zeitnah und nicht wie auch vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) (2001, S. 87) kritisiert, versetzt durchgeführt werden. Mit der Vorgabe, so zeitnah wie möglich zu evaluieren, werden implizit auch aufwendigere Verfahren ausgeschlossen, bei denen die Evaluationsmüdigkeit auf der Seite der Studierenden durch eine vom Evaluationsystem geplante, gleichmäßige Streuung der Erhebungszeitpunkte reduziert werden soll.

### Population

Um eine möglichst hohe Repräsentativität zu erlangen, liegt das statistische Ziel des Evaluationsystems in der Vollerhebung. Um dieses Ziel zu erreichen, werden in der Praxis diverse Sanktionen wie die Verweigerung der Scheinausgabe verhängt, wenn sich Studierende weigern, an der Evaluation teilzunehmen. Auf diese Maßnahmen, die sich jedoch im rechtlichen Graubereich bewegen (vgl. Peus et al. 2005a, S. 2, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) 2001, S. 87) wird unter Abschnitt „Rücklaufquote“ auf den Seiten 36–37 detailliert eingegangen.

Mit zunehmender Anzahl der Erhebungen sinkt jedoch die Bereitschaft zur Beantwortung und damit die Antwortrate, was Peus et al. (2005b, S. 5) als eine Form der Evaluationsmüdigkeit beschreiben. Deshalb wird in der Literatur eine Beschränkung des Ausmaßes der Erhebungen auf ein vernünftiges Maß gefordert (vgl. Wissenschaftsrat 2005b, S. 99, Rindermann 2003, S. 94). Um zu verhindern, dass der Ermüdungseffekt zu Verzerrungen führt,



müssen Erhebungen auch auf eine Stichprobe beschränkt werden können. Wie unter Abschnitt „Störvariablen“ auf den Seiten 16–17 dargestellt, werden in der Literatur eine Reihe potentieller Störvariablen diskutiert. Um die damit zusammenhängenden, potentiell verzerrenden Effekte bereits bei der Stichprobenauswahl zu berücksichtigen, sollte die Stichprobe nach Quoten ermittelt werden können.

### **Fremd- und Eigenbildentwicklung**

Hornbostel (2001, S. 98–100) zeigt, dass eine hohe Übereinstimmung zwischen den Urteilen der Studierenden und der Dozenten über die Dozenten herrscht, wenn die Dozenten die Fragen beantworten sollen, die auch den Studierenden gestellt wurden. In der Literatur wird deshalb gefordert, nicht nur Studierenden, sondern auch Dozenten zur Eigenwahrnehmung (aus der Perspektive der Studenten) zu befragen und die beiden Perspektiven gegenüberzustellen (vgl. Kromrey 2001a, S. 45, Mandl et al. 1999, S. 29 ff., Balk 2000, S. 116 f.). Den Zweck dieses Instruments sieht Herzwurm (2000, S. 400) in der „[...] Schärfung des Bewusstseins der verantwortlichen Dozenten oder zur Motivation bei der erstmaligen Verwendung des Befragungsinstruments.“ Eine alternative Ausprägung bei der Studierende und Dozenten die Fragen jeweils aus ihrer Sicht beantworten sollen, stellen Mandl et al. (1999, S. 39 f.) dar.

Eine Gegenüberstellung der beiden Bilder wird als wertvoll erachtet, um den Dozenten Hinweise auf Verbesserungsmöglichkeiten geben zu können.<sup>25</sup> Beispiele einer solcher Auswertung zeigen Mandl et al. (1999, S. 39, 43 und 45).

### **Online-Erhebung**

Nippert (2003, S. 59) leitet aus den Ergebnissen seiner Umfrage ab, dass der Anteil der Studierenden, die an der Evaluation teilnehmen, bei Online-Erhebungen deutlich ansteigt. Er führt dies darauf zurück, dass die Abgabezeitpunkte im Gegensatz zur sofortigen Abgabe am Ende eines Termins (siehe Fußnote 17 auf Seite 14) deutlich entzerrt sind und das Medium Internet eine höhere Anonymität vermittelt.<sup>26</sup> Für den Rahmen dieser Arbeit soll unter einer Online-Erhebung der Versand, die Darstellung und die Beantwortung eines Fragebogens über das Internet verstanden werden. Auch der Wissenschaftsrat (2005b, S. 98 f.) begrüßt die Online-Erhebung. Über diese methodischen Aspekte hinaus bietet die Online-Erhebung zusätzlich den Vorteil, dass die variablen Kosten für die Erhebung und Auswertung bei Online-Systemen gegen Null gehen und auch Fixkosten weitgehend vernachlässigt werden können. Da der Einsatz von Evaluationssystemen langfristig angelegt ist und dieser strukturelle Kostenvorteil gegenüber papierbasierten Erhebungen mit der Zeit zunimmt, sind Online-Systeme anderen Erhebungsformen auch bezogen auf die Kosten überlegen.

---

<sup>25</sup> Zu beachten ist dabei, dass sich die jeweiligen Sichten durchaus überschneiden können. So gelten bestimmte Fragen für Studenten und Dozenten (z. B. „Die Tutorials haben gefallen“), nur für Studenten (z. B. „Die Tutorials möchte ich nicht mehr besuchen“) oder nur für Dozenten (z. B. „Die Studierenden waren konzentriert bei der Sache“) (vgl. Mandl et al. 1999, S. 39).

<sup>26</sup> Der Aspekt der höheren Anonymität kann unter einem technischen Blickwinkel jedoch allenfalls als ‚vermeintlich‘ bezeichnet werden.

Da Online-Erhebungen die Nebenbedingung einer effizienten Untersuchung bei vernachlässigbaren Grenzkosten erfüllen und das Vorhandensein eines Internet-Zugangs unter den Studierenden der Medizin auf breiter Basis als gegeben angenommen werden kann, soll diese Erhebungsform als zielführend betrachtet werden.

### 2.1.5 Auswertung

Ziel der Auswertung ist es, die Ergebnisse darzustellen. Da eine Evaluation nicht zum Selbstzweck durchgeführt wird, sondern nur dann sinnvoll ist, wenn an die Erhebungs- und Auswertungsphase konkrete Konsequenzen anknüpfen, soll die Auswertung eine Qualität aufweisen, die zur Ableitung von Handlungsmaßnahmen geeignet ist.<sup>27</sup> Die Qualität bezieht sich in diesem Zusammenhang auf die korrekte Anwendung statistischer Instrumente und eine zielgruppenspezifische Darstellung der Ergebnisse, die eine schnelle und zugleich fundierte Interpretation der Daten erlaubt.<sup>28</sup>

Statistische Grundregeln für die in einem Evaluationssystem sinnvoll implementierbaren Instrumente der univariaten deskriptiven Statistik zeigt Weiß (2002). Für eine umfassendere Verarbeitung fordert Kromrey (2001a, S. 54 ff.) eine Analyse in dedizierten Statistik-Programmen.

### Berichte

Die einfachste Auswertungsform neben dem Export der Rohdaten ist die Verdichtung der Erhebungsergebnisse in einfachen Berichten. Charakteristisch für die vom Evaluationssystem generierten Berichte ist die Tatsache, dass sie für sich genommen keine Konsequenzen verursachen. Die Konsequenzen ergeben sich rein aus der Interpretation der in den Berichten dargestellten Informationen. Gerade deshalb, sind Berichte jedoch so wichtig: sie verdichten die Daten und erlauben damit die Ableitung von Handlungsvorgaben.<sup>29</sup>

---

<sup>27</sup> Auf die Probleme der zulässigen Interpretation studentischer Urteile wird in dieser Arbeit nicht weiter eingegangen, da sie nicht im Fokus der Anforderungen an ein Evaluationssystem stehen. Es sei an dieser Stelle dennoch auf die Hinweise von Kromrey (2005) verwiesen, der eine Darstellung der Ergebnisse dann als zulässig betrachtet, wenn „die Auskunft gebenden Studierenden zugleich auch Untersuchungsobjekte“ (ebd., S. 63) sind, aber eine Interpretation zumindest als problematisch und angreifbar erachtet, wenn eine Qualität gemessen wird bei der Veranstaltungen und Lehrende die Untersuchungsobjekte sind und die Studierenden als Sachverständige herangezogen werden. Für eine derartige Interpretation ließe sich auch nur schwerlich auf die gesetzlichen Grundlagen zurückgreifen, da diese lediglich eine Beteiligung der Studierenden an einer Evaluation fordern ohne konkret zu bestimmen, in welcher Richtung die so erhobenen Ergebnisse ausgewertet werden sollen.

<sup>28</sup> An dieser Stelle sei darauf hingewiesen, dass keine noch so fundierte und aufbereitete Darstellung von statistischen Berechnungen vor einer Fehlinterpretation gefeit ist (siehe dazu auch Beck-Bornholdt & Dubben 2005).

<sup>29</sup> So beschreibt Putz (2003, S. 114) intern in Dienstbesprechungen veröffentlichte Berichte als „effektives Führungsinstrument“.

Umfassende Evaluationsberichte sollen grundsätzlich für die Evaluationsbeauftragten zur Verfügung stehen, so dass diese die Berichte sowohl für eine interne<sup>30</sup> Darstellung wie auch für eine mögliche externe Darstellung aufbereitet werden können. Damit Dozenten jedoch die Möglichkeit haben, sich zu verbessern, sollten die Ergebnisse der personenbezogenen Evaluation in individualisierten Berichten allen evaluierten Dozenten zur Verfügung stehen. Diese individualisierten Berichte dürfen aus Gründen des Datenschutzes dabei nur die Antworten berücksichtigen, die sich auf den Dozenten beziehen, der den Bericht erhält. Eine transparente Darstellung der Ergebnisse insbesondere für die Studierenden wird dabei von nationalen und internationalen Gremien kontinuierlich gefordert (vgl. Wissenschaftsrat 2006a, S. 87, Wissenschaftsrat 2006c, S. 93, Wissenschaftsrat 2005b, S. 99, Wissenschaftsrat 1992, S. 66 f., European Association for Quality Assurance in Higher Education 2005, S. 19, European Ministers of Education 2003, S. 3). Eine Darstellung für die breite Öffentlichkeit bleibt dagegen optional, wobei sie als Werkzeug gesehen werden kann, welches zur Erreichung der von der Öffentlichkeit immer wieder geforderten transparenteren Darstellung erbrachter Leistungen und damit zu dem von Meffert (1998, S. 1180 f.) beschriebenen Ziel, einer verbesserten Kundenorientierung beitragen kann. Dies soll letztendlich zu einer Erhöhung des Vertrauens führen. Ein solcher Einsatz der Evaluationsergebnisse als Marketing-Instrument wird auch von Reissert & Konnerth (2001, S. 190) beschrieben.

### **Leistungsindikatoren**

Indikatoren zum öffentlichen und internen Vergleich der Lehrleistungen werden insbesondere im Kontext des Neuen Steuerungsmodells (NSM) (vgl. Nickel & Zechlin 2000, S. 154 ff.) diskutiert (vgl. Titscher 2006 vs. Schimank 2006, Kromrey 2001a, S. 32 f.). Leistungsindikatoren können dabei auch Rückmeldungen der Studierenden enthalten (vgl. Weber 2006, S. 278).

**Rankings** Eine im Sinne der Konsequenzen ‚milde‘ Ausprägung von Leistungsindikatoren stellen sog. „Rankings“ dar. Unter diesem Begriff wird laut Duden die „Einordnung in eine Rangliste“ verstanden. Rankings können sich damit auf unterschiedlichste Entitäten wie Dozenten, Instituten, Fakultäten oder ganzen Universitäten<sup>31</sup> beziehen. Der Wissenschaftsrat (2004a, S. 86) bemängelt die noch nicht ausreichend ausgeprägte Entwicklung der Anreizstrukturen für gute Lehrleistungen und sieht das Ranking als ein Ziel, für welches die Evaluation eingesetzt werden soll (vgl. Wissenschaftsrat 1996, S. 12 f.). Rankings selbst werden in der Praxis<sup>32</sup> wiederum als Mittel zum Zweck der Vergabe von Lehrpreisen eingesetzt, bei denen die am höchsten bewerteten Dozenten bestimmte Preise erhalten. Die Lehrleistung soll aber

---

<sup>30</sup> Der Wissenschaftsrat (1996, S. 53 ff.) zeigt einen „Leitfaden zur Evaluation von Studium und Lehre“ der als umfassender Rahmen gesehen werden kann und zu dem der Bericht eines Evaluationssystem relevante Inhalte beitragen kann.

<sup>31</sup> vgl. Hornbostel (2001) für eine Darstellung des universitätsübergreifenden Studienführers des Centrums für Hochschulentwicklung (CHE).

<sup>32</sup> Nippert (2003, S. 58–60) zeigt in seiner Studie, dass 35 % der befragten medizinischen Fakultäten „ein Prämierungssystem für anerkannt gute Lehre eingerichtet haben, das in Form von Lehrpreisen vergeben wird“ (ebd. S. 60). Konkret werden Lehrpreise derzeit beispielsweise an der medizinischen Fakultät der Uni-

neben den finanziellen auch Anreize für die Karriere haben und bei Berufungen eine wichtigere Rolle spielen (vgl. Wissenschaftsrat 2004a, S. 87). Die Vergabe von Lehrpreisen und das Ranking sollen damit letztendlich der von Putz (2003) allgemein formulierten Forderung dienen, nach der das „Engagement einzelner Hochschullehrer [...] am besten durch öffentliche Anerkennung gewürdigt werden [kann]“ (ebd. S. 111 f.).

Wenn Leistungsindikatoren eingesetzt werden, dann dürfen sie jedoch nicht dazu führen, dass gute Dozenten immer besser und schlechte immer schlechter werden (vgl. Schimank 2006, S. 217). Deshalb soll an die reine Darstellung der Ranglisten auch eine individuelle Beratung zur Verbesserung der Schwächen anknüpfen.

**Leistungsorientierte Mittelvergabe** Empfindlichere Konsequenzen nach sich ziehend ist die leistungsorientierte Mittelvergabe (LOM), bei der Entscheidungen über die staatliche<sup>33</sup> und die fakultätsinterne Zuweisung von Finanzmitteln auf der Basis von Leistungsindikatoren getroffen werden. Nach Küpper & Tropp (2004) wird in Zeiten knapper Kassen dabei eine stärkere Output-Orientierung diskutiert, die sich „[...] weg vom kameralistischen System mit dem Fokus auf historisch gewachsene Strukturen, hin zur Schaffung von Transparenz und Effizienz [bewegt]“ (ebd. S. 88 f.). Der Wissenschaftsrat fordert grundsätzlich (vgl. Wissenschaftsrat 2004a, S. 86, 111, Wissenschaftsrat 1996, S. 31) und kontinuierlich in seinen Stellungnahmen<sup>34</sup> zur weiteren Entwicklung der medizinischen Einrichtungen immer wieder die Einbeziehung der Lehrleistung in die LOM. Im Speziellen empfiehlt der Wissenschaftsrat (2006b, S. 75) die Ergebnisse der studentischen Lehrevaluation in der LOM abzubilden, wobei er die vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.) (2001, S. 99–101) erarbeiteten Kriterien<sup>35</sup> als beispielhaft bezeichnet (vgl. Wissenschaftsrat 2004a, S. 87) und damit die Ergebnisse der studentischen Evaluation als Teil eines umfassenden Kriterienkatalogs sieht (vgl. ebd. S. 100).

Die Ermittlung der Verteilungsschlüssel bei der LOM wie auch die Berechnung von Rangplätzen beim Ranking (und damit auch die Vergabe von Lehrpreisen) kann durch eine Verdichtung der Ergebnisse der Lehrevaluation zu Kennzahlen erfolgen.<sup>36</sup> Die dabei algorithm-

---

versität zu Lübeck, der Johannes-Gutenberg-Universität Mainz oder der Friedrich-Alexander-Universität Erlangen-Nürnberg vergeben.

<sup>33</sup> In § 5 Satz 1 HRG wie folgt kodifiziert: „Die staatliche Finanzierung der Hochschulen orientiert sich an den in Forschung und Lehre sowie bei der Förderung des wissenschaftlichen Nachwuchses erbrachten Leistungen.“ Eine Übersicht aktueller Verfahren der LOM zeigen Leszczensky & Orr (2004).

<sup>34</sup> Vgl. Wissenschaftsrat (2006a, S. 87), Wissenschaftsrat (2006c, S. 93), Wissenschaftsrat (2006d, S. 74), Wissenschaftsrat (2005b, S. 99), Wissenschaftsrat (2003a, S. 73), Wissenschaftsrat (2003b, S. 88), Wissenschaftsrat (2002a, S. 83) und Wissenschaftsrat (2002b, S. 78) um nur die neuesten Publikationen zu nennen.

<sup>35</sup> Bei diesen Kriterien werden die Evaluationsergebnisse definierter Fragen zusammen mit externen und fakultätsspezifischen Parametern im Verhältnis 2:2:1 gewichtet. In diesem Zusammenhang misst der Wissenschaftsrat der fakultätsinternen Lehrevaluation eine besondere Bedeutung zu und empfiehlt, die Ergebnisse der Lehrevaluation mit der leistungsorientierten Mittelzuweisung zu verknüpfen (vgl. Wissenschaftsrat 2004b, S. 52).

<sup>36</sup> Explizite Empfehlungen, welche Merkmale für die Berechnung einer für das Ranking verwendeten Kennzahl sinnvoll sind, lassen sich in der Literatur nicht finden. Nach Bayer (2004, S. 175) muss die Menge der zu

misch<sup>37</sup> grundsätzlich sehr einfache Anforderung, Maßzahlen zu berechnen und weiter zu verdichten, steht in der Literatur jedoch in der Kritik. So weisen Engel & Krekeler (2001, S. 173), darauf hin, dass der „mit der Transformation von Mittelwert- in Rangplatzunterschiede einhergehende Informationsverlust als eine potentielle Schwäche von Ranglisten zu erkennen“ ist. Um dieses Problem zu umgehen, ist es sinnvoll, neben der reinen Darstellung der Rangplätze zusätzlich auch die zugrunde gelegte Kennzahl darstellen, um der eigentlich ordinalskalierten Rangfolge einen intervallskalierten Charakter zu verleihen. Um diesem Charakter auch in der tatsächlichen Wirkung Rechnung zu tragen, könnten diese intervallskalierten Kennzahlen dann auch als Schlüssel bei der Berechnung der zu vergebenden Mittel verwendet werden. Diese Maßnahme setzt jedoch voraus, dass für die Berechnung der Kennzahl lediglich intervall- oder verhältnisskalierte Merkmale zugrunde gelegt werden dürfen. Nach Engel & Krekeler (2001, S. 125 f.) und Bayer (2004, S. 198) müssen sich die Kriterien, die zur Berechnung der Ranking-Kennzahl eingesetzt werden, individuell<sup>38</sup> zusammenstellen lassen. Unbestritten ist, dass wenn (auf den Ergebnissen der Lehrevaluation aufbauende) Leistungsindikatoren berechnet werden, sie wegen ihrer reputativen bzw. finanziellen Konsequenzen standardisiert, aktuell und unstrittig<sup>39</sup> sein müssen (vgl. Titscher 2006, S. 210 f., Gerke et al. 2006, S. 1).

Auf die Bedeutung von Störvariablen wurde bereits unter Abschnitt „Störvariablen“ auf den Seiten 16–17 eingegangen. Ihr (potentieller) Einfluss zeigt sich bei der Berechnung der Leistungsindikatoren wegen dem hohen Maß der Informationsverdichtung jedoch noch stärker. Kromrey (1994, S. 111) formuliert in diesem Zusammenhang plakativ, dass die Berechnung von Leistungsindikatoren ohne die Berücksichtigung von Störvariablen „methodischer Unsinn“ wäre. Um diese Angriffsfläche zu minimieren, müssen Mechanismen implementiert werden, um den Effekt der Störvariablen bei der Berechnung der Kennzahlen zu minimieren und „die den Qualitätsurteilen zugrunde liegenden sachbezogenen und insofern rational begründeten von verzerrenden (,Bias-variablen‘) und/oder subjektivistischen, d.h. nicht hinreichend objektivierbaren, Faktoren zu trennen“ (Engel & Krekeler 2001, S. 124).

---

berücksichtigenden Merkmale jedoch „vollständig, problemadäquat und operational sein“. Nach Engel & Krekeler (2001, S. 124 ff., 174) müssen insbesondere die Variablen von der Berechnung ausgenommen werden, auf welche die Bewerteten keinen Einfluss haben.

<sup>37</sup> Eine beispielhafte Umsetzung wird unter Abschnitt „Rankings“ auf den Seiten 69–75 gezeigt.

<sup>38</sup> Nach Bayer (2004, S. 22, 28) müsste darüber hinaus eine Differenzierung des Rankings nach Ranking-Zielgruppen (z. B. Entscheidungsträger, die „bezüglich eines Ranking-Ziels homogene Entscheidungsprämissen besitzen“) vorgenommen werden. Die Auswahl der bei der Berechnung der Kennzahlen zu berücksichtigenden Kriterien müsste also (so wie beispielsweise beim CHE Ranking) flexibel definierbar sein. Im Rahmen dieser Arbeit wird das Ranking jedoch auf eine Perspektive fixiert. Der Aussage von Kromrey (2000) (zitiert nach Bayer 2004, S. 28) entsprechend, dass „mit der Festlegung, aus wessen Perspektive Kriterien und Standards ausgewählt und formuliert werden, [...] dann bereits eine wesentliche Vorentscheidung getroffen [ist]“, wird eine weitere Differenzierung damit als nicht notwendig erachtet.

<sup>39</sup> Neben den erwähnten Aspekten unterstreicht diese Forderung insbesondere die bereits dargestellte Anforderung einer dozenten-spezifischen Evaluation. Ist das Evaluationssystem nicht in der Lage, die Dozenten eindeutig zu identifizieren und zu evaluieren, sind alle Kennzahlen und daraus abgeleitete Konsequenzen in höchstem Maße angreifbar.

### 2.1.6 Effizienz

Die Durchführung der Evaluation soll aus Mitteln der Hochschule finanzierbar sein (vgl. Wissenschaftsrat 1996, S. 36), in Abwägung der Kosten sowie der benötigten Zeit insbesondere bei der Informations- und Datensammlung pragmatisch durchgeführt werden können und organisatorische und personelle Ressourcen so wenig wie möglich binden (vgl. Reissert & Konnerth 2001, S. 190 f., Landfried 2003, S. 13).

## 2.2 Besondere Aspekte des medizinischen Curriculums

Im vorangegangenen Abschnitt wurde das theoretische Spannungsfeld dargestellt, in welchem sich die Lehrevaluation allgemein befindet. Der vorliegende Abschnitt erweitert und konkretisiert diese Grundlagen um besondere Charakteristika des medizinischen Curriculums, wobei der Fokus auf den humanmedizinischen<sup>40</sup> Ausprägungen liegt.

### 2.2.1 Neue Lehrformen

Ziel und berufliches Leitbild des Studiengangs Humanmedizin ist laut ÄAppO<sup>41</sup> und Wissenschaftsrat (2004a, S. 18) der praktizierende Arzt. Aufgabe der Lehre ist es also, nicht nur theoretisches Wissen, sondern auch praktisches Können zu vermitteln um so eine möglichst hohe Ebene auf der Miller'schen Pyramide zu erreichen (vgl. Miller 1990, Mandl et al. 1999, S. 11, Scherpbier 2003, S. 68–70). Zudem ist die Medizin von einem andauernden, enormen Wissenszuwachs geprägt (vgl. Wissenschaftsrat 2005a, S. 36). Diese Tatsachen führen kombiniert mit den Anforderungen der ÄAppO<sup>42</sup>, dass die Lehre „auf der Basis des aktuellen Forschungsstandes“ zu vermitteln ist und den Empfehlungen des Wissenschaftsrats (vgl. Wissenschaftsrat 1992, S. 69) nach einer vertikalen Integration mehrerer Fächer zwischen vorklinischen und klinischen Lehrveranstaltungen zu einem sehr diversifizierten Lehrangebot und neuen Lehrkonzepten, die ständigen Änderungen unterliegen. Diese Tatsachen erfordern eine kontinuierliche und häufig weit reichende<sup>43</sup> Anpassung des Curriculums (vgl. Burger 2003, S. 27, Wissenschaftsrat 1992, S. 69 ff.)

---

<sup>40</sup> Da die humanmedizinischen Curricula einen hohen Grad an Komplexität aufweisen, lassen sich die Erkenntnisse dieses Abschnitts auch auf die hier nicht explizit berücksichtigten Curricula der Zahnmedizin sowie der neu eingeführten Medizin-Studiengänge mit Bachelor- bzw. Masterabschluss übertragen. Somit können die Erkenntnisse dieser Arbeit durchaus auch bei der laut Wissenschaftsrat (2004a, S. 17) dringend ausstehenden Reform der Zahnärztlichen Approbationsordnung Berücksichtigung finden.

<sup>41</sup> Vgl. § 1 Abs. 1 Satz 1 ÄAppO.

<sup>42</sup> Vgl. § 1 Abs. 1 Satz 4 ÄAppO.

<sup>43</sup> Als Beispiel sei hier die Zusammenarbeit zwischen der Medizinischen Fakultät der Ludwig-Maximilians-Universität München mit der Harvard Medical School (Boston) in der „Alliance for Medical Education“ genannt, deren Aufgabe die Neugestaltung des medizinischen Curriculums mit dem Ziel ist, „aufbauend auf einer systematischen Wissensvermittlung, den Studenten durch die Methode des problemorientierten Lernens die Möglichkeit zu bieten, sich anhand authentischer Falldarstellungen die Inhalte eines Stoffgebiets selbständig und praxisnah zu erarbeiten“ (Mandl et al. 1999, S. 18 ff.).

Beispiele für neue Konzepte sind das Problemorientierte Lernen (POL) als Weiterentwicklung des klassischen Unterrichts am Krankenbett weg von der Vermittlung reinen Faktenwissens hin zu Lösungsstrategien komplexer Problemstellungen aus dem beruflichen Alltag eines Arztes oder das E-Learning wie beispielsweise im Rahmen der Virtuellen Hochschule Bayern (vhb). Hinzu kommt, dass diese Anpassungen sowohl inter- wie auch intrauniversitär in der Regel nicht einheitlich vorgenommen werden. So unterscheiden sich die humanmedizinischen Curricula beispielsweise der privaten Universität Witten/Herdecke, der Humboldt-Universität zu Berlin und der Ludwig-Maximilians-Universität München deutlich (vgl. Wissenschaftsrat 2004a, S. 22–26).

Sowohl die Inhalte der Lehrveranstaltungen wie auch die Formen der Lehrkonzepte selbst unterliegen damit einem steten Wandel. Wie unter Abschnitt „Untersuchung“ auf den Seiten 18–20 dargestellt, soll die Evaluation jedoch langfristig ausgelegt werden, so dass sich die Ergebnisse auch über längere Perioden hinweg vergleichen lassen. Weiterhin ist der Vergleich der Evaluationsergebnisse ein zentrales Instrument<sup>44</sup>, um herauszufinden, ob die kontinuierlichen Anpassungen die gewünschten Effekte erzielen.

### 2.2.2 Modellstudiengänge

Neben der kontinuierlichen Anpassung des Curriculums ermöglicht § 41 ÄAppO den medizinischen Fakultäten, unter bestimmten Voraussetzungen einen sog. Modellstudiengang anzubieten. Im Rahmen eines solchen Modellstudiengangs, der das Physikum aushebelt, können die Fakultäten das Curriculum unter bestimmten Rahmenbedingungen abweichend vom Regelstudiengang selbst gestalten. Diese Möglichkeit haben bereits mehrere Fakultäten wie z. B. die der Charité Berlin, der Universität Witten/Herdecke oder der Technischen Universität Dresden umgesetzt. Gemäß § 41 Abs. 2 Nr. 4 ÄAppO wird für die Zulassung als Modellstudiengang vorausgesetzt, „dass eine sachgerechte begleitende und abschließende Evaluation des Modellstudiengangs gewährleistet ist.“ Da § 41 Abs. 2 Nr. 5 ÄAppO darüber hinaus festlegt, dass „[...] Verlängerungsanträge anhand von Evaluationsergebnissen zu begründen sind“, haben Fakultäten, die einen Modellstudiengang anbieten und beabsichtigen, diesen zu verlängern, ein großes Interesse daran, die Evaluationsergebnisse der beiden Studiengänge miteinander zu vergleichen um einen entsprechenden Antrag Normen konform begründen zu können.

### 2.2.3 Fragmentierte Veranstaltungen

Im Fachbereich der Humanmedizin zeigt jedoch nicht nur das Curriculum selbst durch stete Anpassungen und Heterogenität innerhalb und zwischen den Universitäten einen höheren Grad an Komplexität, als die Curricula anderer Fachbereiche. Vielmehr sind auch die Veranstaltungen innerhalb des Curriculums von Charakteristiken gekennzeichnet, die sich in ihrer Komplexität deutlich von den Veranstaltungen anderer Curricula unterscheiden.

---

<sup>44</sup> Vgl. Evaluationsstudien zum Problemorientierten Lernen von Mandl et al. (1999, S. 17).

Bei Curricula wie beispielsweise denen der wirtschaftswissenschaftlichen Fachbereiche, welche grundsätzlich von einem Dozenten gehalten werden, können die Veranstaltungen und die Dozenten durch einfache Umfragesysteme gut evaluiert werden. In diesem Fall kann es beispielsweise durch Krankheit oder sonstigen Abwesenheitsgründen zwar zu Abweichungen zwischen dem geplanten und dem tatsächlichen Verlauf der Veranstaltung kommen. In diesen Fällen wird der für die Veranstaltung vorgesehene Dozent jedoch in der Regel für einen oder wenige Termine von einem anderen Dozenten vertreten. Wird eine Veranstaltung ganz überwiegend von einem Dozenten gehalten, so lassen sich seltene und kurze Vertretungen bei der personenbezogenen Evaluation vernachlässigen, wenn die Studierenden bei der Evaluation darauf hingewiesen werden, dass solche Ausnahmen bei der Beurteilung nicht berücksichtigt werden sollen. In diesen Bereichen lässt sich die personenbezogene mit der veranstaltungsbezogenen Evaluation approximieren.

In der Medizin dagegen führen die spezialisierte Wissensvermittlung und die interdisziplinäre Orientierung an übergeordneten Strukturen wie Organen oder Problemen (vgl. Wissenschaftsrat 1992, S. 69) anstelle von Fächern dazu, dass eine Veranstaltung bereits von vorne herein mit mehreren Dozenten geplant wird. Da insbesondere beim Unterricht in Lehrkrankenhäusern und -praxen die Dozenten neben der Lehrtätigkeit auch ihren ärztlichen Pflichten nachkommen müssen, ist die Abweichung der geplanten von der tatsächlichen Zuweisung der Dozenten in der Medizin darüber hinaus ein konstanter Faktor. Dieses Prinzip, nach dem eine Veranstaltung auf mehrere Dozenten verteilt ist, soll nachfolgend unter dem Begriff „Fragmentierung“ subsumiert werden. Hinzu kommt, dass aufgrund der großen Studentenzahl, die oftmals einem begrenzten Platzangebot – beispielsweise am Krankentbett – gegenüberstehen, Veranstaltungen häufig in Gruppen unterteilt werden müssen (vgl. Wissenschaftsrat 2005a, S. 36, Wissenschaftsrat 2004a, S. 18 f.).

Deshalb benötigt das medizinische Curriculum eine besondere Betrachtung, die den positiven Nebeneffekt hat, dass die Ergebnisse auch für einfacher strukturierte Curricula verwendet werden können.

### **2.2.4 Komplexe Organisationsstruktur**

Neben der universitätsinternen Lehre wird im medizinischen Curriculum auch an externen Einrichtungen wie Lehrkrankenhäusern und Lehrpraxen gelehrt, wobei für alle organisatorischen Einheiten sichergestellt werden muss, dass die Ausbildung auf entsprechendem Niveau durchgeführt wird (vgl. Wissenschaftsrat 2005a, S. 46 sowie § 6 HRG). Wie von Neuser (2003) dargestellt, kann es jedoch nicht die Aufgabe aller organisatorischer Einheiten sein, sich die für eine Evaluation nötigen Kenntnisse und Ausstattungen anzueignen. Zudem müssen Evaluationsdaten in einer Form vorliegen, welche es erlaubt, die Einheiten untereinander vergleichen zu können. Es liegt deshalb nahe, die Evaluation zu zentralisieren. Das Ziel der Zentralisierung gewinnt auch deshalb an Gewicht, da der Einsatz eines zentralen Online-Systems wegen positiver Skaleneffekte zu geringeren Kosten als der Einsatz mehrerer dezentraler Systeme führt.



traler Systeme führt.<sup>45</sup> Das Vorhaben der Zentralisierung steht jedoch der Herausforderung gegenüber, Zugriff auf alle benötigten Informationen zu erhalten. Zu diesen Daten zählen vor allem die Personaldaten und die Informationen darüber, welcher Dozent welche Termine welcher Veranstaltungen vor welchen Studierenden tatsächlich gehalten hat. Der Grad der Komplexität steigt noch weiter an, falls neben den Dozenten wie unter Abschnitt „Räumliche Ressourcen“ auf den Seiten 14–15 gefordert, auch Räume evaluiert werden sollen. In diesem Fall müssten zusätzlich die Raumdaten sowie die Informationen darüber erfasst werden, welche Termine in welchen Räumen tatsächlich gehalten wurden. Während sich Räume bei inneruniversitären Veranstaltungen durchaus erfassen lassen, werden Veranstaltungen in Lehrkrankenhäusern und -praxen jedoch häufig nur mit einem Treffpunkt angegeben. Der tatsächliche Raum konkretisiert sich in diesen Fällen dann erst bei dem jeweiligen Termin.

### 2.2.5 Homogene Prüfungsstruktur

Neben den bisher dargestellten Komplexitätsaspekten weist das medizinische Curriculum jedoch im Vergleich zu anderen Curricula eine relativ homogene Prüfungsstruktur auf. So ist das Hauptstudium im medizinischen Curriculum von einem sehr geringen Grad an Diversifikation geprägt. Auch wenn der Fragmentierungsgrad der Veranstaltungen selbst sehr hoch ist, so führt dies zu einer geringen Anzahl unterschiedlicher Veranstaltungen. Durch die relativ hohe Anzahl an Studenten die nur eine geringe Anzahl an unterschiedlichen Veranstaltungen besuchen, liegen sehr gute Voraussetzungen für die Ziehung von Stichproben bei der veranstaltungsbezogenen Evaluation vor.

Weiterhin wird die medizinische Ausbildung durch Staatsexamina geprüft, was in Deutschland nur noch bei der juristischen Ausbildung der Fall ist. Nachdem die Ergebnisse der Staatsexamina bundesweit vergleichbar sind, ist es in der medizinischen Ausbildung damit möglich, eine universitätsübergreifende Korrelation zwischen den Evaluationsergebnissen und den Ergebnissen der Staatsexamina herzustellen. Es ließe sich damit insbesondere ein Beitrag zur Beantwortung der Frage leisten, ob die in den Evaluationsergebnissen ausgedrückte Zufriedenheit der Studierenden mit deren Abschlussnoten zusammenhängt.

## 2.3 Evaluationspraxis an den medizinischen Fakultäten in Deutschland

In den vorherigen beiden Abschnitten wurde der methodische Rahmen dargestellt, in welchem sich die Lehrevaluation allgemein und im speziellen Fall der medizinischen Lehre bewegt. Da die Evaluation letztendlich von den Fakultäten getragen und umgesetzt werden muss, sollen die bisher erlangten Erkenntnisse im vorliegenden Abschnitt um die Perspektive der

---

<sup>45</sup> Die positiven Skaleneffekte bei Online-Systemen ergeben sich aus dem unter Abschnitt „Online-Erhebung“ auf den Seiten 19–20 bereits beschriebenen Zusammenhang, nachdem Online-Systeme zu vernachlässigende variable Kosten aufweisen. Die verbleibenden Kosten (insbesondere die Anschaffungskosten) sind bei einem Zentralsystem mit großer Verarbeitungskapazität in der Regel geringer, als die Summe der Anschaffungskosten mehrerer dezentraler Systeme mit geringer Verarbeitungskapazität.

Evaluationspraxis erweitert werden. Nachdem aus den bisher veröffentlichten empirischen Studien nur wenige konkrete Anforderungen für ein Evaluationssystem abgeleitet werden können (vgl. Nippert 2003), wurde im Rahmen der vorliegenden Arbeit im Zeitraum von Oktober 2006 bis März 2007 eine neue Umfrage durchgeführt. Dabei wurden die 35 medizinischen Fakultäten in Deutschland nach unterschiedlichen Aspekten ihrer Evaluationspraxis befragt. Die Rücklaufquote lag mit 31 antwortenden Fakultäten bei 89 %. Die dabei gestellten Fragen sind im Anhang A auf den Seiten 97–98 aufgeführt. Diese Fragen sollten einerseits das Evaluationsverfahren ganz allgemein erheben, ohne dabei konkrete Strukturen vorzugeben, die zu einem Verlust wertvoller Informationen hätten führen können. Zudem sollten spezifische Fragen die Umsetzung bestimmter, in der Literatur vorgegebener Aspekte untersuchen sowie die summativen Erfahrungen aus der täglichen Evaluationspraxis erheben. Ziel der Umfrage war es also, die derzeitigen Verfahren darzustellen und daraus konkrete Anforderungen an ein Evaluationssystem aus der Praxis ableiten zu können.

### 2.3.1 Erhebungsform

An erster Stelle des Interesses stand die Frage, in wieweit die Lehrevaluation derzeit bereits vollständig online<sup>46</sup> durchgeführt wird und wie weit der Einzug der IT bereits vorangeschritten ist.<sup>47</sup> Abbildung 2.1 auf der nächsten Seite zeigt dabei die Entwicklung der letzten Jahre durch eine Gegenüberstellung mit den Ergebnissen von Nippert (2003, S. 54).<sup>48</sup> Während im Jahre 2003 mit 79 % der überwiegende Teil der Fakultäten die Erhebung noch rein schriftlich durchführte, ist dieser Anteil in der aktuellen Studie auf 32 % gesunken. Derzeit setzen mit 48 % knapp die Hälfte der Fakultäten bereits vollständig auf eine Online-Erhebung. Mehr als Zweidrittel der Fakultäten (67 %) evaluieren vollständig oder in bestimmten Bereichen online. Es lässt sich also ein Trend zur Umstellung der schriftlichen hin zur Online-Erhebung feststellen, der auch vom Wissenschaftsrat begrüßt wird (vgl. Wissenschaftsrat 2005b, S. 98 f.). Diese Entwicklung ist eine wichtige Voraussetzung um bestimmte, in den vorherigen Abschnitten hergeleitete Anforderungen unter der Nebenbedingung der Effizienz in der Praxis umsetzen zu können. So ließe sich beispielsweise eine dozentenbezogene, formative Evaluation mit Berücksichtigung spezieller Fragen für die jeweiligen Lehreinheiten schriftlich nur unter hohem administrativen Aufwand realisieren.

---

<sup>46</sup> Unter der Online-Erhebung wird eine Befragung in elektronischer Form über einen Web-Browser verstanden. Unter der schriftlichen Erhebung wird die Befragung mit Papier-Fragebögen verstanden, die auch „Paper & Pencil“-Verfahren genannt wird.

<sup>47</sup> Zwar lassen sich auch schriftliche Verfahren durch die IT unterstützen. Eine umfassende Abdeckung der Aufgabenbereiche durch die IT ist jedoch nur bei Online-Verfahren möglich.

<sup>48</sup> Die Studie von Nippert führte die Erhebung ebenfalls unter den medizinischen Fakultäten Deutschlands durch, weshalb sich deren Ergebnisse mit den Ergebnissen der vorliegenden Studie vergleichen lassen. Die Rücklaufquote der Studie von Nippert lag mit 24 antwortenden Fakultäten bei 65 %.

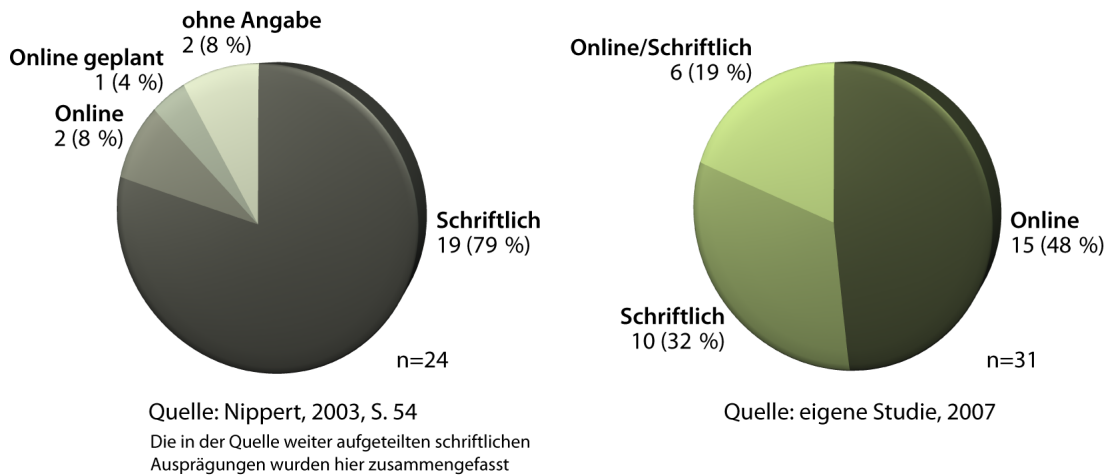


Abbildung 2.1: Entwicklung der Erhebungsformen

### 2.3.2 Externe Evaluation

Alle aktuell befragten Fakultäten beteiligen die Studierenden an der Evaluation in Form von Umfragen. Damit folgend den Fakultäten grundsätzlich den oben dargestellten Empfehlungen und Forderungen. Der in der Literatur häufig zu findende Forderung nach einer mehrstufigen Evaluation durch eine Beteiligung externer Evaluationsagenturen kommt die Praxis jedoch nur eingeschränkt nach. Wie in Abbildung 2.2 dargestellt, gaben mit 68 % mehr als Zweidrittel der befragten Fakultäten an, keine externe Evaluation durchzuführen. Mit 8 Fakultäten (26 %) führt lediglich gut ein Viertel der Befragten derzeit eine externe Evaluation durch.

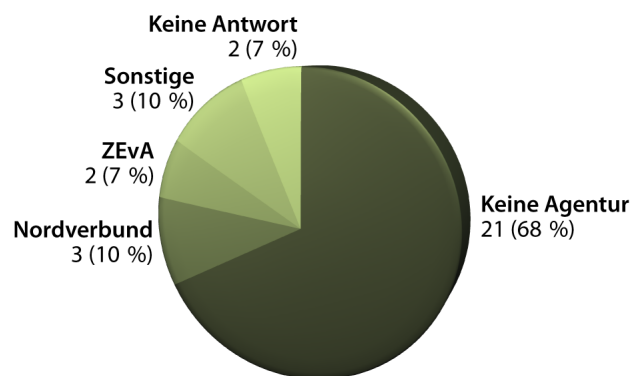


Abbildung 2.2: Externe Evaluation

Sollen die Evaluationsergebnisse im Rahmen einer mehrstufigen Evaluation eingesetzt werden, kann sich die Anforderung ergeben, die Rohdaten der Umfragen zu exportieren, um

externen Analysen alle Freiheitsgrade der Auswertung zu überlassen und dabei Medienbrüche zu vermeiden.

### 2.3.3 Leistungsindikatoren

Die nachfolgenden drei Abschnitte stellen die Verwendung von Leistungsindikatoren durch unterschiedliche Ausprägungen dar. Dabei wird jeweils aufgezeigt, ob für die Berechnung dieser Indikatoren die Evaluationsergebnisse miteinbezogen werden. Diese Erkenntnis ist aus zweierlei Blickwinkeln interessant. Zum einen stellt sie das Spannungsfeld zwischen den methodischen Restriktionen und der praktischen Umsetzung dar. Zum anderen erlaubt sie die Ableitung weiterer Anforderungen an ein Evaluationssystem, solche Indikatoren berechnen zu müssen.

#### Leistungsorientierte Mittelvergabe

Auch wenn der Umgang mit und die Interpretation von Leistungsindikatoren bei Rankings, Lehrpreisen und der LOM wie unter Abschnitt „Leistungsindikatoren“ auf den Seiten 21–23 beschrieben, in der Literatur kontrovers diskutiert wird, so hat insbesondere die LOM bereits ihren Einzug in die Praxis gefunden. Der Erhebung von Nippert (2003, S. 58 f.) entsprechend, setzten im Jahre 2003 bereits 46 % der 24 ausgewerteten medizinischen Fakultäten in Deutschland LOM-Verfahren ein.<sup>49</sup> Der Anteil in der damaligen Studie steigt auf 54 %<sup>50</sup> an, wenn auch die geplanten LOM-Implementierungen berücksichtigt werden. Den aktuellen Stand zum Einsatz von LOM-Verfahren zeigt Abbildung 2.3 auf der nächsten Seite.

Bei der vorliegenden Studie gaben bereits 23 Fakultäten (74 %) an, Mittel leistungsorientiert zu vergeben. Mit 16 Fakultäten (52 %) führt mehr als die Hälfte aller Befragten die LOM unter Einbeziehung der Evaluationsergebnisse in die Berechnung der Verteilung durch. 6 Fakultäten (19 %) berücksichtigen die Evaluationsergebnisse bei der LOM nicht. Die Ursache dafür könnte darin liegen, dass eine Vielzahl der Fakultäten keine dozentenbezogene Evaluation durchführt und damit die Leistungen zumindest nicht personenbezogen vergeben kann.

#### Lehrpreise

Während die LOM einen nach innen gerichteten Charakter aufweist, zielt die Vergabe von Lehrpreisen auf eine Anerkennung von Leistungen mit Außenwirkung. Wie in der Abbildung 2.4 auf der nächsten Seite gezeigt, vergeben derzeit mit 18 antwortenden (58 %) ein Großteil der Fakultäten Lehrpreise. Jeweils 9 Fakultäten (29 %) beziehen die Evaluationsergebnisse bei den Berechnungen für die Vergabe der Lehrpreise ein, bzw. unterlassen dies.

---

<sup>49</sup> Dabei gaben 17 % an, die LOM-Verfahren exklusiv für die Lehre durchzuführen. 29 % gaben an, die Verfahren für Forschung und Lehre kombiniert einzusetzen.

<sup>50</sup> Wegen eines Rechenfehlers von Nippert (Addition einer absoluten anstelle einer relativen Zahl) wurde der von ihm in der Quelle angegebene Anteil von 48 % (vgl. ebd. S. 59) vom Autor der vorliegenden Arbeit hier auf 54 % korrigiert.

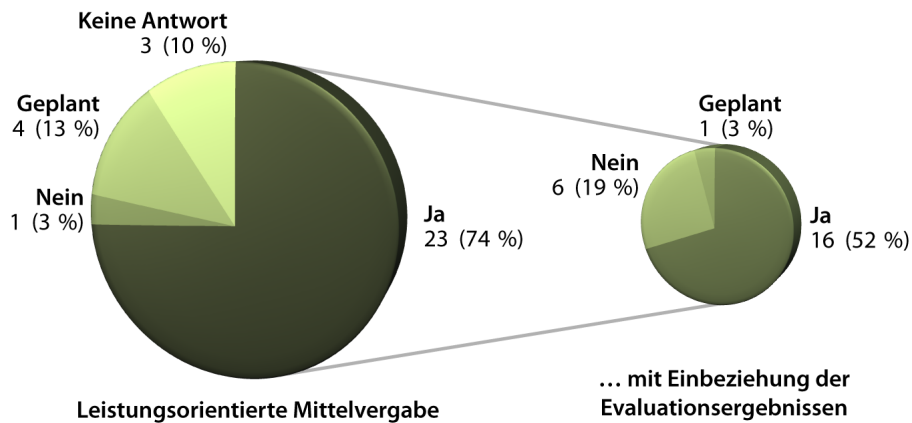


Abbildung 2.3: Leistungsorientierte Mittelvergabe

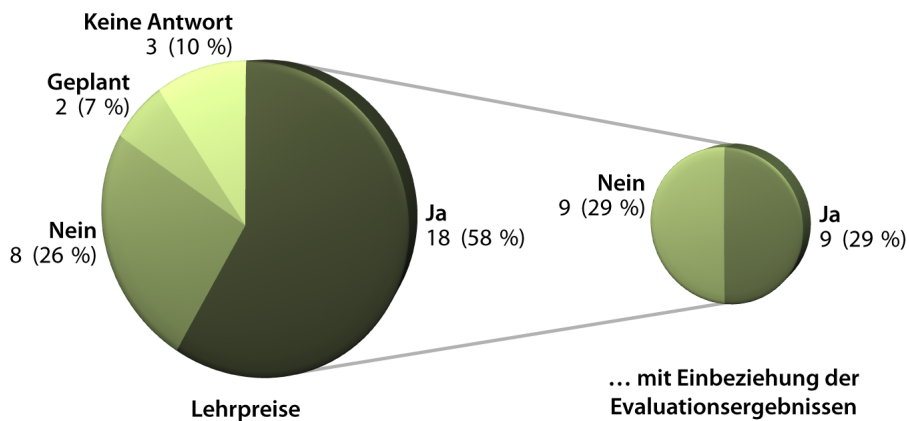


Abbildung 2.4: Lehrpreise

## Dozenten-Rankings

Ein ebenfalls nach außen gerichteter Einsatz von Leistungsindikatoren sind Dozenten-Rankings, deren methodische Grundlagen ebenfalls unter Abschnitt „Leistungsindikatoren“ auf den Seiten 21–23 bereits dargestellt wurden. Wie in der Abbildung 2.5 auf der nächsten Seite gezeigt, ergab sich aus der Umfrage, dass nur 5 Fakultäten (16 %) ein Dozenten-Ranking durchführen. 24 Fakultäten (77 %) dagegen setzen ein solches Instrument nicht ein, wobei 3 dieser 24 Fakultäten ein Veranstaltungs-Ranking anbieten.<sup>51</sup>

Speziell unter diesem Aspekt lässt sich die Vermutung anstellen, dass dozentenbezogene Leistungsindikatoren, nach denen in diesen drei Abschnitten gefragt wurde, auch deshalb

<sup>51</sup> Da diese Information in den Antworten der Fakultäten als Zusatzhinweis gegeben wurde und nach dieser Option nicht explizit gefragt wurde, kann nicht ausgeschlossen werden, dass Veranstaltungs-Rankings noch von weiteren Fakultäten durchgeführt werden.

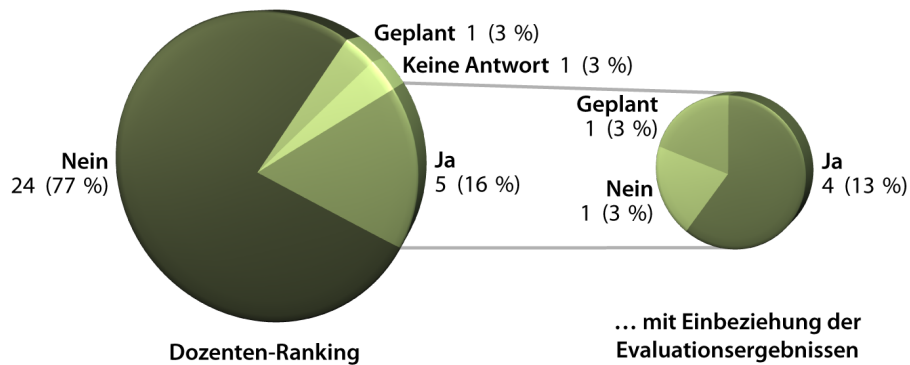


Abbildung 2.5: Dozenten-Ranking

nicht eingesetzt bzw. nicht auf die Evaluationsergebnisse gestützt werden, da die benötigten Daten mangels der Implementation einer dozentenbezogene Evaluation gar nicht vorhanden sind.

### 2.3.4 Berücksichtigung von Störvariablen

Nachdem die Berücksichtigung von Störvariablen in der Literatur insbesondere bei der Berechnung von Leistungsindikatoren vehement (wenn auch im Detail uneinheitlich) gefordert wird (siehe Abschnitt „Störvariablen“ auf den Seiten 16–17) und dies eine für ein Evaluationssystem prädestinierte Funktion darstellt, sollte durch die Umfrage ermittelt werden, ob die Evaluationsergebnisse in der Praxis um Störvariablen bereinigt werden. Die Abbildung 2.6 zeigt die dabei erhaltenen Ergebnisse.

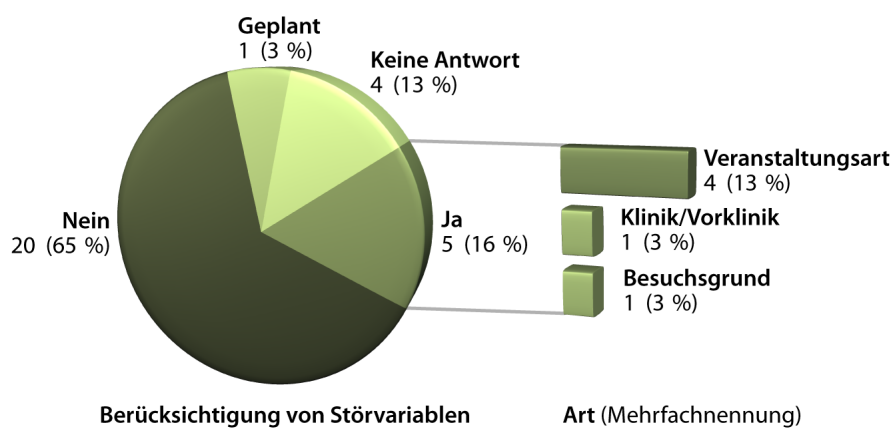


Abbildung 2.6: Berücksichtigung von Störvariablen

Auf diese Frage hin, gaben nur 5 Fakultäten (16 %) an, die Evaluationsergebnisse um Störvariablen zu bereinigen. Konkret werden dabei mit 4 Nennungen vorrangig die Veranstaltungsart sowie jeweils einmal die Unterscheidung zwischen Vorklinik und Klinik sowie der Besuchsgrund berücksichtigt.

### 2.3.5 E-Learning

Auch wenn die Euphorie der 1990er Jahre um das E-Learning einer Ernüchterung um den benötigten Aufwand für die Erstellung der Inhalte und die Etablierung im Curriculum gewichen ist, so sollte der aktuelle Stand des Angebots von E-Learning-Kursen dennoch erhoben werden, da der Wissenschaftsrat (2005b, S. 98) diese Lehrform ausdrücklich begrüßt. Wären E-Learning-Kurse weit verbreitet – so die Überlegung – dann müssten diese ebenso wie von Dozenten persönlich gehaltene Veranstaltungen evaluiert werden. Die Abbildung 2.7 zeigt die Ergebnisse, denen zufolge mit 23 Fakultäten (74 %) bereits knapp Dreiviertel aller Fakultäten E-Learning-Kurse anbieten.

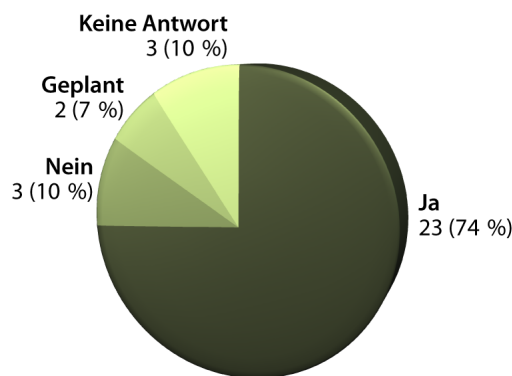


Abbildung 2.7: Angebot von E-Learning-Kursen

Nur noch 3 Fakultäten (10 %) bieten ausdrücklich keine E-Learning-Kurse an und planen dies auch nicht. E-Learning ist also unter den medizinischen Fakultäten in Deutschland weit verbreitet und muss deshalb auch von einem Evaluationssystem berücksichtigt werden.

### 2.3.6 Zufriedenheit

Schließlich wurden die Fakultäten gebeten, die Aspekte herauszustellen, mit denen sie in ihrem bestehenden Verfahren besonders zufrieden sind. Die Antworten erfolgten in offener Form ohne vorgegebenen Kategorien wobei sowohl nach einer besonderen Zufriedenheit mit den Aspekten der eingesetzten Systeme wie auch mit dem Evaluationsverfahren gefragt wurde. Die daraus abgeleiteten Ergebnisse stellen die folgenden beiden Abschnitte dar.

**Technische Aspekte** Wie in Abbildung 2.8 dargestellt, sehen 12 Fakultäten (39 %) ihre Zufriedenheit mit den bestehenden Systemen besonders in der damit erzielten Effizienz begründet. Nachdem auch die Systeme, die für die schriftliche Erhebung zum Einsatz kommen, Funktionen für die effiziente Auswertung anbieten, ist es nicht verwunderlich, dass die effiziente Auswertung der Evaluationsergebnisse an erster Stelle steht. Konkret führten die Befragten unter diesem Aspekt an, dass sie mit den eingesetzten Systemen in der Lage waren, die erhobenen Daten in Form von Berichten aufzubereiten und diese an die Dozenten für ein individuelles Feedback weiterzuleiten. Systeme, bei denen diese beiden Prozessschritte teil- oder vollautomatisch durchgeführt werden können, wurden besonders hervorgehoben.

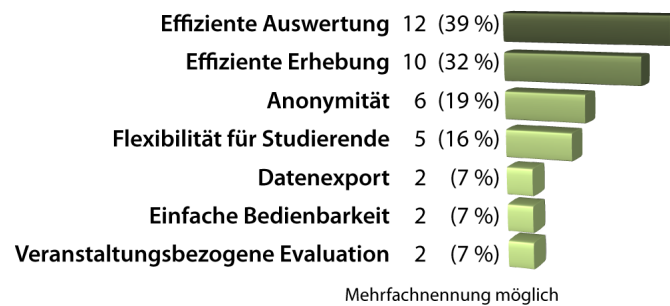


Abbildung 2.8: Gründe der Zufriedenheit (technische Aspekte)

Nachdem nicht alle Fakultäten ihre Erhebung vollständig online durchführen, wurde die effiziente Erhebung nicht mit derselben Häufigkeit wie die effiziente Auswertung genannt. Eine besondere Zufriedenheit mit diesem Aspekt gaben 10 Fakultäten (32 %) an. Die bezogen auf das Verhältnis zwischen online und schriftlichen Erhebungen relativ häufige Nennung dieses Aspekts lag den Antworten entsprechend darin begründet, dass die Systeme, mit denen eine schriftliche Erhebung durchgeführt wird, inzwischen zuverlässige Verfahren für das Einlesen der Fragebögen im Stapelbetrieb ermöglichen, die deutlich effizienter als eine vollständig manuelle Datenerfassung sind.

Den Aspekten, welche die Effizienz betreffen und deren Auswirkungen primär fakultäts-intern positive Effekte verursachen, folgen Aspekte, welche konkrete Auswirkungen auf die Befragten haben.<sup>52</sup> Es gaben 6 Fakultäten (19 %) an, mit der Anonymität für die Studierenden besonders zufrieden zu sein; 5 Fakultäten (16 %) waren mit der Flexibilität für die Studierenden besonders zufrieden. Unter der Flexibilität wurde die zeitliche und räumliche Freiheit der Studierenden verstanden, in der sie die Fragebögen online ausfüllen können. Vor dem Hintergrund des von Nippert (2003, S. 59) dargestellten Zusammenhangs, nach dem die für eine Erhebung so wichtige Rücklaufquote mit der Etablierung von Online-Verfahren schlagartig ansteigt, lässt sich vermuten, dass die Erkenntnis einer auf dieser Flexibilität aufbauenden ‚Win-Win-Situation‘ bei den Fakultäten Einzug hält: Mit der Etablierung von

<sup>52</sup> Zu berücksichtigen ist dabei, dass die Umfrage von Vertretern der Fakultät beantwortet wurde. Die Sicht der Studierenden wurde deshalb immer aus der Perspektive der Fakultätsvertreter wiedergegeben.



Online-Verfahren steigt die Flexibilität für die Studierenden, mit der Flexibilität steigt die Bereitschaft zur Beantwortung und damit die Rücklaufquote, die schließlich die Ergebnisse der Evaluation für die Fakultäten aufwertet. Eine analoge Sichtweise trifft für die Anonymität zu. Neben den bereits dargestellten gesetzlichen Vorgaben kann von einer ernsthaften Teilnahme der Studierenden an der Evaluation nur bei einer umfassenden und glaubhaft dargestellten Anonymität der Studierenden ausgegangen werden. Die Sicherstellung dieser Vorgabe liegt deshalb letztendlich auch wieder im Interesse der Fakultäten. Mit jeweils 2 Antworten (7 %) relativ selten genannt wurde die Zufriedenheit mit dem Datenexport, der einfachen Bedienbarkeit der Systeme und der Möglichkeit einer veranstaltungsbezogenen Evaluation.

Nur jeweils einmal genannt wurde die Zufriedenheit mit folgenden Punkten.

- Ubiquitäre Erreichbarkeit des Systems für die Dozenten
- Funktion, dozentenspezifische Fragen anzulegen und zu erheben
- Einschränkung der Evaluationsmöglichkeit auf berechnete Studierende
- Möglichkeit für die Studierenden, eigene Bewertungen mit dem Gesamtergebnis vergleichen zu können
- Flexibilität bei der Bestimmung der Erhebungseinheiten
- Zuverlässigkeit des Systems

**Allgemeine Aspekte** Die Abbildung 2.9 auf der nächsten Seite zeigt alle von den Befragten genannten Aspekte, die sich auf eine besondere Zufriedenheit mit dem bestehenden Evaluationsverfahren beziehen. Mit 7 Antworten (23 %) nannten knapp ein Viertel der befragten Fakultäten an erster Stelle die Bedeutung der Evaluationsergebnisse als Datenbasis für eine fundierte Ableitung von Konsequenzen aus der Evaluation. Während der Einsatz der Evaluationsergebnisse für ein individuelles Feedback an die Dozenten in der Literatur als zulässig angesehen wird, handelt es sich bei der Ableitung konkreter Konsequenzen um genau den Punkt, der in der Literatur am kontroversesten diskutiert wird: Können konkrete Handlungsvorgaben auf der Basis von Evaluationsergebnissen abgeleitet werden, die unter den Studierenden erhoben wurden und falls ja, mit der Abfrage welcher Merkmale lassen sich welche Aussagen valide fundieren? 6 Fakultäten (19 %) gaben an, mit der Funktion der Evaluation besonders zufriedenen zu sein, Rückmeldungen für die Dozenten zur Verfügung stellen zu können, über welche die Dozenten konkrete Hinweise über positive und negative Aspekte ihrer Lehre erhalten. Dabei wurde insbesondere die Bedeutung der Freitextantworten hervorgehoben. Nur 4 Fakultäten (13 %) sahen sich dazu veranlasst, eine besondere Zufriedenheit mit der Rücklaufquote zu äußern. Diese Tatsache deckt sich mit der Vermutung von Nippert (2003, S. 58 f.), dass „hinsichtlich der Evaluierungsbereitschaft auf Seiten der Studierenden vermutlich ein Problem existiert“, nachdem in seiner Erhebung 71 % der Befragten keine Angaben zu der Antwortbereitschaft der Studierenden machten. 3 Fakultäten (10 %) waren mit der Akzeptanz der Verfahren und 2 Fakultäten (7 %) mit dem Einsatz der

Evaluationsergebnisse als Grundlage für Besprechungen mit den Studierenden besonders zufrieden.

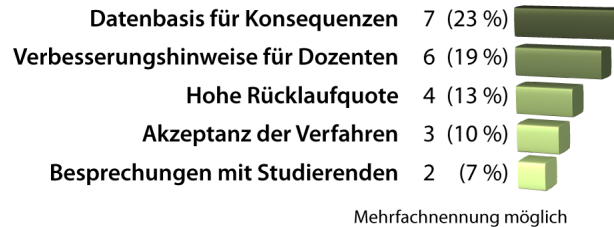


Abbildung 2.9: Gründe der Zufriedenheit (allgemeine Aspekte)

### 2.3.7 Rücklaufquote

Da die Zufriedenheit mit der Rücklaufquote wie im vorherigen Abschnitt gezeigt, im Verhältnis zu deren Bedeutung sehr gering ist, wurde den Mechanismen zur Erhöhung der Rücklaufquote eine eigene Frage gewidmet, deren Auswertung in diesem Abschnitt dargestellt wird. Die Abbildung 2.10 zeigt die Ergebnisse auf die Frage, ob die befragten Fakultäten die Studierenden derzeit zur Teilnahme an der Evaluation verpflichten. Nachdem manche Fakultäten die Konsequenzen in unterschiedlichen Bereichen unterschiedlich handhaben, war es ihnen möglich, mehrere Antworten zu geben.

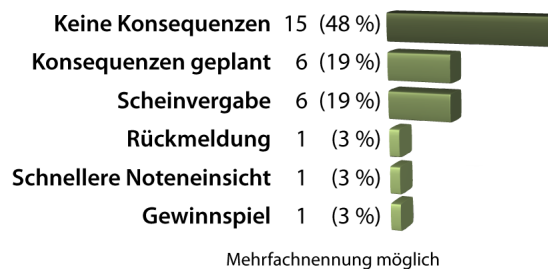


Abbildung 2.10: Konsequenzen aus der Teilnahme bzw. Nicht-Teilnahme an der Evaluation

Tatsächliche, negative Konsequenzen hat die Nicht-Teilnahme an der Evaluation derzeit (zumindest in Teilbereichen) bei insgesamt 8 der befragten Fakultäten (26 %). Wenn sie an der Evaluation nicht teilgenommen haben, erhalten Studierende an 6 Fakultäten (19 %) keinen Schein, an einer Fakultät (3 %) können sie sich nicht rückmelden und an einer weiteren erhalten sie in diesem Fall einen verzögerten Einblick in die Noten. Die Technische Universität München kehrt das Prinzip um und verlost unter allen an der Evaluation teilnehmenden

Studierenden als Anreizinstrument Büchergutscheine. Weitere<sup>53</sup> 6 Fakultäten (19 %) planen, Konsequenzen bei der Nicht-Teilnahme einzuführen.

### 2.3.8 Notwendige Verbesserungen

Bei der Frage nach den Aspekten, die einer Verbesserung bedürfen (vgl. Abbildung 2.11), gab mit 10 Antworten (32 %) knapp ein Drittel der Befragten an, dass die Evaluation stärker für die Ableitung konkreter Konsequenzen eingesetzt werden müsse. Ein Grundkonsens bei diesen Befragten herrschte darüber, dass die Sanktionen für besonders negativ bewertete Dozenten verstärkt werden müssen. Die konkrete Ausprägung solcher Sanktionen reichte dabei von einer Aussprache mit dem (Studien-)dekan zur Erarbeitung von Verbesserungsmaßnahmen bis zur Reduzierung der Mittel im Rahmen der LOM. Jedoch auch die Gratifikation besonders positiv bewerteter Dozenten sollte den Aussagen zufolge erhöht bzw. überhaupt eingeführt werden. Auch hierbei wurde primär die Berücksichtigung der Evaluationsergebnisse bei der LOM als probates Mittel angeführt.

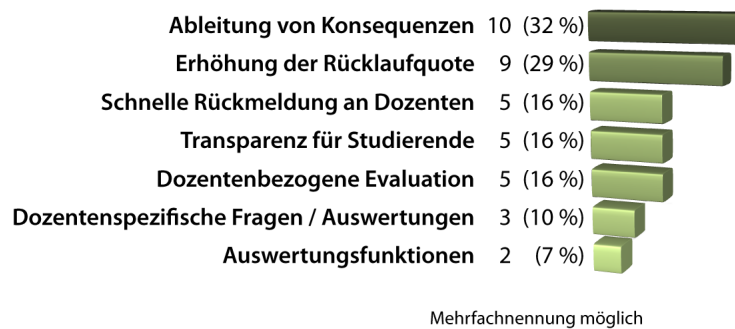


Abbildung 2.11: Notwendige Verbesserungen

Eine besonders interessante Form der Gratifikation wurde von Dr. Gustav Quade am Institut für Medizinische Statistik, Dokumentation und Datenverarbeitung der Rheinischen Friedrich-Wilhelm-Universität Bonn entwickelt. Bei diesem Verfahren, das in Kürze eingeführt werden soll, erhalten die Dozenten eine virtuelle Bezahlung durch die Studierenden. Dabei erhält ein Student ein virtuelles Budget von beispielsweise 50 Euro. Aus diesem Budget kann der Student Teilbeträge bis zu einem vorgegebenen Maximalwert (z. B. 20 Euro) an die einzelnen Dozenten vergeben – sie also faktisch bezahlen. Jeder Dozent erhält den ihm zugewiesenen Betrag am Ende einer Berechnungsperiode (z. B. ein Semester) als reales Budget um damit Leistungen für seine Forschungs- und Lehrtätigkeit (wie z. B. wissenschaftliche Projekte oder Kongressbesuche) zu bezahlen.

Als nahezu genauso wichtig wie die Ableitung von Konsequenzen wird die Erhöhung der Rücklaufquote erachtet. Wie in Abbildung 2.9 auf der vorherigen Seite dargestellt, sind nur

<sup>53</sup> Die beiden speziellen Teilmengen der befragten Fakultäten, die bereits Konsequenzen eingeführt haben und die, die Konsequenzen planen sind trotz der Option zur Mehrfachnennung disjunkt.

4 der Befragten (13 %) mit der Rücklaufquote zufrieden, gleichzeitig zeigt Abbildung 2.11 auf der vorherigen Seite, dass 9 Fakultäten (29 %) die Erhöhung der Rücklaufquote als notwendige Verbesserung ansehen. Die zur Umsetzung dieser Forderung angeführten Maßnahmen reichen von der an die Studierenden gerichtete Bitte zur verstärkten Teilnahme bis hin zu handfesten Sanktionen wie die Verweigerung der Scheinvergabe oder der Bekanntgabe der Noten bei einer Nichtteilnahme an der Evaluation. Wie die Ergebnisse auf den Seiten 36–37 gezeigt haben, werden solche Sanktionen bereits jetzt durchgeführt. Diese sind juristisch jedoch äußerst problematisch, weshalb die Aussagen der Befragten zu diesem Thema wohl so zu verstehen sind, dass solche Maßnahmen neben der grundsätzlichen Einführung auch eine juristische Absicherung finden sollen.

Für die oben geforderte Steigerung der Ableitung von konkreten Konsequenzen, die sich auf die Dozenten beziehen, ist eine dozentenbezogene Evaluation unabdingbar. Die Einführung der dozentenbezogenen Evaluation wird von 5 der Befragten (16 %) explizit gefordert. Von ebenso vielen Fakultäten gefordert wird eine schnellere Rückmeldung der Ergebnisse an die Dozenten und eine transparentere Rückmeldung der Ergebnisse an die Studierenden. 3 der Befragten (10 %) sahen eine notwendige Verbesserung darin, den Dozenten die Möglichkeit zu geben, individuelle Fragen einzurichten und auf sie zugeschnittene persönliche Auswertungen zu erhalten. 2 der Befragten (7 %) gaben weiterhin an, dass die Auswertungsfunktionen verbessert werden müssen. Jeweils einmal genannt wurden folgende Aspekte:

- Erhöhung der für die Evaluation zur Verfügung gestellten Ressourcen
- Verbesserung des Beratungs- und Weiterbildungsangebots für Lehrende
- Umstellung von der schriftlichen auf eine Online-Evaluation
- Integration der Entwicklung der Studienleistungen in die Auswertung (als Vergleich zwischen dem aktuellen und dem Vorjahr)
- Funktion für die Studierenden, einen teilweise beantworteten Fragebogen zwischenspeichern, um diesen später weiter ausfüllen zu können
- Senkung des Evaluationsvolumens
- Verbesserung der Vergleichbarkeit der Evaluationsergebnisse mit denen anderer Universitäten
- Zentralisierung der Evaluation innerhalb der Fakultät auf ein System
- Möglichkeit, veranstaltungsspezifische Erhebungen durchführen zu können
- Erhöhung der Rechtssicherheit (insbesondere beim Datenschutz) bei gleichzeitiger Praktikabilität
- Verbesserung der Stabilität des Evaluationssystems

- Verstärkte Auseinandersetzung der Dozenten mit den Evaluationsergebnissen
- Weiterentwicklung der Erhebungsinstrumente

Speziell bei der schriftlichen Evaluation wurde weiterhin der Umgang (Transport, Vervielfältigung und Entsorgung) mit den großen Papiermassen als verbesserungswürdig beschrieben.

### 2.3.9 Anforderungen an ein ideales Online-Evaluationssystem

Als konkreteste Grundlage für das folgende Kapitel der Ergebnisse wurden schließlich die Anforderungen an ein ideales Lehrevaluationssystem explizit erhoben. Während die vorangegangenen Fragen die allgemeinen Verfahren mit den unterschiedlichen Erhebungsformen berücksichtigten, war diese Frage rein auf ein Online-System ausgerichtet.<sup>54</sup> Abbildung 2.12 zeigt die Ergebnisse.

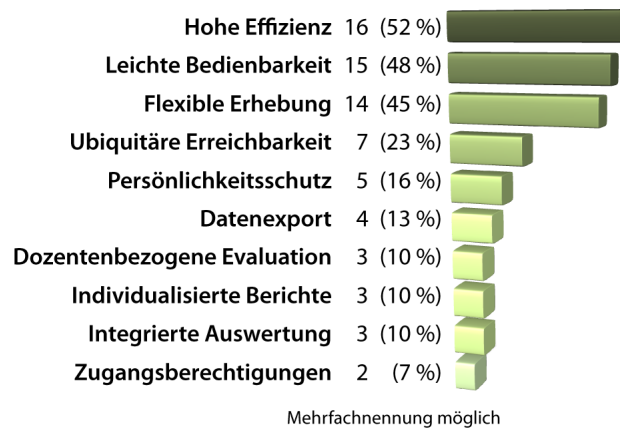


Abbildung 2.12: Anforderungen an ein ideales Online-Evaluationssystem

Primär wurden dabei die hohe Effizienz (16 bzw. 52 % der Nennungen), die leichte Bedienbarkeit (15 bzw. 48 % der Nennungen) und die flexible Erhebung (14 bzw. 45 % der Nennungen) angeführt. Die Effizienz wurde teilweise abstrakt genannt, jedoch auch wie folgt weiter konkretisiert. Im Vordergrund standen dabei die automatische Auswertung der Umfragen und die ebenfalls automatische Rückmeldung der Ergebnisse an die Dozenten. Weiterhin gefordert wurde in diesem Zusammenhang die automatische Benachrichtigung der Studierenden per E-Mail. Die auf die Effizienz ausgerichteten Anforderungen waren von dem Wunsch nach möglichst geringem administrativen Aufwand geprägt.

Im Gegensatz zu den anderen Aspekten wurde die Anforderung nach einer leichten Bedienbarkeit nur in wenigen Fällen genauer beschrieben. Als Grundkonsens lässt sich die

<sup>54</sup> Antworten, die sich nicht auf ein System bezogen, werden dementsprechend hier nicht dargestellt.

Anforderung dahingehend konkretisieren, dass alle Anwender – insbesondere auch Laien – das System mit geringem Aufwand und ohne Frustration bedienen können müssen.

Dicht gefolgt wurde an dritter Stelle eine hohe Flexibilität bei der Erhebung gefordert. Der Grundtenor lautete dabei, dass ein Online-System dieselbe Flexibilität bei der Erstellung der Fragebögen bieten muss, wie es bei der schriftlichen Erhebung der Fall ist. Mehr als die Hälfte der Stimmen für eine flexible Erhebung (8 von 14) forderten dabei insbesondere die Möglichkeit, Freitextantworten verwenden zu können, um die Stimmungen und Hinweise der Studierenden qualitativ aufnehmen zu können. Weiterhin gefordert wurde die Flexibilität bei der Bestimmung der Erhebungseinheiten im Sinne der Beschreibung unter Abschnitt „Merkmalsträger“ auf den Seiten 13–15 sowie die Möglichkeit, die Erhebungszeitpunkte flexibel bestimmen zu können und Umfragen direkt nach dem Ende einer Erhebungseinheit durchführen zu können.

Mit 7 Nennungen äußerten knapp ein Viertel aller Befragten (23 %) die Anforderung nach einer ubiquitären Erreichbarkeit des Systems. Das System soll demnach zeitlich und räumlich unbeschränkt zur Verfügung stehen. Dozenten und Studierende sollen von beliebigen Orten – also auch von zu Hause oder vom Urlaub aus – Zugriff auf das Online-System haben. Das System soll zudem zu jeder Zeit erreichbar sein und dabei auch in Stoßzeiten eine schnelle Reaktionszeit aufweisen. Insbesondere bei der Anzeige des Fragebogens soll es keine merkbare Verzögerung geben.<sup>55</sup>

Mit 16 % der Antworten stellten 5 der Befragten den Persönlichkeitsschutz als wichtige Anforderung heraus. Darunter wurde sowohl die Anonymität der Studierenden aber auch der Schutz der Dozenten vor einem ‚Evaluations-Mobbing‘ verstanden. So wurde gefordert, dass besonders schlecht bewertete Dozenten entweder nicht veröffentlicht werden sollen, oder sie vor einer Bekanntgabe zumindest die Möglichkeit haben sollen, sich in einer internen Aussprache zu verteidigen und bei der Entscheidung über die Veröffentlichung beteiligt zu werden. Mit 13 % forderten 4 Fakultäten, die erhobenen Daten in einem gängigen Austauschformat exportieren zu können, so dass sie in externen Statistik-Paketen umfassender verwertet werden können.

Von jeweils 3 Fakultäten (10 %) wurden die folgenden drei Anforderungen geäußert. Sie forderten zum einen die Möglichkeit, eine dozentenbezogene Evaluation durchführen zu können. Weiterhin forderten Sie, Berichte generieren zu können, die auf die jeweiligen Dozenten zugeschnitten und grafisch aufbereitet sind, so dass die Dozenten daraus schnell persönliche Informationen ziehen und die Berichte bei Bedarf auch direkt in den Veranstaltungen zur

---

<sup>55</sup> Da diese Anforderung nur schwer quantifizierbar ist, soll an dieser Stelle auf die Studie von JupiterResearch (2006, S. 5) aus dem E-Commerce-Bereich hingewiesen werden, welche die Abbruchquote beim Einkaufsverhalten im Internet analysiert. Diese Studie zeigt, dass 19 % der Benutzer mit Wählverbindung (schmalbandige Verbindung) und 33 % der Benutzer mit Breitbandanbindung nicht bereit sind, länger als 4 Sekunden auf die Darstellung der Web-Seite zu warten. Insgesamt 32 % (schmalbandige Anbindung) bzw. 63 % (breitbandige Anbindung) der Benutzer sind nicht bereit, länger als 6 Sekunden zu warten. Auch wenn eine direkte Übertragung auf die Online-Evaluation nicht belegt ist, so lassen diese Zahlen dennoch auf ein grundlegendes Benutzerverhalten schließen, nachdem die Wahrscheinlichkeit groß ist, dass ein erheblicher Anteil der Population bereits nach rund 5 Sekunden Wartezeit nicht mehr an der Erhebung teilnehmen wird.

Besprechung mit den Studierenden präsentieren können. Zudem gefordert wurde die Möglichkeit, eine grundlegende statistische Auswertung auch innerhalb des Systems durchführen zu können.

2 Nennungen (7 %) erhielt die Forderung nach der Sicherstellung, dass nur zur Evaluation berechnigte Studierende genau einmal an der Erhebung teilnehmen dürfen. Konkret wurde dabei beschrieben, dass das System Fragebögen nur an die Studierenden verschicken darf, die tatsächlich an der Veranstaltung teilgenommen haben und mithilfe von Transaktionsnummern (TANs) sichergestellt werden soll, dass bei Gewährleistung der Anonymität eine zur Evaluation berechnigte Person auch nur genau einen Fragebogen zurücksenden darf.

Jeweils einmal genannt wurden schließlich folgende Anforderungen:

- Darstellung eines Rankings (inklusive einer Entwicklung der Rankings über die Zeit hinweg)
- Funktion, einen Fragebogen teilausgefüllt zwischenspeichern zu können
- Sichere Übermittlung der Evaluationsergebnisse
- Rückmeldung über die Teilnahme eines Studierenden an einer bestimmten Evaluation
- Möglichkeit, formative Evaluationen durchzuführen
- Statistische Korrektur von Störvariablen zur Vermeidung „einseitiger Urteile der Studenten“
- Durchführung von TED-Umfragen
- Darstellung des Stands der Evaluation in zentralen Übersichten
- Mehrsprachige Fragebögen (insbesondere in englischer Sprache)





## 3 Ergebnisse

Das vorliegende dritte Kapitel verdichtet die im zweiten Kapitel hergeleiteten Anforderungen<sup>56</sup> an ein Evaluationssystem und zeigt die Umsetzung ausgewählter Anforderungen anhand eines im Rahmen dieser Arbeit entwickelten Evaluationssystems, einschließlich dessen Piloteinsatz an der Medizinischen Fakultät der Ludwig-Maximilians-Universität München.

### 3.1 Technische Anforderungen

Dieser Abschnitt fasst die Anforderungen zusammen, die aus der Literatur und der Evaluationspraxis der medizinischen Fakultäten in Deutschland hergeleitet wurden und beleuchtet sie unter den technischen Aspekten der Umsetzung zu einem Evaluationssystem. Während dieser Abschnitt den gesamten Rahmen darstellt, in welchem sich ein anforderungs- und praxisgerechtes Evaluationssystem bewegen sollte, werden ausgewählte Konzepte unter Abschnitt „Implementation“ auf den Seiten 58–80 anhand des konkreten Evaluationssystems veranschaulicht.

#### 3.1.1 Rahmenbedingungen

##### Leistungsfähigkeit

Grundlegendes Element der Anforderungsanalyse ist die Bestimmung der Leistungsparameter, die das Zielsystem erfüllen muss. Da diese Rahmenbedingungen die Systemarchitektur maßgeblich beeinflussen, wird ihre Darstellung allen anderen Anforderungen vorangestellt. Zur Bestimmung der benötigten Leistungsfähigkeit sind in erster Linie die Anzahl der Anfragen an das System sowie die zu speichernde Datenmenge von primärer Relevanz. Aus diesen Variablen lassen sich insbesondere die Anforderungen an die Leistungsfähigkeit der Hardware und der Datenbank ableiten, die von Einsatzbereich zu Einsatzbereich (z. B. Fakultät, Universität etc.) verschieden sind und deshalb nicht als Konstanten dargestellt werden können.

**Antwortverhalten** Da das Evaluationssystem wie unter Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27 beschrieben, tendenziell als Zentralsystem für eine ganze Fakultät oder gar eine ganze Universität eingesetzt wird, muss das gesamte System – hier verstanden als Hard- und Software – in der Lage sein, eine große Menge an Anfragen zu

---

<sup>56</sup> Der Ausrichtung dieser Arbeit entsprechend werden die Anforderungen möglichst allgemein verständlich dargestellt. Für Ausführungen über die Grundlagen der Anforderungsanalyse sei auf Jacobson et al. (1999, S. 111–172), Robertson & Robertson (2006) und Schienmann (2002) verwiesen.

verarbeiten. Da sich eine Streuung der Evaluationszeitpunkte wie unter Abschnitt „Evaluationszeitpunkte“ auf Seite 18 beschrieben kaum realisieren lässt, muss das System zudem in der Lage sein, große Anfragespitzen verarbeiten zu können ohne dabei die unter Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf Seite 40 dargestellten maximalen Latenzzeiten zu überschreiten. Die mit Abstand größte Komponente der Anfragen sind die Aufrufe der Fragebögen, deren Anzahl sich für den Fall der dozentenbezogenen Evaluation wie folgt berechnen lässt.

$$Fb|_{t_0}^{t_1} = \underbrace{\sum_{i=1}^n D_i S_i}_{\text{Lehrveranstaltungen}} + \underbrace{\sum_{j=1}^m S_j}_{\text{sonstige Lehreinheiten}} \quad (3.1)$$

Die Anzahl aller Fragebögen  $Fb$  besteht aus der Summe der Fragebögen für die Lehrveranstaltungen und aller sonstigen Lehreinheiten (siehe Abschnitt „Lehreinheiten“ auf Seite 14) wie beispielsweise Semesterevaluationen, die im gegebenen Zeitraum  $[t_0; t_1]$  abgeschlossen werden.<sup>57</sup> Für die Lehrveranstaltungen wird das Produkt aus der Anzahl der zu evaluierenden Dozenten  $D$  und der Anzahl der Studierenden  $S$  jeweils bezogen auf die  $i$ -te Veranstaltung für alle  $n$  Veranstaltungen des Zeitraums aufsummiert. Für die sonstigen Lehreinheiten wird die Anzahl der Studierenden  $S$  jeweils bezogen auf die  $j$ -te Veranstaltung für alle  $m$  Lehreinheiten des Zeitraums aufsummiert.<sup>58</sup>

Unter der Vernachlässigung zahlenmäßig weitaus geringerer Bereiche aus denen Anfragen zu erwarten sind – wie beispielsweise für die Erstellung von Berichten, lässt sich die maximale Anzahl der Anfragen  $Anf_{\max}$  für einen gegebenen Zeitraum  $[t_0; t_1]$  mit der Anzahl der in diesem Zeitraum erstellten Fragebögen approximieren.

$$Anf|_{t_0}^{t_1} \approx Fb|_{t_0}^{t_1} \quad (3.3)$$

Es muss auch deshalb von einer Approximation gesprochen werden, da der mehrfache Aufruf eines Fragebogens (beispielsweise bei einer Teilbeantwortung) nicht berücksichtigt wird.

**Leistungsfähigkeit der Datenbank** Aus der Berechnung der Anfragen lassen sich nicht nur die Anforderungen an die Reaktionsgeschwindigkeit des Systems ableiten. Unter der Vorgabe, individuelle Berichte mit variablen Auswertungen erstellen zu können, für welche

<sup>57</sup> Der Zeitraum wird hier als Intervall (und nicht nur als  $\delta_t$ ) angegeben, da bei der Evaluation nicht nur die Dauer sondern auch die Start- und End-Zeitpunkte  $t_0$  und  $t_1$  für die Berechnung relevant sind.

<sup>58</sup> Im Fall der veranstaltungsbezogenen Evaluation wird von einem Dozenten ausgegangen. Da in diesem Fall  $D_i = 1$  und die Lehrveranstaltung eine konkrete Ausprägung einer abstrakten Lehreinheit ist, vereinfacht sich die Berechnung wie folgt.

$$Fb|_{t_0}^{t_1} = \sum_{j=1}^m S_j \quad (3.2)$$

der Zugriff auf die Rohdaten benötigt wird, lässt sich auch die Menge der in der Datenbank zu speichernden Daten berechnen. Zu beachten ist dabei, dass diese Menge der Datensätze negativ mit der Reaktionsgeschwindigkeit korreliert.<sup>59</sup> Hierbei soll angenommen werden, dass die Antworten auf die Fragebögen sowie die Stammdaten im Referenzzeitraum den mit Abstand größten Teil zu der zu speichernden Daten ausmachen. Weiterhin soll ein Datenbank-Modell angenommen werden, das für jede Antwort einen eigenen Datensatz anlegt. Die maximale Anzahl der zu speichernden Datensätze  $Ds_{\max}^{t_0; t_1}$  für den gegebenen Zeitraum  $[t_0; t_1]$  kann deshalb als Produkt aus der unter den Gleichungen 3.1 bis 3.2 auf der vorherigen Seite berechneten Anzahl der Fragebögen  $Fb$  und der erwarteten<sup>60</sup> mittleren Anzahl der pro Fragebogen verwendeten Fragen  $\bar{F}$  zuzüglich der Anzahl der Datensätze für die Stammdaten  $Sd$  wie folgt approximiert werden.

$$Ds_{\max}^{t_0; t_1} \approx \underbrace{Fb}_{\text{Antworten}}^{t_0; t_1} \bar{F} + \underbrace{Sd}_{\text{Stammdaten}}^{t_0; t_1} \quad (3.4)$$

Um dem Ziel zu entsprechen, das Maximum zu berechnen, geht diese Gleichung von dem aus der Sicht der Datenhaltung ungünstigen Fall aus, dass die Stammdaten entgegen der Empfehlung unter Abschnitt „Stammdatenbezug“ auf Seite 48 nicht fremdbezogen werden können und weiterhin die unter Abschnitt „Zeitpunkte“ auf Seite 53 beschriebenen Längsschnittanalysen durchführbar sein sollen.<sup>61</sup>

Die Länge des Zeitraums ist so zu wählen, dass er alle gewünschten Auswertung der rückwirkenden Darstellung von Parameter-Entwicklungen erlaubt; die Spannweite  $t_1 - t_0$  wird somit sicherlich mehrere Jahre betragen. Um eine Approximation handelt es sich in diesem Fall, da die Anzahl der Fragen pro Fragebogen als Erwartungswert angegeben wird und

<sup>59</sup> Dies gilt insbesondere bei der geforderten Möglichkeit, die Beantwortung eines teilausgefüllten Fragebogens wieder aufnehmen zu können. Im Gegensatz zur Anzeige eines noch nicht ausgefüllten Fragebogens muss das System in diesem Fall die bereits gegebenen Antworten aus allen vorhandenen herausfiltern.

<sup>60</sup> Unter der Annahme, dass das System dozenten-spezifische Fragen zulässt, kann die Anzahl der zu erwartenden Fragen nicht ex ante bestimmt werden. Da diese Größe jedoch für die Kapazitätsplanung eben im Vorhinein benötigt wird, soll mit  $\bar{F}$  der erwartete Wert verwendet werden.

<sup>61</sup> Können die Stammdaten dagegen vollständig fremdreferenziert werden, so ist  $Sd_{t_0}^{t_0; t_1} = 0$  weshalb die Anzahl der zu speichernden Datensätze unabhängig von der Möglichkeit zur Durchführung von Längsschnittanalysen wie folgt eingeschränkt werden kann:

$$Ds_{\max}^{t_0; t_1} \approx Fb_{t_0}^{t_0; t_1} \bar{F} \quad (3.5)$$

Wenn die Stammdaten nicht fremdbezogen werden können, jedoch die Anforderung nach Längsschnittanalysen entfallen kann, so würde sich die Anzahl der zu speichernden Datensätze unter der Annahme diskreter Zeiträume zumindest wie folgt reduzieren:

$$Ds_{\max}^{t_0; t_1} \approx Fb_{t_0}^{t_0; t_1} \bar{F} + Sd_{t_1}^{t_0; t_1} \text{Stud} + Sd_{t_0}^{t_0; t_1} \text{Sonst} \quad (3.6)$$

Da in diesem Fall die für Längsschnittanalysen notwendige Referenzierung der Studierenden über die Perioden hinweg entfällt, müssten die Stammdaten der Studierenden  $Sd_{\text{stud}}$  nur noch für die aktuelle Periode  $[t_1; t_1]$  gespeichert werden. Es verblieben damit die Datensätze für alle anderen Stammdaten  $Sd_{\text{sonst}}$ , die für den gesamten Zeitraum  $[t_0; t_1]$  zu speichern sind.

Datensätze für weitere Informationen (insbesondere für die Abbildung der „n:m“-Relationen) nicht berücksichtigt werden.

### **Effizienz, Benutzerfreundlichkeit und Flexibilität**

Hohe Effizienz, gute Bedienbarkeit und große Flexibilität waren in der empirischen Studie (siehe Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41) die wichtigsten Anforderungen an ein Evaluationssystem. Unter einer hohen Effizienz wird hier die Zielerreichung bei gegebener Qualität und minimalem Aufwand verstanden. Unter der Prämisse, die Anschaffungskosten nicht weiter zu berücksichtigen, lässt sich der Aufwand auf den personellen Aufwand für den Betrieb und die Pflege des Systems einschränken. Der Ansatz lautet dabei: je umfassender sich die primären Funktionen des Systems – also die Einrichtung und Pflege der Fragen, die Durchführung der Erhebung und die Auswertung der Ergebnisse – automatisieren lassen, desto geringer ist ceteris paribus der personelle Aufwand und damit der Effizienzgrad. Neben der expliziten Forderung nach der guten Bedienbarkeit im Rahmen der Studie impliziert die Forderung nach einer hohen Effizienz selbst ein hohes Maß an Benutzerfreundlichkeit, die nicht nur ein Lippenbekenntnis sein darf. Vielmehr muss ein Evaluationssystem so ausgelegt sein, dass alle Benutzer ihre Aufgaben darin mit minimalem Aufwand durchführen können. Weiterhin muss es so gestaltet sein, dass die Benutzer eine intrinsisch motivierte Bereitschaft zeigen, überhaupt mit dem System zu arbeiten. Die Forderung nach einer großen Flexibilität wurde im Rahmen der Studie so beschrieben, dass das Evaluationssystem bei der Online-Erhebung dieselben Möglichkeiten bieten muss, wie bei der schriftlichen Erhebung.

Die Umsetzung dieses abstrakten Verständnisses von Effizienz und Flexibilität wird in den folgenden Punkten weiter konkretisiert. Für die Ausprägung einer benutzerfreundlichen Schnittstelle wird auf die Bildschirmfotos in den Abbildungen 3.3 bis 3.19 auf den Seiten 60–79 unter Abschnitt 3.2 verwiesen.

### **Ubiquitäre Erreichbarkeit**

Wie mehrfach hergeleitet, soll die Evaluation online durchgeführt werden (siehe Abschnitt „Online-Erhebung“ auf den Seiten 19–20, Abschnitt „Erhebungsform“ auf Seite 28, Abschnitt „Zufriedenheit“ auf den Seiten 33–36, Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41). Um eine breite Zielgruppe ansprechen zu können, sollten die Anforderungen an die technischen Voraussetzungen bei den Benutzern dafür minimal sein. Dies bedeutet, dass ein Benutzer das Evaluationssystem auch mit einem mehrere Jahre alten Computer ohne Probleme benutzen können sollte, ohne dazu Anpassungen wie z. B. die Installation von zusätzlicher Software an seinem System vornehmen zu müssen. Die Benutzung eines Evaluationssystems erfolgt unter diesen Aspekten bevorzugt über einen Web-Browser. Hinweise und Erinnerungen sollten per E-Mail verschickt werden, wobei bei der technischen Gestaltung der E-Mails darauf zu achten ist, dass die E-Mail-Programme der Empfänger diese nicht als

unerwünschte Nachrichten (sog. „Spam-Mails“) klassifizieren und sie den Empfänger damit nie erreichen. Andernfalls könnte dies zu einem Non-Responder-Bias führen.

Ein besonderer Aspekt ergibt sich bei der Berücksichtigung des Effizienzkriteriums. Im Gegensatz zur Verwendung anderer Erhebungsmedien wie insbesondere bei der schriftlichen Erhebung, ist bei der Implementierung der Online-Erhebung ein schneller Zugang zum Fragebogen sicherzustellen. Bei der schriftlichen Erhebung obliegt es den Veranstaltern der Evaluation, den Fragebogen zur Verfügung zu stellen. Wurde er einmal verteilt, kann ein papierbasierter Fragebogen sofort ausgefüllt werden. Im Gegensatz dazu gelangen die Befragten bei der Online-Evaluation von der Benachrichtigung in der Regel nur über Umwege zum eigentlichen Fragebogen. Dieser Weg muss so einfach und schnell wie möglich gestaltet werden, so dass der Aufwand nicht wesentlich größer ist, als bei schriftlichen Erhebungen. Ein solch minimaler Aufwand bedeutet in diesem Zusammenhang, eine minimale Interaktion mit dem Computer (beispielsweise ein einziger Maus-Klick) und die Einhaltung einer aus der Sicht der Befragten maximal akzeptierten Antwortzeit des Evaluationssystems bei der Darstellung des Fragebogens. Aus den Erkenntnissen unter Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf Seite 40 lässt sich ableiten, dass ein Evaluationssystem den gewünschten Fragebogen bei der Online-Erhebung binnen weniger Sekunden zur Verfügung stellen muss. Dies gilt unabhängig davon, ob es sich um einen neuen oder bereits teilausgefüllten Fragebogen handelt.

### 3.1.2 Stammdatenbezug

Wie unter Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27 beschrieben, steht ein zentrales Evaluationssystem vor der Herausforderung, alle benötigten Daten aus unterschiedlichsten Bereichen und Systemen abrufen zu können, zu denen insbesondere die folgenden zählen:

- Personenbezogene Daten  
(Studierende, ehemalige Studierende und Dozenten)
- Veranstaltungsdaten  
(Bei der dozentenbezogenen Evaluation werden zudem auch die Daten der einzelnen Termine benötigt)
- Raumdaten  
(Falls diese auch evaluiert werden sollen)
- Zuweisungsdaten  
(Informationen darüber, welcher Dozent welche Termine welcher Veranstaltungen vor welchen Studierenden – ggf. in welchem Raum – tatsächlich gehalten hat)

Eine besondere Berücksichtigung benötigt die Erfassung der unter Abschnitt „Störvariablen“ auf den Seiten 16–17 beschriebenen impliziten Störvariablen. Während explizite Störvariablen durch eigene Fragen über den Fragebogen direkt erhoben werden, müssen die Ausprägungen

impliziter Störvariablen aus unterschiedlichen externen Systemen bezogen werden. Wo immer möglich, ist zur Vermeidung von Ermüdungseffekten durch zu viele Fragen die implizite der expliziten Erfassung vorzuziehen.

In der Praxis gestaltet es sich als äußerst schwierig, die oben aufgeführten Daten aufgrund datenschutzrechtlicher Fragen, wegen fehlender elektronischer Erfassung und mangels Schnittstellen in den Fremdsystemen für den Austausch der (benötigten) Daten in einem zentralen Evaluationssystem verwenden zu können. Können diese Hürden jedoch überwunden werden, so muss ein zentrales Evaluationssystem bei der Datenabfrage so flexibel wie möglich gestaltet werden, um sicherzustellen, dass die vorhandenen Informationen auch tatsächlich abgerufen werden können.<sup>62</sup> Dabei stehen insbesondere folgende Optionen zur Verfügung, die in der hier dargestellten Reihenfolge zu bevorzugen sind.

1. Fremdzugriff
  - a) Verzeichnisdienste<sup>63</sup>
  - b) Web-Dienste<sup>64</sup>
  - c) Externer Datenbankzugriff über logische Relationen<sup>65</sup>
2. Datenübernahme
  - a) Datenimport
  - b) Manuelle Dateneingabe (impliziert einen Medienbruch)

Um Redundanzen sowie Synchronisationsprobleme zwischen bestehenden und neuen Daten zu vermeiden, um den administrativen Aufwand zu reduzieren und um eine maximale Aktualität sicherzustellen, ist der Fremdzugriff gegenüber der Datenübernahme deutlich zu bevorzugen.

### 3.1.3 Fragen

#### Fragenarten

Wie von der Praxis gefordert (siehe Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41), muss ein Evaluationssystem beim Anlegen der Fragen dieselben Möglichkeiten bieten, wie bei der schriftlichen Befragung. Das Evaluationssystem muss damit alle Fragenarten zur Verfügung stellen, die unter Abschnitt „Fragearten“ auf Seite 15 beschrieben wurden. Dazu zählen offene (Freitextantworten) und geschlossene Fragen (Multiple-Choice- und Single-Choice-Fragen – insbesondere auch bipolare Single-Choice-Fragen). Nachdem es

---

<sup>62</sup> Werden benötigte Daten jedoch erst gar nicht erfasst und ist eine solche Erfassung auch nicht abzusehen, muss geprüft werden, ob es sinnvoll ist, die Erfassung durch das Evaluationssystem vorzunehmen (siehe Abschnitt „Lehrzettel“ auf den Seiten 75–76).

<sup>63</sup> derzeit vor allem auf dem LDAP-Standard (Lightweight Directory Access Protocol) aufbauend

<sup>64</sup> engl.: „Web Services“

<sup>65</sup> engl.: „Views“

in der Literatur unterschiedliche Empfehlungen gibt,<sup>66</sup> muss die Anzahl der Antwortmöglichkeiten sowie deren Formulierung bei geschlossenen Fragen frei definierbar sein. Weiterhin sollte das Evaluationssystem bei der Erhebung automatisch zu jeder Frage eine sog. „Item-Non-Response“-Option anzeigen, durch deren Auswahl der Befragte deutlich machen kann, dazu keine Aussage machen zu wollen oder zu können. Wie in der Literatur gefordert (siehe Abschnitt „Fragearten“ auf den Seiten 15–16), sollte ein Evaluationssystem zu jeder Frage neben der Ausprägung auch die<sup>67</sup> persönliche Gewichtung einer Frage erheben können, um diese Informationen bei der Verdichtung zu Kennzahlen berücksichtigen zu können.

Um statistisch korrekte Auswertungen sicherzustellen, sollte ein Evaluationssystem weiterhin bereits beim Anlegen der Fragen das betreffende Skalenniveau vom Benutzer abfragen und damit von vorne herein festlegen, welche statistischen Instrumente bei der Auswertung zulässig sind. So kann das Evaluationssystem bei der Auswertung verhindern, dass die Ergebnisse geschlossener Fragen mit unzulässigen Instrumenten verdichtet werden.

### **Dozentenspezifische Fragen**

Um wie sowohl methodisch (siehe Abschnitt „Dozenten“ auf den Seiten 11–12 und Abschnitt „Evaluationsverantwortliche“ auf Seite 12) wie auch von der Praxis (siehe Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41) gefordert, die Dozenten besser an die Evaluation zu binden und ihnen die Möglichkeit zu geben, Informationen über ganz individuelle Aspekte ihrer Lehre zu erhalten, soll ein Evaluationssystem in der Lage sein, dozentenspezifische Fragen aufnehmen zu können. Diese Fragen sollen nur auf Fragebögen erscheinen, die sich auf Veranstaltungen der entsprechenden Dozenten beziehen.

#### **3.1.4 Untersuchungseinheiten**

##### **Dozentenbezogene Evaluation**

Nachdem neben den dozentenbezogenen Fragen im Speziellen auch eine dozentenbezogene Evaluation im Allgemeinen sowohl in der Literatur (siehe Abschnitt „Merkmalsträger“ auf den Seiten 13–15) wie auch in der Praxis (siehe Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41) gefordert wird, soll ein Evaluationssystem in der Lage sein, die Erhebung auch direkt auf Dozenten beziehen zu können. Diese Funktionalität ist ganz allgemein dann

---

<sup>66</sup> Besonders bei bipolaren Fragen wird ganz grundlegend zwischen Empfehlungen zu einer geraden oder einer ungeraden Anzahl an Ausprägungen unterschieden. Eine ungerade Anzahl an Ausprägungen erlaubt es den Befragten, sich exakt in der Mitte festzulegen, während eine gerade Anzahl die Befragten dazu zwingt, sich für eine Tendenz zu entscheiden (vgl. Mandl et al. 1999, S. 29).

<sup>67</sup> Theoretisch ließen sich mehrere Gewichtungen für unterschiedliche Ziele erfassen. Da die studentische Evaluation nur für die Erfassung der studentischen Zufriedenheit weitgehend anerkannt ist und die Erhebung mehrerer Gewichtungen kaum praktikabel ist, beschränkt sich die Anforderung in diesem Rahmen auf eine einzige Gewichtung verstanden als Relevanz der Frage für die Zufriedenheit.

notwendig, wenn an einer beliebigen Stelle im Evaluationsprozess Aussagen über einen Dozenten getroffen werden sollen. Im Speziellen trifft dies auf die leistungsorientierte Mittelvergabe, auf Lehrpreise und auf Dozenten-Rankings zu.

Dieser grundsätzlich leicht verständliche Zusammenhang gestaltet sich bei der Konzeption eines Evaluationssystems jedoch als durchaus anspruchsvoll, da ein gängiges Prinzip der Evaluationspraxis – nämlich die Gleichsetzung einer Veranstaltung mit einem Dozenten – damit nur noch dann zulässig ist, wenn jeder Termin einer Veranstaltung von genau ein und demselben Dozenten gehalten wird. Speziell im medizinischen Curriculum muss diese Annahme jedoch verworfen werden (siehe Abschnitt „Fragmentierte Veranstaltungen“ auf den Seiten 25–26 und Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27). Das Evaluationssystem muss damit so ausgelegt sein, dass ein Fragebogen nicht nur Veranstaltungen, sondern auch die einzelnen Termine einer Veranstaltung referenzieren kann. Wurde ein Termin von mehreren Dozenten gehalten, so muss das Evaluationssystem die Erhebung weiterhin auch auf jeden dieser Dozenten einzeln beziehen können.

Hinzukommt, dass die für einen Termin geplanten Dozenten – wie unter Abschnitt „Fragmentierte Veranstaltungen“ auf den Seiten 25–26 beschrieben – häufig nicht denen entsprechen, die ihn tatsächlich gehalten haben. Das Evaluationssystem muss deshalb weiterhin die Möglichkeit bieten, zum Zeitpunkt der Erhebung auf Daten zugreifen zu können, die den tatsächlichen Stand widerspiegeln.

Um zu verhindern, dass die Evaluation eines jeden einzelnen Termins zu einer Überflutung mit Fragebögen und damit zu einer schnellen Evaluationsmüdigkeit bei den Befragten führt, sollte das Evaluationssystem alle Termine einer Veranstaltung zusammenfassen,<sup>68</sup> die sich auf denselben Dozenten beziehen. Letztendlich geht es ja nicht darum, jeden einzelnen Termin zu evaluieren, sondern den Dozenten. In jedem Fall soll ein Fragebogen eindeutig darüber informieren, auf welche Termine welcher Veranstaltung bei welchem Dozenten er sich bezieht.

#### **Evaluation beliebiger Lehreinheiten**

Der Bezug der Evaluation auf Dozenten alleine reicht jedoch nicht aus. Ein Evaluationssystem soll zudem in der Lage sein, Erhebungen auf ganze Lehreinheiten wie Veranstaltungen, Module oder Semester beziehen zu können. Konkret muss ein Evaluationssystem dabei die Funktion anbieten, bestimmte Fragen nur für bestimmte Lehreinheiten zu erheben, da nicht alle Fragen für alle Lehreinheiten zutreffen. Nachdem sich der Aufbau des Curriculums von Fakultät zu Fakultät unterscheiden kann, muss ein Evaluationssystem bei der Berücksichtigung der Lehreinheiten umfassende Freiheiten bieten. Es soll Erhebungen auf unterschiedliche Lehrformen beziehen können und dabei sowohl innerhalb wie auch zwischen den Universitäten vergleichbare Ergebnisse liefern können. Die Menge der unterschiedlichen Lehrformen wird sich nie endgültig fixieren lassen, weshalb ein Evaluationssystem ständig auf neue Lehrformen

---

<sup>68</sup> Bei der formativen Evaluation werden mehrere Erhebungen parallel zum Veranstaltungsverlauf vorgenommen. Die Bezugsgröße ist in diesem Fall also nicht die ganze Veranstaltung selbst, sondern die einzelnen Segmente (siehe Abschnitt „Zeitpunkte“ auf Seite 53), in welche die Veranstaltung durch die mehrfache Erhebung unterteilt wurde.



adaptierbar sein muss. Idealerweise soll dies zur Laufzeit innerhalb des Systems und ohne zusätzlichen Erweiterungsaufwand realisierbar sein.

Letztendlich muss es mit dem Evaluationssystem also möglich sein, unterschiedliche Fragebögen zu erhalten, die sich auf unterschiedliche Lehreinheiten beziehen. Ferner soll es möglich sein, die Auswertung der Evaluationsergebnisse so anzupassen, dass sich unterschiedliche Lehreinheiten (siehe Abschnitt „Merkmalsträger“ auf den Seiten 13–15), Lehrformen (siehe Abschnitt „Neue Lehrformen“ auf den Seiten 24–25) oder ganze Studiengänge (vgl. Abschnitt „Modellstudiengänge“ auf Seite 25) gegenüberstellen lassen. Wie unter Abschnitt „E-Learning“ auf Seite 33 gezeigt, ist das Angebot von E-Learning-Kursen in der Praxis weit verbreitet und muss deshalb ebenfalls als Ausprägung einer Erhebungseinheit berücksichtigt werden können.

### **3.1.5 Erhebung**

#### **Zielgruppen**

Wie unter Abschnitt „Rechtsnormen“ auf den Seiten 9–10 und Abschnitt „Studierende, Absolventen und Studienabbrecher“ auf den Seiten 10–11 hergeleitet, sollen Studierende in die Evaluation einbezogen werden. Ein Evaluationssystem soll also Erhebungen unter den Studierenden durchführen können. Wie unter Abschnitt „Studierende, Absolventen und Studienabbrecher“ auf den Seiten 10–11 zudem beschrieben, soll ein Evaluationssystem weiterhin in der Lage sein, Umfragen auch unter Absolventen und Studienabbrechern durchführen zu können. Kann das Evaluationssystem die Informationen über diese Zielgruppen nicht von einem Zentralsystem beziehen, muss es die Daten der Studierenden über einen langen Zeitraum speichern und Absolventen bzw. Studienabbrecher über den Immatrikulationsstatus von einem Zentralsystem identifizieren.

#### **Population**

Nachdem insbesondere bei der Kombination von dozentenbezogener und formativer Evaluation in fragmentierten Veranstaltungen des medizinischen Curriculums eine hohe Anzahl an auszufüllenden Fragebögen zu erwarten ist, soll ein Evaluationssystem – wie unter Abschnitt „Population“ auf den Seiten 18–19 beschrieben – sowohl eine Vollerhebung wie auch eine Erhebungen in Stichproben durchführen können. Die einfachste Methode, um eine Stichprobe im Rahmen eines elektronischen Systems zu erhalten, ist die Listenauswahl als Ausprägung einer einfachen Zufallsstichprobe. Bei diesem Verfahren, wählt das System die Elemente zufällig aus der Grundgesamtheit aus. Dabei muss sichergestellt werden, dass jedes Element mit einer Wahrscheinlichkeit größer Null in die Stichprobe eingehen kann. Um potentielle Störvariablen zu berücksichtigen, die bei einer einfachen Zufallsstichprobe zu einer Verzerrung der Repräsentativität führen können, soll das System zudem auch eine Quotenauswahl mit möglichst frei definierbarer Angabe der zu berücksichtigen Quotenparameter durchführen können.

## **Anonymität**

Um die unter Abschnitt „Rechtsnormen“ auf den Seiten 9–10 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41 geforderte Anonymität zu gewährleisten, muss ein Evaluationssystem die Antworten auf einen Fragebogen grundsätzlich ohne Bezug zu der beantwortenden Person speichern. Soll das System jedoch auch Längsschnittanalysen durchführen, so muss es bei der Beantwortung eines Fragebogens auch eine Relation zu der befragten Person herstellen. Um die Prämissen des Datenschutzes dabei nicht zu verletzen, muss das Evaluationssystem die Auflösung dieser Relation von außen unterbinden. Es muss also sichergestellt werden, dass die Relation der Antworten zu den beantwortenden Personen tatsächlich nur für die Zuordnung der Antworten über den Zeitraum der Längsschnittanalyse hinweg systemintern aufgelöst wird und die personenbezogenen Daten dabei nicht preisgegeben werden. Wie bei der Darstellung der Rechtsnormen unter Abschnitt „Rechtsnormen“ auf Seite 10 bereits erwähnt, muss das Evaluationssystem dabei die Zustimmung der Betroffenen zur Speicherung der Daten einholen und genaue Auskunft über die Form der Nutzung und Verarbeitung geben. Da das System persönliche Daten speichert, muss es neben der strukturellen Anforderung an das Datenbank-Schema als Ganzes umfassend vor unbefugten Zugriffen geschützt werden. Dazu zählen sowohl die Zugriffe per Software wie auch der physische Zugriff auf die Geräte.

## **Publikation des Fragebogens**

Für den Zugang zu einem Fragebogen kann ein System entweder der Person die Möglichkeit zur Verfügung stellen, die zu evaluierenden Fragebögen über einen Benutzerzugang zentral anzuzeigen (passives Verfahren)<sup>69</sup>. Alternativ dazu kann das Evaluationssystem eine Nachricht an den Empfänger verschicken, die ihn dazu auffordert, einen Fragebogen auszufüllen (aktives Verfahren). Kann die unter Abschnitt „Ubiquitäre Erreichbarkeit“ auf den Seiten 46–47 beschriebene Gefahr des Untergangs von E-Mails minimiert werden, so ist das aktive Verfahren aus folgenden Gründen zu bevorzugen. Die zu befragende Person muss sich nicht selbst darum kümmern, wann welcher Fragebogen auszufüllen ist. Über eine entsprechende Verknüpfung in der E-Mail kann der Fragebogen durch einen einzigen Maus-Klick angezeigt werden. Zudem entfällt ein Abgleich von Zugangsdaten mit anderen System oder gar die Vergabe neuer und damit redundanter Zugangsdaten. Dieser höhere Komfort kann zu einer höheren Rücklaufquote beitragen, weil der für die Teilnahme an der Evaluation benötigte Aufwand minimiert wird. Das passive Verfahren kann jedoch ergänzend zum aktiven angeboten werden, um die Konsequenz des Untergangs von E-Mails durch einen alternativen Zugang zu umgehen.

---

<sup>69</sup> aus der Sicht des Evaluationssystems

## **Berechtigung**

Unabhängig von dem Verfahren zur Publikation des Fragebogens muss sichergestellt werden, dass nur eine berechtigte Person den Fragebogen genau einmal ausfüllen darf. Beim passiven Verfahren ergibt sich die Zugangsberechtigung durch die Anmeldung am System, wobei jedoch – entgegen der derzeit zu beobachtenden Praxis<sup>70</sup> – sichergestellt werden muss, dass nur eine tatsächlich berechtigte Person die entsprechenden Zugangsdaten erhält. Beim aktiven Verfahren muss das Evaluationssystem für jeden Fragebogen eine Transaktionsnummer erstellen, welche es der berechtigten Person erlaubt, diesen auszufüllen. Diese Transaktionsnummer sollte dabei in der Verknüpfung zum Aufruf des Fragebogens in der E-Mail eingebunden sein, so dass der Fragebogen trotz Zugangsberechtigung mit nur einem einzigen Maus-Klick und ohne weitere manuelle Dateneingabe aufgerufen werden kann. Unabhängig vom Verfahren muss das Evaluationssystem sicherstellen, dass ein vollständig ausgefüllter Fragebogen nicht erneut ausgefüllt werden kann. Wie von der Praxis gefordert (siehe Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41) soll es jedoch möglich sein, einen Fragebogen nur teilweise auszufüllen und den Vorgang zu einem späteren Zeitpunkt wieder fortzusetzen.

## **Zeitpunkte**

Grundsätzlich soll ein Evaluationssystem sowohl die formative (begleitende) wie auch die summative (zusammenfassende) Evaluation ermöglichen. Bei der formativen Evaluation muss das Evaluationssystem die Erhebungseinheit in mehrere Zeitabschnitte unterteilen, am Ende derer es die Erhebung jeweils durchführt. Bei der summativen Evaluation entspricht dieser Zeitabschnitt der gesamten Erhebungseinheit. Jede Erhebung erfolgt innerhalb eines Zeitfensters, das mit dem Ende eines Zeitabschnitts beginnt und nach der Dauer der Verfügbarkeit des Fragebogens endet. Das Evaluationssystem sollte den Fragebogen wie unter Abschnitt „Evaluationszeitpunkte“ auf Seite 18 dargestellt, möglichst am Anfang eines solchen Zeitfensters zur Verfügung stellen. Um unnötige Ausfälle zu vermeiden, muss die Frist für die Abgabe eines Fragebogens den Studierenden deutlich mitgeteilt werden. Da Erhebungseinheiten semesterunabhängig sind – eine Veranstaltung z. B. kann bereits wenige Wochen nach Semesterbeginn beendet sein – reicht eine pauschale Erhebung am Ende eines Semesters (bzw. Trimesters) nicht aus.

## **Rückkopplung für die Studierenden**

Um den Studierenden eine sofortige Rückkopplung geben zu können und damit die Bindung an die Evaluation zu erhöhen, könnte ihnen ein Evaluationssystem direkt im Anschluss an das Ausfüllen des Fragebogens die gemittelten Ergebnisse aller bisherigen Antworten zu

---

<sup>70</sup> In der Praxis wird diese Anforderung nicht immer erfüllt, da einige Fakultäten sog. Semesterkennwörter oder gar einen allgemeinen Zugang veröffentlichen, über die beliebige Personen nach kurzer Recherche aktuelle Veranstaltungen evaluieren können. Dieser Mangel wurde vom Autor dieser Arbeit an mehreren medizinischen Fakultäten festgestellt.

dieser Erhebungseinheit anzeigen. Sollen diese Ergebnisse den aktuellen Stand widerspiegeln, entstehen hohe Anforderungen an die Verarbeitungsgeschwindigkeit des Systems, da es insbesondere nach längerem Einsatz die relevanten Antworten aus einer sehr großen Anzahl von Datensätzen herausuchen und verarbeiten muss. Um diese Funktion mit deutlich geringeren Anforderungen anzubieten, könnte das System in Zeiten niedriger Belastung diese Ergebnisse vorab berechnen. Die so ermittelten Werte sind dann zwar nicht maximal aktuell, können den Studierenden jedoch mit minimaler Belastung des Systems sofort angezeigt werden. In diesem Fall liefe die Funktion nicht Gefahr sich wegen der unter Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf Seite 40 beschriebenen Anforderung an die Reaktionsgeschwindigkeit des Systems selbst ad absurdum zu führen.

Zusätzlich zu der sofortigen Rückkopplung soll den Studierenden wie unter Abschnitt „Rechtsnormen“ auf den Seiten 9–10 dargelegt, Einblick in die allgemeinen Evaluationsergebnisse gewährt werden. Entsprechende Berichte können zwar wie unter Abschnitt „Berichte“ auf den Seiten 65–69 beschrieben vom System vorbereitet werden, müssen jedoch nicht zwangsweise von diesem zur Verfügung gestellt werden, sondern können auch über statische Verknüpfungen auf den Web-Seiten der Fakultäten publiziert werden.

#### **Fremd-/Eigenbildentwicklung**

Zur Umsetzung der unter Abschnitt „Fremd- und Eigenbildentwicklung“ auf Seite 19 beschriebenen Vorschläge sollte ein Evaluationssystem eine Umfrage sowohl an die Studierenden wie auch an den bzw. die Dozenten einer Veranstaltung richten können. Um die Antworten in den Berichten gegenüberstellen zu können, muss das System dabei speichern, ob ein Fragebogen von einem Dozenten oder von einem Studierenden beantwortet wurde. Da hier keine Mutationsanomalien zu erwarten sind, reicht es entgegen dem Normalisierungsansatz aus, diese Information als Status und nicht als Relation zu den Personendaten zu speichern. Dies verhindert neben einer erhöhten Verarbeitungsgeschwindigkeit das Aufkommen datenschutzrechtlicher Aspekte.

#### **3.1.6 Auswertung**

Wie unter Abschnitt „Breite Öffentlichkeit“ auf Seite 13, Abschnitt „Auswertung“ auf Seite 20, Abschnitt „Externe Evaluation“ auf den Seiten 29–30, Abschnitt „Zufriedenheit“ auf den Seiten 33–36, Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41 hergeleitet, soll ein Evaluationssystem die Ergebnisse der Umfragen so verdichten können, dass sie durch Instrumente der deskriptiven Statistik grundlegend dargestellt und als Kennzahlen für weitere Transformationen verwendet werden können. Die statistischen Instrumente sollen in Berichten zusammengefasst und aufbereitet werden. Die Kennzahlen sollen insbesondere für die Berechnung von Rangplätzen beim Ranking (und damit auch für die Vergabe von Lehrpreisen) verwendet werden, sowie in Formeln zur Berechnung von Verteilungsschlüsseln bei der LOM einfließen können. Da die Fragen während der Laufzeit angelegt und ggf. modifiziert

werden, soll es das Evaluationssystem erlauben, die Menge der dabei zu berücksichtigenden Fragen auszuwählen und wie unter Abschnitt „Fragearten“ auf den Seiten 15–16 beschrieben optional zu gewichten. Sowohl für die Generierung von Berichten wie auch für die Berechnung von Kennzahlen soll es das Evaluationssystem erlauben, Störvariablen zu benennen und sie bei den Ergebnissen entweder darzustellen, oder ihrer verzerrenden Wirkung rechnerisch entgegenwirken zu können (siehe Abschnitt „Störvariablen“ auf den Seiten 16–17).

Nachdem anspruchsvolle Statistik-Pakete wie beispielsweise SPSS oder SAS zur Verfügung stehen, deren Nachentwicklung höchst unökonomisch wäre, soll das System neben der statistischen Auswertung der Ergebnisse innerhalb des Evaluationssystems die Rohdaten für eine umfassendere Auswertung weiterhin so ausgeben können, dass diese von solchen externen Statistik-Paketen eingelesen und verwertet werden können (siehe Abschnitt „Auswertung“ auf Seite 20). Dabei soll das System ggf. erfasste Störvariablen mit übergeben, um sie in den externen Systemen bei Bedarf statistisch kontrollieren zu können.

## Berichte

Die in einem einfachen<sup>71</sup> Bericht dargestellten Instrumente der univariaten Datenbeschreibung müssen auf die Art der Fragestellung abgestimmt sein und einen konkreten Nutzen für die Zielgruppe schaffen. Die unter Abschnitt „Auswertung“ auf Seite 20 erwähnten Grundregeln sollen in Anlehnung an Weiß (2002, S. 63) durch folgende Instrumente<sup>72</sup> explizit berücksichtigt werden, da ihre Bedeutung und damit auch die Interpretierbarkeit als weitgehend bekannt vorausgesetzt werden kann.

Tabelle 3.1: Instrumente der univariaten Datenbeschreibung

| Skala                     | Lagemaße                      | Streuungsmaße                    | Graphische Darstellung  |
|---------------------------|-------------------------------|----------------------------------|---|
| Nominal                   | Modus                         |                                  | Kreisdiagramm   |
| Ordinal                   | Modus<br>Median               | Spannweite                       | Balkendiagramm  |
| Intervall &<br>Verhältnis | Modus<br>Median<br>Mittelwert | Spannweite<br>Standardabweichung | Diskrete Daten:<br>Balkendiagramm<br>Stetige Daten:<br>Häufigkeitspolygon <sup>73</sup> |

<sup>71</sup> Die hier dargestellten Instrumente beziehen sich dabei auf eine rein summative Fremdbild-Auswertung und überlassen den Vergleich einzelner Merkmale im Rahmen der normativen Evaluation und die Gegenüberstellung einer Fremd- und Eigenbildentwicklung den externen Statistik-Paketen.

<sup>72</sup> Auf die Erwähnung der Darstellung der Rohdaten in Listenform, die für alle Formen der Skalierung zulässig ist, wurde hier verzichtet.

<sup>73</sup> Auf unkorrekte Verwendung von Häufigkeitspolygonen und Histogrammen geht Bortz (2005, S. 32 f.) ein.

Bei der Darstellung dieser Instrumente soll ein Evaluationssystem die Ergebnisse grafisch so aufbereiten, dass die Informationen effizient aufgenommen werden können. Durch die Erfassung der Skalierung der Fragen bereits beim Anlegen (siehe Abschnitt „Fragenarten“ auf Seite 49), kann das Evaluationssystem die für eine Frage zulässigen Instrumente selbständig zusammenstellen.

Um sowohl Gesamtberichte wie auch Berichte zu bestimmten Themen generieren zu können, sollte ein Evaluationssystem den Benutzern umfassende Einschränkungskriterien wie die Bestimmung der Items, des Zeitraums oder der Erhebungseinheiten zur Verfügung stellen. Dabei sollten neben den Evaluationsbeauftragten, die eine uneingeschränkte Auswertung durchführen können auch die Dozenten die Möglichkeit erhalten, eigene Berichte zu generieren. Aus datenschutzrechtlichen Gründen darf das Evaluationssystem bei der Erstellung dozentenspezifischer Berichte jedoch nur die Antworten auf Fragebögen der Veranstaltungen berücksichtigen, die von dem anfragenden Dozenten selbst gehalten wurden.

Um den administrativen Aufwand minimal zu halten und den Forderungen der Praxis (siehe Abschnitt „Zufriedenheit“ auf den Seiten 33–36, Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41) nachzukommen, sollte ein Evaluationssystem die Berichte zeitnah sowie automatisch generieren und versenden.

## Rankings

Grundsätzlich sind die Rankings den Evaluationsbeauftragten zur Verfügung zu stellen, so dass es einer menschlichen Entscheidung überlassen bleibt, welche Informationen daraus in welcher Form weiter veröffentlicht werden, bzw. welche weiteren Konsequenzen daraus zu ziehen sind. Um den Dozenten jedoch ein individuelles Feedback über die Bewertung ihrer Lehre bieten zu können, soll das Evaluationssystem individualisierte Rankings auch den Dozenten zur Verfügung stellen. Dabei soll ein Dozent in der Rangliste nur seinen eigenen Namen sehen dürfen um die relative Bewertung seiner Lehrleistung erkennen zu können. Die Namen aller anderen Dozenten soll das Evaluationssystem ausblenden.

Damit Rankings nicht nur zu bloßen ‚Hitlisten‘ verkommen, sondern einen profunden Beitrag zur Umsetzung der in der Literatur geforderten Verbesserung der Lehre leisten, soll ein Evaluationssystem die zur Berechnung der Rangplätze verwendeten Merkmale und deren verdichtete Werte einzeln auflisten. Den Evaluationsbeauftragten können diese Werte sowie die Namen gesamthaft zur Verfügung stehen, während ein einzelner Dozent aus Gründen des Datenschutzes nur seine eigenen Werte einsehen darf. Somit erhalten die Evaluationsbeauftragten aufgeschlüsselte Auskünfte über die Stärken und Schwächen der Dozenten, die sie individuellen Gesprächen zugrunde legen können. Interessierte Dozenten erhalten ebenfalls Einblick in ihre ganz persönlichen Stärken und Schwächen und können daraus Konsequenzen für die weitere Entwicklung ihrer Lehre ziehen. Um die Dozenten dabei jedoch nicht alleine zu lassen und der Forderung nach einer individuellen Beratung zur Verbesserung der Schwächen (siehe Abschnitt „Leistungsindikatoren“ auf den Seiten 21–23) nachzukommen, kann das Evaluationssystem beim Auftreten von Schwächen (also beim Absinken bestimmter Kennzahlen)

die Evaluationsbeauftragten darüber informieren und so individuelle Beratungsgespräche initiieren. Eine solche Funktion kann insbesondere bei größeren Fakultäten hilfreich sein.

### 3.1.7 Integration

Wie unter Abschnitt „Hochschulleitung“ auf den Seiten 12–13 und Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27 dargestellt, hat die Hochschulleitung über die systeminterne Effizienz hinaus ein besonderes Interesse daran, dass sich ein als strategisch einzustufendes Evaluationssystem nahtlos in die IT-Landschaft integrieren lässt. Deshalb soll ein Evaluationssystem so aufgebaut sein, dass es innerhalb der technischen Rahmenbedingungen üblicher universitärer Rechenzentren und ohne überdimensionierte Anforderungen an die benötigte Hardware betrieben werden kann. Weiterhin soll es flexible und effiziente Schnittstellen für den Bezug und die Publikation von Daten an externe Systeme zur Verfügung stellen. Bezogen werden müssen dabei die unter Abschnitt „Stammdatenbezug“ auf den Seiten 47–48 beschriebenen Daten. Publiziert werden sollen neben dem Export der rohen Erhebungsdaten für die externe statistische Analyse folgende Informationen.

**Rückmeldung der Teilnahme** Wie unter Abschnitt „Rücklaufquote“ auf den Seiten 36–37, Abschnitt „Zufriedenheit“ auf den Seiten 33–36, Abschnitt „Notwendige Verbesserungen“ auf den Seiten 37–39 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41 beschrieben, werden in der Praxis durchaus (positive, wie auch negative) Konsequenzen an die Beteiligung an der Evaluation geknüpft, um nicht zuletzt die Rücklaufquote zu erhöhen. Diese Konsequenzen wie beispielsweise die Vergabe von Scheinen oder die vorzeitige Noteneinsicht müssen jedoch von externen Systemen umgesetzt werden. Deshalb muss ein Evaluationssystem eine Schnittstelle anbieten, um einem externen System mitteilen zu können, ob ein Studierender eine Erhebungseinheit evaluiert hat oder nicht. Darüber hinaus muss das Evaluationssystem unterschiedliche Definitionen einer Teilnahme kommunizieren können. So kann die Teilnahme bedeuten, dass nur ein Fragebogen ausgefüllt wurde, dass alle Fragebögen geöffnet wurden oder alle Fragen aller Fragebögen beantwortet wurden. Diese Informationen dienen dem externen Verwaltungssystem zur Entscheidung über die Einleitung einer entsprechenden Konsequenz.

**E-Learning** Da E-Learning-Kurse von externen Systemen angeboten werden und diese wie unter Abschnitt „E-Learning“ auf Seite 33 dargestellt, ebenfalls evaluiert werden sollen, muss ein System weiterhin Schnittstellen zur Verfügung stellen, um die Evaluation mit dem externen E-Learning-System zu koordinieren. Idealerweise kann ein Evaluationssystem die Informationen wahlweise selbständig an Zielsysteme übermitteln, sobald sie verfügbar sind, oder eine Schnittstelle anbieten, über welche externe Systeme diese Informationen bei Bedarf abrufen können. Für die Kommunikation bieten sich Web Services als Stand der Technik an.

## 3.2 Implementation

Dieser Abschnitt zeigt die Umsetzung bestimmter Anforderungen, die im vorherigen Abschnitt dargestellt wurden. Die Implementation wird anhand des Systems mit dem Projektnamen „Vilfredo“<sup>74</sup> (das System) dargestellt, das vom Autor im Rahmen der vorliegenden Arbeit entwickelt wurde.<sup>75</sup> Dazu werden in diesem Abschnitt neben Schemazeichnungen auch Bildschirmfotos des Systems zur Veranschaulichung der komplexen Zusammenhänge sowie UML-Diagramme<sup>76</sup> zur formalen Beschreibung verwendet. Die Auswahl der Anforderungen und deren Umsetzung wurden unter der Bedingung vorgenommen, einen technischen Beitrag zur Umsetzung der Evaluationsziele leisten zu können. Die Diskussion dieses Beitrags ist eine zentrale Aufgabe des vierten Kapitels.

Die folgenden Abschnitte konzentrieren sich auf die Darstellung der neuen Möglichkeiten. Auf eine Darstellung der diesem System zugrunde liegenden dreischichtigen Client-Server-Architektur (Three Tier Architecture) wurde aus Platzgründen und mangels Innovation verzichtet. Auf Weiterentwicklungspotenziale an dem System wird ebenfalls in der Diskussion unter Abschnitt „Schwachstellen und Weiterentwicklungspotenziale“ auf den Seiten 90–92 eingegangen.

### 3.2.1 Fragebogen

Ein wichtiges architektonisches Grundprinzip des Systems ist die Berücksichtigung der unter Abschnitt „Lehreinheiten“ auf Seite 14 hergeleiteten Erkenntnis, dass bei der Evaluation – vereinfacht ausgedrückt – das Ganze anders bewertet wird, als die Summe seiner Teile. Aus diesem Grund erlaubt es das System wie in Abbildung 3.1 auf der nächsten Seite verdeutlicht, die Erhebungen auf beliebige Elemente des Curriculums beziehen zu können.

So kann das System beispielsweise einen einzigen Fragebogen der ganzen organisatorischen Einheit „Modul 2“ aber auch einen Fragebogen der von dem Dozenten „Prof. Boerne“ im Rahmen der Vorlesung „Innere Medizin“ gehaltenen Termine widmen. Mit diesem Konzept lässt sich insbesondere die Anforderung nach der dozentenbezogenen Evaluation erfüllen. Abbil-

---

<sup>74</sup> Der Projektname wurde von dem italienischen Ökonomen und Soziologen Vilfredo Pareto abgeleitet, da das nach ihm benannte Pareto-Optimum den Zustand beschreibt, in welchen das Lehrevaluationssystem die Lehrevaluation bringen soll.

<sup>75</sup> Die Rechte an diesem System liegen vollständig beim Autor der vorliegenden Arbeit.

<sup>76</sup> Die Unified Modeling Language (UML) ist eine standardisierte und in der Software-Entwicklung weit verbreitete Notation zur Visualisierung. Für eine weiterführende Beschreibung sei auf Oestereich (2006) und Object Management Group (2006) verwiesen. Um eine leichte Lesbarkeit zu ermöglichen, wurden sämtliche Namen in den Modellen in der für die Modellierung ansonsten unüblichen deutschen Sprache dargestellt. Um die Persistenz sicherzustellen, gleichzeitig jedoch auf zusätzliche Entity Relationship Modelle (ERM) verzichten zu können, wurden die hier gezeigten Fachklassen bereits so modelliert, dass sie sich ohne wesentliche Umwandlungen in solche Datenbank-Modelle übertragen lassen. Aus diesem Grund wurde insbesondere auf die Modellierung mit Interfaces verzichtet, die ohne diese Nebenbedingung an mehreren Stellen eine elegantere Umsetzung erlaubt hätte. Das bei der Modellierung eingesetzte Instrument der Vererbung wurde im Datenbank-Modell per Single-Table-Inheritance mit einem Type-Discriminator zur Ableitung der unterschiedlichen Klassen aus der jeweils zugrunde liegenden Tabelle realisiert.



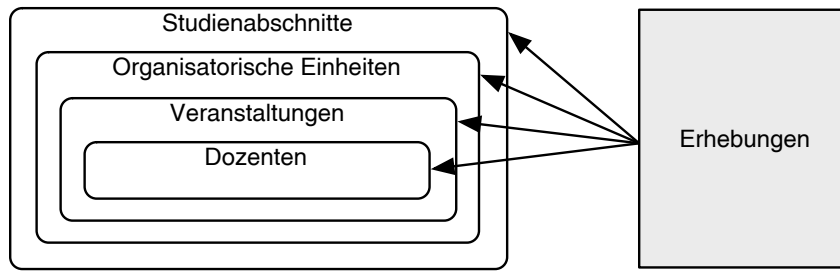


Abbildung 3.1: Bezug der Erhebungen auf unterschiedliche Einheiten

Abbildung 3.2 zeigt ein Klassendiagramm, welches das Konzept der dozentenbezogenen Erhebung formalisiert. Eine Verallgemeinerung dieses Modells, das der Darstellung in Abbildung 3.1 entspricht zeigt Abbildung 3.5 auf Seite 62.

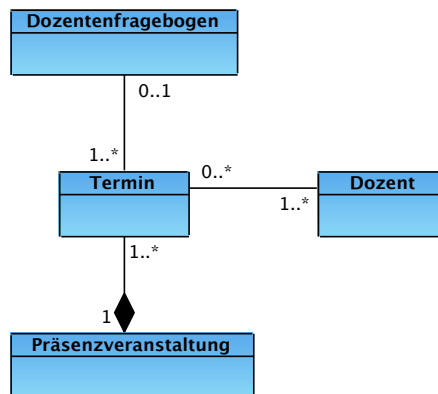


Abbildung 3.2: Klassendiagramm der dozentenbezogenen Evaluation

Abbildung 3.3 auf der nächsten Seite zeigt einen von Vilfredo generierten Fragebogen, der sich auf den im Titelbereich angezeigten Termin des genannten Dozenten bezieht. Zur Orientierung zeigt der Fragebogen dabei alle relevanten Informationen wie den Titel der Veranstaltung, den Dozenten sowie das Datum und die Zeit des Termins an. Um eine Evaluationsmüdigkeit unter den Studierenden zu verhindern, die dann zu erwarten wäre, wenn für jeden Einzeltermin ein eigener Fragebogen erstellt wird, bezieht das System einen Fragebogen automatisch auf mehrere Einzeltermine, die von demselben Dozenten gehalten wurden. In diesem Fall gibt der Titelbereich des Fragebogens den Zeitraum vom ersten bis zum letzten Termin an und erlaubt am Ende des Fragebogens eine detaillierte Auflistung aller Einzeltermine für eine genaue Übersicht.

Den Zugang zu einem Fragebogen gewährt das System über eine automatisch erstellte E-Mail (siehe Abbildung 3.4 auf Seite 61), über welche sich der Fragebogen durch einen einzigen Maus-Klick öffnen lässt. Die Übermittlung der dabei verwendeten Transaktionsnummer

The screenshot shows a web browser window with the URL <http://vilfredo/cgi-bin/WebObjects/Vilfredo.woa/wa/showQuestionnaire?tan=pf56u23srovheqnlaw90>. The page title is 'Vilfredo Evaluation'. The main heading is 'Evaluation der Veranstaltung Innere Medizin: Methodenseminar - Endoskopie bei Prof. Dr. med. Karl-Friedrich Boerne am 04. August 2006 von 14:00 bis 15:30'. Logos for LMU (Ludwig-Maximilians-Universität München) and Vilfredo are visible in the top left. The form is divided into two main sections: 'Fragen zur Veranstaltung' and 'Fragen zur Dozentin/zum Dozenten'. Each section contains several statements with radio buttons for 'Stimme nicht zu' and 'Stimme voll zu'. The first section also includes a 6-point Likert scale for 'Der Veranstaltung gebe ich folgende Schulnote' and three text input fields for 'Allgemeine Kommentare zur Veranstaltung', 'An der Veranstaltung hat mir besonders gut gefallen', and 'Folgendes sollte unbedingt verbessert werden'. The second section contains three statements with radio buttons for 'Stimme nicht zu' and 'Stimme voll zu'.

Abbildung 3.3: Dozentenbezogener Fragebogen für einen Termin

stellt sicher, dass die angeschriebene Person die Berechtigung zur Teilnahme hat und ein Fragebogen nur einmal ausgefüllt werden kann. Die Fragebögen können auch teilweise ausgefüllt und die Beantwortung der Fragen zu einem späteren Zeitpunkt fortgesetzt werden.

Eine verallgemeinerte Formalisierung des unter Abschnitt „Evaluation beliebiger Lehreinheiten“ auf den Seiten 50–51 beschriebenen und in Abbildung 3.1 auf der vorherigen Seite dargestellten Prinzips nach dem sich Erhebungen auf unterschiedliche Einheiten beziehen können, zeigt Abbildung 3.5 auf Seite 62.

### 3.2.2 Fragen

Der Versuch, das weit verbreitete Konzept eines Standardfragebogens auf diese Architektur zu übertragen würde wegen der Vielzahl der möglichen Ausprägungen zwangsläufig zu einem



Abbildung 3.4: Zugang zu einem neuen Fragebogen per E-Mail

unüberschaubaren administrativem Aufwand für die Einrichtung, Pflege und Auswertung der Fragebögen führen. Da dies jedoch der Anforderung der Effizienz klar widerspricht, wurde der Fokus auf den Fragebogen durch den Fokus auf die Fragen abgelöst. In dem System werden einmalig unterschiedliche Fragen angelegt, wobei für jede Frage definiert werden kann, auf welche konkreten Einheiten sie sich beziehen soll. Aufgrund dieser Informationen erstellt das System am Ende einer Erhebungseinheit (z. B. eine Veranstaltung oder ein Semester) einen Fragebogen, mit den dieser Einheit zugewiesenen Fragen. Dieser Zusammenhang ist in Abbildung 3.6 auf Seite 63 schematisiert.

Bei der Gestaltung der Fragen stellt das System den Anforderungen aus Abschnitt „Fragearten“ auf Seite 15 und Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41 entsprechend alle gängigen Typen zur Verfügung:

- Offene Fragen
- Multiple-Choice-Fragen
- Single-Choice-Fragen
- Bipolare Fragen (als spezielle Ausprägung der Single-Choice-Fragen)

Die Umsetzung der Fragenarten stellt Abbildung 3.7 auf Seite 63 formal dar. Zentrales Element ist dabei die Beziehung zwischen Frage und Fragebogen. Diese Beziehung legt nicht nur das gängige Grundprinzip fest, nachdem ein Fragebogen eine oder mehrere Fragen beinhalten kann, sondern auch die spezifische Erweiterung, dass eine Frage auch auf mehreren Fragebögen erscheinen kann. In diesem Modell ebenfalls dargestellt sind die unterschiedlichen Konkretisierungen der abstrakten Frage zu den oben beschriebenen Typen.

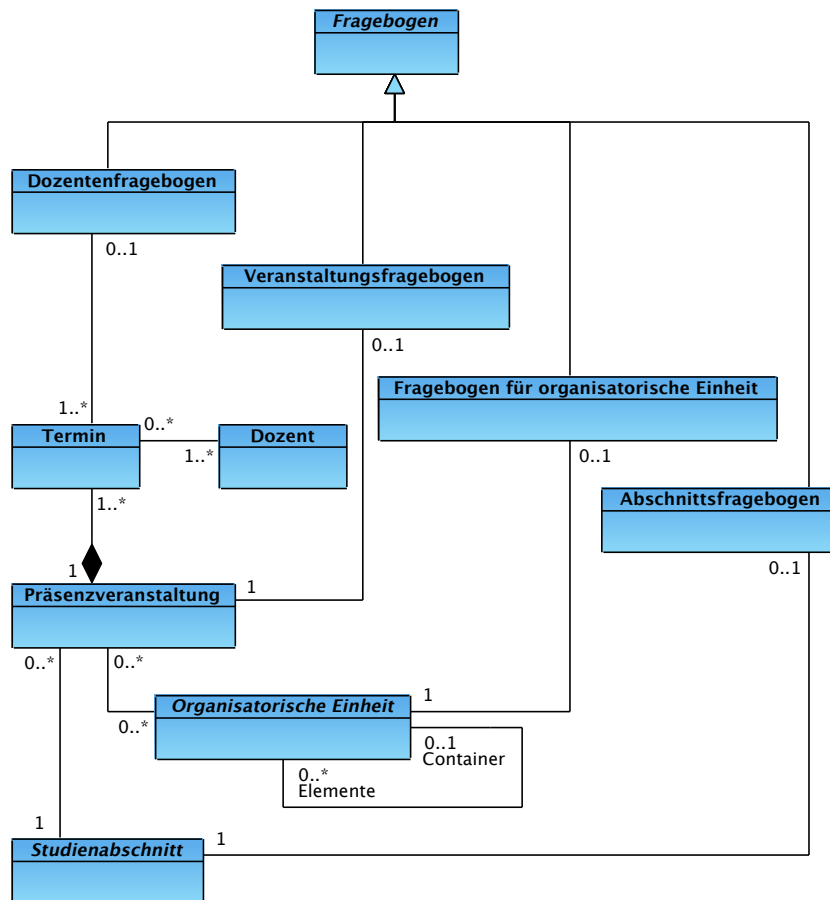


Abbildung 3.5: Klassendiagramm des Bezugs der Erhebungen auf unterschiedliche Einheiten

Beim Anlegen einer Frage fordert das System vom Benutzer die Angabe des Skalenniveaus und ggf. des Datentyps. Mithilfe dieser Informationen kann das System die Auswertungen mit den korrekten statistischen Instrumenten durchführen. Im Beispiel der Abbildung 3.8 auf Seite 64 soll eine bipolare Frage angelegt werden, für welche das Skalenniveau der Intervall-Skalierung<sup>77</sup> bereits implizit feststeht. Um schnell mit denselben Skalierungen arbeiten zu können, stellt das System die bereits verwendeten Ausprägungen der Skalierung (hier eine fünfstufige Likert-Skala) für eine direkte Übernahme zur Verfügung. Gibt der Benutzer

<sup>77</sup> Um dem Wunsch der Praxis nachzukommen, hier auch das Lagemaß des Mittelwerts berechnen zu können und damit die Vergleichbarkeit der Ergebnisse mit den Berichten anderer Systeme beizubehalten, wurde für bipolare Fragen grundsätzlich eine Intervall-Skalierung unterstellt. Diese Annahme wurde getroffen, obwohl bipolare Fragen in der Regel lediglich ordinal skaliert sind und Mittelwerte deshalb statistisch gesehen nicht sinnvoll interpretiert werden können. In diesem Zusammenhang soll darauf hingewiesen werden, dass sich der Mittelwert zwar großer Beliebtheit erfreut. Demgegenüber bietet der Median den Vorzug, gegenüber Ausreißern robust zu sein und auch für ordinalskalierte Daten sinnvoll interpretiert werden zu können.

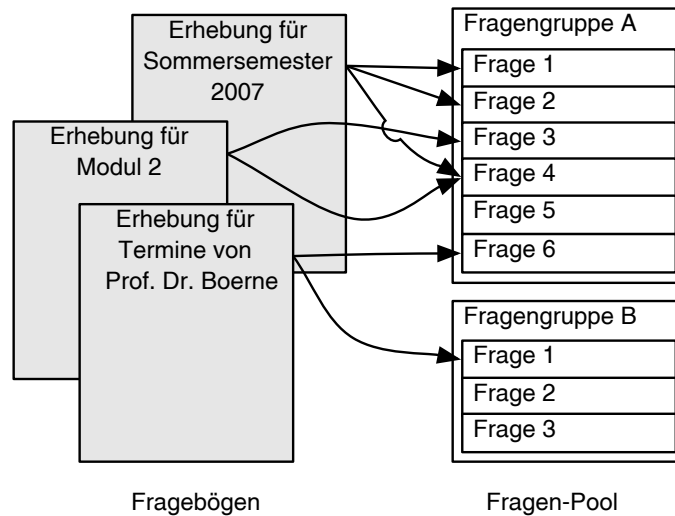


Abbildung 3.6: Dynamischer Bezug der Fragebögen auf die Fragen

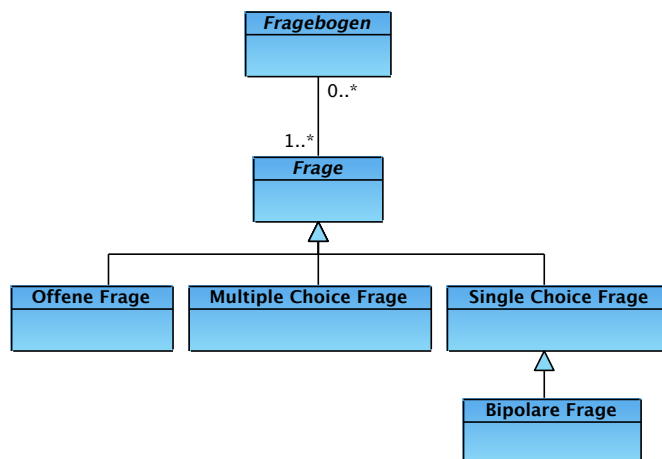


Abbildung 3.7: Klassendiagramm des dynamischen Fragebogens

weiterhin die Sortierung an, so bietet das System die Option an, die Frage in die Berechnung der Rankings einzubeziehen. Durch die Angabe der Sortierung können sowohl positive (z. B. „Die Veranstaltung war interessant“) wie auch negative (z. B. „Der Dozent kam zu spät“) Interpretationen der Skalierung berücksichtigt werden.

Im Falle einer offenen Frage erlaubt das System zudem die optionale Angabe eines Datentyps sowie eines zulässigen Wertebereichs. So kann beispielsweise festgelegt werden, dass die Frage „Wie viel Zeit (in Minuten) haben Sie durchschnittlich zur Nachbereitung jedes Termins dieser Veranstaltung aufgebracht?“ nur ganzzahlige Antworten im Bereich von 0 bis

300 Minuten zulässt. Alleine durch die Angabe des Datentyps kann das System so auch bei offenen Fragen entscheiden, welche statistischen Instrumente für die Auswertung zulässig sind. Durch die Angabe und Überprüfung des zulässigen Wertebereichs können genauere Erkenntnisse als bei geschlossenen Fragen erzielt werden ohne dabei auf den Vorteil der Kategorisierbarkeit (hier in Form eines Wertebereichs) in Gänze verzichten zu müssen.

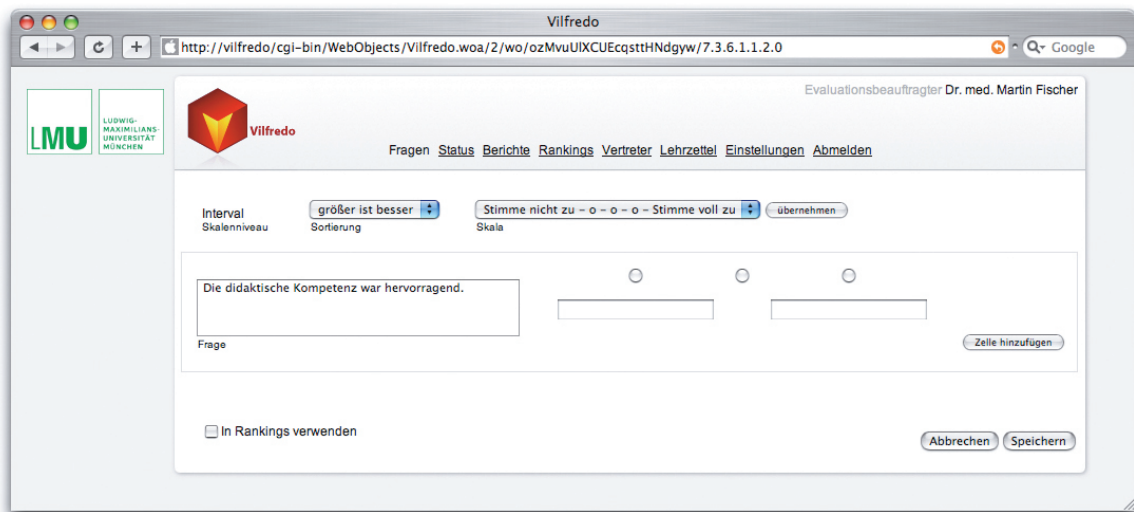


Abbildung 3.8: Anlegen einer neuen Frage

Grundsätzlich werden die Fragen von einem der im System definierten Evaluationsbeauftragten angelegt. Um es den Dozenten zu erlauben, Informationen aus der Evaluation ableiten zu können, die durch die Standard-Fragen möglicherweise nicht abgedeckt sind, wurde darüber hinaus eine Funktion implementiert, mit der Dozenten eigene Fragen ergänzend zu den von den Evaluationsbeauftragten angelegten direkt im System erstellen können. Solche von Dozenten angelegte Fragen müssen ein Review-Verfahren durchlaufen, bei dem sie ein Evaluationsbeauftragter begutachtet und daraufhin löschen, mit einem Kommentar zur Überarbeitung zurückgeben oder direkt freigeben kann. Um Medienbrüche zu vermeiden und damit der Nebenbedingung der Effizienz zu entsprechen, wird das gesamte Review-Verfahren komplett im System abgebildet. Freigegebene Fragen verwendet das System automatisch auf allen, ab dem Zeitpunkt der Freigabe neu erstellten Fragebögen, die sich auf Veranstaltungen des entsprechenden Dozenten beziehen. Diese Möglichkeit stellt ein weiteres Instrument dar, um die Dozenten stärker in die Evaluation einzubinden und den Wert des persönlichen Feedbacks durch die Evaluation zu erhöhen. Durch das integrierte Review-Verfahren soll dabei die Qualität der Fragen (und damit der Evaluation an sich) bei minimalem administrativen Aufwand sichergestellt werden. Eine Formalisierung des implementierten Review-Verfahrens zeigt Abbildung 3.9 auf der nächsten Seite anhand eines entsprechenden UML-Zustandsautomaten.

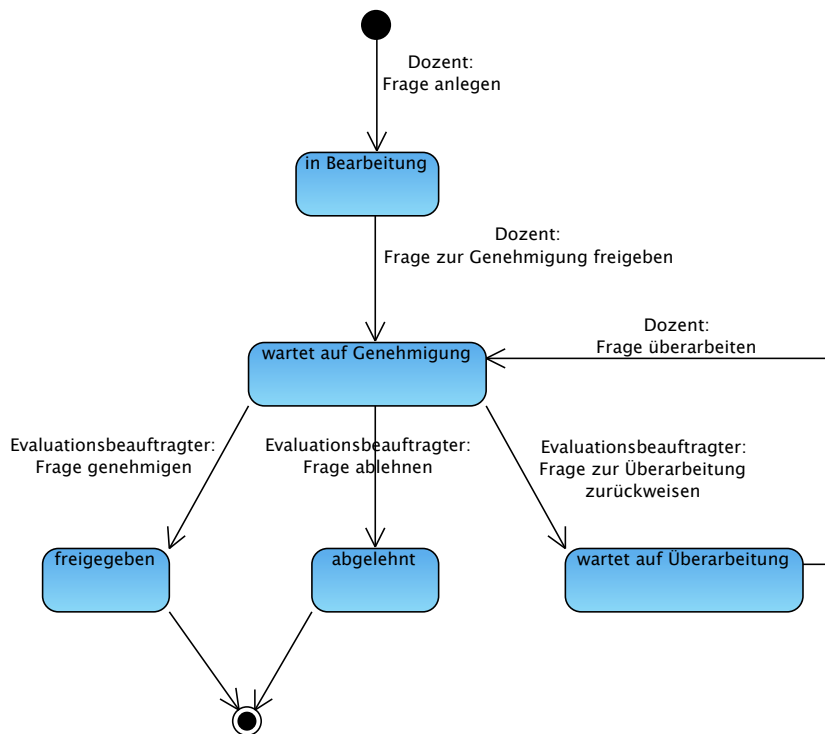


Abbildung 3.9: Zustandsautomat einer Frage im Review-Verfahren

### 3.2.3 Berichte

Analog zum Anlegen der Fragen wird die Auswertung der Erhebungen grundsätzlich von den Evaluationsbeauftragten angestoßen. Sie können dabei die Menge der zu berücksichtigten Antworten auf eine frei definierbare Menge an Fragen durch folgende, miteinander kombinierbare<sup>78</sup> Filter einschränken, um individuelle Berichte zu erhalten:

- Zeitraum
- Organisatorische Einheit
- Dozent
- Veranstaltung

<sup>78</sup> Die Kombination wird in diesem Modell ausschließlich durch eine Konjunktion (logisches Und) hergestellt. Jeder hinzugefügte Filter führt also zu einer Erhaltung oder Einschränkung, nicht aber zu einer Erweiterung der Ergebnismenge.

Durch diese Filter kann sich ein Evaluationsbeauftragter also beispielsweise jeweils einen Bericht für einen der folgenden Bereiche erstellen lassen:<sup>79</sup>

- Fragen 1 und 7–10 für alle Veranstaltungen im Wintersemester 2006/2007 in Modul 2
- Frage nach der Pünktlichkeit der Dozenten in den Veranstaltungen des Themenbereichs „Innere Medizin“ vom Sommersemester 2006 bis Wintersemester 2006/2007
- Fragen 14–16 für alle Termine die vom Dozenten „Prof. Boerne“ im Wintersemester 2006/2007 gehalten wurden

Aufgrund der in Abbildung 3.5 auf Seite 62 gezeigten Relationen von Termin, Präsenzveranstaltung, organisatorischer Einheit und Studienabschnitt zu den jeweiligen Konkretisierungen des abstrakten Fragebogens ist es dem System möglich, die Filter unabhängig von der Art der Fragebögen einzusetzen, auf denen die ausgewählten Fragen verwendet wurden. Die zu berücksichtigenden Antworten werden also rein durch die vom Benutzer angegebenen Filter eingeschränkt, nicht aber durch das dem System zugrunde liegende Datenmodell. Trotz der unter Abschnitt „Fragen“ auf Seite 61 beschriebenen Mehrfachverwendung einer Frage auf unterschiedlichen Arten von Fragebögen können Berichte damit auf Wunsch alle Einsatzbereiche einer Frage berücksichtigen. Es ist dabei unerheblich, ob die Frage auf einem Dozenten-, Veranstaltungs-, Abschnitts- oder einem Fragebogen für organisatorische Einheiten verwendet wurde.

Auch bei der Erstellung von Berichten muss die Nebenbedingung der Effizienz gewahrt bleiben. Aus dem Prinzip des dynamischen Bezugs der Fragebögen auf die Fragen ergibt sich neben der Möglichkeit, die Einheiten individuell evaluieren zu können, der positive Nebeneffekt, dass eine Frage im System tatsächlich nur einmal vorhanden ist und deshalb auch eindeutig ausgewertet werden kann. Eine nachträgliche manuelle (und damit ineffiziente) Zusammenführung von Fragen mit gleicher Bedeutung bei der Auswertung entfällt damit. Zudem kann das System zu jeder Frage automatisch die statistisch korrekten Auswertungsinstrumente bestimmen und im Bericht darstellen, da es beim Anlegen einer Frage die dafür benötigten Informationen von Skalenniveau und Datentyp bereits erfasst hat (siehe Abschnitt „Fragen“ auf den Seiten 61–64).

Aus diesen Informationen berechnet das System einen Bericht der als PDF-Dokument<sup>80</sup> ausgegeben wird. Durch die Ausgabe in diesem weit verbreiteten Format kann sichergestellt werden, dass der Bericht in grafisch ansprechender Form erstellt und betrachtet werden kann. Die grafische Aufbereitung ist eine Grundlage zur Erfüllung der Forderung nach Effizienz und Benutzerfreundlichkeit. Um jedoch gleichzeitig die Anforderung des Persönlichkeitsschutzes zu wahren, stellt das System die Option zur Verfügung, den Bericht mit einem Kennwort zu

---

<sup>79</sup> Die Fragen werden hier teilweise über Nummern referenziert, um die Übersichtlichkeit bei der Angabe des vollständigen Titels nicht einzuschränken. Innerhalb des Systems lassen sich die Fragen jedoch direkt auswählen.

<sup>80</sup> PDF = Portable Document Format. Ein plattformunabhängiges Dateiformat zur einheitlichen Darstellung insbesondere von grafisch aufbereiteten Dokumenten.



verschlüsseln. Wird diese Option ausgewählt, kann der Bericht nur mit dem System-Kennwort des Benutzers eingesehen werden. Unabhängig von dieser Option wird der Bericht immer in unveränderbarer Form abgespeichert, so dass er nicht manipulierbar ist.<sup>81</sup>

Abbildung 3.10 auf der nächsten Seite zeigt einen typischen Ausschnitt aus einem vom System generierten Bericht mit der Auswertung unterschiedlicher Fragetypen. Zur Erstellung dieses Berichts mussten lediglich die auszuwertenden Fragen und der Zeitraum angegeben werden. Auf der dargestellten Seite des Berichts werden drei Fragen unterschiedlichen Typs ausgewertet. Es handelt sich dabei um eine Single-Choice-Frage, eine bipolare Frage und eine offene Frage. Wie ersichtlich, hat das System für jede Frage automatisch nur die in Tabelle 3.1 auf Seite 55 aufgeführten statistisch zulässigen Instrumente dargestellt. So wurde die erste Frage – eine Single-Choice-Frage – nominal skaliert angelegt, weshalb nur der Modus als Lagemaß und das Kreisdiagramm als graphische Darstellung angezeigt wird. Da der zweiten Frage eine Intervall-Skalierung unterstellt wurde (siehe Fußnote 77 auf Seite 62), steht dem Bericht ein umfassenderes Instrumentarium zur Verfügung. Nachdem bei der Anlage dieser Frage auch die Sortierung definiert wurde, berechnet das System automatisch einen normierten Score<sup>82</sup>. Dadurch kann das Ergebnis auch mit anderen Fragen verglichen werden; auch wenn diese andere Merkmalsausprägungen, eine andere Skalierung oder gar eine andere Fragenart aufweisen. Neben der Darstellung des reinen Score-Werts trägt der Bericht die relative Position des Scores zur Visualisierung farblich kodiert in der normierten Skala von 0 bis 100 ein und gibt dazu eine verbale Interpretation – hier „eher gut“ – an. Die letzte hier dargestellte Frage ist eine offene, weshalb der Bericht die Antworten nicht weiter verdichtet sondern rein in Form einer Liste darstellt.

Um das zentrale Thema des von der Praxis geforderten individuellen Feedbacks zu unterstreichen, steht den Dozenten bei der Erstellung eines persönlichen Berichts dasselbe Instrumentarium wie den Evaluationsbeauftragten zur Verfügung. Zur Wahrung des ebenfalls geforderten Persönlichkeitsschutzes, bezieht das System jedoch automatisch nur die Antworten mit ein, die sich auf Fragebögen von Veranstaltungen des Dozenten beziehen, der den Bericht erstellt. Technisch gesehen legt das System in diesem Fall implizit immer einen entsprechenden, nicht entfernbaren Dozentenfilter an, um die bestehende Logik des Systems wieder verwenden zu können.

Eine formalisierte Darstellung des Aufbaus eines Berichts mit den frei wählbaren Fragen und Filtern sowie deren Auswertung zeigt Abbildung 3.11 auf Seite 69.<sup>83</sup>

In diesem Klassendiagramm implizit enthalten ist die Datenstruktur der Logik, über welche das System (wie unter Abschnitt „Fragen“ auf den Seiten 61–64 eingeführt) die korrekten Auswertungsinstrumente aus den Eigenschaften einer auszuwertenden Frage übernehmen kann. So bezieht die Logik der Frageauswertung zuerst die notwendigen Informationen wie

<sup>81</sup> Es sei jedoch darauf hingewiesen, dass es – wie in allen Bereichen – mit einem gewissen Maß an krimineller Energie möglich ist, den Schutz der PDF-Dokumente zu umgehen.

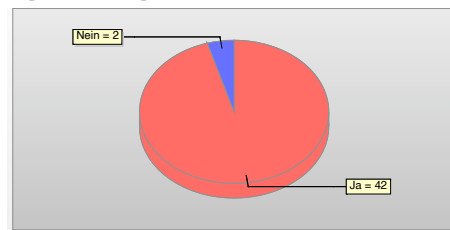
<sup>82</sup> Eine ausführliche Darstellung dieses Instruments wird unter Abschnitt „Rankings“ auf den Seiten 69–75 dargestellt.

<sup>83</sup> Auf die Modellierung des Berichts als Sub-Klasse der abstrakten Klasse „Analyse“ wird bei der Darstellung der Modellierung des Ranking unter Abschnitt „Rankings“ auf den Seiten 72–75 eingegangen.

## Fragen zur Evaluation

Haben Sie bemerkt, dass Sie für diese Veranstaltung einen anderen Fragebogen als sonst erhalten haben?

Single Choice Frage (n=44)



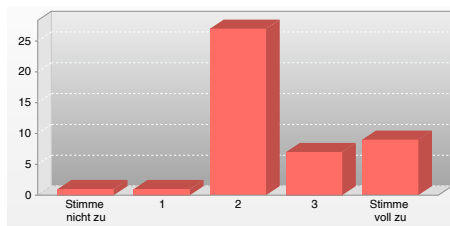
Lagemaße

Modus Ja

Kuchendiagramm

Dieser Fragebogen gefällt mir besser als die sonst üblichen

Bipolare Single Choice Frage (n=45)



Lagemaße

Mittelwert 2,49

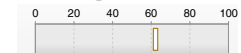
Median 2

Modus 2

Streuungsmaße

Spannweite 4

Bewertung



Score 62,22

eher gut

Balkendiagramm

Was gefällt Ihnen an diesem neuen Fragebogen besonders gut?

Offene Frage (n=14)

1. Nichts wesentlich anderes.
2. Genaue Auflistung der Veranstaltung/en, die es zu evaluieren gilt.
3. Möglichkeit einzelne Dozenten/Veranstaltungen zu bewerten.
4. man wird persönlich angeschrieben
5. übersichtlicher
6. übersichtlicher
7. übersichtlicher
8. gut strukturiert
9. übersichtlicher, das fordert unbewusst auf, eher bei der evaluation mitzumachen, als zuvor

Abbildung 3.10: Auszug eines automatisch generierten Berichts

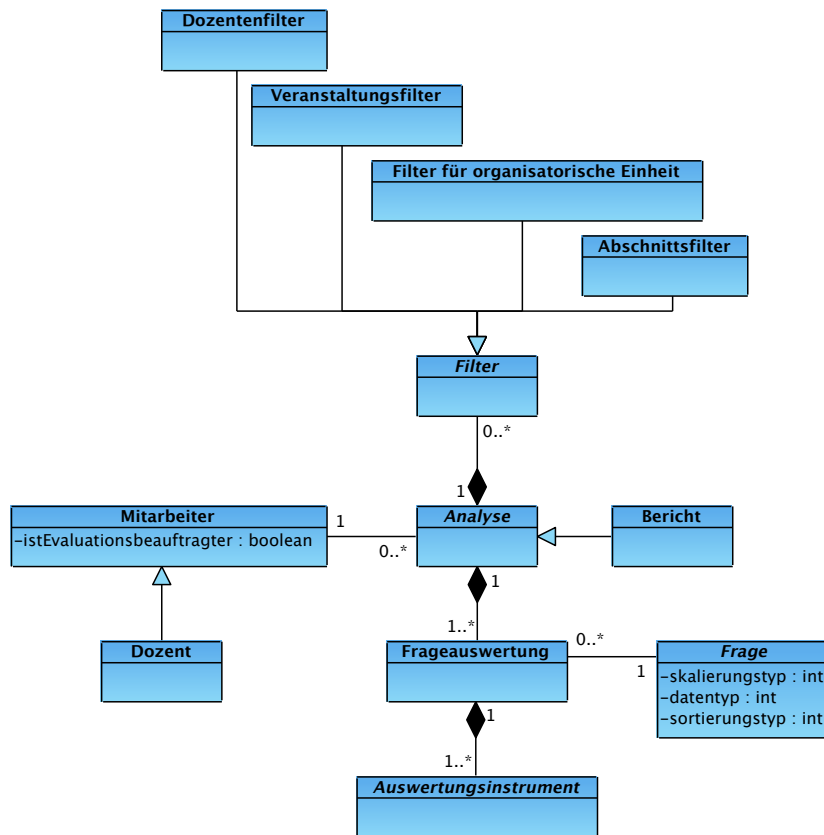


Abbildung 3.11: Klassendiagramm des Berichts

Skalierung und Datentyp von der auszuwertenden Frage und legt daraufhin die zulässigen Auswertungsinstrumente an.

Da die Bereinigung der Evaluationsergebnisse um Störvariablen in der Literatur zwar gefordert wird, konkrete Umsetzungshinweise jedoch nicht beschrieben werden (siehe Abschnitt „Störvariablen“ auf den Seiten 16–17), dieses Instrument auch in der Praxis kaum zum Einsatz kommt (siehe Abschnitt „Berücksichtigung von Störvariablen“ auf den Seiten 32–33) und auch nicht als vorrangig relevant gefordert wird (siehe Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41), wurden entsprechende Verfahren in dem hier beschriebenen Evaluationssystem nicht implementiert.

### 3.2.4 Rankings

Zusätzlich zu den Berichten bietet das System eine weitere Form des individuellen Feedbacks an. So erzeugt das System für jedes Semester ein eigenes Dozenten-Ranking aus den Fragen, welche den folgenden benutzerdefinierten Kriterien entsprechen, die beim Anlegen der Frage definiert werden (siehe Abbildung 3.8 auf Seite 64):

- Benötigtes Mindest-Skalenniveau ist vorhanden<sup>84</sup>
- Sortierung (die ‚Besser-Richtung‘) ist definiert
- Die Frage wurde zur Verwendung im Ranking freigegeben

Zur Wahrung der Effizienz übernimmt das System die Erstellung und Aktualisierung der Rankings vollkommen selbständig. Das System berücksichtigt dabei automatisch nur die Antworten, die sich auf dozentenspezifische Fragebögen beziehen. Aus der so eingeschränkten Menge an Antworten berechnet das System einen auf den Wertebereich von 0 bis 100 normierten Score  $S$  für eine  $j$ -te Frage wie folgt:

$$S_j = \begin{cases} \frac{1}{n} \sum_{i=1}^n \frac{A_{ij} - F_{\min j}}{F_{\max j} - F_{\min j}} \cdot 100 & \text{für } F_{\min j} \prec F_{\max j} \\ \frac{1}{n} \sum_{i=1}^n \frac{F_{\min j} - A_{ij}}{F_{\min j} - F_{\max j}} \cdot 100 & \text{für } F_{\min j} \succ F_{\max j} \end{cases} \quad (3.7)$$

Jeder der beiden Fälle dieser Gleichung berechnet das arithmetische Mittel aller um den individuellen Wertebereich  $[F_{\min j}; F_{\max j}]$  einer  $j$ -ten Frage bereinigten  $n$  Antwortwerte  $A$  auf diese Frage und multipliziert es mit 100 um einen fragespezifischen Score  $S_j$  im Intervall  $[0; 100]$  zu erhalten, der auch als natürlich Zahl dargestellt noch ausreichend Differenzierungskraft besitzt. Null stellt den minimalen, 100 stellt den maximalen Wert dar. Die Fallunterscheidung berücksichtigt dabei sowohl aufsteigend (‚größer ist besser‘ bzw.  $F_{\min j} \prec F_{\max j}$ ) wie auch absteigend (‚kleiner ist besser‘ bzw.  $F_{\min j} \succ F_{\max j}$ ) definierte Präferenzen.<sup>85</sup>

Der Dozenten-Score  $S$  wird als arithmetisches Mittel der Scores aller berücksichtigten  $m$  Fragen berechnet:

$$S = \frac{1}{m} \sum_{j=1}^m S_j \quad (3.8)$$

Somit weisen sowohl der fragespezifische Score  $S_j$  wie auch der Dozenten-Score  $S$  denselben Wertebereich von 0 bis 100 auf. Diese Normierung ermöglicht sowohl einen direkten Vergleich zwischen den Dozenten wie auch zwischen den Fragen. Diese Vergleichsmöglichkeit ist wiederum die Grundlage für die nachfolgend beschriebene Stärken-/Schwächenanalyse.

Durch die Bereinigung des fragespezifischen Wertebereichs ist es dem System möglich, nicht nur bipolare Fragen mit unterschiedlichen Skalen zu berücksichtigen, sondern auch einen Score für andere geschlossene und offene Fragen zu berechnen, welche die oben genannten Anforderungen erfüllen. Dieser normierte Score ist also zwischen den Fragen beliebig vergleichbar, auch wenn die Fragen beispielsweise bipolare mit 4, 5 und 10 Einheiten oder

<sup>84</sup> Neben geschlossenen Fragen mit entsprechender Skalierung können durch die Angabe eines Datentyps (und ggf. eines Wertebereichs) auch offene Fragen berücksichtigt werden.

<sup>85</sup> Durch die Berücksichtigung der Präferenzen in der Fallunterscheidung können die Ausprägungen geschlossener Fragen einheitlich mit aufsteigenden Werten kodiert und angezeigt werden.

offen mit einem Wertebereich von 20 bis 70 abgefragt wurden. Durch die Fallunterscheidung können die Fragen beliebig formuliert werden um beispielsweise den Score der Frage „Der Dozent verfügte über ein hohes Fachwissen“ mit der Frage „Der Dozent kam häufig zu spät“ direkt vergleichen zu können. Diese Eigenschaften sind die Voraussetzung dafür, die Scores aller berücksichtigten Fragen für einen Dozenten (durch das arithmetische Mittel) zu einem Dozenten-Score verrechnen zu können. Die Rangplätze der Dozenten ergeben sich schließlich durch die absteigende Sortierung nach dem jeweiligen Dozenten-Score. Ein Beispiel eines Rankings zeigt Abbildung 3.12 auf der nächsten Seite mit fiktiven Dozenten-Namen.

Das Ranking ist in die drei Bereiche überdurchschnittlich und unterdurchschnittlich sowie nicht bewertete Dozenten unterteilt. Die Einteilung in über- und unterdurchschnittliche Bewertungen erfolgt anhand eines rankingbezogenen Durchschnittswerts (im Beispiel beträgt dieser 77,21), der analog zur Berechnung in Gleichung 3.8 auf der vorherigen Seite als arithmetisches Mittel aller Scores gebildet wird und gleichzeitig als Score des gesamten Rankings gesehen werden kann.

Diese Sicht auf das Ranking, welche die Namen aller Dozenten offenlegt, erhält nur ein Evaluationsbeauftragter, um mit den Dozenten fundiert individuelle Gespräche führen zu können. Darüber hinaus stellt das System wie in Abbildung 3.13 auf Seite 73 gezeigt für jeden im System vorhandenen Dozenten ein persönliches und gegenüber den anderen Dozenten anonymisiertes Ranking zur Verfügung, so dass jeder Dozent seine persönliche Einstufung einsehen kann.

Um den unter Abschnitt „Leistungsorientierte Mittelvergabe“ auf Seite 23 beschriebenen Informationsverlust zu vermeiden, der sich bei Rankings ergibt, die nur einen ordinalen Rangplatz angeben, zeigt das Ranking zusätzlich den Dozenten-Score an, der eine höhere Aussagekraft als die reine Rangzahl hat. Darüber hinaus ermöglicht es das System wie in Abbildung 3.14 auf Seite 74 gezeigt, die Teil-Scores der einzelnen Fragen aufzuschlüsseln, aus denen sich der Score berechnet. Dozenten können der Anonymisierung entsprechend nur ihren Score aufschlüsseln, Evaluationsbeauftragte können die Scores aller Dozenten aufschlüsseln.

Die Detailansicht zeigt dabei sowohl eine grafische wie auch eine tabellarische Darstellung des dozentenspezifischen Scores in Relation zu Minimum, Durchschnitt und Maximum der jeweiligen Frage bezogen auf die Werte aller Dozenten. Beide Darstellungen sind dabei nach der Stärke des Dozenten absteigend sortiert. An erster Stelle steht also die am besten, an letzter Stelle die am schlechtesten bewertete Frage. Dieses Beispiel zeigt, dass der Dozent „Prof. Boerne“ nach der Meinung der Studenten besonders pünktlich war und über ein besonders hohes Fachwissen verfügt, jedoch Schwächen bei der Strukturierung des Stoffs und der Aufbereitung der Unterrichtsmaterialien hat. Bei den Schwächen (also Fragen mit unterdurchschnittlichem Score), zeigt das System zusätzlich eine Verknüpfung zu Informationsseiten („Wie kann ich mich verbessern“) auf denen speziell für die jeweilige Frage Hinweise zur Verbesserung der entsprechenden Schwachstelle enthalten sind. Über diesen schnellen und vor allem anonymen Einstieg, können die Dozenten erste Hinweise zur Verbesserung abrufen. Darüber hinaus kann ein Evaluationsbeauftragter die einzelnen Dozenten auf Basis der Er-

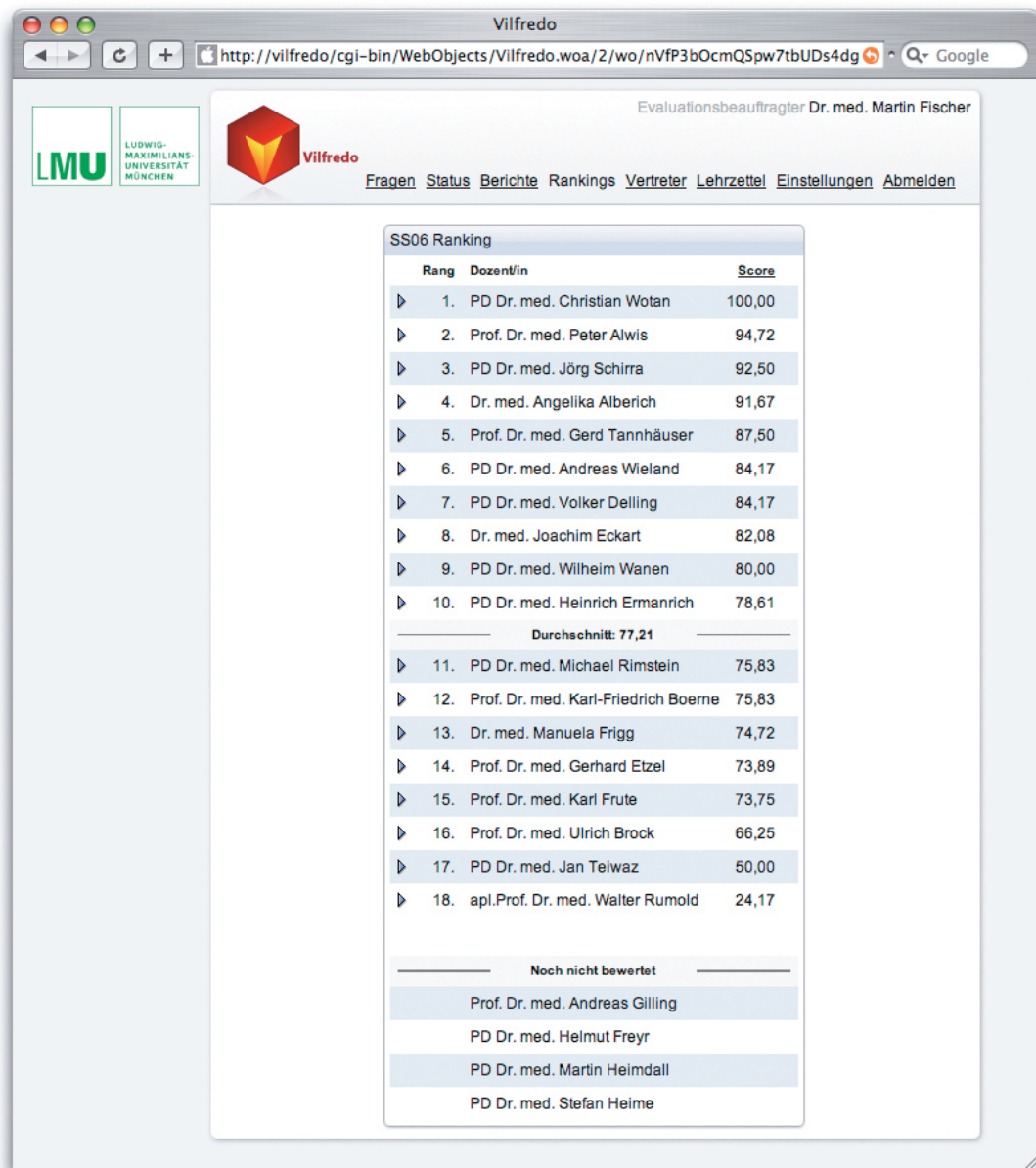


Abbildung 3.12: Automatisch erstelltes Ranking für den Evaluationsbeauftragten

kenntnisse aus den Rankings auch persönlich ansprechen, um individuelle Maßnahmen zu erörtern.

Technisch gesehen, ist das Ranking dem Bericht sehr ähnlich. Es bezieht sich auf eine definierte Menge an Fragen, deren Antworten aus einem bestimmten Zeitraum mit einem bestimmten Werkzeug ausgewertet werden müssen. Das Auswertungswerkzeug – die Berech-

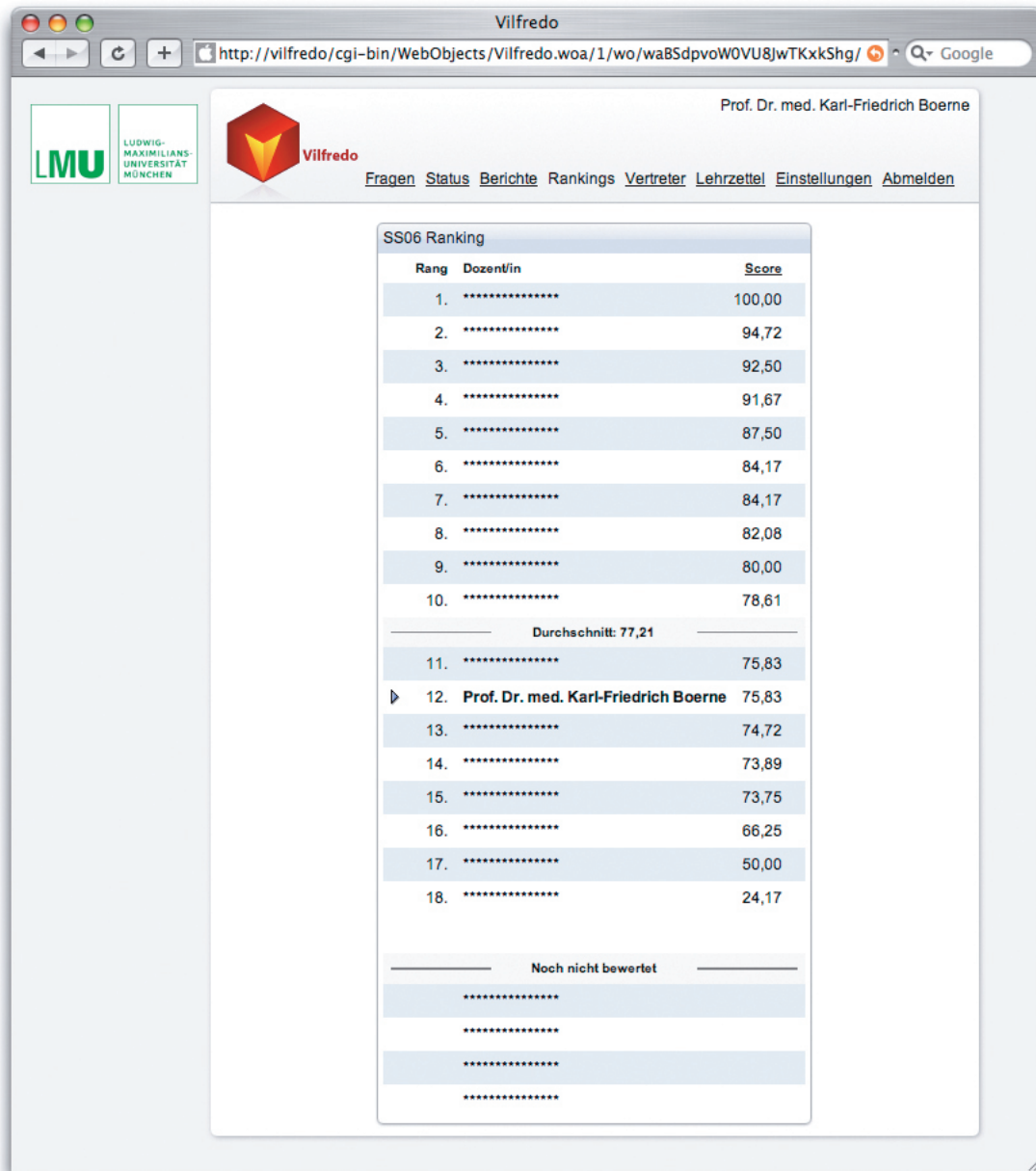


Abbildung 3.13: Individuelles Ranking für einen Dozenten

nung des Scores – wird sowohl beim Bericht (siehe Abbildung 3.10 auf Seite 68) wie auch beim Ranking (siehe insbesondere Abbildung 3.14 auf der nächsten Seite) eingesetzt. Aus diesem Grund wurde im Sinne der Objektorientierung eine effiziente<sup>86</sup> Modellierung vorgenommen,

<sup>86</sup> Eine gewisse Redundanz ergibt sich lediglich aus der parallelen Speicherung der Gesamtscores sowohl im „Rangplatz“ wie auch im spezifischen Auswertungsinstrument.

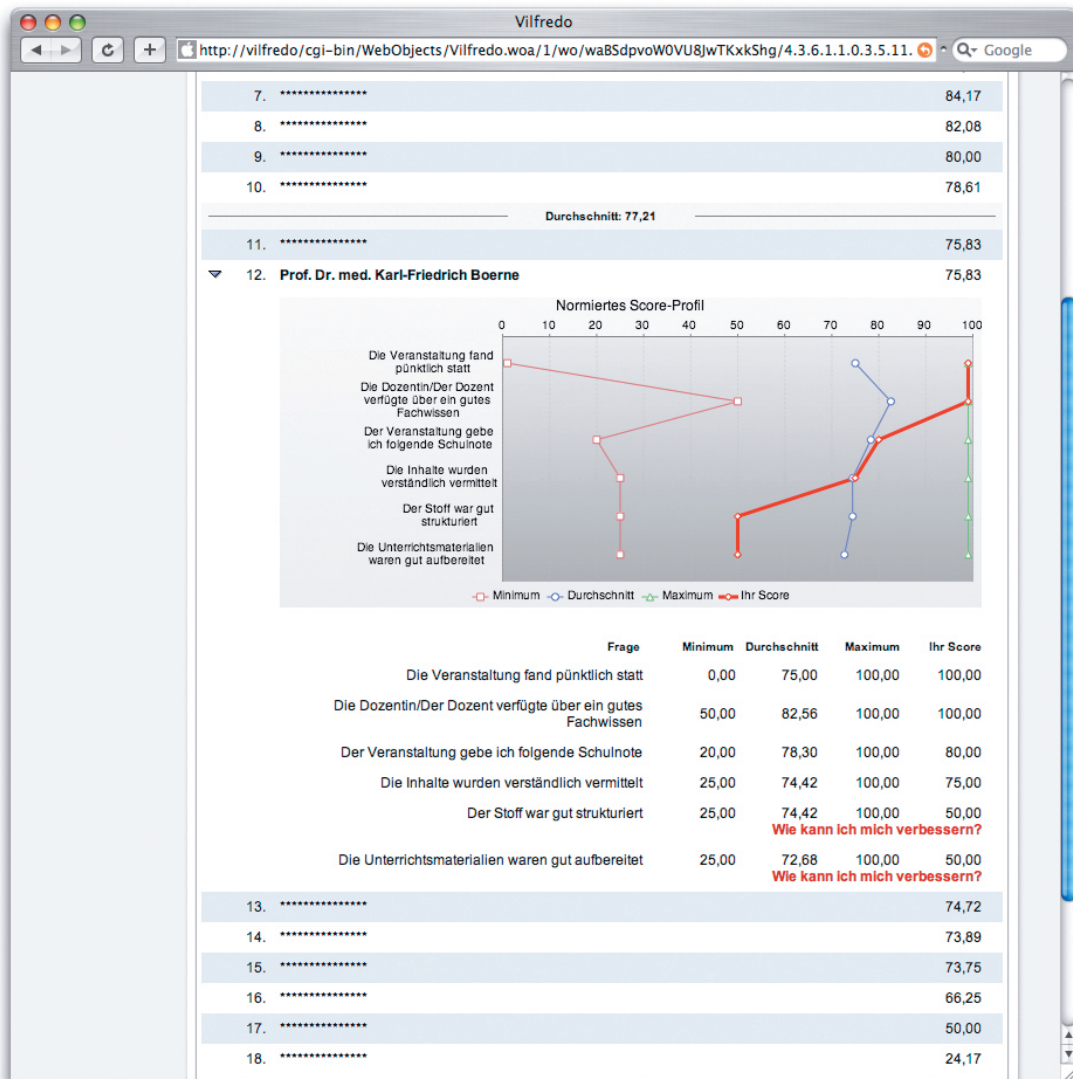


Abbildung 3.14: Individuelles Ranking für einen Dozenten mit Stärken-/Schwächenanalyse

bei der die abstrakte Super-Klasse „Analyse“ alle Relationen (und einen Teil der Logik) hält und diese an die Sub-Klassen „Bericht“ und „Ranking“ vererbt. Aus diesem Grund lässt sich das Ranking unter Bezugnahme auf die Modellierung des Berichts (siehe Abbildung 3.11 auf Seite 69) wie in Abbildung 3.15 auf der nächsten Seite dargestellt, formal sehr einfach modellieren.

Ebenfalls in dem Modell abgebildet ist die Grundlage für die Stärken-/Schwächenanalyse, die sich über die Hilfsrelation „Bewertung“ zwischen dem Rangplatz und der Frage herstellen lässt. Diese Analyse kann das System wie in Abbildung 3.14 gezeigt sehr schnell durch eine ab-



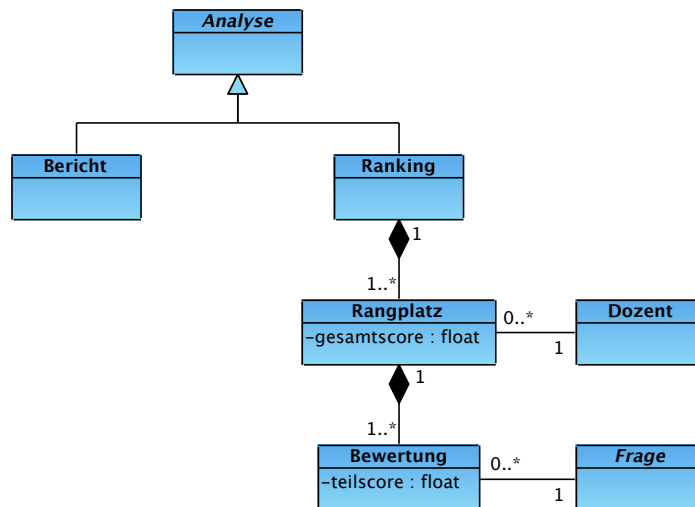


Abbildung 3.15: Klassendiagramm des Rankings

steigende Sortierung der Fragen nach dem in der Bewertung enthaltenen Teil-Score darstellen.

### 3.2.5 Unterstützende Funktionen

Nach der Darstellung ausgewählter Kernfunktionen sollen nun zwei unterstützende Funktionen vorgestellt werden, welche den administrativen Aufwand bei der Benutzung des Systems weiter reduzieren und damit der Anforderung einer hohen Effizienz nachkommen.

#### Lehrzettel

Das System muss über möglichst aktuelle und detaillierte Daten verfügen, um eine dozentenbezogene Evaluation sinnvoll durchführen zu können. Da die geplanten Termine in der Regel in elektronischer Form vorliegen (aus irgendeiner Datenquelle muss das Vorlesungsverzeichnis ja erstellt werden), sollte in jedem Fall versucht werden, auf diese Daten nach den unter Abschnitt „Stammdatenbezug“ auf den Seiten 47–48 beschriebenen Vorgaben, aus dem Evaluationssystem heraus zugreifen zu können.

Schwieriger gestaltet sich in der Praxis jedoch der Zugriff auf die Ist-Daten – also die Informationen darüber, welche Termine von welchen Dozenten tatsächlich gehalten wurden. Wie unter Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27 bereits beschrieben, werden Termine insbesondere im medizinischen Curriculum häufig von anderen Dozenten gehalten, als ursprünglich geplant. Diese ex-post-Sicht wird oftmals nicht strukturiert und vor allem nicht zentral verfügbar festgehalten. Aus diesem Grund implementiert das System die Funktion elektronischer und dozentenspezifischer Lehrzettel. Ein Lehrzettel in diesem Verständnis ist eine Liste aller Termine, für die ein Dozent ursprünglich eingeplant wurde.

Das System verwaltet Lehrzettel für jeden registrierten Dozenten. Abbildung 3.16 zeigt einen solchen Lehrzettel für einen Dozenten.

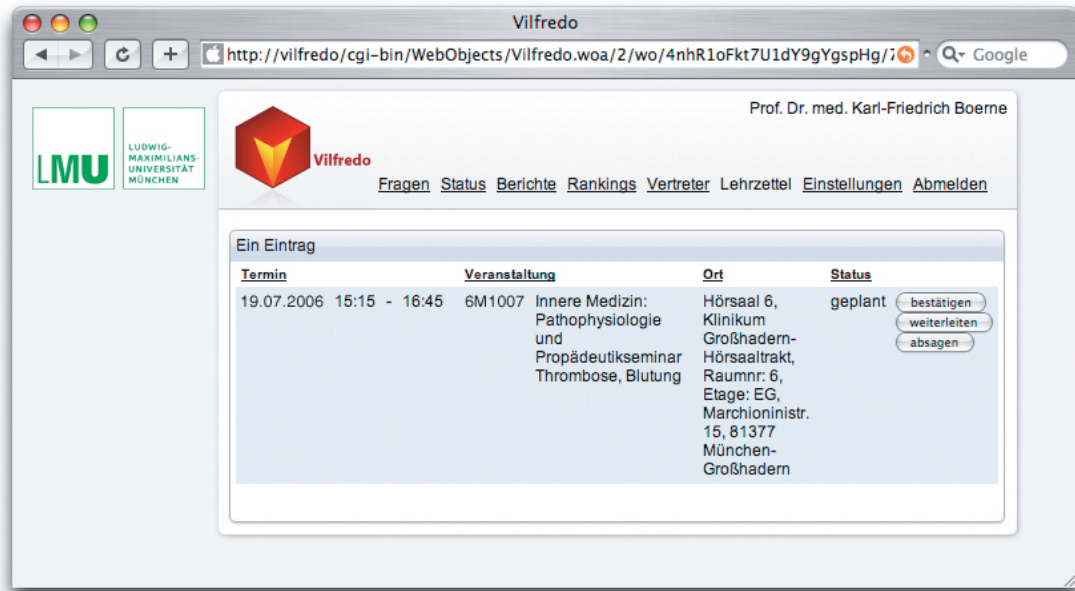


Abbildung 3.16: Darstellung eines Termins im Lehrzettel

Dieser Lehrzettel zeigt einen einzigen Termin, für welchen der Dozent eingeplant wurde, wobei alle relevanten Informationen angegeben werden, die der Dozent benötigt, um den Termin eindeutig identifizieren zu können. Hat der Dozent diesen Termin tatsächlich gehalten oder wurde er abgesagt, so kann er dies dem System durch einen einfachen Maus-Klick auf den jeweiligen Knopf mitteilen. Hat der Dozent diesen Termin nicht gehalten, kann er ihn an einen anderen Dozenten weiterleiten. In diesem Fall bietet ihm das System die Liste aller Dozenten zur Auswahl an um – falls bekannt – den Dozenten zu bestimmen, der diesen Termin tatsächlich gehalten hat. Ein Dozent an den ein Termin übertragen wurde, erhält diesen Termin automatisch in seinem Lehrzettel, auf dem ihm wieder dieselben Optionen zur Bestätigung, Weiterleitung und Absage zur Verfügung stehen.

Bei der Erhebung berücksichtigt das System alle geplanten und bestätigten Termine. Termine die abgesagt wurden, werden von der Erhebung ausgeschlossen. Übertragene Termine werden so lange nicht berücksichtigt, bis sie von einem Dozenten angenommen wurden. Abbildung 3.17 auf der nächsten Seite stellt die Status eines Termins und deren Übergänge anhand eines Zustandsautomaten formal dar.

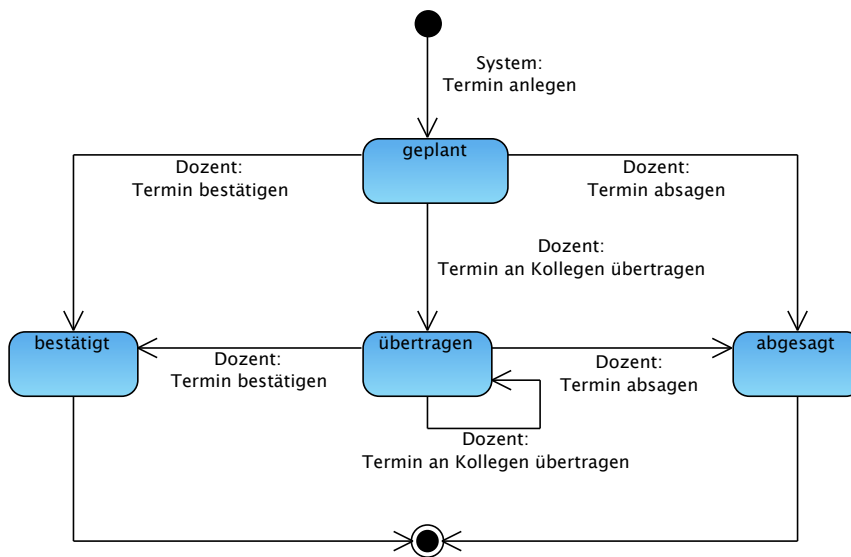


Abbildung 3.17: Zustandsautomat eines Termins

## Vertreter

Eine Funktion, welche den Umgang mit dem System erleichtert und die Anwendungsfälle praxisgerechter macht ist die Möglichkeit, Vertreter einrichten zu können. Vertreter im Sinne des Systems sind Mitarbeiter, die einen eigenen Zugang zum System haben und denen bestimmte Funktionen so freigeschaltet werden können, als würden sie diese unter der Anmeldung des Vertretenen durchführen. Damit leistet auch diese Funktion einen Beitrag zum Persönlichkeitsschutz, weil keine Zugangsdaten mehr (beispielsweise an die Sekretärin) weitergegeben werden müssen.

Abbildung 3.18 auf der nächsten Seite zeigt die Einrichtung einer Vertreterin, welche Berichte anlegen und einsehen, diese jedoch nicht löschen darf.

Meldet sich die von „Prof. Boerne“ eingerichtete Vertreterin „Alberich“ an, so kann sie – wie in Abbildung 3.19 auf Seite 79 gezeigt – über ein Menü am oberen Rand auswählen, ob sie in eigenem Namen oder für einen Vertretenen im System arbeiten möchte.

Wie in der Abbildung zu sehen, kann die Vertreterin den ihr vom Vertretenen zugewiesenen Rechten entsprechend Berichte anlegen und herunterladen, sie jedoch nicht löschen. Um der Vorgabe einer einfachen Bedienung zu entsprechen, kann eine Person die Vertretung für mehrere Personen – jeweils mit unterschiedlichen Rechten – übernehmen und die Vertretung über das entsprechende Menü jederzeit beliebig wechseln.

Eine formale Darstellung der technischen Umsetzung zeigt das Klassendiagramm der Abbildung 3.20 auf Seite 79.

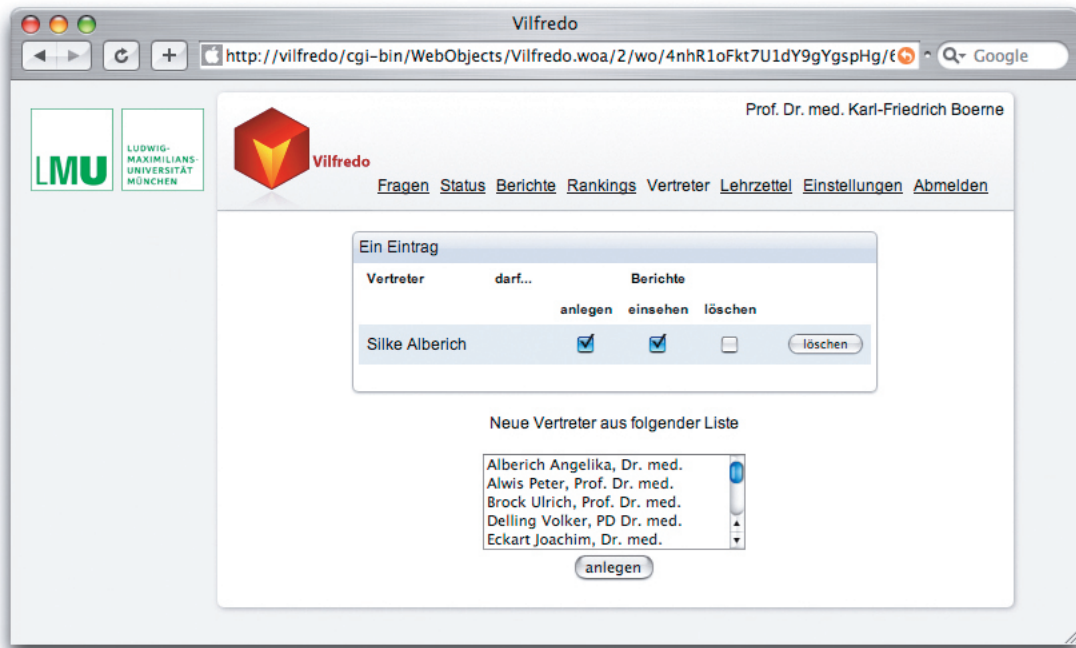


Abbildung 3.18: Einrichtung einer Vertretung

### 3.2.6 E-Learning

Wie in der Studie unter Abschnitt „E-Learning“ auf Seite 33 gezeigt, bieten 23 Fakultäten (74 %) E-Learning-Kurse an. Insbesondere wenn diese Kurse Bestandteil des Curriculums sind, sollten diese auch evaluiert werden. Da für die Evaluation von E-Learning-Kursen sicherlich (auch) andere Fragen gestellt werden als für die von Dozenten gehaltenen Veranstaltungen, stellt das System die unter Abschnitt „Fragen“ auf den Seiten 60–61 beschriebene Funktion zur Verfügung, einzelne Fragen für unterschiedliche Erhebungseinheiten freizugeben. Die in Abbildung 3.1 auf Seite 59 gezeigte Erhebungseinheit einer abstrakten Veranstaltung kann im System auch in Form eines E-Learning-Kurses konkretisiert werden. Abbildung 3.21 auf Seite 80 zeigt die formale Modellierung dieses Zusammenhangs, nach dem eine Veranstaltung auch ein E-Learning-Kurs sein kann, auf den sich ein entsprechend konkretisierter Fragebogen beziehen kann.

Nachdem E-Learning-Kurse in der Regel nicht zu festen Zeiten abgehalten werden – dies würde die Vorteile des selbst gesteuerten Lernens ad absurdum führen – kann das System die Freigabe eines Fragebogens für E-Learning-Kurse im Gegensatz zu den von Dozenten gehaltenen Veranstaltungen jedoch nicht selbst bestimmen. Eine auf einen E-Learning-Kurs bezogene Erhebung sollte dann angestoßen werden, wenn der oder die Teilnehmer einen solchen Kurs beendet haben. Diesen Zeitpunkt kennt jedoch nur das E-Learning-System. Aus

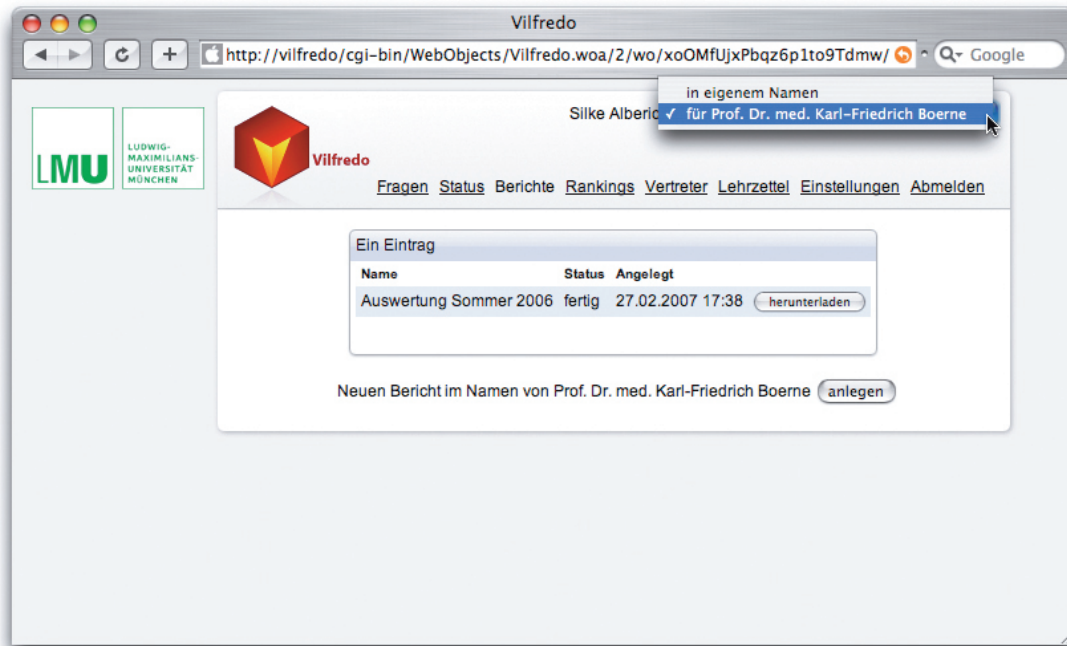


Abbildung 3.19: Nutzung einer Vertretung

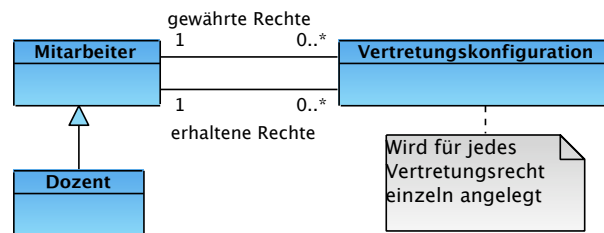


Abbildung 3.20: Klassendiagramm der Vertretung

diesem Grund stellt das System externen E-Learning-Systemen wie beispielsweise CASUS<sup>87</sup> eine Schnittstelle zur Verfügung, über welche sie die Generierung eines Fragebogens in Vilfredo beauftragen können. Über diese Schnittstelle teilt das E-Learning-System Vilfredo mit, welche Studierende welchen E-Learning-Kurs bearbeitet haben. Vilfredo generiert daraufhin die entsprechenden Fragebögen und antwortet dem beauftragenden E-Learning-System mit einer Liste von Objekten, die neben weiteren unterstützenden Informationen die jeweiligen URLs<sup>88</sup> enthalten, über welche die Studierenden ihren persönlichen Fragebogen aufrufen

<sup>87</sup> Weitere Informationen zu diesem System sind unter <http://www.casus.net> zu finden.

<sup>88</sup> URL = Uniform Resource Locator. Eine eindeutige Internet-Adresse. Eine URL kann insbesondere auf eine Web-Seite verweisen.

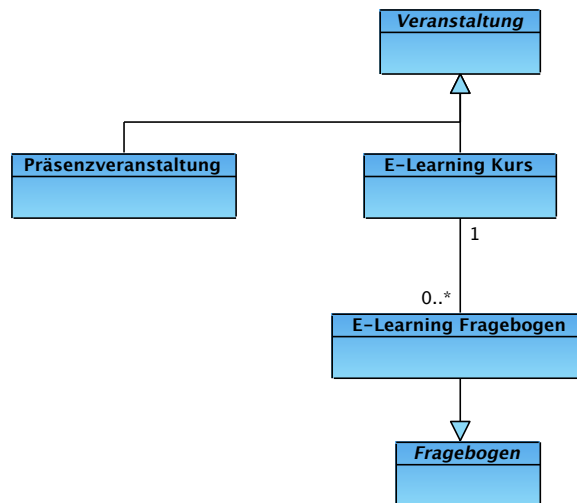


Abbildung 3.21: Klassendiagramm der Evaluation von E-Learning-Kursen

können. Eine solche URL enthält bereits die Transaktionsnummer um auch bei diesem Zugriff die Berechtigung bei der Einhaltung der Anonymität sicherzustellen und dem E-Learning-System gleichzeitig ebenfalls einen direkten Zugriff auf den Fragebogen zu ermöglichen. Das E-Learning-System kann diese URLs dann in beliebiger Form an die Kursteilnehmer übermitteln; es kann sie also beispielsweise dazu verwenden, um die Fragebögen gleich im Anschluss an die Bearbeitung des Kurses direkt anzuzeigen, oder, um sie ebenfalls per E-Mail an die Teilnehmer zu verschicken.

Die Schnittstelle für diese Dienste – auch API<sup>89</sup> genannt – wurde dem Stand der Technik entsprechend als Web Service implementiert, der über unspezifische Datentypen kommuniziert. Diese Schnittstelle, die im Anhang C auf den Seiten 113–115 technisch beschrieben ist, erlaubt damit der unter Abschnitt „Effizienz, Benutzerfreundlichkeit und Flexibilität“ auf Seite 46 aufgestellten Anforderung entsprechend eine maximale Flexibilität für die Kommunikation mit externen E-Learning-Systemen, die in unterschiedlichsten Programmiersprachen erstellt wurden.

### 3.3 Piloteinsatz

Im vorherigen Abschnitt wurde die Umsetzung ausgewählter Anforderungen durch ein konkretes System gezeigt. Dieser Abschnitt beschreibt den ersten Einsatz dieses Systems und die dabei gewonnenen Erfahrungen im Umgang mit den neuen Funktionen an der Medizinischen Fakultät der Ludwig-Maximilians-Universität München im Sommersemester 2006. Die Planung des Piloteinsatzes erfolgte in enger Zusammenarbeit mit dem Evaluationsbeauftragten

<sup>89</sup> API = Application Programming Interface

der Fakultät, Dr. Martin Fischer. Der Fokus des Einsatzes lag dem Thema der Arbeit entsprechend nicht auf pädagogisch-psychologischen Aspekten wie der Validierung der Fragen, sondern auf der Überprüfung, ob und wie die oben dargestellten neuen Funktionen in der täglichen Praxis eingesetzt werden können. Dieser Testlauf war wichtig, da das System einige neue Abläufe und Voraussetzungen erfordert um die in der Diskussion im vierten Kapitel auf den Seiten 85–90 weiter ausgeführten Vorteile realisieren zu können.

### 3.3.1 Einrichtung

Das System wurde am Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE) betrieben. Aufgrund der schlanken Architektur reichte dabei ein einziger Rechner mit einer – bezogen auf den Stand der Technik – sehr geringen Ausstattung aus.<sup>90</sup> Dieser Rechner übernahm damit die Funktionen eines Datenbank<sup>91</sup>-, Applikations<sup>92</sup>- und Web-Servers<sup>93</sup>. Als Testgruppe wurden die jeweils letzten Termine aus einer Vorlesung und zwei Seminaren der Inneren Medizin im Sommersemester 2006 ausgewählt.<sup>94</sup>

Da die Vorlesung alle Studierenden dieses Abschnitts einbezog, die Studierenden auf die beiden Seminare jedoch jeweils in 12 Gruppen aufgeteilt wurden, ergaben sich insgesamt 25 zu evaluierende Termine, die von 22 unterschiedlichen Dozenten gehalten wurden. Diesen Terminen waren insgesamt 181 Studierende zugewiesen.

Da zum Zeitpunkt des Piloteinsatzes noch keine direkte Anbindung zum Bezug der Daten aus einem Zentralsystem zur Verfügung stand, wurde das Import-Modul des Evaluationsystems so angepasst, dass die vom Dekanat erstellten Listen im Microsoft-Excel-Format eingelesen werden konnten. Es lag für jede der oben genannten Veranstaltungen jeweils eine Liste vor, in welcher die Termine mit den jeweiligen Themen, die Dozenten, die Studierenden sowie die Gruppen enthalten waren. Jede Veranstaltung konnte über eine Veranstaltungsnummer und jeder Student über die Matrikelnummer eindeutig referenziert werden. Die Termine wurden über einen zusammengesetzten Primärschlüssel aus Datum, Uhrzeit und Gruppe in Verbindung mit der Veranstaltung eindeutig referenziert. Die Zuweisungen der Dozenten auf die Termine waren im Zentralsystem zwar vorhanden, konnten in diesen Listen

<sup>90</sup> Verwendet wurde ein herkömmlicher Apple Macintosh G4 Rechner, mit einer Taktfrequenz von 800 MHz, 256 MB Hauptspeicher und 120 GB Festplattenspeicher, der unter dem Betriebssystem Mac OS X Server 10.4 betrieben wurde. Da die Applikation lediglich eine Java-Umgebung (ab Version 1.4) für den Betrieb benötigt, hätte das System jedoch auch problemlos auf gängigen Sun-, Linux- oder Windows-Systemen betrieben werden können.

<sup>91</sup> PostgreSQL 8.2 (<http://www.postgresql.org>)

<sup>92</sup> WebObjects 5.3 (<http://www.apple.com/webobjects>)

<sup>93</sup> Apache 1.3 (<http://httpd.apache.org>)

<sup>94</sup> Zur Sicherstellung einer umfassenden Anonymisierung der in dieser Arbeit dargestellten Ergebnisse des Piloteinsatzes wird neben der Anonymisierung der Personennamen an dieser Stelle auf die explizite Angabe der Veranstaltungen verzichtet.

Dem Titel der Arbeit entsprechend bestand das Ziel des Piloteinsatzes darin, die Verwendbarkeit der neuen Ansätze von Vilfredo in der Praxis zu zeigen. Aus diesem Grund wurde auf die Einrichtung einer Kontrollgruppe für den Vergleich der Umfrageergebnisse zwischen unterschiedlichen Systemen verzichtet und dient als Anknüpfungspunkt für weitere Arbeiten.

jedoch nicht mit ausgegeben werden. Deshalb mussten diese Informationen händisch in den Listen nachgetragen werden. Zur eindeutigen Referenzierung der Dozenten wurde deren E-Mail-Adresse verwendet. Da damit alle Entitäten (Veranstaltungen, Termine, Dozenten und Studierende) eindeutig referenziert werden konnten, war das System in der Lage, Änderungen durch erneutes Einlesen der Listen auf den internen Datenbestand zu übertragen.

Da kein direkter Zugriff auf die Datenquelle des Zentralsystems möglich war, konnten die Probleme der Datenredundanz im Evaluationssystem und der nicht permanenten Datenaktualität nicht vermieden werden (siehe Abschnitt „Stammdatenbezug“ auf den Seiten 47–48). Über die eindeutige Objektreferenzierung wurde jedoch sichergestellt, dass alle benötigten Daten initial eingelesen und Änderungen über denselben Weg übernommen werden konnten.

Die Fragen wurden über den im System eingebauten Editor in den folgenden drei Fragebereichen manuell<sup>95</sup> eingegeben.

1. Fragen zur Veranstaltung
2. Fragen zum Dozenten
3. Fragen zur Evaluation

Die Fragen zur Veranstaltung und zum Dozenten wurden aus dem bestehenden Evaluationssystem übernommen. Hinzugenommen wurden für den Piloteinsatz weiterhin fünf Fragen zur Evaluation selbst. Die Ergebnisse dieser Fragen zur Evaluation finden sich in Form eines vom System generierten Berichts in Anhang B auf den Seiten 99–111.

### **3.3.2 Durchführung**

Nach der Einrichtung der Daten und der Festlegung der Zeitspanne, für die eine Umfrage zur Beantwortung verfügbar ist, konnte das System die Erhebung vollkommen selbständig durchführen. Dabei prüfte es im Intervall von fünf Minuten ob Termine, die noch nicht evaluiert wurden, seit der letzten Prüfung beendet wurden. Für die so ermittelten Termine erstellte das System wie unter Abschnitt „Fragebogen“ auf den Seiten 58–60 beschrieben, einen dozentenbezogenen Fragebogen und verschickte jeweils eine E-Mail mit der Zugangsberechtigung an alle einem Termin zugewiesenen Studierende. Parallel dazu aktualisierte das System täglich das automatisch erstellte Dozenten-Ranking. Das mit fiktiven Namen anonymisierte Endergebnis des Rankings im Piloteinsatz wurde bereits in Abbildung 3.12 auf Seite 72 gezeigt.

Der Piloteinsatz verlief fehlerfrei.

---

<sup>95</sup> Da nur eine verhältnismäßig geringe Anzahl an Fragen zu übernehmen war, wurde auf die Implementation eines Moduls zur automatischen Übernahme der Fragen verzichtet.



### 3.3.3 Bewertung

Die Medizinische Fakultät setzt für die Evaluation bislang das Open-Source-System „PHPSurveyor“<sup>96</sup> ein, das auf die Generierung und Auswertung von Umfragen spezialisiert ist. Gegenüber dem bisherigen System haben sich beim Piloteinsatz von Vilfredo insbesondere folgende Vorteile gezeigt.

- **Proaktive Evaluation mit Prüfung der Berechtigung**  
Bei der Evaluation mit PHPSurveyor mussten die Studierenden die zu evaluierenden Veranstaltungen selbst auswählen. Das System prüfte dabei weder die grundsätzliche Berechtigung, noch verhinderte es eine Mehrfachteilnahme. Vilfredo dagegen schrieb die zur Teilnahme an der Evaluation berechtigten Studierenden an und stellte den Zugang sowie die Beschränkung auf die Abgabe eines Fragebogens über eine TAN sicher.
- **Dozentenbezogene Evaluation**  
Im Gegensatz zum bestehenden System war mit Vilfredo eine dozentenbezogene Evaluation möglich, die wiederum dozentenspezifische Berichte sowie das Dozenten-Ranking ermöglichte.
- **Aufbereitete Berichte**  
Die grafische Aufbereitung der Berichte und die als PDF-Dokument standardisierte Ausgabe unter Vilfredo ermöglichte eine schnellere Übersicht und erlaubte es, die vom System generierten Dokumente auch an andere Personen weiterzugeben, ohne sie zusätzlich aufbereiten zu müssen. Durch den hohen Grad der Automatisierung ließen sich die Berichte sehr einfach erstellen. Mithilfe der Filter waren die Berichte dennoch individuell anpassbar.

Folgende Punkte stellten sich als nachteilig heraus.

- **Dateneinrichtung**  
Da das externe System, in welchem die benötigten Stammdaten hinterlegt sind, keine Schnittstelle zur Verfügung stellen konnte, mussten die benötigten Daten durch den Import der Dekanatslisten durchgeführt werden. Auch wenn die Ausrichtung des Import-Moduls von Vilfredo einen zügigen Ablauf ermöglichte, so mussten die Listen dennoch um die Dozentendaten manuell erweitert werden.
- **Keine konditionalen<sup>97</sup> Fragen**  
Während es das bestehende System erlaubte, konditionale Fragen einzurichten, ist diese Möglichkeit in Vilfredo nicht gegeben. Letztendlich zeigte sich jedoch, dass die konditionalen Fragen in Vilfredo ohnehin nicht mehr nötig waren, da sie in PHPSurveyor

<sup>96</sup> Weitere Informationen zu diesem System sind unter <http://www.phpsurveyor.org> zu finden.

<sup>97</sup> Bei konditionalen Fragen handelt es sich um Fragen, die in Abhängigkeit der Ergebnisse anderer Fragen gestellt bzw. angepasst werden.

nur für die Unterscheidung der Veranstaltungsarten eingesetzt wurden, die Vilfredo automatisch vornimmt.

Diese Gegenüberstellung von Vor- und Nachteilen bezieht sich rein auf die tatsächlichen Erkenntnisse aus dem Piloteinsatz. Eine umfassende Darstellung der Möglichkeiten und Grenzen wird im folgenden vierten Kapitel diskutiert.

## 4 Diskussion

### 4.1 Auswirkungen

Im vorangegangenen dritten Kapitel wurden die Anforderungen an ein Lehrevaluationssystem sowie deren Umsetzung anhand eines konkreten Systems und dessen Piloteinsatz an der Medizinischen Fakultäten der LMU München dargestellt. Der vorliegende Abschnitt erweitert diese Sichtweise und zeigt, wie sich diese Erkenntnisse auf andere Fakultäten übertragen lassen, welche besonderen Aspekte dabei zu berücksichtigen sind und welche Effekte sich dabei einstellen können. Da die unter Abschnitt „Technische Anforderungen“ auf den Seiten 43–57 dargestellten Anforderungen bereits aus einer Analyse der Literatur und der Praxis im zweiten Kapitel hergeleitet wurden, wird an dieser Stelle auf einen Abgleich der Erkenntnisse mit eben diesen Quellen verzichtet.

Zunächst soll auf einen ganz grundlegenden Aspekt hingewiesen werden – wenngleich dieser auf den ersten Blick trivial erscheinen mag. Da die Ergebnisse im dritten Kapitel in die Verdichtung der hergeleiteten Anforderungen (siehe Abschnitt „Technische Anforderungen“ auf den Seiten 43–57) und die darauf aufbauende Implementation eines konkreten Systems (siehe Abschnitt „Implementation“ auf den Seiten 58–80) unterteilt wurden, können die Erkenntnisse dieser Arbeit sowohl für die Beschaffung neuer Systeme (Anforderungen) wie auch für die Weiterentwicklung bestehender Systeme (Implementation) verwendet werden.

#### 4.1.1 Dozentenbezogene Evaluation

Zentrales Thema der im dritten Kapitel dargestellten Funktionen ist die dozentenbezogene Evaluation. Diese architektonische Grundlage eröffnet weit reichende Möglichkeiten. Sie hat gleichzeitig jedoch auch tief greifende Anforderungen an die System-Umwelt zur Folge. Die anspruchsvollste dieser Anforderungen ist der Zugriff des Evaluationssystems auf die aktuellen Daten darüber, welcher Dozent welchen Termin vor welchen Studierenden tatsächlich gehalten hat. Es muss ganz grundsätzlich eine Datenquelle existieren, in welcher die Zuweisung eines oder mehrerer Dozenten zu einem Termin gespeichert ist. Diese Datenquelle muss so strukturiert sein, dass sie von einem System ausgelesen werden kann. Eine reine Speicherung dieser Zuweisungen in einem Text-Dokument (z. B. als Microsoft-Word-Dokument) reicht dafür nicht aus. Aktuelle Daten werden benötigt, da diese insbesondere im Medizinischen Curriculum aus den unter Abschnitt „Fragmentierte Veranstaltungen“ auf den Seiten 25–26 und Abschnitt „Komplexe Organisationsstruktur“ auf den Seiten 26–27 dargestellten Gründen, häufig von den geplanten Daten abweichen. Wurde eine Veranstaltung nicht von dem Dozenten gehalten, der dafür eingeplant wurde und liegt diese Änderung dem

Evaluationssystem nicht vor, so bezieht sich die Erhebung auf den falschen Dozenten. Dies führt die dozentenbezogene Evaluation letztendlich ad absurdum.

Können diese Anforderungen nicht in letzter Konsequenz erfüllt werden, so gibt es Optionen zur Lösung dieser Einschränkungen. Insbesondere die Forderung nach aktuellen Daten kann dabei wie folgt beschrieben approximiert werden, sofern eine zentrale Datenquelle diese Aktualität nicht leisten kann, oder die Erfassung dieser Informationen über schriftlich geführte Anwesenheitslisten erst gar nicht (vollständig) möglich ist. Mit der unter Abschnitt „Lehrzettel“ auf den Seiten 75–76 beschriebenen Funktion der Lehrzettel lässt sich die Aufgabe, die Daten aktuell zu halten, auf die Dozenten verlagern. Einen ähnlichen Weg geht das System „Tuevalon“<sup>98</sup>, über das alle Dozenten, die für einen Termin eingeteilt wurden, per E-Mail danach gefragt werden, ob sie den Termin tatsächlich gehalten haben. In diesem Fall obliegt es der Verwaltung, den Dozenten die Reaktion auf diese Funktion vorzuschreiben. Die in Vilfredo dargestellte Funktion der Lehrzettel ließe sich so erweitern, dass ein Dozent den Termin nur bestätigen oder ablehnen kann. Zur Pflege der abgewiesenen Termine könnte das System einer Gruppe von Hilfskräften einen Zugang gewähren, um den tatsächlichen Dozenten zu ermitteln und dem System mitzuteilen. Zu beachten ist dabei, dass die Erhebung erst dann durchgeführt werden kann, wenn ein Termin vom Dozenten bestätigt wurde oder der Termin vom Dozenten abgewiesen wurde und der tatsächliche Dozent von einer Hilfskraft ermittelt wurde. Insofern entstünde auch bei dieser Weiterentwicklung die von der Verwaltung durchzusetzende Anforderung an die Dozenten, ihre elektronischen Lehrzettel ständig zu pflegen, da ansonsten die unter Abschnitt „Evaluationszeitpunkte“ auf Seite 18 geforderte zeitnahe Evaluation nicht möglich ist.

Der Hauptkritikpunkt an diesen Optionen liegt darin, dass die Dozenten kaum dazu verpflichtet werden können, dem System die benötigten Informationen immer und zeitnah mitzuteilen. Damit kann ein solches Verfahren allenfalls eine Approximation darstellen.<sup>99</sup> Dieser Mangel kann dadurch geheilt werden, dass eine Organisationseinheit (beispielsweise eine Gruppe studentischer Hilfskräfte) für jeden Termin prüft, von welchem Dozenten er gehalten wurde und damit für die Aktualität der Daten verantwortlich ist. Dieser Weg steht jedoch in Konflikt mit der Anforderung, den administrativen Aufwand möglichst gering zu halten (siehe Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41). Eine Lösung, welche mit der Minimierung des administrativen Aufwands kompatibel ist, wäre eine automatische Vor-Ort-Identifizierung von Personen mit Transpondern beispielsweise über eine Funkwellenidentifikation wie die RFID<sup>100</sup>. Die RFID-gestützte

---

<sup>98</sup> Zwar wird das System der Universität Tübingen zugeschrieben. Eine eindeutige Verantwortlichkeit ist jedoch nicht erkennbar, weshalb auf die Web-Site unter der Adresse <http://www.tuevalon.de> verwiesen wird.

<sup>99</sup> Es ist zu beachten, dass diese Approximation einen verzerrenden Effekt auf die Ergebnisse (z. B. Berichte oder Rankings) haben kann, wenn die geplanten nicht mit den tatsächlichen Dozenten übereinstimmen und sie von den Studierenden in den erhobenen Items unterschiedlich bewertet werden würden. Vor diesem Hintergrund lässt sich die Hypothese aufstellen, dass die Qualität der Lehrevaluation mit der Qualität bzw. Aktualität der Stammdaten positiv korreliert.

<sup>100</sup> RFID = Radio Frequency Identification

Identifizierung wird an der Medizinischen Fakultät der Technischen Universität München bereits für die Anwesenheitskontrolle der Studierenden geplant. Eine Identifizierung von Dozenten ist mit einem solchen Verfahren grundsätzlich genauso möglich und sollte bei der Diskussion um die Einführung von Klinikausweisen für die Mitarbeiter als mögliche Umsetzung bei der Einhaltung der rechtlichen Rahmenbedingungen berücksichtigt werden.

Kann die eindeutige und aktuelle Verknüpfung eines Termins mit dem bzw. den haltenden Dozenten jedoch sichergestellt werden, so ist es möglich, die Dozenten individuell über ihre Einschätzung durch die Studierenden zu informieren. Beispiele dafür sind die unter Abschnitt „Berichte“ auf den Seiten 65–69 vorgestellten, individualisierbaren Berichte, sowie die unter Abschnitt „Rankings“ auf den Seiten 69–75 gezeigte Ausprägung eines anonymisierten und nicht-öffentlichen Dozenten-Rankings. Durch diese Möglichkeit, personenbezogene Aussagen treffen zu können, ist es schließlich möglich, die Dozenten als ausführendes Organ der Lehre persönlich anzusprechen und individuelle Hinweise zur Verbesserung ihrer Lehrtätigkeit geben zu können. Damit trägt diese Funktion mittelbar zur Verbesserung der Lehre bei und erfüllt somit eines der Hauptziele der Lehrevaluation.

#### 4.1.2 Integration der Dozenten

Mit den individualisierten Berichten, den anonymisierten Dozenten-Rankings und der Möglichkeit, durch die Eingabe eigener Fragen Einfluss auf die Evaluation zu nehmen, steht ein wichtiger Funktionsbereich zur Verfügung, der einen Beitrag zur Verbesserung der Lehrevaluation leisten kann. Die Möglichkeit, Fragen selbst einzurichten und die Ergebnisse individuell auswerten zu können, stellt eine Grundlage dar, die Dozenten stärker in die Evaluation zu integrieren. Sie werden damit von passiven Empfängern der Ergebnisse (so sie diese überhaupt erhalten), zu aktiven Teilnehmern, die zur Gestaltung der Lehrevaluation beitragen können. Die Kombination aus individuellen Fragen und individuellen Berichten erlaubt es den Dozenten, persönliche Messpunkte einzurichten um herauszufinden, wie ganz spezifische Änderungen in ihrer Lehre von den Studierenden eingeschätzt werden. Die Darstellung der Ergebnisse in der hier gezeigten Form anonymisierter Dozenten-Rankings schließlich kann den Anreiz steigern, sich insgesamt zu verbessern ohne auf dem Weg dorthin kompromittiert zu werden. Auf eine grafisch übersichtliche und verständliche Aufbereitung wurde dabei großen Wert gelegt, da Rindermann (2001a, S. 80) dies als wichtigen Faktor bei der Rückmeldung beschreibt. Die anonym abrufbaren Hinweise zu Verbesserungen bei unterdurchschnittlich bewerteten Items stellen konkrete Hilfestellungen dar. Umfassende und persönliche Hilfestellungen können die Evaluationsbeauftragten über die vollständige Sicht auf das Ranking dadurch leisten, dass ihnen die Stärken und Schwächen der Dozenten fundiert und aufgeschlüsselt dargestellt werden. Gleichzeitig lässt die Aufschlüsselung wie von Titscher (2006, S. 208, 218) gefordert Platz für unorthodoxe Ideen und trägt damit dazu bei, ein ähnliches Dilemma wie bei der Diskussion um die reine Shareholder-Value-Orientierung in der Wirtschaft zu vermeiden.<sup>101</sup>

---

<sup>101</sup> Dieses Dilemma entsteht vor folgendem Hintergrund. Zum einen sollen Entscheidungsträger anhand von Kennzahlen entlohnt werden, die das Aktionärsvermögen (den Shareholder-Value) messen. Andererseits

Diese Formen der Einbeziehung der Dozenten stellen den Grundstein dar, auf dem individuelle Verbesserungen aufgebaut werden können, die in der Summe letztendlich zu einer gesamthaftern Verbesserung der Qualität der Lehre führen können.

### 4.1.3 Integration der Studierenden

Neben der Integration der Dozenten gilt es jedoch auch die Studierenden in die Evaluation möglichst stark zu integrieren (siehe Abschnitt „Rechtsnormen“ auf den Seiten 9–10, Abschnitt „Interessensgruppen“ auf den Seiten 10–13 sowie Bargel 2003, S. 41 f.). Aus der Sicht der Evaluation gilt es, das gravierende Problem der ausreichend hohen Rücklaufquote zu lösen und damit schließlich die Aussagekraft der Evaluation zu erhöhen. Aus Sicht der Studierenden geht es jedoch viel stärker darum, mit der Teilnahme an der Evaluation eine tatsächliche Verbesserung der Lehre herbeiführen zu können und damit zur Erfüllung des durchaus legitimen<sup>102</sup> Ziels beizutragen, die eigene Situation zu verbessern.

Zur Steigerung der Rücklaufquote sind die im dritten Kapitel dargestellten Funktionen darauf ausgelegt, die Teilnahme an der Evaluation so sicher, schnell und komfortabel wie möglich zu gestalten. Zentrales Element dabei ist der unter Abschnitt „Fragebogen“ auf den Seiten 58–60 dargestellte einfache Zugang zu einem Fragebogen. Dieser Zugang wird per E-Mail automatisch versendet. Die Studierenden müssen selbst nicht aktiv werden, um sich an der Evaluation zu beteiligen. Weiterhin enthält die E-Mail einen Verweis (Link), über den ein Studierender durch einen einzigen Maus-Klick direkten Zugang zum Fragebogen erhält. Dieser Zugang erfolgt durch eine Transaktionsnummer, durch die sowohl die Anonymität wie auch die Berechtigung zur Teilnahme sichergestellt werden. Der Aufbau des Fragebogens nach dem Maus-Klick auf den Verweis erfolgt mit einer Latenzzeit von rund einer Sekunde (dem Eindruck nach also sofort), weshalb auch an dieser Stelle Ausfälle wegen zu langer Wartezeiten vermieden werden. Der Fragebogen selbst ist nach eigenem Bekunden der Studierenden sehr übersichtlich gestaltet (siehe Anhang B auf den Seiten 110–111), was die Motivation zur Teilnahme weiter unterstützt. Reicht die so erzielte Rücklaufquote nicht aus, steht eine Schnittstelle zur Verfügung, über welche externe Systeme die Teilnahme automatisiert prüfen können um – falls zulässig – weitere Konsequenzen wie die Scheinvergabe oder die Teilnahme an Verlosungen anknüpfen zu können.

---

sollen Entscheidungen getroffen werden, die eine Unternehmung langfristig erfolgreich machen bzw. halten und alle Stakeholder berücksichtigen. Der Shareholder-Value ist jedoch auf die Perspektive der Aktionäre eingeschränkt und misst nur den aktuellen Marktwert. Andere Interessensgruppen und langfristige Strategien werden jedoch nur dann berücksichtigt, wenn sie sich in der Marktkapitalisierung wieder finden. Dieses Dilemma kann im wissenschaftlichen ‚Unternehmen Universität‘ bei zu starker Gewichtung der indikatorgesteuerten Ressourcenallokation ebenfalls zur Gefahr werden, wenn Maßnahmen nur noch an der Verbesserung der Indikatorwerte ausgerichtet werden und das zu verbessernde Gesamtbild in den Hintergrund tritt.

<sup>102</sup> Das Ziel ist dann legitim, wenn es die Qualifikation der Studierenden erhöht und letztendlich zur Steigerung des Volkseinkommens beiträgt. Für die Schaffung dieser zukünftigen Werte investieren die Steuerzahler (und zunehmend die Studierenden selbst) in die Universitäten und finanzieren damit auch die Gehälter der Dozenten.

Die Berücksichtigung der Interessen der Studierenden an der Evaluation bleibt bei dem oben beschriebenen Verständnis entsprechend den Evaluationsverantwortlichen und den Dozenten insofern überlassen, konkrete Konsequenzen aus den Erkenntnissen der Evaluation zu ziehen und sie zu einer Verbesserung der Lehre umzusetzen. Würden die Umsetzung bzw. die Gründe für die Ablehnung der Umsetzung den Studierenden transparent kommuniziert, so würde sich vermutlich auch das Problem der Rücklaufquote auf einer nicht-technischen Ebene lindern lassen.

#### 4.1.4 Effizienz

Wie durch die Aussagen der Praxis unterstrichen (siehe Abschnitt „Anforderungen an ein ideales Online-Evaluationssystem“ auf den Seiten 39–41), nimmt die Anforderung nach einer effizienten Umsetzung der Funktionen eine dominante Rolle ein. Das vorrangige Instrument für die Erreichung einer wie unter Abschnitt „Effizienz“ auf Seite 24 verstandenen Effizienz, die auf die Minimierung des personellen Aufwands abzielt, ist die Online-Erhebung, deren wesentlicher Effekt die Vermeidung von Medienbrüchen darstellt. Medienbrüche entstehen an der Schnittstelle zwischen unterschiedlichen Medien wie z. B. einem Papier-Fragebogen und einer elektronischen Auswertung, da die Informationen von dem einen in das andere Medium übertragen werden müssen. Bei der papierbasierten Erhebung fällt personeller und finanzieller Aufwand für die Lagerung, Verteilung, Einsammlung, Erfassung<sup>103</sup> und Entsorgung der Papierfragebögen an, der bei der Online-Erhebung nicht erforderlich ist. Da bei der Online-Erhebung auch kein sonstiger Aufwand zu identifizieren ist, der bei der papierbasierten Erhebung nicht in vergleichbarem Maß anfällt, ist die Online-Erhebung der papierbasierten aus dieser Perspektive der Effizienz überlegen.

Kann von einer Online-Erhebung ausgegangen werden, so lassen sich darauf eine Reihe von Funktionen aufbauen, die ebenfalls zur Effizienz beitragen und die mit der papierbasierten Erhebung praktisch nicht realisierbar sind. An erster Stelle ist dabei die unter Abschnitt „Fragebogen“ auf den Seiten 58–60 beschriebene Möglichkeit anzuführen, Fragebögen abhängig von der Erhebungseinheit vollkommen individuell zusammenstellen zu lassen. Diese Funktion eines dynamischen Fragebogens ist wiederum die Grundlage dafür, dozentenspezifische Fragen zu ermöglichen, deren integrative Wirkung unter Abschnitt „Integration der Dozenten“ auf den Seiten 87–88 dargestellt wurde. Effizienz fördernd ist die ebenfalls unter Abschnitt „Fragebogen“ auf den Seiten 58–60 beschriebene Funktion des Systems, die dynamischen Fragebögen automatisch zu generieren und die Zugangsberechtigungen ebenso automatisch zu versenden. Kombiniert mit den Funktionen, Berichte (siehe Abschnitt „Berichte“ auf den Seiten 65–69) und Rankings (siehe Abschnitt „Rankings“ auf den Seiten 69–75) automatisiert erstellen zu können, ist das System damit in der Lage, eine Evaluation nach der Einrichtung

---

<sup>103</sup> Die Erfassung erfolgt zwar in der Regel durch stapelverarbeitende Scanner. Während dieser Prozess landläufig als ‚automatisch‘ bezeichnet wird, fällt dennoch personeller Aufwand für das Starten des Prozesses, das Nachlegen der Stapel sowie für die Korrektur nicht lesbarer Bögen an, weshalb der Prozess tatsächlich allenfalls als ‚halbautomatisch‘ bezeichnet werden kann.

der Fragen und mit einer Anbindung an eine entsprechende Datenquelle, automatisch und ohne personellen Aufwand durchführen zu können.

#### **4.1.5 Methodische Konformität**

Die bisher dargestellten Auswirkungen haben durchwegs positive Effekte, sind jedoch wertlos, wenn die Nebenbedingung dabei nicht eingehalten wird, den statistischen Methoden zu entsprechen. Ein zentraler Ansatz zur Umsetzung dieser Anforderung besteht darin, dass das System bei der Auswertung der angelegten Fragen durch ihre statistische Klassifizierung in den Berichten automatisch nur die aus statistischer Sicht zulässigen Instrumente verwendet. Da die Klassifizierung (insbesondere die Festlegung des Skalenniveaus) beim Anlegen der Frage vorgenommen wird und Fragen nur durch Evaluationsbeauftragte selbst bzw. von ihnen überprüft in das System gelangen, wird verhindert, dass Laien unzulässige Auswertungen durchführen können. Zudem stellt das System aufgrund der Klassifizierung sicher, dass nur entsprechend auswertbare Fragen in die Dozenten-Rankings aufgenommen werden können.

Zu einer mittelbaren Sicherstellung der methodischen Konformität trägt das System durch folgende Instrumente bei. Mit dem unter Abschnitt „Integration der Studierenden“ auf den Seiten 88–89 beschriebenen Beitrag zur Erhöhung der Rücklaufquote wird die Repräsentativität gestützt. Durch die Vergabe von Transaktionsnummern zur Sicherstellung der Berechtigung können Verzerrungen durch mehrfache Stimmabgabe verhindert werden. Schließlich konzentriert sich der Einsatz des Systems auf die Darstellung der studentischen Zufriedenheit. Eine Verdichtung dieser Ergebnisse zur Ableitung von Konsequenzen wurde nicht implementiert, um das Gesamtsystem nicht der diesbezüglichen Kritik auszusetzen, die unter Abschnitt „Leistungsindikatoren“ auf den Seiten 21–23 beschrieben wurde. Auch das in dem System implementierte Dozenten-Ranking beschränkt sich durch die Anonymität und die Aufschlüsselung der individuellen Stärken- und Schwächen-Analyse auf das persönliche Feedback an die Dozenten und die Informationsgrundlage für weiterführende Gespräche mit den Evaluationsbeauftragten. Es wurde also versucht, das System so methodenkonform wie möglich zu gestalten, so dass sich die Einhaltung der statistischen Gütekriterien rein auf die inhaltliche Gestaltung der Fragen konzentrieren kann. Durch diesen Rahmen soll letztendlich sichergestellt werden, dass mit dem System eine effiziente, integrative und von methodischer Kritik möglichst ungetrübte Lehrvaluation möglich wird.

## **4.2 Schwachstellen und Weiterentwicklungspotenziale**

Während der vorangegangene Abschnitt die Auswirkungen und Möglichkeiten im positiven Sinne dargestellt hat, unterzieht der vorliegende Abschnitt die Erkenntnisse einer kritischen Sicht um Schwachstellen und Weiterentwicklungspotenziale aufzuzeigen.



### 4.2.1 Störvariablen

Wie unter Abschnitt „Störvariablen“ auf den Seiten 16–17 dargestellt, wird in der Literatur die Berücksichtigung von Störvariablen gefordert. In der Praxis wird dieser Forderung nur sehr eingeschränkt nachgekommen (siehe Abschnitt „Berücksichtigung von Störvariablen“ auf den Seiten 32–33). Das hier beschriebene System verzichtet auf die explizite Berücksichtigung von Störvariablen aus mehreren Gründen. Zum einen zeigt sich die Auswahl der zu berücksichtigenden Störvariablen in der Literatur wie ebd. gezeigt sehr uneinheitlich. Die Berücksichtigung einer fixen Menge an Störvariablen erscheint daher nicht sinnvoll. Eine zur Laufzeit variabel definierbare Menge an Störvariablen und statistischer Instrumente zur Bereinigung ist auf der anderen Seite nur sehr schwer praktikabel. Aus diesem Grund verzichtet das System auf die Berechnung von Leistungsindikatoren, die (im Falle den Dozenten-Rankings) über den Zweck eines individuellen Feedbacks hinausgehen. Sollten Leistungsindikatoren für weiterreichende Konsequenzen wie beispielsweise Lehrpreise oder gar die leistungsorientierte Mittelvergabe eingesetzt werden, so muss – neben der Klärung weiterer Kritikpunkte wie die grundsätzliche Zulässigkeit studentischer Urteile für den jeweiligen Einsatzzweck – die Bereinigung von Störvariablen integriert werden. Eine Umsetzung dieser Vorgabe stellt einen Anknüpfungspunkt für weitere Arbeiten dar.

Alternativ dazu besteht die Möglichkeit, Leistungsindikatoren durch die Übernahme der Rohdaten aus dem Evaluationssystem in externen Statistik-Paketen zu berechnen. Wie unter Abschnitt „Störvariablen“ auf Seite 17 und Abschnitt „Stammdatenbezug“ auf den Seiten 47–48 beschrieben, sollte das System dazu neben den explizit in Fragen erfassbaren Variablen (z. B. der Frage nach der Motivation) auch implizite Variablen (z. B. die Veranstaltungsart) aus externen Systemen beziehen, so dass diese zur umfassenden Bereinigung mit exportiert werden können. Während eine externe Berechnung von Leistungsindikatoren eine große Flexibilität erlaubt, erfordert diese Alternative ein hohes Maß an statistischem Wissen. Eine Unterstützung und Implikation dieses Wissens wie im Fall der automatischen Berücksichtigung der statistisch zulässigen Auswertungsinstrumente bei den systeminternen Berichten ist bei dieser Alternative nicht mehr möglich.

### 4.2.2 Statistische Instrumente

Neben dem globalen Aspekt der Berücksichtigung von Störvariablen existieren unter der statistischen Perspektive zwei weitere, jedoch deutlich kleinere Verbesserungspotentiale. So könnten die Auswertungsinstrumente mindestens intervallskalierter Items neben der Spannweite als bisher einziges Streuungsmaß um die Angabe und grafische Darstellung der Standardabweichung erweitert werden, um Aufschluss über die Homogenität der Antworten zu erhalten. Im Speziellen könnte die Standardabweichung sowohl im Bericht bei den Fragen selbst wie auch bei den Polaritätenprofilen und im Ranking beim gesamten Score sowie bei der Aufschlüsselung der Einzelscores dargestellt werden.

Beim Ranking könnte anstelle des arithmetischen Mittels der Median zur Berechnung der Scores verwendet werden, um gegenüber Ausreißern robustere Ergebnisse zu erhalten.

Zudem könnte die Einführung einer Gewichtung der Fragen (siehe Abschnitt „Fragenarten“ auf Seite 49) bei der Berechnung zu ausgewogeneren Scores führen.

### 4.2.3 Rankings

Neben den statistischen Optimierungen ließen sich die Rankings auf zwei weitere Arten verbessern. Während sich die Rankings in der hier vorgestellten Form immer auf ein ganzes Semester beziehen, könnten analog zu den Berichten auch individuelle Rankings angelegt werden. So könnten beispielsweise Institute oder Veranstaltungsarten durch Rankings bewertet werden. Im Gegensatz zu den Berichten sollten individuelle Rankings nur von Evaluationsbeauftragten angelegt, von allen in den Rankings bewerteten Dozenten jedoch anonymisiert eingesehen werden können. Das System muss dabei prüfen, ob die Anzahl der im Ranking berücksichtigten Dozenten groß genug ist, um die Anonymität sicherzustellen. Ist dies nicht der Fall, muss das System den Zugriff durch die Dozenten unterbinden. Zudem sollte das System die Evaluationsbeauftragten generell automatisch über Veränderungen der Ranking-Leistungen einzelner Dozenten per E-Mail informieren, wenn die Werte bzw. deren Veränderung bestimmte Schwellwerte überschreiten. Damit könnte das System die Evaluationsbeauftragten frühzeitig auf eine absinkende Lehrqualität aufmerksam machen. Die Funktion der automatischen Benachrichtigung ist insbesondere dann sinnvoll, wenn eine große Anzahl an Dozenten einer kleinen Anzahl an Evaluationsbeauftragten gegenüber steht. In der Kombination aus individuellen Rankings und automatischer Benachrichtigung muss das System jedoch darauf achten, keine redundanten Informationen zu kommunizieren, um den Evaluationsbeauftragten nicht unnötig Aufmerksamkeit abzuverlangen.

## 4.3 Interpretation

Die umfassende Automatisierung und hohe Effizienz des Evaluationssystems ist gleichzeitig dessen potenzielle Schwäche. Wie unter Abschnitt „Effizienz“ auf den Seiten 89–90 beschrieben, kann das Evaluationssystem nach einer initialen Einrichtung nahezu ohne personellen Aufwand betrieben werden. Es führt sowohl die Erhebungen wie auch die Auswertungen automatisch durch und kann somit als ‚Evaluationsmaschine‘ begriffen werden. Die potenzielle Schwäche liegt nun darin, diese ‚Evaluationsmaschine‘ mit ‚der Evaluation‘ gleichzusetzen und sich der Illusion hinzugeben, damit bereits genug für die Evaluation getan zu haben.

Das wäre eine fatale Fehleinschätzung, denn Hochschulen sind soziale Systeme ohne eindeutig quantifizierbare Zielsetzungen. Sie können somit nicht nur durch quantitative Evaluations-Ergebnisse gesteuert werden (vgl. Schmidt 2005, S. 40 f.). Ein zu großer Fokus auf die Ergebnisse der quantitativen Evaluation kann dazu führen, dass die qualitativen Aspekte – insbesondere die persönlichen Einstellungen der Interessensgruppen – nicht den angemessenen Stellenwert in der Situationsbeschreibung und der daraus abgeleiteten Handlungsvorgaben finden. Den Verantwortlichen sollte deshalb bewusst sein, dass zumindest ein Teil des mit der gezeigten Effizienz eingesparten Zeit- und Kostenaufwands in die Ableitung und Umsetzung von Verbesserungsmaßnahmen im Sinne eines kontinuierlichen Verbesserungsprozesses

(KVP) (vgl. Kostka & Kostka 2002) investiert werden sollte, um den eigentlichen Zielen der Evaluation – insbesondere der Verbesserung der Lehrqualität (vgl. Wissenschaftsrat 1996) – gerechter zu werden.

Das in dieser Arbeit vorgestellte System ist darauf ausgerichtet, bei der Einhaltung der methodischen Rahmenbedingungen eine effiziente Evaluation zu ermöglichen, welche durch die dozentenbezogene Erhebungen und die Aufbereitung der Ergebnisse in Form von persönlichen Berichten und anonymisierten Rankings eine individuelle Rückmeldung an die Dozenten und Evaluationsbeauftragten ermöglicht. Diese individuelle Rückmeldung erlaubt es den Dozenten, ihre persönlichen Stärken und Schwächen zu erkennen; den Evaluationsbeauftragten gibt sie eine fundierte Informationsgrundlage für individuelle Gespräche mit den Dozenten. Bereits die Evaluationsrückmeldung selbst hat wegen der Individualisierung das Potenzial, zu einer von Balk (2000, S. 137) beschriebenen „substantiellen Veränderung von Handlungsparametern“ zu führen. Die detaillierte Stärken-/Schwächenanalyse des Rankings stellt darüber hinaus eine Grundlage für die im Besonderen geforderte Schulung der Dozenten dar (vgl. Rindermann 2003, S. 92, 94, Landfried 2003, S. 12, Putz 2003, S. 113), die Rindermann (2003, S. 95) als effektiver als das hochschuldidaktische Diskursmodell beschreibt. Sie stellt damit ein Mittel dar, um den Bedenken von Schimank (2006, S. 217) zu entgegnen, dass durch den Einsatz von Leistungsindikatoren gute Dozenten immer besser und schlechte immer schlechter werden.

Von einer in der Literatur geforderten Ableitung weiterreichender Konsequenzen aus den Scores wie beispielsweise der für den Leistungsvergleich zwischen den Universitäten (vgl. Wissenschaftsrat 1996, S. 12 f., Hofer 2003, S. 93, Hilzenbecher 2003, S. 101, Neuser 2003), die Berücksichtigung bei Berufungen und Habilitationen (vgl. Wissenschaftsrat 1992, S. 59) oder die Orientierung der Universitätssteuerung an Grundsätzen des Controlling<sup>104</sup> wird wegen der Kritik an dem Fokus auf die studentischen Urteilen (siehe Abschnitt „Studentische Veranstaltungskritik“ auf den Seiten 4–5) und mangels der Bereinigung um Störvariablen (siehe Abschnitt „Störvariablen“ auf Seite 91) an dieser Stelle ausdrücklich Abstand genommen.<sup>105</sup> Auch wenn diese Möglichkeiten technisch sehr einfach umzusetzen wären, bleibt der Fokus des hier vorgestellten Systems auf der individuellen Rückmeldung an die Dozenten und Evaluationsbeauftragten. Der zu komplexer Verrechnung verleitende Leistungsindikator „Score“ bleibt hier ein Artefakt.

Nachdem die Dozenten das Organ der Lehre sind und die Lehrqualität der Dozenten damit ein essentieller Aspekt der gesamten Lehrqualität ist, kann ein Evaluationssystem der hier vorgestellten Ausprägung dazu *beitragen*, die Qualität der Lehre entkoppelt von der Diskussion um die Abgrenzung des Begriffs „Qualität“ (siehe Abschnitt „Begriffliche Abgrenzung“ auf den Seiten 2–3) zu verbessern. Die vorliegende Arbeit setzt dies mit den von Reissert & Konnerth (2001, S. 193) geforderten „kreativen Ideen und Lösungsvorschlägen“ um,

<sup>104</sup> Wie z. B. von Weber (2006, S. 277) mit der Balanced Scorecard (vgl. Kaplan & Norton 1992, Hahne 2004) im Rahmen des New Public Managements (NPM) gefordert.

<sup>105</sup> Dieser Zusammenhang ist gleichzeitig als Kritik an den Dozenten-Bewertungsportalen und -Rankings im Internet wie beispielsweise MeinProf.de (siehe <http://www.meinprof.de>) zu verstehen.

die wie ebd. gefordert „in der Praxis des Hochschulalltags“ erprobt und umgesetzt werden können.

Dieser Beitrag ist jedoch eingebettet in einen Kontext weiterer Maßnahmen zu sehen, welche das Verständnis einer umfassenden Evaluation vervollständigen. Dabei gilt es insbesondere, die durch ein Evaluationssystem erlangten Informationen über die studentischen Urteile hinaus mit weiteren Erkenntnissen – insbesondere denen der externen Evaluation – zusammenzuführen (vgl. Wissenschaftsrat 2002b, S. 78), auf denen eine fundierte und umfassende Strategie zur Qualitätssicherung aufgebaut werden kann (vgl. Hochschulrektorenkonferenz 2003, S. 17 f., Reissert & Konnerth 2001, S. 191 f.). Aus diesem Zusammenhang wird klar, dass die Ableitung von Konsequenzen ein Teil der Evaluation sein muss. In der Literatur herrscht weitgehende Einigkeit darüber, dass Evaluation nicht Selbstzweck sein kann und darf (vgl. Hochschulrektorenkonferenz 2003, S. 28 f., Landfried 2003, S. 12, Putz 2003, S. 114, el Hage 1996, S. 140–143, Kromrey 2001a, S. 46, 50 sowie Wottawa & Thierau 1998).

Könnten in dieser Diskussion um die Validität studentischer Urteile und die Bereinigung von Störvariablen Fortschritte erzielt werden, wäre es Evaluationssystemen der hier gezeigten Ausprägung mit durchaus überschaubarem Aufwand möglich, den Umfang noch weiter zu steigern, in dem sie zur Effizienz und Qualitätssicherung beitragen können. Die technologischen Grundsteine sind gelegt. Es liegt nun an den Entwicklern der Methoden, die neuen Möglichkeiten zu nutzen.

## 5 Zusammenfassung

Ziel der Arbeit war es, durch ein anforderungsgerechtes Lehrevaluationssystem einen Beitrag zur Verbesserung der Qualität der Lehre zu leisten. Während die pädagogisch-psychologischen Aspekte der Lehrevaluation in der Literatur ausführlich behandelt werden, wird den Aspekten der Informationstechnologie bisher keine ausreichende Beachtung geschenkt, obwohl sie auch in der Lehrevaluation bereits umfassend zum Einsatz kommt. Der Anspruch der Arbeit ist es, diese Forschungslücke zu schließen.

Da in der Literatur keine Arbeiten darüber existieren, wie ein Lehrevaluationssystem ausgeprägt sein soll, bestand die erste Aufgabe der Arbeit darin, solche technischen Anforderungen herzuleiten. Um darzustellen, wie evaluiert werden *soll*, wurden die methodischen Grundlagen dargestellt, denen ein Lehrevaluationssystem entsprechen muss (siehe 2.1). Diese wurden um die besonderen Aspekte des medizinischen Curriculums erweitert (siehe 2.2). Anschließend stellte eine empirische Studie, die im Rahmen dieser Arbeit durchgeführt wurde, die Evaluationspraxis an den medizinischen Fakultäten in Deutschland vor (siehe 2.3), um Aufschluss darüber zu erhalten, wie in der Praxis bereits evaluiert *wird*.

Diese Erkenntnisse waren die Grundlage für die Entwicklung der im weiteren Verlauf benötigten technischen Anforderungen an ein Lehrevaluationssystem (siehe 3.1). Diese abstrakt gehaltenen Anforderungen wurden anschließend durch die Implementation eines im Rahmen dieser Arbeit entwickelten Evaluationssystems konkretisiert und veranschaulicht (siehe 3.2). Die Praxistauglichkeit und die Erfahrungen mit diesem System wurden anhand dessen Pileinsatzes im Sommersemester 2006 an der Medizinischen Fakultät der Ludwig-Maximilians-Universität München aufgezeigt (siehe 3.3). Die Aufteilung in die abstrakte Darstellung der Anforderungen und deren konkrete Implementation diente weiterhin dazu, eine Grundlage für die Beschaffung neuer Systeme zur Verfügung zu stellen, wie auch die Verbesserung bestehender Systeme zu ermöglichen. Aufbauend auf diesen Ergebnissen wurden sowohl die Möglichkeiten (siehe 4.1) wie auch die Schwachstellen (siehe 4.2) eines Evaluationssystems der hier vorgestellten Ausprägung diskutiert und in den Gesamtzusammenhang der Evaluation eingebettet (siehe 4.3).

Die Arbeit zeigte, wie mithilfe eines neuen Systems die Lehrevaluation mit minimalem personellen und finanziellen Aufwand durchgeführt und gleichzeitig alle Interessensgruppen stärker an das Thema „Evaluation“ gebunden werden können. Die hier vorgestellten Verfahren liefern methodisch fundierte Ergebnisse in praxisrelevanter Form, welche den Dozenten und Evaluationsbeauftragten direkte Rückschlüsse über deren Stärken und Schwächen aus der Sicht der Studierenden erlauben. Durch die einfache, aufbereitete und gleichzeitig fundierte Information für die Dozenten als Träger der Lehre und für die Evaluationsbeauftragten als ‚Qualitätssicherer‘ können diese Verfahren mittelbar einen Beitrag zur Verbesserung der Qualität der Lehre leisten.



# Anhang A

## Fragen der Studie

Den in der Studie unter dem Abschnitt „Evaluationspraxis an den medizinischen Fakultäten in Deutschland“ auf den Seiten 27–41 dargestellten Ergebnissen lagen die folgenden Fragen zugrunde.

1. Welche fremdbezogenen oder selbstentwickelten Systeme setzen Sie an Ihrer Fakultät zur studentische Veranstaltungsevaluation ein? Bei selbstentwickelten Systemen: Gibt es eine Beschreibung der Funktionalität und/oder einen Testzugang?
2. Wie läuft die studentische Veranstaltungsevaluation an Ihrer Fakultät ab? Bitte schildern Sie in einem kurzen Überblick die wichtigsten Prozessschritte aus der Sicht der Beteiligten (Evaluationsverantwortliche, Dozenten, Studenten)
3. Mit welchen Aspekten des Evaluationsverfahrens bzw. der -systeme sind die Beteiligten (Evaluationsverantwortliche, Dozenten, Studenten) besonders zufrieden?
4. Welche Punkte müssten aus der Sicht der Beteiligten (Evaluationsverantwortliche, Dozenten, Studenten) verbessert werden?
5. Welche Anforderungen muss ein Lehrevaluationssystem Ihrer Ansicht nach generell vor allem erfüllen? Bitte gewichten Sie diese Anforderungen wenn möglich nach deren Relevanz.
6. Externe Evaluation
  - a) Beziehen Sie neben der studentischen Veranstaltungsevaluation auch externe Evaluationsverbände- bzw. agenturen mit ein?
  - b) Falls ja, welche und in welcher Form?
7. Rückmeldungen
  - a) Setzen Sie bestimmte Verfahren ein, um Studenten zur Teilnahme an der Evaluation zu „verpflichten“?
  - b) Falls ja, welche?
8. LOM
  - a) Existiert an Ihrer Fakultät eine leistungsbezogene Mittelvergabe?
  - b) Falls ja, setzen Sie dabei auch die Ergebnisse der Lehrevaluation ein?

9. Dozenten-Ranking

- a) Existiert an Ihrer Fakultät ein Dozenten-Ranking?
- b) Falls ja, setzen Sie dabei auch die Ergebnisse der Lehrevaluation ein?

10. Lehrpreise

- a) Vergeben Sie an Ihrer Fakultät Lehrpreise?
- b) Falls ja, setzen Sie dabei auch die Ergebnisse der Lehrevaluation ein?

11. Bieten Sie an Ihrer Fakultät E-Learning Kurse an?

12. Störvariablen

- a) Rechnen Sie bei den Ergebnissen der Lehrevaluation Störvariablen heraus (z. B. Alter der Studierenden, Veranstaltungsart, Motivation)?
- b) Falls ja, welche?



## **Anhang B**

### **Vollständiger Evaluationsbericht des Piloteinsatzes**

Die folgenden Seiten zeigen den von Vilfredo für den Piloteinsatz erstellten Evaluationsbericht, der vollständig von Vilfredo erzeugt und nicht manuell nachbearbeitet wurde. Lediglich der namentliche Bezug auf zwei Dozenten im Rahmen der offenen Fragen wurde nachträglich durch „xxx“ anonymisiert.



Demonstrator zur Dissertation  
von Marc Gumpinger



## Evaluationsbericht Vilfredo Piloteinsatz

Datum 20.02.2007  
Autor Marc Gumpinger



## *Erläuterungen*

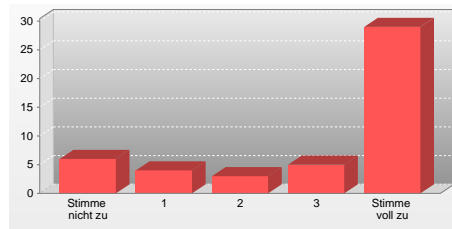
Dieser Evaluationsbericht wertet die Antworten aus, die wie folgt eingeschränkt wurden.

Zeitraum                      01.08.2006 bis 01.09.2006

## Fragen zur Veranstaltung

### Die Veranstaltung fand pünktlich statt

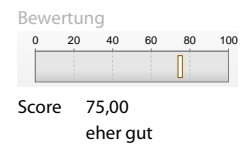
Bipolare Single Choice Frage (n=47)



Balkendiagramm

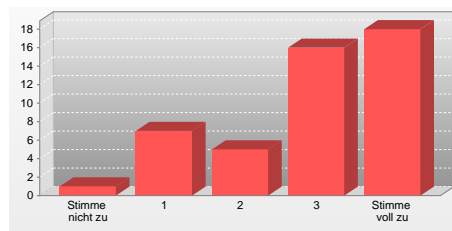
Lagemaße  
Mittelwert 3,00  
Median Stimme voll zu  
Modus Stimme voll zu

Streuungsmaße  
Spannweite 4



### In der Veranstaltung habe ich viel gelernt

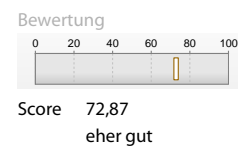
Bipolare Single Choice Frage (n=47)



Balkendiagramm

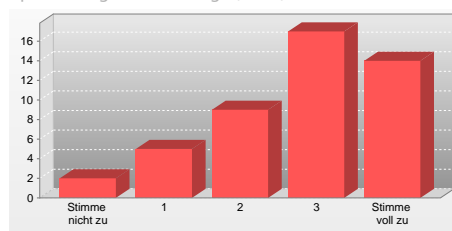
Lagemaße  
Mittelwert 2,91  
Median 3  
Modus Stimme voll zu

Streuungsmaße  
Spannweite 4



### Die Stoffvermittlung war meinem Vorwissen angepasst

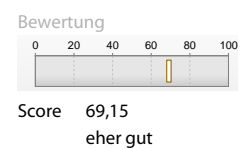
Bipolare Single Choice Frage (n=47)



Balkendiagramm

Lagemaße  
Mittelwert 2,77  
Median 3  
Modus 3

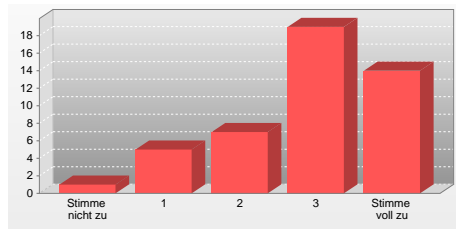
Streuungsmaße  
Spannweite 4



## Fragen zur Veranstaltung

### In der Veranstaltung wurde ich zum weiteren Selbststudium motiviert

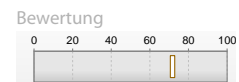
Bipolare Single Choice Frage (n=46)



Balkendiagramm

Lagemaße  
Mittelwert 2,87  
Median 3  
Modus 3

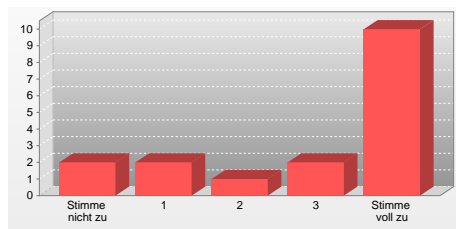
Streuungsmaße  
Spannweite 4



Score 71,74  
eher gut

### Die Einbindung von Patienten war lehrreich

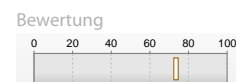
Bipolare Single Choice Frage (n=17)



Balkendiagramm

Lagemaße  
Mittelwert 2,94  
Median Stimme voll zu  
Modus Stimme voll zu

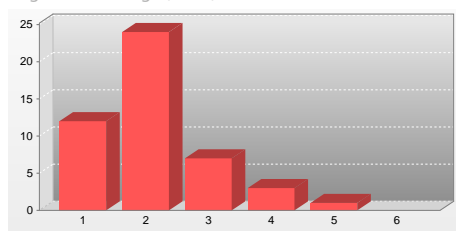
Streuungsmaße  
Spannweite 4



Score 73,53  
eher gut

### Der Veranstaltung gebe ich folgende Schulnote

Single Choice Frage (n=47)



Balkendiagramm

Lagemaße  
Median 2  
Modus 2

Streuungsmaße  
Spannweite 1 bis 5



## Fragen zur Veranstaltung

### Allgemeine Kommentare zur Veranstaltung

Offene Frage (n=16)

1. war nicht anwesend
2. nicht anwesend
3. Oft unpünktliche Dozenten, die nicht sonderlich motiviert waren, dann ihre Powerpoint Präsentation herunterlesen.
4. Man konnte selbst eine Sonographie an seinen Kommilitonen durchführen. xxx erklärte bereitwillig alles.
5. Sehr umfangreiches Thema, das natürlich etwas unter der knappen Zeit litt. Ansonsten sehr gut und humorvoll gemacht von xxx.
6. Absolute Spitzenklasse! Guter, interessanter Stoff mit guten Patienten. Davon bitte mehr! Sehr motivierter und versierter Dozent.
7. Sehr gute Aufbereitung eines relevanten Themas.
8. ich konnte leider nicht teilnehmen und daher auch nichts dazu sagen
9. Ich finde die Kardiologievorlesung sollte eher im Semester stattfinden, da man mit den wichtigsten Themen bereits im Tutorial, Blockpraktikum, L-Kurs,... konfrontiert wurde, und man sich einen Großteil des Vorlesungsstoffes bereits vorher durch Lehrbuchwissen angeeignet hatte.
10. hat früher stattgefunden, als im VI-plan stand, ohne, dass uns vorher jemand bescheid gegeben hätte (z.B auf mecum. ging teilweise zu schnell und die EKG auswertung blieb größtenteils auf der strecke.
11. Die Präsentationen waren schön, leider war unsere Gruppe klein, was gewissermaßen auch ein Vorteil für uns fünf war.
12. Einige Dinge, die wir dieses Semester bereits mehrmals gehört haben und auch schon vorher wussten, hätte man rauslassen können (zB. wie das RAAS funktioniert), auf der anderen Seite hätte ich mir gewünscht dass auf andere Pkte mehr eingegangen wäre wie zB. wie das Reentry denn nun genau funktioniert, hier hätte man sich etwas mehr Zeit lassen können. Auch hätte ich mir mehr Pathophysiologie bei dieser Vorlesung gewünscht.
13. Der DOzent ist nicht wirklich gut mit den Patienten umgegangen.
14. konnte leider nicht erscheinen
15. Die Vorlesung war sehr gut, dass einzige schade war, dass die Vorlesung eine Stunde früher war als sonst und wir dadurch noch eine Stunde mehr frei hatten bis unsere Seminare angefangen haben, also 4 Stunden und das war immer sehr sehr lang.
16. z.t. nicht pünktlich

### An der Veranstaltung hat mir besonders gut gefallen

Offene Frage (n=6)

1. Alles praktische war interessant, leider war die Gruppengröße oft zu groß um wirklich etwas zu machen. Das Prinzip der Veranstaltungen fand ich gut, weiter so!
2. Endlich einmal langsam Sonographie erklärt zu bekommen und selbst praktisch durchzuführen.
3. Gute Strukturierung und Bezug auf das Wesentliche. Mit vielen Anekdoten die wichtigen Inhalte vermittelt, so bleibt sicher mehr "hängen".



## Fragen zur Veranstaltung

4. Klasse Dozent, interessantes und relevantes Gebiet.
5. Wichtiges Thema, kein reiner Vortrag, sondern sehr interaktiv gestaltet. Also ein "echtes" Seminar, von denen es immer noch zu wenige gibt.
6. Man konnte offen und ohne Bedenken Fragen stellen.

## Folgendes sollte unbedingt verbessert werden

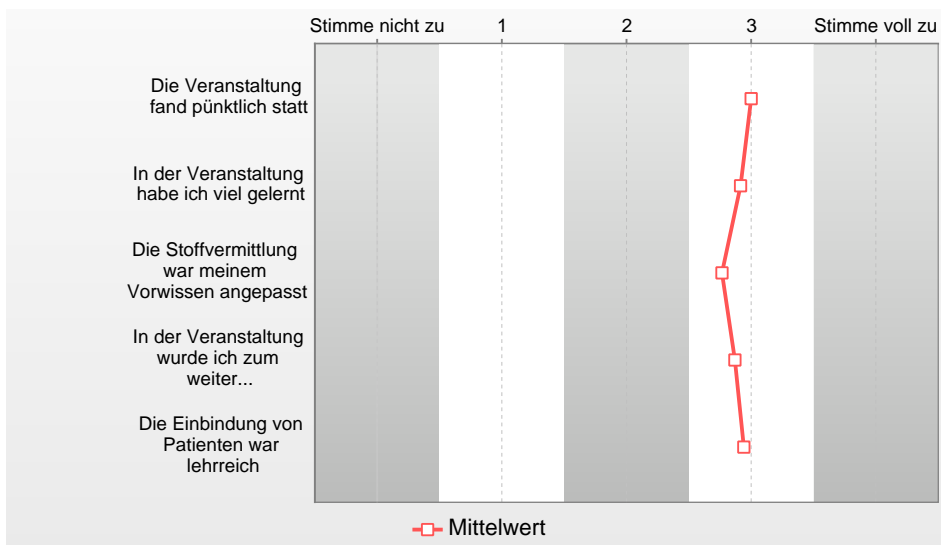
Offene Frage (n=8)

1. Evtl. vorher bekannt geben, wo es die Materialien gibt um sich diese ggf. vorab anzusehn (erschlagende Stoffmenge mit etwas irreführendem Titel - jeder erwartete Endoskopie!).
2. Nicht so oft überziehen, bzw. in den letzten Minuten das Tempo unmöglich verschärfen.
3. Evtl. Unterrichtsmaterialien im Internet bereitstellen.
4. Vorlesung am besten schon vor der Vorlesung ins Internet stellen, damit sich die Studenten vorbereiten können.
5. s.o.
6. Mich würde es noch interessieren, wie man den Katheter nutzt um die Klappenfunktion/Herzfunktion zu beurteilen, sprich Druckmessungen, etc.
7. Die Zeit. Die Mittagspause ist einfach viel zu lang.
8. Vorlesungsfolien ins Internet stellen  
Fragen an Patienten unbedingt selber stellen lassen von den Studenten



## Fragen zur Veranstaltung

### Polaritätenprofil 1



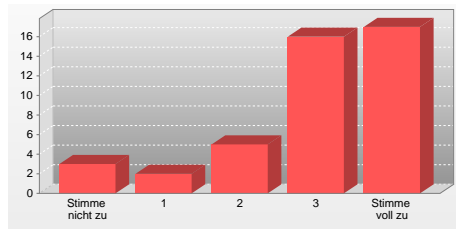




## Fragen zur Dozentin/zum Dozenten

### Die Inhalte wurden verständlich vermittelt

Bipolare Single Choice Frage (n=43)



Balkendiagramm

Lagemaße

Mittelwert 2,98

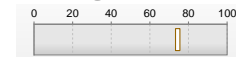
Median 3

Modus Stimme voll zu

Streuungsmaße

Spannweite 4

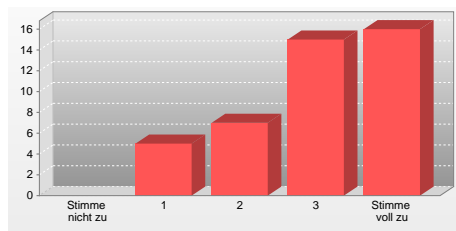
Bewertung



Score 74,42  
eher gut

### Der Stoff war gut strukturiert

Bipolare Single Choice Frage (n=43)



Balkendiagramm

Lagemaße

Mittelwert 2,98

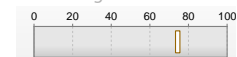
Median 3

Modus Stimme voll zu

Streuungsmaße

Spannweite 3

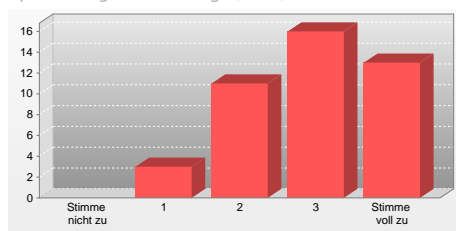
Bewertung



Score 74,42  
eher gut

### Die Unterrichtsmaterialien waren gut aufbereitet

Bipolare Single Choice Frage (n=43)



Balkendiagramm

Lagemaße

Mittelwert 2,91

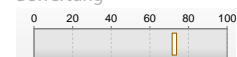
Median 3

Modus 3

Streuungsmaße

Spannweite 3

Bewertung



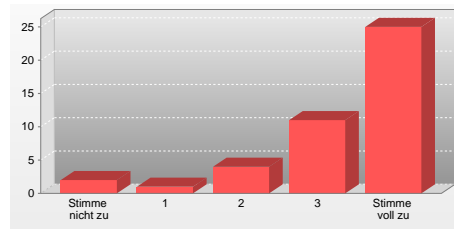
Score 72,68  
eher gut



## Fragen zur Dozentin/zum Dozenten

### Die Dozentin/Der Dozent verfügte über ein gutes Fachwissen

Bipolare Single Choice Frage (n=43)



Balkendiagramm

#### Lagemaße

Mittelwert 3,30

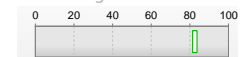
Median Stimme voll zu

Modus Stimme voll zu

#### Streuungsmaße

Spannweite 4

#### Bewertung

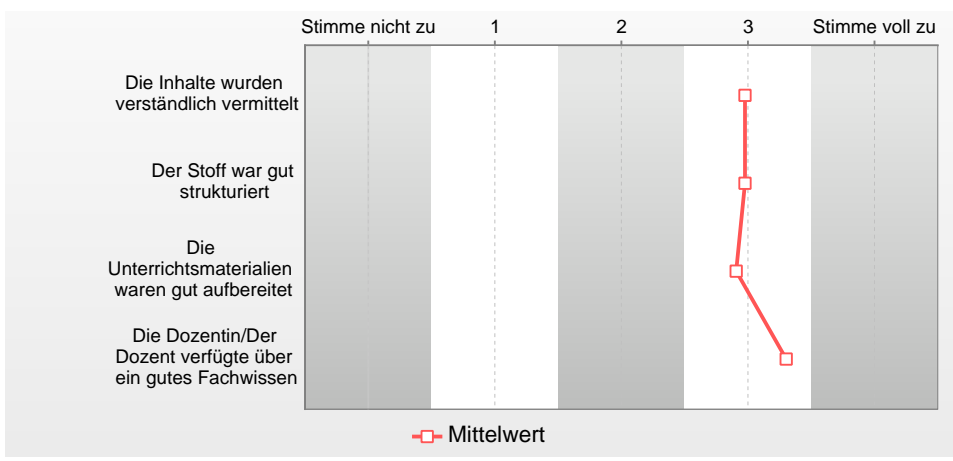


Score 82,56

besonders gut

## Fragen zur Dozentin/zum Dozenten

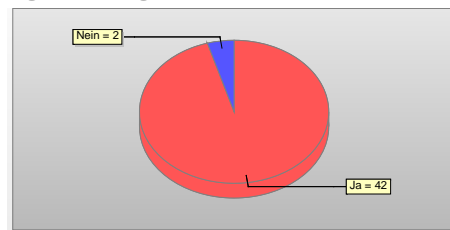
### Polaritätenprofil 1



## Fragen zur Evaluation

Haben Sie bemerkt, dass Sie für diese Veranstaltung einen anderen Fragebogen als sonst erhalten haben?

Single Choice Frage (n=44)



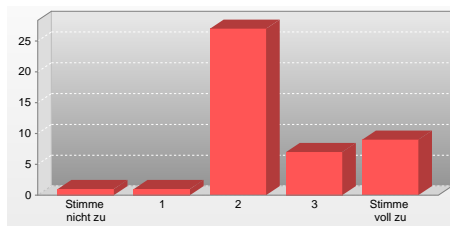
Lagemaße

Modus Ja

Kuchendiagramm

Dieser Fragebogen gefällt mir besser als die sonst üblichen

Bipolare Single Choice Frage (n=45)



Lagemaße

Mittelwert 2,49

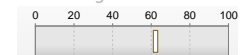
Median 2

Modus 2

Streuungsmaße

Spannweite 4

Bewertung



Score 62,22

eher gut

Balkendiagramm

Was gefällt Ihnen an diesem neuen Fragebogen besonders gut?

Offene Frage (n=14)

1. Nichts wesentlich anderes.
2. Genaue Auflistung der Veranstaltung/en, die es zu evaluieren gilt.
3. Möglichkeit einzelne Dozenten/Veranstaltungen zu bewerten.
4. man wird persönlich angeschrieben
5. übersichtlicher
6. übersichtlicher
7. übersichtlicher
8. gut strukturiert
9. übersichtlicher, das fordert unbewusst auf, eher bei der evaluation mitzumachen, als zuvor



## Fragen zur Evaluation

10. übersichtlicher, farblich ansprechender
11. Übersichtlicher
12. per email gekommen, fein
13. Ich finde es ganz gut, wenn man eine bestimmte Vorlesung evaluiert, denn bei den Wochenevaluationen haben teilweise die Dozenten während der Woche gewechselt und dadurch war es etwas schwierig sie zu bewerten
14. besseres Design, mehr offene Fragen

## Was gefällt Ihnen an diesem neuen Fragebogen gar nicht?

Offene Frage (n=7)

1. Nichts Besonderes.
2. Nichts.
3. Nichts.
4. Keine Möglichkeit um eine Nicht-Teilnahme an einer Veranstaltung anzugeben !!!
5. Weniger Fragen
6. die email sieht wie spam aus ("do\_not\_reply@lmu.de" und diese komische Tan)
7. Ich finde es sehr verwirrend, dass man einerseits Schulnoten vergibt mit 1 beste Note und 6 schlechteste Note und diese 6 ganz rechts steht, man aber andererseits bei stimme voll zu oder stimme nicht zu genau Seitenverkehrt bewertet, d.h. hier steht die schlechteste Bewertung ganz links.

## Was sollte bei der Evaluation generell verbessert werden?

Offene Frage (n=9)

1. Man könnte die Ergebnisse der Evaluationen veröffentlichen. Und dazu sagen, was denn jetzt nun wirklich verbessert wird.
2. Zeitnah und es sollte die Möglichkeit gegeben werden, einzelne Dozenten zu evaluieren, z.B. indem es ein Feld gibt, den Namen anzugeben, denn es kommt vor, dass man die Seminare tauscht und so evtl. eine falsche Zuweisung zu stande kommt.
3. Zeitnah und Möglichkeit, wie hier einzelne Dozenten zu evaluieren.
4. Zeitnahe Bewertung auch einzelner Dozenten.
5. Zeitnahe Evaluation!
6. zeitnaher, da man sich sonst an die Veranstaltung nicht erinnert
7. Evaluation nicht jede woche sondern z.B. nur einmal maonatlich
8. möglichst an den Tag versenden, an dem die Veranstaltung war und in der veranstaltung darauf hinweisen, Biofeedback zurück an die Studierenden wäre nicht schlecht, sozusagen als Rückblende an das Semester
9. Irgendwie muss es doch machbar sein, dass daran mehr leute teilnehmen.



## Anhang C

### Technische Beschreibung der E-Learning-Schnittstelle

Die hier vorgestellte API wurde aus den javadoc-Informationen der Vilfredo-Klassen mithilfe eines speziellen doclets für  $\LaTeX$  erstellt. Der technischen Ausrichtung entsprechend ist die API-Dokumentation in Englisch gehalten. Um die Übersichtlichkeit zu erhalten, wurde auf eine Darstellung der Exceptions verzichtet, deren einzige Besonderheit darin liegt, die Referenz auf das Objekt zu transportieren, welches die Exception verursacht hat.

#### Package `net.gumpinger.vilfredo.webservice`

##### Class *ELearningTanProvider*

Provides TANs for a given eLearning lecture and student matriculation numbers. This class is designed to be used as a web service.

##### Declaration

```
public class ELearningTanProvider extends java.lang.Object
```

##### Constructors

- `public ELearningTanProvider()`
  - Usage
    - \* Default constructor by bean requirement

##### Methods

- `getTans`

```
public ELearningTanProvider.StudentTan[] getTans(  
    java.lang.String clientUsername,  
    java.lang.String clientWebServicePassword,  
    java.lang.String lectureId,  
    java.lang.String[] matriculationNumbers,  
    boolean ignoreStudentsNotFound)
```

  - Usage
    - \* Returns the transaction numbers (TANs) for the specified eLearning lecture and student matriculation numbers.  
This method makes sure, that the TANs refer to a questionnaire which holds the latest set of questions for the specified eLearning lecture. In case no questionnaire has ever been created for the specified lecture or if the available

questionnaires do not contain the latest set of questions, this method will create a new questionnaire with the latest set of questions. It will then create the TANs for the most up to date questionnaire (either newly created or already existing). **Notice:** Due to anonymity there is no way for the system to find out if it already did issue a TAN for one or more of the specified students. It is left up to this method's invoker to make sure, students don't receive multiple TANs. The system can cope with multiple TANs. However, it would be annoying to students and cause potential bias on interim data monitoring. Also notice that students are not associated to eLearning lectures. Therefore, there is no way to check if the students represented by their specified matriculation numbers are eligible to participate or evaluate the specified eLearning lecture.

– **Parameters**

- \* `clientUsername` - The client's username
- \* `clientWebServicePassword` - The client's web service password
- \* `lectureId` - The eLearning lecture's identification
- \* `matriculationNumbers` - An array of Strings each representing a student's matriculation number
- \* `ignoreStudentsNotFound` - If false, a `StudentNotFoundException` is thrown if an unknown matriculation number is encountered

– **Returns** - Array of `StudentTan` objects

– **Exceptions**

- \* `ClientNotFoundException` - Thrown if no client is found for the specified `clientUsername/clientWebServicePassword` combination
- \* `ELearningLectureNotFoundException` - Thrown if no eLearning lecture is found for the specified `lectureId`
- \* `StudentNotFoundException` - Thrown if no student is found for the specified matriculation number **and** `ignoreStudentsNotFound` is false

### **Class `ELearningTanProvider.StudentTan`**

The object that allows for referencing the TAN, URL and requesting student's matriculation number. **Notice:** Storing the TAN and matriculation number in one single object does not break the paradigm of anonymity since Vilfredo does not store a relationship between TAN and matriculation number and the requesting eLearning system cannot resolve a reference between the TAN and its answers.

#### **Declaration**

```
public class ELearningTanProvider.StudentTan extends java.lang.Object
```

#### **Constructors**

- `public ELearningTanProvider.StudentTan()`

– **Usage**



---

\* Default constructor by bean requirement

- `public ELearningTanProvider.StudentTan(  
java.lang.String matriculationNumber,  
java.lang.String TAN,  
java.lang.String url )`

### Methods

- `public String getMatriculationNumber()`
  - **Returns** - The matriculation number of the student this TAN was assigned to.
- `public String getTan()`
  - **Returns** - The raw TAN
- `public String getUrl()`
  - **Returns** - The complete URL that will ask Vilfredo to display the corresponding questionnaire. The `java.net.URL` class has not been used in favour of the `String` class that is easier to use for non-Java systems that wish to access this Web Service.
- `public String toString()`



## Literaturverzeichnis

- Balk, M., 2000. *Evaluation von Lehrveranstaltungen: Die Wirkung von Evaluationsrückmeldungen*. Europäischer Verlag der Wissenschaften Peter Lang, Frankfurt am Main.
- Bargel, E. T., 2003. *Was ist gute Lehre? Studentische Erwartungen und Erfahrungen*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 35–43. Shaker Verlag, Aachen.
- Bayer, C. R., 2004. *Hochschul-Ranking: Vorschlag eines ganzheitlichen Ranking-Verfahrens*. Duncker & Humblot, Berlin.
- Beck-Bornholdt, H.-P., Dubben, H.-H., 2005. *Der Hund, der Eier legt: Erkennen von Fehlinformation durch Querdenken*. Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, 6. Auflage.
- Bortz, J., 2005. *Statistik für Human- und Sozialwissenschaftler*. Springer, Heidelberg, 6., vollständig überarbeitete und aktualisierte Auflage.
- Bortz, J., Döring, N., 2006. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer, Heidelberg, 4., überarbeitete Auflage.
- Bülow-Schramm, M., 2000. *Evaluation als Qualitätsmanagement – ein strategisches Instrument der Hochschulentwicklung? Zur Reformierbarkeit der Hochschulen nach Managementprinzipien*. In Hanft, A., (Hrsg.) *Hochschulen managen?*, 170–190. Luchterhand Verlag, Neuwied, Kriftel, Berlin.
- Burger, W., 2003. *Was ist gute Lehre? Curricula und didaktische Prinzipien*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 23–33. Shaker Verlag, Aachen.
- van den Bussche, H., Weidtmann, K., Kohler, N., Frost, M., Kaduszkiewicz, H., 2006. *Evaluation der ärztlichen Ausbildung, Methodische Probleme der Durchführung und der Interpretation von Ergebnissen*. GMS Z Med Ausbild, 23(2):Doc37.  
URL <http://www.egms.de/pdf/journals/zma/2006-23/zma000256.pdf>
- Deutsche Gesellschaft für Evaluation, 2002. *Standards für Evaluation*.  
URL <http://www.degeval.de/calimero/tools/proxy.php?id=70>
- Diekmann, A., 2005. *Empirische Sozialforschung*. Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, 14 Auflage.
- Ebel-Gabriel, C., 2006. *Begrüßung*. In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*, 11–14. Hochschulrektorenkonferenz (HRK), Bonn.

- Engel, U., Krekeler, G., 2001. *Studienqualität: Über studentische Bewertungen und Rankings von Studienfächern einer Universität*. In Engel, U., (Hrsg.) *Hochschul-Ranking: Zur Qualitätsbewertung von Studium und Lehre*, 121–176. Campus, Frankfurt am Main, New York.
- European Association for Quality Assurance in Higher Education, 2005. *ENQA report on Standards and Guidelines for Quality Assurance in the European Higher Education Area*.  
URL <http://www.enqa.eu/files/ENQA%20Bergen%20Report.pdf>
- European Ministers of Education, 1999. *The Bologna Declaration of 19 June 1999*.  
URL [http://www.bologna-berlin2003.de/pdf/bologna\\_declaration.pdf](http://www.bologna-berlin2003.de/pdf/bologna_declaration.pdf)
- European Ministers of Education, 2003. *Realising the European Higher Education Area: Communiqué of the Conference of Ministers responsible for Higher Education in Berlin on 19 September 2003*.  
URL <http://www.bologna-berlin2003.de/pdf/Communique1.pdf>
- European Ministers of Education, 2005. *The European Higher Education Area Achieving the Goals: Communiqué of the Conference of European Ministers Responsible for Higher Education, Bergen, 19-20 May 2005*.  
URL <http://www.hrk.de/de/download/dateien/Bergen-Communique.pdf>
- Frank, D., 1997. *Theorie und Geschichte der Evaluationsforschung*. In Moosbrugger, H., Frank, D., (Hrsg.) *Möglichkeiten und Grenzen der wissenschaftlichen Evaluation universitärer Lehre*, Heft 6, 1–13. Institut für Psychologie der Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Gerke, W., Breipohl, W., Forster, J., Hahn, E. G., Kraft, H.-G., Öchsner, W., Onur, Ö., Schirlo, C. H., Tullius, M., Wenekes, V., 2006. *Medizinische Ausbildung und der Bologna-Prozess: Ein Positionspapier der GMA*. *GMS Z Med Ausbild*, 23(1):Doc24.  
URL <http://www.egms.de/pdf/journals/zma/2006-23/zma000243.pdf>
- el Hage, N., 1996. *Lehrevaluation und studentische Veranstaltungskritik: Projekte, Instrumente und Grundlagen*. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, Bonn.
- Hahne, A., 2004. *Balanced Scorecard*. In Hanft, A., (Hrsg.) *Grundbegriffe des Hochschulmanagements*, 15–21. UniversitätsVerlagWebler, Neuwied.
- Herzwurm, G., 2000. *Messung der Studierendenzufriedenheit*. In Heinrich, L. J., Häntschel, I., (Hrsg.) *Evaluation und Evaluationsforschung in der Wirtschaftsinformatik: Handbuch für Praxis, Lehre und Forschung*, 395–409. Oldenbourg, München, Wien.
- Hilzenbecher, M., 2003. *Kriterien für die Bewertung der Leistungen in der medizinischen Lehre im Rahmen der leistungsorientierten Mittelvergabe in Baden-Württemberg*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation in der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 97–105. Shaker Verlag, Aachen.
- Hochschulrektorenkonferenz, 2001. *Hochschulgesetzliche Regelungen zur Qualitätssicherung*.  
URL <http://www.hrk.de/de/download/dateien/Hochschulgesetze7-2001.pdf>

- Hochschulrektorenkonferenz, 2003. *Wegweiser 2003 Qualitätssicherung an Hochschulen: Sachstandsbericht und Ergebnisse einer Umfrage des Projektes Qualitätssicherung*. Hochschulrektorenkonferenz Projekt Qualitätssicherung, Bonn.
- Hochschulrektorenkonferenz, 2004. *Metaevaluation Evaluation von Studium und Lehre auf dem Prüfstand: Zwischenbilanz und Konsequenzen für die Zukunft*. Hochschulrektorenkonferenz Projekt Qualitätssicherung, Bonn.
- Hochschulrektorenkonferenz, 2005. *Tagung Potenziale und Grenzen des Europäischen Reformprozesses für die Medizin und die Gesundheitswissenschaften in Deutschland am 14. und 15.10.2005 im Wissenschaftszentrum Bonn*.  
URL [http://www.hrk-bologna.de/bologna/de/download/dateien/Bericht\\_Tagung\\_Med.pdf](http://www.hrk-bologna.de/bologna/de/download/dateien/Bericht_Tagung_Med.pdf)
- Hofer, M., 2003. *Konsequenzen der Lehrevaluation: Neue Karriereperspektiven für die Lehrenden*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation in der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 89–95. Shaker Verlag, Aachen.
- Hopbach, A., 2005. *Hochschule entwickeln, Qualität managen: Studierende als (Mittel)punkt*. In Hochschulrektorenkonferenz, HRK, (Hrsg.) *Die Rolle der Studierenden im Prozess der Qualitätssicherung und -entwicklung*, 13–22. Hochschulrektorenkonferenz Projekt Qualitätssicherung, Bonn.
- Hornbostel, S., 2001. *Der Studienführer des CHE ein multidimensionales Ranking*. In Engel, U., (Hrsg.) *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*, 83–120. Campus, Frankfurt am Main.
- Hornbostel, S., Daniel, H.-D. D., 1994. *Das SPIEGEL-Ranking: Mediensensation oder ein Beitrag zur hochschulvergleichenden Lehrevaluation?* In Mohler, P., (Hrsg.) *Universität und Lehre*, 29–44. Waxmann, Münster.
- Hungerbühler, R., 1997. *Zielexplication und Festlegung von Bewertungskriterien*. In Moosbrugger, H., Frank, D., (Hrsg.) *Möglichkeiten und Grenzen der wissenschaftlichen Evaluation universitärer Lehre*, Heft 6, 30–39. Institut für Psychologie der Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Jacobson, I., Booch, G., Rumbaugh, J., 1999. *The unified software development process*. Addison-Wesley, Reading.
- JupiterResearch, 2006. *Retail Web Site Performance: Consumer Reaction to a Poor Online Shopping Experience*.  
URL <http://www.akamai.com/4seconds>
- Kaplan, R. S., Norton, D. P., 1992. *The Balanced Scorecard: Measures That Drive Performance*. Harvard Business Review, 72–79.
- Kostka, C., Kostka, S., 2002. *Der Kontinuierliche Verbesserungsprozess*. Carl Hanser Verlag, München.

- Koutala, A., 1997. *Professionalisierung der Evaluation künftige Perspektiven*. In Moosbrugger, H., Frank, D., (Hrsg.) *Möglichkeiten und Grenzen der wissenschaftlichen Evaluation universitärer Lehre*, Heft 6, 51–58. Institut für Psychologie der Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Kriz, J., 1994. *Die Wirklichkeit von (Vor-)Urteilen. Über die inhaltlichen und methodischen Hintergründe der STERN-Image-Analyse*. In Mohler, P., (Hrsg.) *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*, 11–28. Waxmann, Münster.
- Kromrey, H., 1994. *Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern*. In Mohler, P., (Hrsg.) *Universität und Lehre: Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*, 91–114. Waxmann, Münster.
- Kromrey, H., 2000. *Qualität und Evaluation im System Hochschule*. In Stockmann, R., (Hrsg.) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, Heft 1, 233–258. Leske + Budrich, Opladen.
- Kromrey, H., 2001a. *Evaluation von Lehre und Studium Anforderungen an Methodik und Design*. In Spiel, C., (Hrsg.) *Evaluation universitärer Lehre zwischen Qualitätsmanagement und Selbstzweck*, 21–59. Waxmann, Münster.
- Kromrey, H., 2001b. *Studierendenbefragung als Evaluation der Lehre? Anforderungen an Methodik und Design*. In Engel, U., (Hrsg.) *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*, 11–47. Campus, Frankfurt am Main.
- Kromrey, H., 2005. *Studierendenbefragung in Lehrveranstaltungen: Instrument der Evaluation oder nur der Qualitätsentwicklung?* In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Hochschule entwickeln, Qualität managen: Studierende als (Mittel)punkt. Die Rolle der Studierenden im Prozess der Qualitätssicherung und -entwicklung*. Hochschulrektorenkonferenz (HRK), Bonn.
- Künzel, R., 2006. *10 Jahre Evaluation von Lehre und Studium Erfahrungen und Zukunftsperspektiven*. In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Qualitätsentwicklung an Hochschulen: Erfahrungen und Lehren aus 10 Jahren Evaluation*, 18–32. Hochschulrektorenkonferenz (HRK), Bonn.
- Küpper, H.-U., Tropp, G., 2004. *Doppik und Kameralistik*. In Hanft, A., (Hrsg.) *Grundbegriffe des Hochschulmanagements*, 86–93. UniversitätsVerlagWebler, Bielefeld.
- Landfried, K., 2003. *Möglichkeiten und Grenzen der Lehrevaluation aus der Sicht der Hochschulen*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 9–14. Shaker Verlag, Aachen.
- Lange, J., 2002. *Evaluation Akkreditierung leistungsbezogene Mittelverteilung*. In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Evaluation, was nun? Erfahrungen mit der Umsetzung von Evaluationsergebnissen*, 7–26. Hochschulrektorenkonferenz (HRK), Bonn.
- Leszczensky, M., Orr, D., 2004. *Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung: Dokumentation und Analyse der Verfahren in 11 Bundesländern*.  
URL [http://www.his.de/pdf/pub\\_kia/kia200402.pdf](http://www.his.de/pdf/pub_kia/kia200402.pdf)

- Mandl, H., Bruckmoser, S., Konschack, J., 1999. *Problemorientiertes Lernen im Münchner Modell der Mediziner Ausbildung. Evaluation des Kardiovaskulären Kurses*. In *Forschungsbericht Nr. 105*. Ludwig-Maximilians-Universität München, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie, München.
- Marsh, H. W., Roche, L. A., 1997. *Making students evaluations of teaching effectiveness effective: The critical issues of validity, bias and utility*. *American Psychologist*, 52(11):1187–1197.
- Meffert, H., 1998. *Marketing*. Gabler, Wiesbaden.
- Mellis, W., 2000. *Lehre als Evaluationsobjekt Einführung und Grundlegung: Handbuch für Praxis, Lehre und Forschung*. In Heinrich, L. J., Häntschel, I., (Hrsg.) *Evaluation und Evaluationsforschung in der Wirtschaftsinformatik*, 383–394. Oldenbourg, München, Wien.
- Miller, G. E., 1990. *The assessment of clinical skills/competence/performance*. *Academic Medicine*, 65(9):63–67.
- Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.), 2001. *Medizinische Ausbildung in Baden-Württemberg: Bericht der Sachverständigenkommission zur Bewertung der Medizinischen Ausbildung (BeMA)*.
- Mittag, S., 2006. *Qualitätssicherung an Hochschulen: Eine Untersuchung zu den Folgen der Evaluation von Studium und Lehre*. Waxmann, Münster.
- Mittag, S., Bornmann, L., Daniel, H.-D., 2003. *Evaluation von Studium und Lehre an Hochschulen: Handbuch zur Durchführung mehrstufiger Evaluationsverfahren*. Waxmann, Münster.
- Moosbrugger, H., Hartig, J., 2001. *Zur Bedeutung von individuellen und institutionellen Studienbedingungen für die vergleichende Evaluation der Lehre*. In Engel, U., (Hrsg.) *Hochschul-Ranking: Zur Qualitätsbewertung von Studium und Lehre*, 49–60. Campus, Frankfurt am Main.
- Naftulin, D. H., Ware, J. E., Donnelly, F. A., 1973. *The Doctor Fox Lecture A Paradigm of Educational Seduction*. *Journal of Medical Education*, 48:630–635.
- Neuser, J., 2003. *Vorwort*. In Neuser, J., Urban, R. J., (Hrsg.) *Evaluation der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, VVI. Shaker Verlag, Aachen.
- Nickel, S., Zechlin, L., 2000. *Zielvereinbarungen als partizipatives Management ein Anwendungsmodell*. In Hanft, A., (Hrsg.) *Hochschulen managen? Zur Reformierbarkeit der Hochschulen nach Managementprinzipien*, 153–169. Luchterhand Verlag, Neuwied, Kriftel, Berlin.
- Nippert, P. R., 2003. *Übersicht über die Praxis der Lehrevaluation an deutschen medizinischen Fakultäten*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation in der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 47–63. Shaker Verlag, Aachen.
- Object Management Group, 2006. *Unified Modeling Language: Superstructure*.  
URL <http://www.omg.org/docs/ptc/06-04-02.pdf>

- Oestereich, B., 2006. *Analyse und Design mit UML 2.1*. Oldenbourg, München, Wien.
- Peus, V., Valerius, G., Schärer, L., Freyer, T., Berger, M., Voderholzer, U., 2005a. *Lehrevaluation an der Medizinischen Fakultät Freiburg, Teil I: Das Evaluationssystem im Überblick*. GMS Z Med Ausbildung, 22(2):Doc18.  
URL <http://www.egms.de/pdf/journals/zma/2005-22/zma000018.pdf>
- Peus, V., Valerius, G., Schärer, L., Freyer, T., Berger, M., Voderholzer, U., 2005b. *Lehrevaluation an der Medizinischen Fakultät Freiburg, Teil II: Formative Lehrveranstaltungsevaluation*. GMS Z Med Ausbildung, 22(2):Doc17.  
URL <http://www.egms.de/pdf/journals/zma/2005-22/zma000017.pdf>
- Postert, S., 2000. *Gestaltungspotenziale eines MSS-gestützten Hochschul-Managements am Beispiel der Universität Osnabrück*. Dissertation, Universität Osnabrück.  
URL [http://elib.ub.uni-osnabrueck.de/publications/diss/E-Diss102\\_thesis.pdf](http://elib.ub.uni-osnabrueck.de/publications/diss/E-Diss102_thesis.pdf)
- Putz, R., 2003. *Konsequenzen der Evaluation der Lehre für die Ressourcenverteilung*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation in der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 107–114. Shaker Verlag, Aachen.
- Radler, D., 1997. *Methode Fallstricke bei der Beurteilung von Hochschullehrern*. In Moosbrugger, H., Frank, D., (Hrsg.) *Möglichkeiten und Grenzen der wissenschaftlichen Evaluation universitärer Lehre*, Heft 6, 69–81. Institut für Psychologie der Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Reissert, R., Konnerth, T., 2001. *Evaluation von Studium und Lehre Ein wirksames Instrument zur Qualitätssicherung?* In Engel, U., (Hrsg.) *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*, 177–194. Campus, Frankfurt am Main.
- Richter, R., 1994. *Qualitätssorge in der Lehre: Leitfaden für die studentische Lehrevaluation*. Luchterhand, Neuwied, Kriftel, Berlin.
- Rindermann, H., 2001a. *Die studentische Beurteilung von Lehrveranstaltungen Forschungsstand und Implikationen*. In Spiel, C., (Hrsg.) *Evaluation universitärer Lehre zwischen Qualitätsmanagement und Selbstzweck*, 61–88. Waxman Verlag, Münster.
- Rindermann, H., 2001b. *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschule mit einem Beitrag zur Evaluation computerbasierten Unterrichts*. Verlag Empirische Pädagogik, Landau.
- Rindermann, H., 2003. *Konsequenzen aus der studentischen Veranstaltungskritik: Bluffen vereinheitlichen profilieren?* In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Evaluation und Akkreditierung*, 83–96. Hochschulrektorenkonferenz Projekt Qualitätssicherung.
- Robertson, S., Robertson, J., 2006. *Mastering the Requirements Process*. Addison-Wesley, Reading.
- Rossi, P. H., Lipsey, M. W., Freeman, H. E., 1999. *Evaluation a systematic approach*. Sage, Thousand Oaks, London.



- Scherpbier, A., 2003. *Evaluation of Clinical Teaching*. In Neuser, J., Urban, R., (Hrsg.) *Evaluation der universitären Lehre in der Medizin: Gegenstände, Methoden, Konsequenzen*, 65–70. Shaker Verlag, Aachen.
- Schienmann, B., 2002. *Kontinuierliches Anforderungsmanagement: Prozesse, Techniken, Werkzeuge*. Addison-Wesley, Reading.
- Schimank, U., 2006. *Contra Leistungsindikatoren*. In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*, 215–218. Hochschulrektorenkonferenz (HRK), Bonn.
- Schmidt, U., 2005. *Qualitätsmanagement und Hochschulentwicklung: Chancen und Modelle der Beteiligung von Studierenden*. In *Hochschule entwickeln, Qualität managen: Studierende als (Mittel)punkt. Die Rolle der Studierenden im Prozess der Qualitätssicherung und -entwicklung*, 39–46. Hochschulrektorenkonferenz Projekt Qualitätssicherung, Bonn.
- Schratz, M., 1992. *Qualität von Universitäten, Evaluation: Impulse für Innovation?* In Altrichter, H., (Hrsg.) *Hochschullehre unter der Lupe. Auf dem Weg zur professionellen Selbstevaluation*, 227–255. Österreichischer StudienVerlag, Innsbruck, Wien, Bozen.
- Scriven, M., 1980. *The Logic of Evaluation*. Edgepress, California.
- Soellner, R., Braun, E., Gusy, B., 2005. *Lehrevaluation aus pädagogisch-psychologischer Sicht. Das Berliner Evaluationsinstrument für Kompetenzen*. In Hochschulrektorenkonferenz (HRK), (Hrsg.) *Hochschule entwickeln, Qualität managen: Studierende als (Mittel)punkt. Die Rolle der Studierenden im Prozess der Qualitätssicherung und -entwicklung*, 72–80. Hochschulrektorenkonferenz Projekt Qualitätssicherung.
- Spiel, C., Engel, U., 2001. *Der differentielle Einfluß von Biasvariablen auf studentische Lehrveranstaltungsbewertungen*. In *Hochschul-Ranking: Zur Qualitätsbewertung von Studium und Lehre*, 61–82. Campus, Frankfurt am Main.
- Spiel, C., Gössler, M., 2001. *Zwischen Selbstzweck und Qualitätsmanagement: Quo vadis, evaluatione?* In Spiel, C., (Hrsg.) *Evaluation universitärer Lehre zwischen Qualitätsmanagement und Selbstzweck*, 9–20. Waxmann Verlag, Münster.
- Stangl, W., 2000. *Die Evaluation universitärer Lehrveranstaltungen*.  
URL <http://paedpsych.jk.uni-linz.ac.at/paedpsych/Evaluation/EvaluationMethoden.html#Materialien%20f%FCr%20Lehrveranstaltungseinheiten>
- Titscher, S., 2006. *Pro Leistungsindikatoren Indikatoren zur Bestimmung der Forschungs- und Lehrleistung*. In Hochschulrektorenkonferenz, (Hrsg.) *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*, 205–212. Hochschulrektorenkonferenz (HRK), Bonn.
- Troschke, von, J., 2006. *Auswirkungen der Neustrukturierung der Hochschulbildung im Kontext des sogenannten Bologna-Prozesses auf das Medizinstudium in Deutschland*. *GMS Z Med Ausbild*,

- 23(1):Doc23.  
URL <http://www.egms.de/pdf/journals/zma/2006-23/zma000242.pdf>
- Weber, H., 2006. *Leistungsmessung in der Lehre*. In Hochschulrektorenkonferenz, HRK, (Hrsg.) *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*, 273–280. Hochschulrektorenkonferenz (HRK), Bonn.
- Weber, M., 1922. *Wissenschaft als Beruf*.  
URL <http://www.textlog.de/2317-2.html>
- Weiß, C., 2002. *Basiswissen Medizinische Statistik*. Springer-Verlag, Berlin, Heidelberg, New York.
- Wissenschaftsrat, 1992. *Leitlinien zur Reform des Medizinstudiums*. Wissenschaftsrat, Berlin.
- Wissenschaftsrat, 1996. *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Wissenschaftsrat, Berlin.
- Wissenschaftsrat, 2002a. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Albert-Ludwigs-Universität Freiburg*.  
URL <http://www.wissenschaftsrat.de/texte/5443-02.pdf>
- Wissenschaftsrat, 2002b. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Universität Ulm*.  
URL <http://www.wissenschaftsrat.de/texte/5444-02.pdf>
- Wissenschaftsrat, 2003a. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Ruprecht-Karls-Universität Heidelberg in Heidelberg*.  
URL <http://www.wissenschaftsrat.de/texte/5517-03.pdf>
- Wissenschaftsrat, 2003b. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Ruprecht-Karls-Universität Heidelberg in Mannheim*.  
URL <http://www.wissenschaftsrat.de/texte/5516-03.pdf>
- Wissenschaftsrat, 2004a. *Empfehlungen zu forschungs- und lehrförderlichen Strukturen in der Universitätsmedizin*.  
URL <http://www.wissenschaftsrat.de/texte/5913-04.pdf>
- Wissenschaftsrat, 2004b. *Standortübergreifende Stellungnahme zur Weiterentwicklung der Universitätsmedizin in Baden-Württemberg*.  
URL <http://www.wissenschaftsrat.de/texte/6196-04.pdf>
- Wissenschaftsrat, 2005a. *Stellungnahme zu Leistungsfähigkeit, Ressourcen und Größe universitätsmedizinischer Einrichtungen*.  
URL <http://www.wissenschaftsrat.de/texte/6913-05.pdf>
- Wissenschaftsrat, 2005b. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Ludwig-Maximilians-Universität München*.  
URL <http://www.wissenschaftsrat.de/texte/6901-05.pdf>

- Wissenschaftsrat, 2006a. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Bayerischen Julius-Maximilians-Universität Würzburg.*  
URL <http://www.wissenschaftsrat.de/texte/7065-06.pdf>
- Wissenschaftsrat, 2006b. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Friedrich-Alexander-Universität Erlangen-Nürnberg.*  
URL <http://www.wissenschaftsrat.de/texte/7244-06.pdf>
- Wissenschaftsrat, 2006c. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Technischen Universität München.*  
URL <http://www.wissenschaftsrat.de/texte/7061-06.pdf>
- Wissenschaftsrat, 2006d. *Stellungnahme zur weiteren Entwicklung der Medizinischen Einrichtungen der Universität Regensburg.*  
URL <http://www.wissenschaftsrat.de/texte/7245-06.pdf>
- Wolf, P., Spiel, C., Pellert, A., 2001. *Entwicklung eines Fragebogens zur globalen Lehrveranstaltungsevaluation ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz.* In Spiel, C., (Hrsg.) *Evaluation universitärer Lehre zwischen Qualitätsmanagement und Selbstzweck*, 89–109. Waxman Verlag, Münster.
- Wottawa, H., 2001. *Qualitätsmanagement durch Zielvereinbarungen.* In Spiel, C., (Hrsg.) *Evaluation universitärer Lehre zwischen Qualitätsmanagement und Selbstzweck*, 151–164. Waxmann Verlag, Münster.
- Wottawa, H., Thierau, H., 1998. *Lehrbuch Evaluation.* Verlag Hans Huber, Bern, Göttingen, Toronto, Seattle.



# Lebenslauf

Marc Gumpinger  
geboren am 3. April 1974 in München

## Ausbildung / wissenschaftliche Tätigkeiten

---

- 1994 – Arbeit am Institut für Medizinische Informationsverarbeitung, Biometrie und  
1997 Epidemiologie (IBE) an der LMU München.
- 
- 1995 – Studium der Betriebswirtschaftslehre. Abschluss: Diplom-Kaufmann.  
2005 Vordiplom: LMU München, Diplom: FernUniversität Hagen.
- 
- 2005 – Promotion zum Doktor der Humanbiologie  
2008 am IBE der Ludwig-Maximilians-Universität (LMU) München.
- 

## Berufliche Tätigkeiten

---

- 1997 – Partner der Gumpinger & Koppany GbR.  
2000 Entwicklung von „plugware.net“. Wurde 2000 von der COSS AG übernommen.
- 
- 2000 – Product Marketing Manager bei COSS Systemtechnik AG.  
2001 Verantwortlich für die Produktentwicklung.
- 
- 2001 – Softwarearchitekt und -entwickler.  
2005 Entwicklung von Systemen vor allem für Handels- und Medienunternehmen.
- 
- seit 2008 Associate bei Target Partners GmbH.  
Investment Analyst im Bereich Venture Capital.
- 

## Veröffentlichungen / Konferenzen

---

- 1995 – Autor von über 60 Artikeln über Softwareentwicklung und neue Technologien  
1999 Veröffentlicht unter anderem in den Fachzeitschriften c't und Macwelt.
- 

Gumpinger, M., 1996. *AppleScript*. Addison Wesley Longman, Bonn.

---

Batschkus, M. M., Gumpinger, M., Dugas, M., 1998. *The Browser Is the Interface: Creating a Multifunctional Workstation for Students with Integrated Access to Online and Offline Media*. Computers in Healthcare Education Symposium and Slice of Life Workshop on Interactive Multimedia.

---

Batschkus, M. M., Gumpinger, M. 1998. *Innovative Lösungen für eine transparente Nutzerführung zu Online- und Offline-Angeboten an Lernstationen für Studenten*. 3. Workshop der AG Lehr- und Lernsysteme der gmds.

---

Gumpinger, M. 2007. *Implementation eines innovativen online Lehrevaluationssystems im medizinischen Curriculum*. 12. Europäische Jahrestagung der Gesellschaft für Medien in der Wissenschaft (GMW07).

---

Gumpinger, M. 2007. *Effiziente Online-Lehrevaluation im Medizinischen Curriculum*. Kongress „Medizin und Gesellschaft“ der gmds.

---