

**Universitätsweite Lehrveranstaltungsevaluation
an der Universität Ulm**

**Abschlussbericht
der Arbeitsgruppe Lehrveranstaltungsevaluation**

Markus Dresel
Seminar für Pädagogik

Karen Tinsner
Studiendekanat Medizin

Claudia Schröck-Pauli
Kommunikations- und Informationszentrum

für die Arbeitsgruppe

Ulm, Januar 2006

Inhaltsverzeichnis

1	Zusammenfassung: Empfehlungen der Arbeitsgruppe	3
2	Auftrag der Arbeitsgruppe	8
3	Zusammensetzung der Arbeitsgruppe	8
4	Überblick über die Tätigkeiten der Arbeitsgruppe	9
5	Ziele der Lehrveranstaltungsevaluation	10
6	Verpflichtende und freiwillige Evaluation.....	11
7	Modellvorstellungen zur Lehrqualität.....	12
7.1	Definition von Lehrqualität	12
7.2	Bedingungen hoher Lehrqualität seitens der Lehrenden	13
7.3	Studentische Bedingungen und Rahmenbedingungen.....	13
7.4	Modell der Lehrqualität	14
8	Instrumentarium	16
8.1	Kriterien für Fragebögen zur Evaluation von Lehrveranstaltungen.....	16
8.2	Fragebogen zur Evaluation von Vorlesungen.....	17
8.3	Empirische Überprüfung des neu konzipierten Fragebogens	20
8.4	Zukünftige Entwicklung von Fragebögen	21
9	Maßstäbe zur Bewertung der Lehrqualität	24
9.1	Bewertung anhand von Durchschnittsergebnissen.....	25
9.2	Bewertung anhand von Qualitätskriterien	25
9.3	Bewertung anhand anderer Teilaspekte derselben Lehrveranstaltung.....	26
9.4	Bewertung anhand früherer Ergebnisse	26
10	Zeitpunkt der Studierendenbefragung.....	27
11	Ergebnisrückmeldung, Berichtslegung und Ergebnisveröffentlichung.....	29
11.1	Rückmeldung der Ergebnisse an die Dozent(inn)en.....	29
11.2	Berichte für Studiendekane, Dekane und den Prorektor für Lehre	30
11.3	Rückmeldung der Ergebnisse an die Studierenden.....	31
11.4	Veröffentlichung der Ergebnisse.....	32
12	Beratungsangebot für Dozent(inn)en	32
13	Form der Datenerhebung.....	33

14	Evaluationsworkflow.....	35
14.1	Anforderungen an den Evaluationsworkflow.....	35
14.2	Testweise Umsetzung eines Evaluationsworkflow.....	36
14.3	Empfohlener Evaluationsworkflow.....	36
14.4	Feldtest des Evaluationssystems <i>EvaSys</i>	40
14.5	Zusammenfassende und ergänzende Analysen.....	42
15	Datenschutz.....	42
16	Erforderliche personelle und sachliche Ressourcen.....	43
17	Literatur.....	45
Anhang A:	Fehler bei der scannergestützten Datenerfassung mit <i>EvaSys</i>	48

1 Zusammenfassung: Empfehlungen der Arbeitsgruppe

Mit dem vorliegenden Bericht legt die vom Senatausschuss Lehre eingesetzte Arbeitsgruppe Lehrveranstaltungsevaluation ein Gesamtkonzept für die Universität vor. Dieses Konzept mündet in 20 Empfehlungen, die im Folgenden aufgeführt sind. Grundlage des Konzepts sind theoretische Überlegungen, ein eingehendes Studium der einschlägigen Fachliteratur zur Lehrveranstaltungsevaluation, Bedürfnisse der Fächer und der Hochschulleitung, Erfahrungen der Fachschaften sowie mehrere umfassende Testevaluationen und deren Auswertung mit empirisch-sozialwissenschaftlichen Methoden.

Empfehlung 1: Ziele der Lehrveranstaltungsevaluation an der Universität Ulm

Das vorrangige Ziel der Lehrveranstaltungsevaluation an der Universität Ulm ist die Bereitstellung von Feedback, das für Dozent(inn)en zur Verbesserung der Lehrqualität nützlich ist. Daneben soll regelmäßig die Gesamtqualität der Lehre an der Universität und in den Fächern bestimmt werden. Schließlich sollen Lehrveranstaltungen identifiziert werden, die definierten Mindestanforderungen an die Lehrqualität nicht genügen.

Empfehlung 2: Verpflichtende und freiwillige Evaluation

Die Arbeitsgruppe empfiehlt, jedes Semester 25% der Lehrveranstaltungen in jedem Fach verpflichtend zu evaluieren. Dabei soll jede regelmäßig angebotene Lehrveranstaltung mindestens alle zwei Jahre und ebenso jede(r) hauptberufliche Dozent(in) mindestens alle zwei Jahre evaluiert werden.

Die Verantwortlichkeit bei der Auswahl der verpflichtend zu evaluierenden Lehrveranstaltungen eines Fachs sollte bei den Studiendekanaten bzw. Studienkommissionen liegen.

Über eine verpflichtende Evaluation hinaus soll allen Fächern die Möglichkeit zur Verfügung stehen, beliebig viele Lehrveranstaltungen (z.B. alle Lehrveranstaltungen) freiwillig zu evaluieren und dabei auf fachkompetente und organisatorische Unterstützung zurückzugreifen.

Empfehlung 3: Akzeptanzsteigerung der Lehrveranstaltungsevaluation

Es sollten Maßnahmen zur Förderung der Akzeptanz der Lehrveranstaltungsevaluation bei Lehrenden und Studierenden ergriffen werden.

Empfehlung 4: Zu berücksichtigende Aspekte der Lehrqualität

Der Lehrveranstaltungsevaluation an der Universität Ulm sollten theoretische Überlegungen zur Lehrqualität zugrunde gelegt werden. Dabei sollte unterschieden werden zwischen dem Lehrerfolg (Produkte guter Lehre: Lernzuwachs, Interessenförderung, Gesamtbeurteilung) als Kriterium der Lehrqualitätsbeurteilung, dem didaktischen Handeln der Do-

zent(inn)en (Prozessmerkmale der Lehre) als Ansatzstelle zur Verbesserung der Lehrqualität sowie Hintergrundvariablen, durch die studentische Veranstaltungsbeurteilungen verzerrt sein können. In der Lehrveranstaltungsevaluation sollten wichtige Merkmale aller drei Merkmalsgruppen erfasst werden.

Das bislang auf Vorlesungen spezifizierte theoretische Modell der Lehrqualität sollte zukünftig im Hinblick auf andere Veranstaltungsformen und den Einsatz von E-Learning in der Lehre erweitert und überprüft werden. Zudem empfiehlt sich, nach der Definition der Modulkataloge konsekutiver Studiengänge zu prüfen, inwieweit der Lehrerfolg ergänzend auch über die darin enthaltenen Lernziele definiert werden können.

Empfehlung 5: Fragebogen zur Evaluation von Vorlesungen

In allen Fächern sollte zur Beurteilung der Lehrqualität von Vorlesungen der von der Arbeitsgruppe entwickelte und psychometrisch getestete „Fragebogen zur Evaluation von Vorlesungen“ eingesetzt werden, der die drei Module „Lehrqualität“, „Didaktisches Handeln“ und „Studentische Hintergrundvariablen“ enthält.

Der obligatorische Teil des Fragebogens (22 geschlossene und 2 offene Fragen) kann dabei fachspezifisch um optionale Fragen ergänzt werden.

Empfehlung 6: Zukünftige Entwicklung von Fragebögen

Es empfiehlt sich die zukünftige Entwicklung von weiteren Fragebögen(-modulen). Diese sollen den formulierten allgemeinen Kriterien für Fragebögen zur Evaluation von Lehrveranstaltungen genügen, vor ihrem Einsatz empirisch überprüft und modular gegliedert werden.

Besonderer Bedarf besteht im Hinblick auf weitere Veranstaltungsformen und den Einsatz von E-Learning in der Lehre. Dabei ist den Charakteristika der Fächer durch die Entwicklung spezifischer Fragebögen Rechnung zu tragen. Daneben sind englische Versionen der Fragebögen erforderlich.

Schließlich besteht laufender Bedarf für die Entwicklung von Fragebögen für spezifische Evaluationsinteressen (z.B. Einführung, Neu- oder Restrukturierung bestehender Studiengänge).

Empfehlung 7: Erforderliche Fachkompetenz zur Entwicklung von Fragebögen

Den Fächern der Universität Ulm sollte die erforderliche Fachkompetenz zur theoretisch fundierten Entwicklung und psychometrischen Überprüfung von Fragebögen zur Verfügung stehen.

Empfehlung 8: Definition von Qualitätskriterien der Lehrqualität

Es sollten universitätsweit einheitliche und anspruchsvolle Qualitätskriterien der Lehrqualität definiert werden, mit denen festgelegt wird, wie gut die Lehrqualität auf Produktebene (Lehrerfolg) mindestens ausgeprägt sein soll. Diese Definition soll auf der Grundlage sachlicher Überlegungen erfolgen.

Die Bewertung der Lehrqualität von Lehrveranstaltungen, Fächern und der Universität soll auf der Grundlage dieser Qualitätskriterien der Lehrprodukte erfolgen. Von einer vergleichenden Bewertung einzelner Lehrveranstaltungen oder Dozent(inn)en anhand von Durchschnittswerten ist abzuraten.

Ergänzend empfiehlt sich zur Identifikation von Verbesserungsmöglichkeiten auch die unverbindliche Definition von Qualitätskriterien im Hinblick auf die Prozessmerkmale guter Lehre (didaktisches Handeln). Diese können Charakteristika der einzelnen Fächern Rechnung tragen.

Empfehlung 9: Zeitpunkt der Befragung

Die Lehrveranstaltungsevaluation soll zur Mitte des Semesters durchgeführt werden, damit Evaluationsergebnisse noch im laufenden Semester zu einer Anpassung der Lehre beitragen können.

Empfehlung 10: Rückmeldung an die Dozent(inn)en

Die Dozent(inn)en sollen möglichst schnell (ideal: innerhalb von zwei Wochen) und in jedem Fall noch im laufenden Semester eine Rückmeldung über die Ergebnisse der Evaluation ihrer Lehrveranstaltungen erhalten.

Diese Rückmeldung soll leicht verständlich und auch für statistische Laien gut interpretierbar sein. Aspekte der Lehre, die nicht den definierten Mindestkriterien an die Lehrqualität entsprechen, sollten gekennzeichnet sein.

Empfehlung 11: Berichte für Studiendekane, Dekane und den Prorektor für Lehre

Studiendekane und Dekane sollten am Ende jedes Semesters einen zusammenfassenden Bericht über die Evaluationsergebnisse ihres Bereichs erhalten.

Ebenfalls am Ende jeden Semesters soll der Prorektor Lehre einen auf höherer Ebene zusammengefassten Bericht erhalten. Dieser Bericht kann auch zur Dokumentation und zur Rechtfertigung der Gesamtlehrqualität der Universität und der Fächer dienen.

Empfehlung 12: Ergebnisrückmeldung an die Studierenden

Die Arbeitsgruppe empfiehlt, dass die Dozent(inn)en die Evaluationsergebnisse in ihren Lehrveranstaltungen vorstellen und Unklarheiten sowie etwaige Veränderungswünsche mit den Studierenden diskutieren. Diese Ergebnispräsentation in den Lehrveranstaltungen sollte den Dozent(en) nachdrücklich empfohlen werden, aber dennoch freiwillig sein.

Dazu sollte die Ergebnismeldung für die Dozent(inn)en eine Vorlage für die Ergebnispräsentation in der Lehrveranstaltung beinhalten.

Empfehlung 13: Veröffentlichung der Ergebnisse

Es empfiehlt sich eine Veröffentlichung der Evaluationsergebnisse, bei der die Ergebnisse der einzelnen Veranstaltungen auf höherer Ebene (z.B. Fächer, Studiengänge) zusammengefasst werden (Aggregation der Daten).

Um eine Demotivierung der Dozent(inn)en bei der Verbesserung ihrer Lehrqualität zu vermeiden, empfiehlt die Arbeitsgruppe, vergleichende Bewertungen auf Ebene der einzelnen Lehrveranstaltungen (z.B. in Form von Rankings) nicht zu veröffentlichen.

Empfehlung 14: Beratungsangebot für Lehrende

Für Lehrende, die Verbesserungsbedarf für ihre didaktischen Handlungsweisen erkennen, sollen Beratungsgespräche zur Interpretation der Evaluationsergebnisse und zur Verbesserung der individuellen Lehrkompetenz angeboten werden.

Empfehlung 15: Form der Datenerhebung

Es wird empfohlen, die Evaluation der Lehrveranstaltungen bis auf weiteres mittels Papierfragebögen durchzuführen. Der mögliche Effizienzgewinn einer Datenerhebung mit einem reinen Onlinesystem wiegt momentan nicht deren Nachteile auf (geringe Beteiligungsquote und Selbstselektionseffekte).

Empfehlung 16: Zentrale Evaluation unter Beteiligung der Fächer und Fachschaften

Es wird das Vorgehen einer zentralen Evaluation empfohlen. Dieses Vorgehen eignet sich besonders für die verpflichtende und in einigen Fächern flächendeckende Lehrveranstaltungsevaluation und bietet die Möglichkeit der Vergleichbarkeit der Ergebnisse. Die organisatorische, technische als auch die inhaltliche Betreuung sollte zentral koordiniert werden. Wichtig ist eine feste Einbindung der jeweiligen Fächern und Fachschaften in den Bereichen der Definition von Evaluationsschwerpunkten der Datensammlung, Datenerfassung und der konzeptionellen Weiterentwicklung von Erhebungsinstrumenten sowie der Rückmeldungen und Interpretation der Ergebnisse.

Die Arbeitsgruppe empfiehlt unabhängig davon, ob Lehrveranstaltungen obligatorisch oder freiwillig evaluiert werden, die Abwicklung über eine zentrale Stelle.

Empfehlung 17: Evaluationssystem

Es wird der universitätsweite Einsatz des Evaluationssystems EvaSys empfohlen. Verwendet werden sollte der in der Testlizenz nicht enthaltene Fragebogendesigner Teleform, der alle erforderlichen Funktionalitäten zur Verfügung stellt.

Daneben ist der Einsatz der XML-Importschnittstelle von LSF nach EvaSys sowohl aufgrund der großen Zahl der zu verarbeitenden Veranstaltungsdaten wie auch zur Vermeidung fehlerhafter Daten durch Medienbrüche zwingend notwendig.

Es wird empfohlen, den technischen Support zentral an das Kommunikations- und Informationszentrum (kiz) anzubinden, das sich dazu bereits bereit erklärt hat.

Empfehlung 18: Ergänzende Auswertungen

Es empfiehlt sich, am Ende jedes evaluierten Semesters zusätzliche, komplexere statistische Auswertungen vorzunehmen, um die Definition der Qualitätskriterien, den Einfluss von Verzerrungsvariablen und die Güte neu entwickelter Fragebögen zu überprüfen sowie die längsschnittliche Entwicklung der Lehrqualität zu bestimmen.

Empfehlung 19: Datenschutz

Bei der Evaluation der Lehrveranstaltungen der Universität Ulm müssen die relevanten Bestimmungen zum Schutz der personbezogener Daten von Studierenden und Dozent(inn)en beachtet werden. Näheres dazu sollte eine Evaluationssatzung regeln. Bei der konkreten Durchführung der Evaluation sollten die Empfehlungen der Zentralen Datenschutzstelle in Baden-Württemberg (ZENDAS) berücksichtigt werden.

Empfehlung 20: Einrichtung einer „Arbeitsstelle Lehrveranstaltungsevaluation“

Die Arbeitsgruppe empfiehlt die Einrichtung einer zentralen „Arbeitsstelle Lehrveranstaltungsevaluation“, die dem Seminar für Pädagogik angegliedert ist.

Diese Arbeitsstelle soll mit einer halben wissenschaftlichen Mitarbeiterstelle (BAT IIa/2) ausgestattet sein. Der/die wissenschaftliche Mitarbeiter(in) soll ein Diplom in Psychologie, Soziologie oder Pädagogik aufweisen und über sehr gute Kenntnisse in den Bereichen empirisch-sozialwissenschaftliche Methoden, Konstruktion und Validierung von Fragebögen sowie Evaluation verfügen. Spezifische Qualifikationen in den Bereichen der Lehrveranstaltungsevaluation, Hochschuldidaktik und Beratung sind wünschenswert. Auf der Mitarbeiterstelle soll die Möglichkeit zur Promotion gegeben sein.

Daneben soll die Arbeitsstelle mit 80 Hilfskraftstunden pro Monat zur organisatorischen und administrativen Unterstützung ausgestattet sein.

Im Hinblick auf die Sachmittel sollte die Arbeitsstelle mit zwei Notebooks mit Dockingstation, einem leistungsfähigen Laserdrucker und laufenden Sachmitteln (Papier, Druck, Softwarelizenzen) ausgestattet sein.

2 Auftrag der Arbeitsgruppe

In der Sitzung vom 12. Januar 2005 des Senatsausschuss Lehre der Universität wurde die Arbeitsgruppe „Universitätsweite Lehrveranstaltungsevaluation“ eingesetzt, um auf der Grundlage der bisherigen Tätigkeiten und Materialien der Fachschaften, des Studiendekans Medizin und des Seminars für Pädagogik ein effektives und praktikables Konzept der Lehrveranstaltungsevaluation für die Universität Ulm zu entwickeln.

Im Einzelnen wurde die Arbeitsgruppe mit folgenden Aufgaben betraut:

- Entwicklung eines universitätsweit einsetzbaren Fragebogens zur Evaluation von Vorlesungen und dessen Überprüfung mit empirisch-sozialwissenschaftlichen Methoden
- Bedarfsanalyse für weitere Fragebogenverfahren
- Erarbeitung eines effizienten Gesamtprozesses der Lehrveranstaltungsevaluation, der universitätsweit auf Dauer umgesetzt werden kann und effektiv im Sinne der mit der Evaluation verfolgten Ziele ist
- Beschreibung der dafür erforderlichen Ressourcen (Personal- und Sachmittel)
- Analyse des Effizienzgewinns und der Limitationen durch eine ausschließlich internetgestützte Evaluation von Lehrveranstaltungen
- Feldtest eines computergestützten Evaluationssystems, mit dem die Anforderungen an den Gesamtprozess realisiert werden können

3 Zusammensetzung der Arbeitsgruppe

Mit der Leitung der Arbeitsgruppe wurden Herr Dr. Markus Dresel (Seminar für Pädagogik) und Herr Hubert Liebhardt M.A. (Studiendekanat Medizin) beauftragt. Als weitere Mitglieder wurden je ein Vertreter pro Fachschaft (Mathematik: Herr Bernhard Wieland, Wirtschaftswissenschaften: Frau Anne-Kathrin Seidelt, Informatik: Herr Christian Sauter, Physik: Herr Stefan Pleus, Chemie: Herr Markus Tonigold, Biologie: Herr Philip Reichel, Elektrotechnik: Frau Adelheid Grob) sowie Frau Dr. Claudia Schröck-Pauli (Kommunikations- und Informationszentrum) und Frau Dipl. Soz. Karen Tinsner (Studiendekanat Medizin) benannt.

Im Laufe der einjährigen Tätigkeit der Arbeitsgruppe wechselten die Vertreter für fünf der sieben Fachschaften (Wirtschaftswissenschaften: Frau Andrea Hefler, Physik: Herr Robert Rosenbach, Chemie: Herr Sebastian Kube, Biologie: Frau Evelyn Böhm, Elektrotechnik: Herr Hermann Schmidt).

4 Überblick über die Tätigkeiten der Arbeitsgruppe

Die Arbeitsgruppe tagte im Berichtszeitraum fünf Mal. Die Protokolle der Sitzungen können bei der Arbeitsgruppenleitung angefordert werden. Darüber hinaus fand eine Reihe von Koordinationsgesprächen in der Arbeitsgruppenleitung und mit dem Prorektor Lehre statt.

Die Arbeitsgruppe führte zwei Testevaluationen durch.

Schwerpunktmäßiger Gegenstand der ersten Testevaluation im Sommersemester 2005, an der Studierende aus zehn Vorlesungen verschiedener Fakultäten teilnahmen (1.292 ausgewertete Fragebögen), war die empirische Überprüfung des in der Arbeitsgruppe entwickelten Fragebogens zur Evaluation von Vorlesungen und der Vergleich zwischen papiergebundener Evaluation und Evaluation mit einem Onlinesystem. Mit Schreiben vom 12.04.2005 wurden die Studiendekane aller Fakultäten darüber informiert und gebeten, Vorlesungen für die Testevaluation zu benennen. Alle beteiligten Dozent(inn)en erhielten eine Rückmeldung über die Ergebnisse ihrer Veranstaltung.

Dem Senatsausschuss Lehre wurde in der Sitzung vom 7. Juli 2005 ein Zwischenbericht über Ergebnisse und Folgerungen der ersten Testevaluation vorgelegt.¹ Die Voten des Ausschusses zur Umsetzung der Lehrveranstaltungsevaluation an der Universität stellten die Grundlage für die zweite Testevaluation und die Erarbeitung der in diesem Bericht aufgeführten Empfehlungen dar.

Schwerpunkt der zweiten Testevaluation im Wintersemester 2005/06 war die Überprüfung der Realisierbarkeit eines auf zeitnahe Rückmeldung ausgelegten Gesamtevaluationsprozesses unter den Bedingungen eines Vollbetriebs sowie der Feldtest des Evaluationssystems *EvaSys* (Fa. Electric Paper), das sich in einer vorab durchgeführten Marktanalyse als dafür am besten geeignet erwies.² Im Rahmen dieses Testlaufs wurden alle Vorlesungen des ersten Semesters aus allen Fächern (Medizin: zusätzlich drittes Semester) verpflichtend und eine große Zahl weiterer Vorlesungen freiwillig evaluiert. Insgesamt wurden bis zum Zeitpunkt der Berichtslegung 102 Vorlesungen evaluiert (7.866 ausgewertete Fragebögen), darunter alle von der Studienkommission Mathematik und ein Großteil der von der Studienkommission Medizin verantworteten Vorlesungen. Mit Schreiben vom 18.10.2005 wurden die Studiendekane aller Fakultäten über den Testlauf informiert. Alle beteiligten Dozent(inn)en erhielten noch im laufenden Semester eine Rückmeldung über die Ergebnisse ihrer Veranstaltung.

¹ Der Zwischenbericht liegt auch in schriftlicher Form vor und enthält eine exakte Darstellung der Ergebnisse der Fragebogenüberprüfung und des Vergleichs von Online- und Papiergebundener Evaluation (Dresel, 2005a). Deshalb werden die Ergebnisse des ersten Testlaufs im vorliegenden Bericht nicht detailliert dargestellt. Der Zwischenbericht ist bei der Arbeitsgruppenleitung und elektronisch verfügbar: http://www.informatik.uni-ulm.de/t3-sfp/fileadmin/user/dresel/Zwischenbericht_AG_LVE_Juli_2005.zip (ca. 553 kB)

² Die Ergebnisse der zweiten Testevaluation sind in dem vorliegenden Bericht in den einzelnen Kapiteln beschreiben.

5 Ziele der Lehrveranstaltungsevaluation

Grundlegend für jedwede Form der Evaluation ist eine exakte Bestimmung der damit verfolgten Ziele (Wottawa & Thierau, 1998). Evaluationsmodell, verwendetes Instrumentarium, Verwertung der Evaluationsergebnisse sowie darauf basierende Folgerungen leiten sich aus den Evaluationszielen ab und in Abhängigkeit von diesen variieren werden.

Die folgende Empfehlung zu den Zielen der Evaluation von Lehrveranstaltungen an der Universität Ulm stellt nicht nur den Konsens der Arbeitsgruppe dar, sondern reflektiert auch die Bedürfnisse der Fächer und der Universitätsleitung.

Empfehlung 1: Ziele der Lehrveranstaltungsevaluation an der Universität Ulm

Das vorrangige Ziel der Lehrveranstaltungsevaluation an der Universität Ulm ist die Bereitstellung von Feedback, das für Dozent(inn)en zur Verbesserung der Lehrqualität nützlich ist. Daneben soll regelmäßig die Gesamtqualität der Lehre an der Universität und in den Fächern bestimmt werden. Schließlich sollen Lehrveranstaltungen identifiziert werden, die definierten Mindestanforderungen an die Lehrqualität nicht genügen.

Mit dieser Zieldefinition stellt das primäre Evaluationsziel die Verbesserung der Lehrqualität dar. Die Lehrveranstaltungsevaluation soll Dozent(inn)en Anhaltspunkte dafür liefern, wie sie ihr didaktisches Handeln verbessern, ihre Lehrkompetenz entwickeln und damit die Qualität ihrer Lehrveranstaltungen steigern und somit zu einem besseren Kompetenzerwerb seitens der Studierenden beitragen können. Eine Einschätzung der Gesamtlehrqualität sollte nur auf aggregierter Ebene der Fächer und der Gesamtuniversität erfolgen und zur Dokumentation und Rechtfertigung der Lehrqualität sowie als Entscheidungsgrundlage für Maßnahmen zur Qualitätssicherung dienen. Von einer vergleichenden Bewertung auf Ebene einzelner Lehrveranstaltungen oder Dozent(inn)en wird explizit abgeraten, da damit (a) zu hohe Anforderungen an die Objektivität von studentischen Veranstaltungsbeurteilungen gestellt werden, (b) Gesamtverbesserungen der Lehrqualität nicht abgebildet werden und (c) Dozent(inn)en häufig eher demotiviert als zu verbessertem Lehrhandeln aktiviert werden (vgl. Abschnitt 9.1).

Stattdessen sollten auf der Grundlage vorab und sachbezogen definierter Mindeststandards der Lehrqualität (sog. Qualitätskriterien) nur jene Lehrveranstaltungen identifiziert werden, die diesen Standards nicht genügen. Die Identifikation dieser Veranstaltungen dient dazu, den verantwortlichen (Studien-)Dekanen Ergebnisse zur Verfügung zu stellen, anhand derer sie mit den verantwortlichen Dozent(inn)en Maßnahmen zur Verbesserung der Lehrqualität erörtern können.

Zusätzlich zu den genannten Zielen wünscht sich die Fachschaft Elektrotechnik, dass mit der Lehrveranstaltungsevaluation Studierenden auch Entscheidungshilfen bei der Auswahl von Wahlpflichtveranstaltungen zur Verfügung gestellt werden.³

6 Verpflichtende und freiwillige Evaluation

Um belastbare, umfassende und aktuelle Indizien zur Lehrqualität zu erhalten und auch Dozent(inn)en, die der Lehrveranstaltungsevaluation skeptisch gegenüber stehen, zu erreichen, ist eine verpflichtende Evaluation erforderlich. In gewissen Intervallen sollen dabei alle hauptberuflichen Dozent(inn)en sowie alle regelmäßig stattfindenden Lehrveranstaltungen einbezogen werden.

Gleichzeitig sollte bei der Durchführung der Lehrveranstaltungsevaluation darauf geachtet werden, Lehrende und Studierende mit der Evaluation und nachfolgenden Aufgaben (Diskussion und Vorstellung der Ergebnisse, Beratungen, Trainings etc.) zeitlich nicht zu überlasten (vgl. Rindermann, 2001a).

Vor dem Hintergrund dieser beiden Randbedingungen gelangte die Arbeitsgruppe im Hinblick auf Umfang und Turnus der verpflichtenden Evaluation von Lehrveranstaltungen zu der folgenden Empfehlung:

Empfehlung 2: Verpflichtende und freiwillige Evaluation

Die Arbeitsgruppe empfiehlt, jedes Semester 25% der Lehrveranstaltungen in jedem Fach verpflichtend zu evaluieren. Dabei soll jede regelmäßig angebotene Lehrveranstaltung mindestens alle zwei Jahre und ebenso jede(r) hauptberufliche Dozent(in) mindestens alle zwei Jahre evaluiert werden.

Die Verantwortlichkeit bei der Auswahl der verpflichtend zu evaluierenden Lehrveranstaltungen eines Fachs sollte bei den Studiendekanaten bzw. Studienkommissionen liegen.

Über eine verpflichtende Evaluation hinaus soll allen Fächern die Möglichkeit zur Verfügung stehen, beliebig viele Lehrveranstaltungen (z.B. alle Lehrveranstaltungen) freiwillig zu evaluieren und dabei auf fachkompetente und organisatorische Unterstützung zurückzugreifen.

Es ist davon auszugehen, dass einige Lehrende der verpflichtenden Evaluation skeptisch gegenüber stehen werden. Dies dürfte insbesondere für diejenigen Fächer gelten, in denen bislang keine Lehrveranstaltungsevaluation durchgeführt wurde. Die Ergebnisse des

³ Dies impliziert die Veröffentlichung der Evaluationsergebnisse auf Ebene der einzelnen Lehrveranstaltungen bzw. Dozent(inn)en. Siehe dazu Abschnitt 15.

Vergleichs zwischen Online- und papiergebundener Evaluation weisen zudem darauf hin, dass für viele Studierende die Sinnhaftigkeit der Lehrveranstaltungsevaluation nicht erkennbar ist (vgl. Abschnitt 13).

Flankierend zur Umsetzung des in Abschnitt 13 beschriebenen Evaluationsworkflow (der mit der zeitnahen Ergebnismeldung an Lehrende und Studierende die Nützlichkeit der Evaluation betont) sollten Maßnahmen zur Steigerung der Akzeptanz der Lehrveranstaltungsevaluation ergriffen werden. Dies kann beispielsweise in der gezielten Informierung der Lehrenden über Ziele und Nutzen der Evaluation sowie dem Ausräumen von Ängsten bestehen. Zur Akzeptanzsteigerung bei den Studierenden sind von den Fachschaften organisierte Vollversammlungen der Fächer oder eine Informierung im Rahmen der Erstsemestereinführungen denkbar.

Empfehlung 3: Akzeptanzsteigerung der Lehrveranstaltungsevaluation

Es sollten Maßnahmen zur Förderung der Akzeptanz der Lehrveranstaltungsevaluation bei Lehrenden und Studierenden ergriffen werden.

7 Modellvorstellungen zur Lehrqualität

7.1 Definition von Lehrqualität

Lehrqualität lässt sich zuvorderst an den Produkten der Lehrbemühungen messen. Anzuführen ist zunächst das wichtigste anzustrebende Lehrprodukt, nämlich der Lernzuwachs seitens der Studierenden (vgl. Spiel, 2004). Lehre ist dann erfolgreich, wenn sie erreicht, dass viele Studierende viel lernen und ein grundlegendes Verständnis des Gegenstands erlangen. Sind die Lernziele von Lehrveranstaltungen explizit definiert (wie etwa in den Modulkatalogen konsekutiver Studiengänge), so kann der Lernzuwachs spezifisch im Hinblick darauf betrachtet werden. Neben dem Lernzuwachs ist auch ein gesteigertes Interesse am Gegenstand ein wichtiges Lehrprodukt und damit ein Indikator für qualitativ hochwertige Lehre. Schließlich ist es ein – wenngleich geringer zu gewichtendes – Ergebnis guter Lehre, wenn Studierende die Veranstaltung insgesamt positiv bewerten. Alle drei Aspekte – Lernzuwachs, Interessenförderung und positive Gesamtbeurteilung – spiegeln damit die *Lehrqualität auf Produktebene* wider, die im Folgenden als *Lehrerfolg* bezeichnet wird (vgl. Rindermann, 2001a).

Anhand des so definierten Lehrerfolgs sollte die Gesamtqualität der Lehre der Universität und der Fächer bestimmt und Lehrveranstaltungen identifiziert werden, die nicht den Mindestanforderungen an die Lehrqualität genügen.

7.2 Bedingungen hoher Lehrqualität seitens der Lehrenden

Um das Evaluationsziel zu erreichen, nützliches Feedback zur Verbesserung der Lehrqualität zur Verfügung zu stellen, ist neben der Berücksichtigung von Lehrprodukten auch eine Betrachtung der Lehrqualität auf Prozessebene erforderlich. Diese bezieht sich auf das konkrete didaktische Handeln der Lehrenden, das zu Lehrerfolg führt. Darunter fallen beispielsweise die Strukturierung des Gegenstands, die Klarheit der Stoffvermittlung, die Interessanztheit der Darstellung aber auch das durch die Lehrenden hergestellte soziale Klima. Diese Prozesselemente qualitativ hochwertiger Lehre sind die im *didaktischen Handeln der Dozent(inn)en* verorteten Bedingungen von qualitativ hochwertigen Lehrprodukten. Im Hinblick auf die Nützlichkeit des Feedbacks zur Verbesserung der Lehrprodukte sollte die Lehrveranstaltungsevaluation jene spezifischen Elemente des didaktischen Handelns erfassen, die (a) hinreichend eng mit dem Lehrerfolg im Zusammenhang stehen und (b) durch die Dozent(inn)en hinreichend beeinflussbar sind. Für die Veranstaltungsform „Vorlesung“ verweist die pädagogisch-psychologische Forschung darauf, dass wichtige relevante und kontrollierbare Faktoren guter Lehre die oben genannten sowie das Engagement der Dozent(inn)en, die durch sie gestellten Anforderungen, die Qualität der zur Wissensvermittlung eingesetzten Präsentationsmedien sowie die Organisation der Lehrveranstaltung sind (Überblicke bei Rindermann, 2001a; Spiel, 2001; siehe auch el Hage, 1999).

7.3 Studentische Bedingungen und Rahmenbedingungen

Der Lehrerfolg wird nicht nur durch das didaktische Handeln der Lehrenden bedingt, sondern hängt – trivialerweise – auch von den Studierenden ab. So ist unmittelbar evident, dass der Lernzuwachs (als wichtigstes Lehrprodukt) auch von der Anstrengung und der Motivation der Studierenden abhängt. Daneben unterscheiden sich Lehrveranstaltungen im Hinblick auf Rahmenbedingungen, die ebenfalls im Zusammenhang mit dem Lehrerfolg stehen können (z.B. Anzahl der Teilnehmer(innen), Veranstaltungsform). Studentische Bedingungen und Rahmenbedingungen des Lehrerfolgs sollten deshalb in der Lehrveranstaltungsevaluation berücksichtigt werden.

Dies ist auch insofern wichtig, als studentische Beurteilungen der Lehrqualität durch eine Reihe von Biasvariablen verzerrt sein können (z. B. Dresel & Rindermann, 2006; Greenwald, 1997; Spiel & Gössler, 2000). Diese Variablen repräsentieren Merkmale, die durch didaktisches Handeln der Dozent(inn)en nicht oder kaum beeinflusst werden können, aber im Zusammenhang mit der Lehrqualität oder der Lehrqualitätsbeurteilung stehen (Rindermann, 2001a). Wichtige Biasvariablen sind die genannten studentischen Variablen, insbesondere Vorwissen, Besuchsgrund (als zusammenfassender Indikator für die Motivation), Anstrengung und Fachsemester, sowie Rahmenbedingungen der Lehrveranstaltung, wie Anzahl der Teilnehmer(innen) und Veranstaltungsform.

Diese potenziellen Verzerrungen verweisen darauf, dass die Urteile, die Studierende im Rahmen der Lehrveranstaltungsevaluation abgeben, zu einem gewissen Grad subjektiver Natur sind. Der aus dieser Subjektivität resultierende Messfehler ist der methodische Grund dafür, warum sie nicht zu vergleichenden Bewertungen auf der Ebene einzelner Lehrveranstaltungen oder Dozent(inn)en herangezogen werden sollten, die sehr hohe Anforderungen an Objektivität und Messgenauigkeit stellen. Zur Bestimmung der Gesamtqualität von Fächern und Universität sollten die studentischen Urteile mit geeigneten statistischen Verfahren (z.B. mehrebenenanalytische Verfahren) für die potenziellen Verzerrungen korrigiert werden (zur Methodik siehe Dresel & Rindermann, 2006).

Ganz allgemein rechtfertigt erst die Berücksichtigung der wesentlichen Verzerrungsvariablen einen Vergleich zwischen Lehrveranstaltungen, unabhängig davon, ob gleiche Lehrveranstaltungen unterschiedlicher Semester, ob unterschiedliche Lehrveranstaltungen gleicher Semester oder ob unterschiedliche Lehrveranstaltungen unterschiedlicher Semester verglichen werden (vgl. Rindermann, 2001b; Dresel & Rindermann, 2006). Dabei muss bedacht werden, dass in der Praxis wohl niemals alle verzerrenden Einflüsse berücksichtigt werden können, weswegen die Aussagekraft von Vergleichen zwischen Lehrveranstaltungen oder gar Dozent(inn)en stets eingeschränkt ist.

Die Subjektivität studentischer Urteile wurde oftmals als Argument gegen ihre Verwendung zur Evaluation von Lehrveranstaltungen ins Feld geführt (siehe Kromrey, 1995). Dem kann entgegnet werden, dass sie sehr wohl dazu geeignet sind, den Lehrerfolg hinreichend inhaltstreu zu messen (hinreichende Validität), relative Stärken und Schwächen von Dozent(inn)en hinreichend messgenau abzubilden und Veranstaltungen mit eindeutig unzureichender Lehrqualität zu identifizieren (zsf. Marsh & Roche, 1997). Zudem kann angeführt werden, dass Studierende die Zielgruppe von Lehrveranstaltungen sind und deshalb per se bei der Beurteilung der Qualität von Lehrveranstaltungen einbezogen werden sollten.

7.4 Modell der Lehrqualität

Die genannten theoretischen und durch empirische Befunde gestützten Ausführungen zur Lehrqualität können in einem Modell der Lehrqualität zusammen gefasst werden, das in Abbildung 1 dargestellt ist (vgl. Rindermann, 2001a). Zu beachten ist dabei, dass die Definition der Lehrqualität auf Produktebene (Lehrerfolg; rechte Spalte) und die potenziellen Verzerrungsvariablen (studentische Bedingungen und Rahmenbedingungen; mittlere Spalte) unabhängig von der Veranstaltungsform sind, die Lehrqualität auf Prozessebene (linke Spalte) aber je nach Veranstaltungsform unterschiedliche Aspekte des didaktischen Handelns umfassen wird. Bislang ausgearbeitet ist lediglich das hier dargestellte Modell für die Lehrqualität von Vorlesungen.

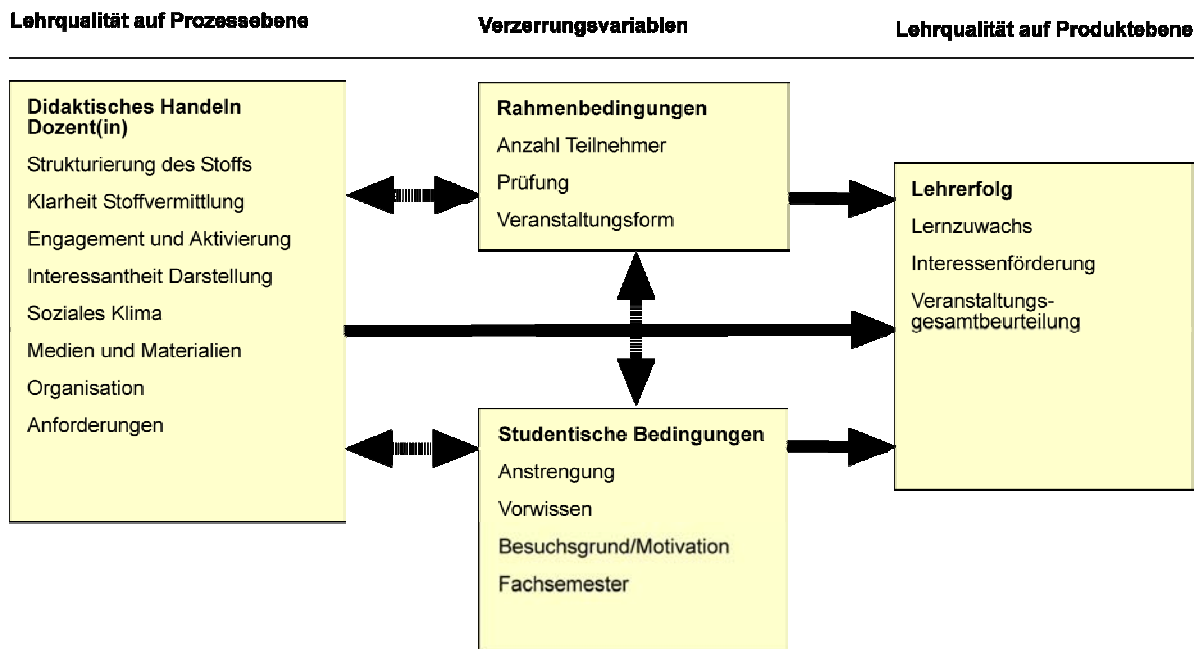


Abbildung 1. Modell der Lehrqualität (modifiziert nach Rindermann, 2001). Die Lehrqualität auf Prozessebene ist bislang auf Vorlesungen beschränkt.

Anzumerken ist, dass sich studentische Bedingungsvariablen, Rahmenbedingungen und didaktisches Handeln der Dozent(inn)en gegenseitig beeinflussen können (z.B. kann der/die Lehrende die Motivation der Studierenden positiv beeinflussen und motivierte Studierende können den/die Lehrende(n) zu einem größeren Engagement veranlassen).

Die theoretischen Überlegungen dieses Abschnitts führen zu folgender Empfehlung:

Empfehlung 4: Zu berücksichtigende Aspekte der Lehrqualität

Der Lehrveranstaltungsevaluation an der Universität Ulm sollten theoretische Überlegungen zur Lehrqualität zugrunde gelegt werden. Dabei sollte unterschieden werden zwischen dem Lehrerfolg (Produkte guter Lehre: Lernzuwachs, Interessenförderung, Gesamtbeurteilung) als Kriterium der Lehrqualitätsbeurteilung, dem didaktischen Handeln der Dozent(inn)en (Prozessmerkmale der Lehre) als Ansatzstelle zur Verbesserung der Lehrqualität sowie Hintergrundvariablen, durch die studentische Veranstaltungsbeurteilungen verzerrt sein können. In der Lehrveranstaltungsevaluation sollten wichtige Merkmale aller drei Merkmalsgruppen erfasst werden.

Das bislang auf Vorlesungen spezifizierte theoretische Modell der Lehrqualität sollte zukünftig im Hinblick auf andere Veranstaltungsformen und den Einsatz von E-Learning in der Lehre erweitert und überprüft werden. Zudem empfiehlt sich, nach der Definition der Modulkataloge konsekutiver Studiengänge zu prüfen, inwieweit der Lehrerfolg ergänzend auch über die darin enthaltenen Lernziele definiert werden können.

8 Instrumentarium

Aufbauend auf den Zielsetzungen der Lehrveranstaltungsevaluation an der Universität Ulm sowie den Modellvorstellungen zur Lehrqualität galt es, dafür geeignete Fragebogenverfahren zu entwickeln.

8.1 Kriterien für Fragebögen zur Evaluation von Lehrveranstaltungen

Zunächst wurden allgemeine Kriterien für Fragebögen formuliert, die im Rahmen der Lehrveranstaltungsevaluation an der Universität Ulm eingesetzt werden sollen. Diese Kriterien sollen unabhängig von der spezifischen Veranstaltungsform (z.B. Vorlesung, Übung, Praktikum, Seminar) und dem spezifischen Fach auch bei der zukünftigen Entwicklung von Fragebögen Anwendung finden.

Die Kriterien für Fragebögen zur Evaluation von Lehrveranstaltungen lauten:

1. Die Fragebögen sollen die Lehrqualität erfassen, und zwar im Hinblick auf die *Produkte guter Lehre* (Lehrerfolg) und die *Prozesselemente gute Lehre* (didaktisches Handeln der Lehrenden). Daneben sollen externe Bedingungen der Lehrqualität erfasst werden (Rahmenbedingungen, studentische Bedingungsfaktoren).
2. Die Fragebögen sollen *relevante Aspekte* der Lehrqualität erfassen und irrelevante Aspekte ausblenden.
3. Mit den Fragebögen sollen Ergebnisse gewonnen werden, die Dozent(inn)en in hohem Maße zweckdienlich dazu sind, ihre Lehre zu verbessern. Um dies zu erreichen, muss die Lehrqualität auf der Ebene *kontrollierbarer und spezifischer Lehrhandlungen* abgebildet werden (z.B. „Der Stoff wird sehr gut mit Beispielen oder Experimenten veranschaulicht“ statt „Der Dozent hat Temperament“).
4. Mit den Fragebögen sollen Ergebnisse gewonnen werden, die für alle Beteiligten einfach interpretierbar dargestellt werden können. Dazu müssen die enthaltenen Fragen in hohem Maße *psychometrischen Anforderungen* genügen, um intuitiv verständliche zusammenfassende Statistiken berechnen zu können.
5. Die Fragebögen sollen geeignet sein, *individuelle Stärken und Schwächen* in Bezug auf verschiedene Aspekte der Lehrqualität abzubilden. Dazu müssen die enthaltenen Fragen hinreichend differenzieren.
6. Die Fragebögen sollen geeignet sein, *individuelle Veränderungen* über die Semester abzubilden.

8.2 Fragebogen zur Evaluation von Vorlesungen

Schwerpunktmäßig hat die Arbeitsgruppe einen Fragebogen zur Evaluation von Vorlesungen entwickelt und empirisch überprüft. Grundlage war der in einzelnen Fächern eingesetzte Fragebogen, der im Sommersemester 2004 einer Testevaluation unterzogen wurde, sowie die darauf basierende Datenanalyse zur Eignung des Fragebogens (Dresel, 2005b). Dabei stellte sich heraus, dass die Fragen dieses Bogens in vielen Fällen psychometrischen Standards nicht genügten, eine Reihe an Verständnis- und Beurteilungsproblemen aufwiesen und theoretisch von größerer Relevanz für die Lehrqualität sein sollten.

Der Neukonzeption des Fragebogens zur Evaluation von Vorlesungen wurde das in Abbildung 1 dargestellte Modell zugrunde gelegt.

In Tabelle 1 sind Erläuterungen zu den einzelnen Aspekten sowie die konkreten Fragen aufgeführt, die zur Messung der Lehrqualität in den Fragebogen aufgenommen wurden. Diese Aufstellung spiegelt das endgültige Ergebnis mehrerer Testevaluationen wider und enthält neben obligatorisch einzusetzenden Fragen auch einige weniger relevante Fragen, die optional verwendet werden können.

Für besonders wichtige Faktoren wurden mehreren Fragen integriert. Dies ist erforderlich, um die notwendige Breite herzustellen, die eine differenzierte und damit nützliche Rückmeldung an die Dozent(inn)en erfordert, und um diese Dimensionen mit einer möglichst hohen Messgenauigkeit zu erfassen (kommen mehrere Items zum Einsatz, „mitteln“ sich Fehlereinflüsse „heraus“). Die Erfassung mit mehreren Aspekten hat zudem den Vorteil, dass inhaltsgültige zusammenfassende Maße berechnet werden können, die in der Ergebnismeldung einen besonderen Stellenwert erhalten, da sie besonders aussagekräftig und leicht interpretierbar sind.

Rahmenbedingungen der Lehrveranstaltung wurden nicht in den Studierendenfragebogen aufgenommen, da sie ökonomischer mit einem „Veranstaltungserfassungsblatt“ erfasst werden können, das von den Dozent(inn)en ausgefüllt wird.

Table 1. Aufstellung der im Fragebogen zur Evaluation von Vorlesungen verwendeten Faktoren und Einzelfragen

Faktor ^a	Erläuterung	Fragen ^b
Lehrerfolg (Lehrqualität auf Produktebene)		
Lernzuwachs	Selbsteinschätzung der Studierenden ihres quantitativen und qualitativen Lernzuwachses	„Ich lerne viel in der Veranstaltung.“ „Ich gewinne ein tiefes Verständnis des Stoffs in der Veranstaltung.“
Interessenförderung	Ausmaß, in dem bei Studierenden das Interesse am Gegenstand gefördert wird	„Die Veranstaltung fördert mein Interesse am Stoffgebiet.“
Gesamtbeurteilung	Allgemeine, unspezifische Gesamtbewertung der Lehrveranstaltung	„Der Besuch der Veranstaltung lohnt sich sehr.“ „Alles in allem ist die Veranstaltung sehr gut.“
Didaktisches Handeln (Lehrqualität auf Prozessebene)		
Strukturierung des Stoffs	Ausmaß, in dem der/die Dozent(in), den Stoff lernförderlich strukturiert, also Wichtiges hervorhebt, den Stoff vernetzt und Querbezüge herstellt sowie den Stoff nachvollziehbar gliedert.	„Der Dozent gibt häufig hilfreiche Zusammenfassungen und Stoffübersichten.“ „Der Dozent stellt häufig Querbezüge her.“ „Die Gliederung des Stoffes (roter Faden) ist stets nachvollziehbar.“
Klarheit der Stoffvermittlung	Ausmaß, in dem der/die Dozent(in), den Stoff anhand von Beispielen, Experimenten und Praxisbezügen konkretisiert, schwierige Sachverhalte verständlich macht und damit zur Klarheit bei den Studierenden beiträgt.	„Es werden viele Bezüge zur Praxis und konkrete Erfahrungen vermittelt.“ „Der Stoff wird sehr gut mit Beispielen oder Experimenten veranschaulicht.“ „Schwierige Sachverhalte werden verständlich erklärt.“
Engagement und Aktivierung	Ausmaß, in dem der/die Dozent(in) signalisiert, dass er/sie den Lernerfolg der Studierenden wichtig nimmt und diese zu einer aktiven Teilnahme motiviert.	„Man merkt, dass dem Dozenten der Lernerfolg der Studierenden wichtig ist.“ „Der Dozent motiviert zu einer aktiven Teilnahme.“
Interessantheit der Darstellung	Ausmaß, in dem die Lehrveranstaltung als interessant und abwechslungsreich erlebt wird.	„Der Dozent hält die Veranstaltung in interessanter Form.“
Soziales Klima	Ausmaß, in dem der/die Dozent(in) im Umgang mit den Studierenden eine freundliche Atmosphäre herstellt.	„Der Dozent ist sehr freundlich im Umgang mit den Studierenden.“
Medien/Materialien	Beurteilung der eingesetzten Präsentationsmaterialien und –medien (z.B. Folien, Tafelbild) in Bezug darauf, ob sie zum Verständnis beitragen.	„Die eingesetzten Medien und Materialien sind eine große Hilfe beim Verstehen des Inhalts.“

Faktor^a	Erläuterung	Fragen^b
Medien/Materialien (ergänzende Aspekte)	Beurteilung der Materialien und –medien darauf hin, ob sie gut strukturiert und lesbar sind. Außerdem wird eine Beurteilung etwaiger Begleitmaterialien/Begleitmedien (z.B. Skript, Internetseiten) erfasst.	„Struktur und Lesbarkeit von Tafelbild, Folien bzw. Präsentation sind sehr gut.“ „Der Support zur Veranstaltung ist vorbildlich (z.B. Download, begleitende Internetseiten).“ „Das Skript zur Veranstaltung ist vorbildlich.“ (Vorangestellte Filterfrage: „Gibt es ein Skript zur Veranstaltung?“)
Anforderungen	Beurteilung der Höhe der Anforderungen, explizit im Hinblick auf zu hohe Anforderungen bzw. zu hohes Tempo.	„Das Tempo der Stoffvermittlung empfinde ich als zu schnell.“
Organisation	Allgemeinbeurteilung der Organisation.	„Die Veranstaltung ist sehr gut organisiert.“
Klarstellung von Anforderungen	Klarstellung von Schein- und Prüfungsanforderungen.	„Schein- und Prüfungsanforderungen wurden klargestellt.“
Studentische Bedingungsvariablen		
Vorwissen	Erfasst, ob die Studierenden über ausreichend Vorwissen zum Verständnis des Stoffs verfügen.	„Ich verfüge über ausreichend Vorwissen, um die Inhalte gut verstehen zu können.“
Anstrengung	Erfasst die wöchentliche Nachbereitungszeit in Stunden.	„Wie viel Nachbearbeitungszeit bringen Sie für diese Veranstaltung auf?“ („keine Nachbereitung“, „bis zu 1 Std./Woche“, ..., „bis zu 7 Std./Woche“, „mehr“)
Motivation	Erfasst als globalen Indikator der Motivation den Grund des Veranstaltungsbesuchs.	„Besuchsgrund (mehrere Angaben möglich): „Pflicht“, „Interesse“, „Dozent“, „Prüfung“, „Relevantes Thema“, „Scheinklausur“
Fachsemester		„In welchem Fachsemester studieren Sie?“ (1, ..., 11, >11)
Studiengang		„Im Zusammenhang welchen Studiengangs besuchen Sie diese Lehrveranstaltung?“ (bis zu fünf fachspezifische Antwortoptionen)

Anmerkungen: ^a Obligatorische Faktoren sind fett gedruckt. ^b Sofern keine Antwortoptionen angegeben sind wird der Grad der Zustimmung mit einer sechsstufigen Antwortskala mit den Polen „trifft gar nicht zu“ (1) und „trifft völlig zu“ (6) sowie der zusätzlichen Antwortoption „weiß nicht“ erfasst.

Neben geschlossenen Fragen wurden zur Erfassung spezifischer und beeinflussbarer Aspekte der Lehrqualität offene Fragen formuliert, die erfahrungsgemäß besonders konkrete Hinweise zur Verbesserung der Lehre liefern und Dozent(inn)en in besonderer Weise zu diesen Verbesserungen motivieren können. Die Arbeitsgruppe misst der Integration von offenen Fragen aufgrund ihres hohen Nutzwerts bei der Verbesserung der Lehrqualität besonderes Gewicht bei. Ihr Einsatz sollte nicht in Frage gestellt werden (z.B. im Rahmen von Effizienzüberlegungen oder datenschutzrechtlichen Bedenken).

Die zwei obligatorischen offenen Fragen thematisieren besonders positiv und besonders negativ hervorstechende Aspekte der Lehrveranstaltung, einschließlich der Aufforderung ggf. Verbesserungsvorschläge zu nennen. Optional können zwei weitere offene Fragen in den Fragebogen aufgenommen werden, mit denen die Stoffvermittlung als zentrales Anliegen fokussiert wird. Die offenen Fragen lauten:

- **Obligatorische Fragen:**
 - „Was ist besonders gut an der Veranstaltung?“
 - „Was ist besonders schlecht an der Veranstaltung (ggf. Verbesserungsvorschläge nennen)?“
- **Optionale Fragen:**
 - „Welche Inhalte wurden besonders verständlich vermittelt?“
 - „Welche Inhalte blieben bisher unverstanden (ggf. Verbesserungsvorschläge nennen)?“

8.3 Empirische Überprüfung des neu konzipierten Fragebogens

Die Eignung von Fragebogenverfahren, die zugrunde liegenden Aspekte inhaltsgetreu und messgenau zu messen, muss stets empirisch geprüft werden. Deshalb wurde der neu konzipierte Fragebogen zur Evaluation von Vorlesungen im Sommersemester 2005 in zehn Vorlesungen verschiedener Fakultäten einer empirischen Prüfung unterzogen. Eine ausführliche Darstellung der Ergebnisse findet sich im schriftlichen Zwischenbericht der Arbeitsgruppe (Dresel, 2005a). Hier soll lediglich eine kurze Zusammenfassung der Befunde dargestellt werden:

- Die überwiegende Mehrheit der Fragen des neu konzipierten Fragebogens ist ohne Verständnis- oder Beurteilungsschwierigkeiten zu beantworten, differenziert hinreichend zwischen Lehrveranstaltungen und verfügt über sehr gute Verteilungseigenschaften. Sie sind deshalb sehr gut zur Erstellung verständlicher und leicht interpretierbarer Rückmeldungen geeignet. Die verbleibenden wenig gewichtigen Probleme bei drei Fragen wurden in der nun vorliegenden Fragebogenversion behoben (vgl. Tabelle 1)
- Alle im Fragebogen enthaltenen Items erfassen Aspekte, die relevant für die Beurteilung des Lehrerfolgs sind und damit Anhaltspunkte für die Verbesserung der Lehre liefern können. Allerdings variieren die Items durchaus in ihrer Relevanz. Beispielsweise ist die Klarstellung von Schein- und Prüfungsanforderungen von eher geringer Rele-

vanz, ebenso wie der Support zur Veranstaltung. Fragen, die in geringem Zusammenhang mit dem Lehrerfolg stehen, wurden aus dem obligatorischen Fragebogenteil entfernt, können aber fachspezifisch eingesetzt werden.

- Die Bildung von übergreifenden Skalenwerten für jene Faktoren aus Tabelle 1, für die mehrere Fragen eingesetzt wurden, ist möglich und sinnvoll. Die resultierenden Skalen weisen gute bis sehr gute Verteilungseigenschaften, eine hinreichende Messgenauigkeit (Reliabilität) und eine sehr zufrieden stellende Relevanz (Validität) auf.

Insgesamt kann als Ergebnis der Fragebogenüberprüfung festgehalten werden, dass alle Fragen und Skalen gute psychometrische Eigenschaften aufweisen und von Relevanz für die Beurteilung der Lehrqualität sind. Insgesamt liegt ein inhaltsgültiges (valides) und messgenaues (reliables) Messinstrument zur Erfassung von studentischen Vorlesungsbeurteilungen vor. Die wenigen noch vorhandenen Schwierigkeiten wurden in dem nun vorliegenden Fragebogen ausgeräumt.

Die theoretische Fundierung und empirische Bewährung des Fragebogens zur Evaluation von Vorlesungen rechtfertigt die folgende Empfehlung:

Empfehlung 5: Fragebogen zur Evaluation von Vorlesungen

In allen Fächern sollte zur Beurteilung der Lehrqualität von Vorlesungen der von der Arbeitsgruppe entwickelte und psychometrisch getestete „Fragebogen zur Evaluation von Vorlesungen“ eingesetzt werden, der die drei Module „Lehrqualität“, „Didaktisches Handeln“ und „Studentische Hintergrundvariablen“ enthält.

Der obligatorische Teil des Fragebogens (22 geschlossene und 2 offene Fragen) kann dabei fachspezifisch um optionale Fragen ergänzt werden.

8.4 Zukünftige Entwicklung von Fragebögen

Der neu konzipierte Fragebogen ist zwar sehr gut zur Evaluation von Vorlesungen geeignet ist, aber für andere Veranstaltungsformen, wie z.B. Übungen, Praktika oder Seminare, nur eingeschränkt verwendbar. Zwar sind die Fragen zur Lehrqualität auf Produktebene (Lehrerfolg) und den studentischen Bedingungen uneingeschränkt auf alle Veranstaltungsformen anwendbar. Allerdings zielen die enthaltenen Fragen zum didaktischen Handeln der Dozent(inn)en ausschließlich auf Vorlesungen ab.

Zur Qualität und Lerneffektivität anderer, insbesondere aktiverer Lernformen (z.B. studentisches Referieren, Experimentieren, Diskutieren, eigenständiges Erarbeiten, Nachvollziehen und Entwickeln, Gruppenarbeiten) und deren Anleitung sowie Betreuung durch Dozent(inn)en (z.B. Hilfestellungen, Feedback, Vorbesprechung, Abstimmung von Übungen auf die Vorlesung, Lehrqualität von Übungsleiter(inn)en und Tutor(inn)en) liegen derzeit

keine theoretisch fundierten und empirisch überprüften Fragen vor.⁴ Ebenfalls nicht abgedeckt sind derzeit Spezifika auf der Ebene von Fächern und Studiengängen.

Die Arbeitsgruppe führte deshalb eine Bedarfsanalyse im Hinblick auf weitere erforderliche Fragebögen bzw. Fragebogenmodule durch. Dabei zeigte sich, dass zukünftig vor allem für die Veranstaltungsformen „Übung/Tutorium“, „Praktikum“ und „Seminar“ Fragebögen entwickelt werden sollten. Weiterer Bedarf besteht für die Entwicklung eines Fragebogenmoduls zum Einsatz von E-Learning in der Lehre. Dabei ist den Charakteristika der unterschiedlichen Fächer durch die Entwicklung spezifischer Fragebögen Rechnung zu tragen.

Für die Evaluation der Lehrveranstaltungen in den an der Universität Ulm angebotenen internationalen Studiengängen sind zudem englische Versionen der relevanten Fragebögen erforderlich. Zur Herstellung der notwendigen Vergleichbarkeit mit den deutschen Sprachversionen sind bei der Übersetzung besondere Vorkehrungen zu treffen (doppelte Übersetzung, Rückübersetzung zur Kontrolle).

Generell empfiehlt sich bei der zukünftigen Fragebogenentwicklung die Anwendung der modularen Fragebogenstruktur, die in Abbildung 2 dargestellt ist. Darin sind die beiden veranstaltungsform- und fachunabhängigen Module „Lehrerfolg“ und „Studentische Hintergrundvariablen“ zu erkennen, die stets in der identischen Form eingesetzt werden sollten. Für die Veranstaltungsform „Vorlesung“ sind zwei Module zum didaktischen Handeln vorhanden, eines mit 13 obligatorischen und eines mit 7 optionalen Fragen. Übungen bzw. Tutorien sind meist vorlesungsbegleitend und keine eigenständigen Lehrveranstaltungen. Deshalb sollten die entsprechenden Fragen in einem optionalen Modul „Didaktisches Handeln/Zusatzaspekte“ zusammengefasst werden. Auch für andere Veranstaltungsformen empfiehlt sich eine Unterscheidung in ein zentrales Modul „Didaktisches Handeln“ und ein ergänzendes, optional einzusetzendes Modul mit Zusatzaspekten.

Weiterhin zeigte sich, dass auch zukünftig laufender Bedarf für die Entwicklung spezifischer Fragebögen für spezifische Evaluationsinteressen besteht. Jüngstes Beispiel ist die Evaluation des Ulmer Trainingscamps, denkbar ist aber auch die Evaluation der Einführung neuer Studiengänge oder die Evaluation spezifischer Innovationen in einzelnen Fächern oder Lehrveranstaltungen. Deshalb sollte dauerhaft die Möglichkeit bestehen, spezifische Evaluationsinteressen zu berücksichtigen und die dafür erforderlichen Instrumentarien dem in Abschnitt 14 ausgeführten Evaluationsworkflow zuzuführen.

⁴ Im Rahmen der Testevaluation im Wintersemester 2005/06 wurde der von der Fachschaft Mathematik genutzte und 16 Fragen umfassende Ergänzungsfragebogen für vorlesungsbegleitende Übungen/Tutoren in 32 Lehrveranstaltungen eingesetzt und überprüft (1.563 ausgewertete Fragebögen). Dabei zeigte sich, dass acht der Fragen psychometrische Probleme aufweisen und entweder getilgt oder modifiziert werden sollten. Die übrigen Fragen stellen dennoch eine gute Basis für die Entwicklung eines validen Ergänzungsfragebogens für Übungen/Tutorien dar, wobei vor allem die theoretische Fundierung noch verbesserungsbedürftig ist.

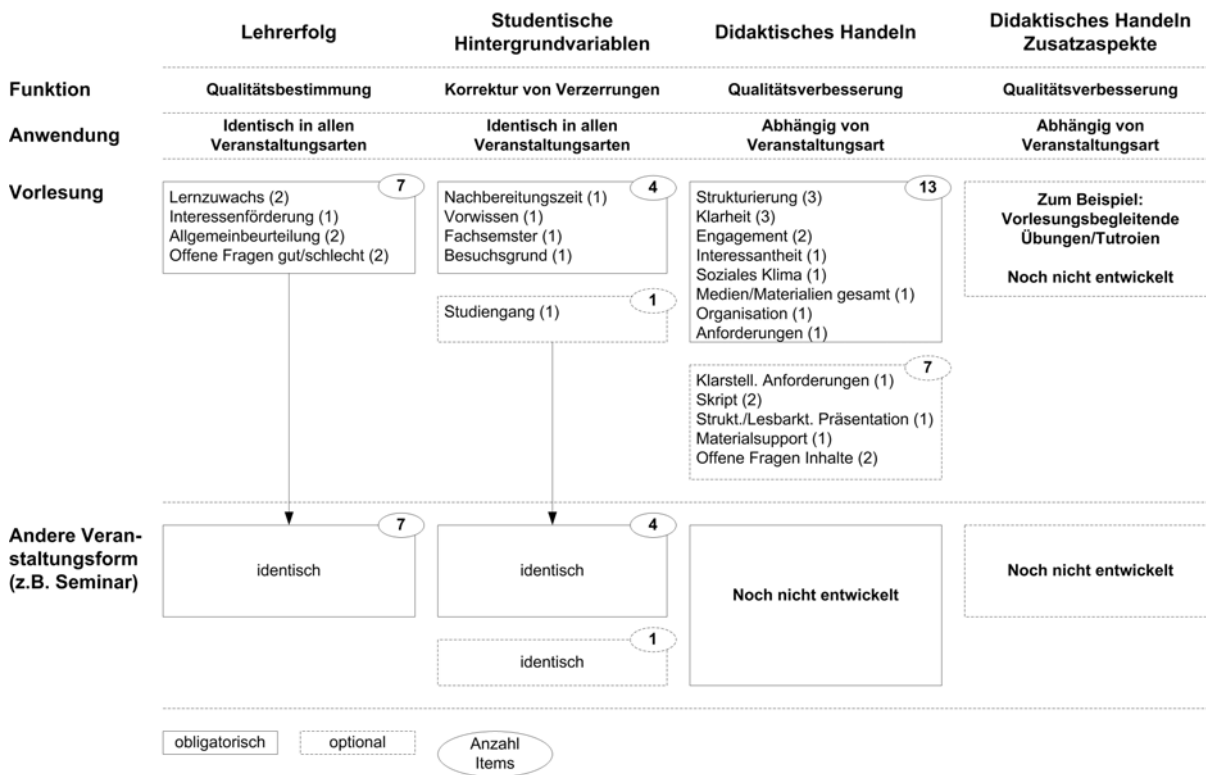


Abbildung 2. Modularer Aufbau von Fragebögen zur Evaluation von Lehrveranstaltungen

Die in der Arbeitsgruppe durchgeführte Bedarfsanalyse lässt sich in der folgenden Empfehlung zusammenfassen:

Empfehlung 6: Zukünftige Entwicklung von Fragebögen

Es empfiehlt sich die zukünftige Entwicklung von weiteren Fragebögen(-modulen). Diese sollen den formulierten allgemeinen Kriterien für Fragebögen zur Evaluation von Lehrveranstaltungen genügen, vor ihrem Einsatz empirisch überprüft und modular gegliedert werden.

Besonderer Bedarf besteht im Hinblick auf weitere Veranstaltungsformen und den Einsatz von E-Learning in der Lehre. Dabei ist den Charakteristika der Fächer durch die Entwicklung spezifischer Fragebögen Rechnung zu tragen. Daneben sind englische Versionen der Fragebögen erforderlich.

Schließlich besteht laufender Bedarf für die Entwicklung von Fragebögen für spezifische Evaluationsinteressen (z.B. Einführung, Neu- oder Restrukturierung bestehender Studiengänge).

Wie die Vielzahl der Aspekte und Faktoren zeigt, die bei der Entwicklung von Fragebogenverfahren zu berücksichtigen sind (z.B. Bortz & Döring, 2002; Krauth, 1995), stellt dies keineswegs eine triviale Aufgabe dar. Zudem sollten Fragebögen stets theoretisch fundiert und empirisch geprüft werden. Fachkompetenz ist hier somit dringend erforderlich.

Empfehlung 7: Erforderliche Fachkompetenz zur Entwicklung von Fragebögen

Den Fächern der Universität Ulm sollte die erforderliche Fachkompetenz zur theoretisch fundierten Entwicklung und psychometrischen Überprüfung von Fragebögen zur Verfügung stehen.

9 Maßstäbe zur Bewertung der Lehrqualität

Die Daten aus studentischen Veranstaltungsbeurteilungen müssen prinzipiell an einen Bewertungsmaßstab relativiert werden, um daraus Aussagen zur Lehrqualität ableiten zu können. Ohne sie mit einem Bewertungsmaßstab zu vergleichen, können die reinen Ergebniswerte zu den einzelnen, in Evaluationsfragebögen enthaltenen Fragen nur schwer interpretiert werden, da sie nicht nur von der tatsächlichen Lehrqualität, sondern auch von Spezifika der Fragenformulierung abhängen.

Prinzipiell sind in der Lehrveranstaltungsevaluation die folgenden Bewertungsmaßstäbe denkbar:

1. Durchschnittsergebnisse des Fachs oder der Universität
2. Sachlogisch definierte Qualitätskriterien (festgelegte Mindestqualität in Bezug auf die einzelnen Aspekte der Lehrqualität)
3. Ergebnisse anderer Teilaspekte der Lehrqualität in derselben Lehrveranstaltung
4. Ergebnisse derselben Lehrveranstaltung bzw. desselben Lehrenden in früheren Semestern

Im Folgenden werden die einzelnen Bewertungsmaßstäbe und ihre Implikationen etwas näher ausgeführt. Dabei wird sich zeigen, dass der erste Bewertungsmaßstab überwiegend negative Folgen hat und nicht eingesetzt werden sollte, die drei übrigen Maßstäbe (allen voran sachlogisch definierte Qualitätskriterien) dagegen gut für Bewertung der Lehrqualität geeignet sind.

9.1 Bewertung anhand von Durchschnittsergebnissen

Lehrqualität wird bei Anwendung dieses Maßstabs „sozial“ definiert, da die Ergebnisse anderen Dozent(inn)en die Vergleichsbasis darstellen. Dieser Bewertungsmaßstab liegt der vergleichenden Bewertung auf Ebene einzelnen Lehrveranstaltungen oder Dozent(inn)en zugrunde. Es resultieren Qualitätsbewertungen wie „überdurchschnittliche Lehrqualität“, „durchschnittliche Lehrqualität“ oder „unterdurchschnittliche Lehrqualität“.

Bei sozialen Vergleichsmaßstäben überwiegen die Nachteile (vgl. Cashin, 1996; Marsh & Roche, 1997; McKeachie, 1996): Sie resultieren stets in demselben Anteil an Lehrveranstaltungen mit hoher und niedriger Lehrqualität und spiegeln Verbesserungen der Gesamtqualität der Fächer oder der Universität nicht wider sondern verschleiern sie vielmehr (die „Rangreihe“ der Dozent(inn)en bleibt unverändert, obwohl sich alle verbessern). Es resultiert typischerweise eine zu hohe Stabilität der Bewertungen (Rheinberg, 1998), d.h. gut bewertete Dozent(inn)en werden dauerhaft gut bewertet, schlecht bewertete Dozent(inn)en werden dauerhaft schlecht bewertet. Eine breite Befundlage deutet zudem darauf hin, dass bei Anwendung sozialer Vergleichsmaßstäbe ungünstige Zielsetzungen seitens der Handelnden resultieren: Da der Blick mit sozialen Vergleichsmaßstäben auf unveränderlich erscheinende Personmerkmale und nicht auf spezifische und kontrollierbare Lehrhandlungen gelenkt wird, geht es vorrangig darum, gut zu *wirken* und nicht darum, gut zu *werden* (zsf. Pintrich & Schunk, 2002). In der Folge dieser Aspekte werden Dozent(inn)en bei der Bewertung des Lehrerfolgs anhand von Durchschnittsergebnissen, wie es bei der vergleichenden Bewertung auf Ebene der einzelnen Dozent(inn)en bzw. Lehrveranstaltungen der Fall ist, häufig eher demotiviert als zu verbessertem Lehrhandeln aktiviert: Die „Schlechten“ resignieren oft, die „Guten“ „lehnen“ sich oft sich „zurück“. Bei der Veröffentlichung von Evaluationsergebnissen auf der Ebene einzelner Veranstaltungen oder Dozent(inn)en (z.B. in Form von Rankings) dürften diese negativen Effekte sogar noch verstärkt auftreten (vgl. Rheinberg, 1998).

9.2 Bewertung anhand von Qualitätskriterien

Kommen sachlogisch definierte Qualitätskriterien im Sinne der zu erreichenden Mindestqualität in Bezug auf die einzelnen Aspekte der Lehrqualität zum Einsatz, wird Lehrqualität „kriterial“ in Relation zu diesen Mindestanforderungen definiert. Diese werden für die einzelnen im Evaluationsbogen enthaltenen Fragen in Form von absoluten Grenzwerten vorab festgelegt. Es resultieren Qualitätsbewertungen wie „hohe Lehrqualität“, „ausreichende Lehrqualität“ oder „Lehrqualität, die nicht den Mindestanforderungen genügt“.

Die Bewertung der Lehrqualität anhand dieser Qualitätskriterien ermöglicht, individuelle wie institutionelle Verbesserungen abzubilden und Dozent(inn)en zu Verbesserungen zu motivieren (vgl. Rheinberg, 1998). Die in sachlichen Überlegungen wurzelnden Qualitätskriterien sind zudem am besten dazu geeignet, Lehrveranstaltungen zu identifizieren, die universitätsweiten oder fachspezifischen Standards an die Lehrqualität nicht genügen.

Aufgrund dieser Vorteile sollte die Lehrqualitätsbewertung primär anhand von Qualitätskriterien erfolgen. Dies gilt sowohl für die Bewertung einzelnen Lehrveranstaltungen als auch der Gesamtqualität der Universität und der Fächer. Zur Bewertung der Gesamtqualität sollten dabei ausschließlich Qualitätskriterien im Hinblick auf die Lehrqualität auf Produktebene (Lehrerfolg) Verwendung finden. Qualitätskriterien in Bezug auf das didaktische Handeln sollten nicht verbindlich sein, da Freiheit in Bezug auf die Wahl der didaktischen Mittel zur Erreichung des Lehrerfolgs besteht. Sie können allerdings hilfreiche Indikatoren für Verbesserungsmöglichkeiten darstellen.

Die motivationspsychologische Literatur zeigt einhellig, dass anspruchsvolle, aber erreichbare Zielsetzungen die besten Auswirkungen im Hinblick auf die Motivierung zu Verbesserungen haben (vgl. Locke & Latham, 1990). Deshalb sollten die Qualitätskriterien nicht zu „weich“ definiert werden.⁵

9.3 Bewertung anhand anderer Teilaspekte derselben Lehrveranstaltung

Wie in Abschnitt 6 beschrieben, ist Lehrqualität multifaktoriell und sollte auch so erfasst werden. Dies eröffnet die Möglichkeit, die Ergebnisse zu den einzelnen Teilaspekten zueinander in Beziehung zu setzen und damit relative Stärken und Schwächen abzubilden (Profilanalyse). Hierbei wird Lehrqualität „individuell“ definiert. Es resultieren Qualitätsbewertungen wie „relativ starker Teilaspekt der Lehrqualität“ und „relativ schwacher Teilaspekt der Lehrqualität“.

Diese relative Bewertung der einzelnen Aspekte ist insofern günstig, als sie positive wie negative Aspekte betont, den Fokus auf spezifische Verbesserungsmöglichkeiten legt und damit erreichbare und handlungsnaher Verbesserungsziele impliziert (vgl. Locke & Latham, 1990). Sie kann sehr gut mit der Bewertung anhand von Qualitätskriterien kombiniert werden (z.B.: „hohe Lehrqualität bei Teilaspekt A und nicht ausreichende Lehrqualität bei Teilaspekt B“).

9.4 Bewertung anhand früherer Ergebnisse

Bei Anwendung dieses Maßstabs im Hinblick auf einzelne Dozent(inn)en bzw. Lehrveranstaltungen werden Veränderungen der Lehrqualität mit der Zeit abgebildet. Auch hier wird Lehrqualität „individuell“ definiert. Es resultieren Qualitätsbewertungen wie „verbesserte Lehrqualität“, „gleichbleibende Lehrqualität“ und „verschlechterte Lehrqualität“. Auch dieser Bewertungsmaßstab hat günstige Auswirkungen: Verschlechterungen lenken den Blick

⁵ Im Rahmen der Testevaluation im Wintersemester 2005/06 wurden testhalber moderat definierte Qualitätskriterien eingesetzt. Als Mindestqualität wurde dabei bei allen fünf Fragen zur Lehrqualität auf Produktebene ein Wert von 3.5 angesetzt. Dies entspricht der „Mitte“ der sechsstufigen und von 1 bis 6 reichenden Antwortskala. Als die Mindestqualität wurde damit zumindest eine leichte Zustimmung zu den entsprechenden Aussagen festgelegt. Von den 39 einbezogenen Vorlesungen des ersten Semesters erreichten 27 die angesetzte Mindestqualität zu 100%. Neun Veranstaltungen lagen in einem unterhalb der Mindestqualität angesetzten Toleranzbereich (Breite: 1 Skaleneinheit) und bei drei Veranstaltungen lag die Lehrqualität auch unterhalb dieses Toleranzbereichs.

darauf, dass man etwas schon einmal besser gemacht hat und motivieren zu einer Ursachenanalyse sowie einer Rückkehr zum vorangegangenen Niveau; Verbesserungen sind eine Bestätigung guten Lehrverhaltens und motivieren zu weiteren Anstrengungen. Allerdings ist dieser Bewertungsmaßstab datentechnisch nur relativ schwer zu realisieren (Datenhaltung über längere Zeit, erforderliche Korrektur von potenziellen Verzerrungen beim Vergleich unterschiedlicher Lehrveranstaltungen desselben Lehrenden). Er sollte deshalb nicht standardmäßig angewendet werden, sondern den einzelnen Dozent(inn)en auf Nachfrage zur Verfügung gestellt werden.

Die im vorliegenden Kapitel ausgeführten Überlegungen zu Bewertungsmaßstäben der Lehrqualität münden in der folgenden Empfehlung:

Empfehlung 8: Definition von Qualitätskriterien der Lehrqualität

Es sollten universitätsweit einheitliche und anspruchsvolle Qualitätskriterien der Lehrqualität definiert werden, mit denen festgelegt wird, wie gut die Lehrqualität auf Produktebene (Lehrerfolg) mindestens ausgeprägt sein soll. Diese Definition soll auf der Grundlage sachlicher Überlegungen erfolgen.

Die Bewertung der Lehrqualität von Lehrveranstaltungen, Fächern und der Universität soll auf der Grundlage dieser Qualitätskriterien der Lehrprodukte erfolgen. Von einer vergleichenden Bewertung einzelner Lehrveranstaltungen oder Dozent(inn)en anhand von Durchschnittswerten ist abzuraten.

Ergänzend empfiehlt sich zur Identifikation von Verbesserungsmöglichkeiten auch die unverbindliche Definition von Qualitätskriterien im Hinblick auf die Prozessmerkmale guter Lehre (didaktisches Handeln). Diese können Charakteristika der einzelnen Fächern Rechnung tragen.

10 Zeitpunkt der Studierendenbefragung

Das vorrangige Ziel, dem die Lehrveranstaltungsevaluation verpflichtet ist, besteht in der Optimierung des Lehr- und Lernprozesses durch Kommunikation zwischen Lehrenden und Lernenden über die Qualität der Lehre und etwaige Probleme in deren konkreten Umsetzung (vgl. Abschnitt 5). Eine unabdingliche Voraussetzung hierfür ist eine zeitnahe Rückmeldung der Ergebnisse, um den Lehrenden die Gelegenheit zu geben noch in der laufenden Veranstaltung ggf. ihre Lehre anzupassen und gegenüber den Studierenden Stellung zu den Evaluationsergebnissen beziehen zu können. Für die Arbeitsgruppe hat das

Ziel der zeitnahen Ergebnismrückmeldung höchste Priorität, um das übergeordnete Evaluationsziel, die Verbesserung der Lehrqualität und die Entwicklung der Lehrkompetenz auf Ebene der individuellen Dozent(inn)en, erreichen zu können. Erhalten die Dozent(inn)en die Evaluationsergebnisse erst nach dem Ende der betreffenden Lehrveranstaltung, dürften (a) diese Ergebnisse von geringerer Relevanz sein (da die Veranstaltung bereits beendet ist) und (b) Transferprobleme bei der Umsetzung der gewonnenen Erkenntnisse auftreten (da im nachfolgenden Semester oftmals andere Lehrveranstaltungen abgehalten werden). Veranschaulicht werden kann dies als nicht geschlossene „Feedbackschleife“ (Abbildung 3).

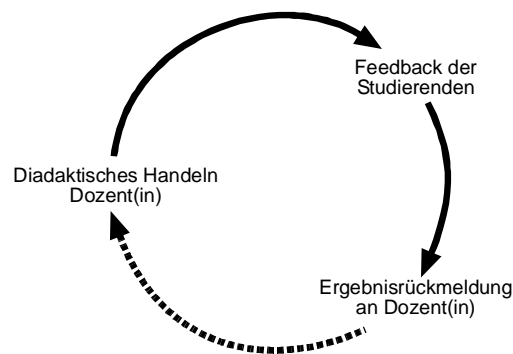


Abbildung 3. Feedbackschleife der Lehrveranstaltungsevaluation: Dozent(inn)en müssen unmittelbare Gelegenheiten haben, ihre Lehre anzupassen. Die Studierenden müssen erkennen, dass ihre Beurteilungen Effekte auf das didaktische Handeln der Dozent(in) haben, anderenfalls sinkt die Bereitschaft, an der Evaluation teilzunehmen. Evaluationsrückmeldungen nach Abschluss der betreffenden Lehrveranstaltungen erschweren beide Aspekte wesentlich

Auch aus Sicht der Studierenden ist eine zeitnahe Ergebnismrückmeldung wichtig. Nur wenn die Dozent(inn)en die Ergebnisse noch im laufenden Semester erhalten, kann ein für die Studierenden sichtbarer Effekt der Lehrveranstaltungsevaluation eintreten oder der/die Dozent(in) kann ihnen signalisieren, dass das Feedback angekommen ist. Ein häufiges Problem von Evaluationen, nicht nur von Lehrveranstaltungsevaluationen, ist, dass für diejenigen, die Evaluationsrückmeldungen geben, nicht erkennbar ist, dass diese Effekte zeigen (obwohl sie es oftmals dennoch tun). Mit anderen Worten: Für sie erscheint das Geben von Feedback sinnlos (vgl. Abbildung 3).

In der Folge kann die Lehrveranstaltungsevaluation nur dann dauerhaft akzeptiert werden und nachhaltige Effekte zeigen, wenn diese Feedbackschleife geschlossen ist (vgl. Abschnitt 6).

Empfehlung 9: Zeitpunkt der Befragung

Die Lehrveranstaltungsevaluation soll zur Mitte des Semesters durchgeführt werden, damit Evaluationsergebnisse noch im laufenden Semester zu einer Anpassung der Lehre beitragen können.

11 Ergebnismrückmeldung, Berichtslegung und Ergebnisveröffentlichung

11.1 Rückmeldung der Ergebnisse an die Dozent(inn)en

Die Ergebnisse der studentischen Befragung sollen den Lehrenden einen Eindruck von der Sichtweise der Teilnehmer(innen) vermitteln und der Ausgangspunkt für Verbesserungen der Lehrqualität und die Diskussion zwischen Lehrenden und Studierenden über die Lehre sein. Hierfür bedarf es verständliche und interpretierbare Ergebnismrückmeldungen.

Um eine schnelle Rückmeldung zu gewährleisten, ist eine automatisierte Verarbeitung der Fragebögen erforderlich. Bei der Darstellung im Ergebnisbericht empfiehlt sich eine Gliederung vom Ergebnisüberblick zu den Detailergebnissen und von den geschlossenen zu den offenen Antworten. Weiterhin empfiehlt sich die Darstellung von statistischen Kennwerten, die auch von „statistischen Laien“ einfach interpretiert werden können (vor allem der arithmetische Mittelwert).⁶

Im Rahmen der Testevaluation im Wintersemester 2005/06 wurde dazu das System *Eva-Sys* der Firma *Electric Paper* getestet (genauer dazu: Abschnitt 14.4). Dieses System hat sich zur unmittelbaren Erstellung leicht interpretierbarer Ergebnismrückmeldungen weitgehend bewährt.

Der automatisch erstellte Ergebnisbericht gliederte sich in die folgenden Teile:

- Zusammengefasste Indikatoren der Lehrqualität auf Produkt- und Prozessebene: Dies entspricht den in Tabelle 1 aufgeführten Faktoren (linke Spalte). Diese zusammengefassten Indikatoren („Globalwerte“) eignen sich besonders gut, um die Lehrqualität im Überblick zu beurteilen und relative Stärken und Schwächen zu erkennen.
- Ergebnisse zu den einzelnen geschlossenen Fragen (graphische Darstellung; Häufigkeiten, arithmetische Mittelwerte, Standardabweichungen). Die graphische Darstellung hat den Vorteil, dass sie leicht interpretierbar ist und Informationen über die Homogenität vs. Heterogenität der studentischen Urteile liefert.
- Zusätzlich zu den deskriptiven Ergebnissen ist für die einzelnen Fragen in Form von Ampelsymbolen dargestellt, ob die vorab definierten Mindestanforderungen an die Lehrqualität erreicht wurden (Qualitätsrichtlinien; vgl. Abschnitt 9.2). Anhand dieser Symbole können Verbesserungsmöglichkeiten des Lehrhandelns identifiziert werden. Insbesondere bei kritisch beurteilten Aspekten im Bereich des didaktischen Handelns besteht für die Lehrenden konkreter Anpassungs- bzw. Verbesserungsspielraum.
- Profillinie zur vergleichenden Veranschaulichung der Mittelwerte der einzelnen Fragen. Damit können fein granuliert relative Stärken und Schwächen identifiziert werden.

⁶ Dies setzt bestimmte psychometrische Eigenschaften der Fragen voraus, was deshalb eines der wichtigsten Kriterien bei der Fragebogenkonstruktion war (vgl. Abschnitt 8.1).

- Antworten auf die offenen Fragen.⁷

Es wird empfohlen den Ergebnisbereich auch zukünftig so zu gestalten. Weiterhin empfiehlt sich, den Ergebnismeldungen zusätzliche Erläuterungen beizufügen, in denen Hinweise zur Interpretation der Evaluationsergebnisse sowie Empfehlungen zur Bewertung der Ergebnisse und zu möglichen Folgerungen für die Verbesserung der Lehrqualität enthalten sind. Die Arbeitsgruppe hat derartige Erläuterungen bereits im Rahmen der Testevaluation erstellt.

Die Reaktionen der an der Testevaluation beteiligten Dozent(inn)en zur Ergebnismeldungen waren durchwegs positiv.

Empfehlung 10: Rückmeldung an die Dozent(inn)en

Die Dozent(inn)en sollen möglichst schnell (ideal: innerhalb von zwei Wochen) und in jedem Fall noch im laufenden Semester eine Rückmeldung über die Ergebnisse der Evaluation ihrer Lehrveranstaltungen erhalten.

Diese Rückmeldung soll leicht verständlich und auch für statistische Laien gut interpretierbar sein. Aspekte der Lehre, die nicht den definierten Mindestkriterien an die Lehrqualität entsprechen, sollten gekennzeichnet sein.

11.2 Berichte für Studiendekane, Dekane und den Prorektor für Lehre

Zur Bewertung der Gesamtqualität der Fächer und der Universität sollten Studiendekane, Dekane und der Prorektor für Lehre am Ende einer Evaluationsperiode zusammenfassende Berichte über die Evaluationsergebnisse der in ihren Verantwortungsbereich fallenden Lehrveranstaltungen erhalten. Diese Berichte sollen einerseits die Daten auf aggregiertem Niveau zusammenfassen und andererseits in einem ergänzenden Teil so beschaffen sein, dass auch Lehrveranstaltungen identifiziert werden können, die den Qualitätsansprüchen an die Lehre nicht genügen (vgl. Abschnitt 5).

Die im Rahmen der Testevaluation im Wintersemester 2005/06 erstellten Berichte, genügten diesen Anforderungen nur teilweise. Das verwendete System *EvaSys* bietet allerdings die Möglichkeit, die Berichte für Studiendekane, Dekane und den Prorektor für Lehre noch besser den Anforderungen anzupassen.

Falls eine vergleichende Ergebnisdarstellung in Bezug auf verschiedene Teilbereiche der Universität vorgenommen werden soll (z.B. Fakultäten, Fächer, Abteilungen), sollte nur der Lehrerfolg (Lehrqualität auf Produktebene) herangezogen werden und nicht das didak-

⁷ Bei zehn oder weniger ausgewerteten Fragebögen empfiehlt sich zur Sicherung der Anonymität der Studierenden, die Antworten händisch zu transkribieren. Bei mehr als zehn Fragebögen werden die Antworten in der Original-Handschrift gelistet. Vergleiche zu weiteren Aspekten des Datenschutzes (siehe Abschnitt 15).

tische Handeln der Lehrenden (Lehrqualität auf Prozessebene). Da Freiheit in der Wahl der (didaktischen) Mittel besteht, ist das didaktische Handeln für einen Vergleich kaum geeignet. Weiterhin sollten dabei potenzielle Verzerrungen durch studentische Bedingungen und Rahmenbedingungen mit manuellen Datenanalysen korrigiert (d.h. der Einfluss von Semester, Vorwissen, Besuchsgrund, Veranstaltungsform, Veranstaltungsgröße etc. herausgerechnet) werden.

Auf der Grundlage dieser Berichte können Zielvereinbarungsgespräche zwischen Vertretern der Universitätsleitung und den verantwortlichen Fakultäts- bzw. Fachvertretern geführt werden. In solchen Zielvereinbarungsgesprächen können Konsequenzen der Lehrveranstaltungsevaluation sowie Maßnahmen zur Verbesserung der Lehrqualität festgelegt werden. Dieses Instrument hat sich an anderen Universitäten bereits bewährt (Hochschulrektorenkonferenz, 2003).

Empfehlung 11: Berichte für Studiendekane, Dekane und den Prorektor für Lehre

Studiendekane und Dekane sollten am Ende jedes Semesters einen zusammenfassenden Bericht über die Evaluationsergebnisse ihres Bereichs erhalten.

Ebenfalls am Ende jeden Semesters soll der Prorektor Lehre einen auf höherer Ebene zusammengefassten Bericht erhalten. Dieser Bericht kann auch zur Dokumentation und zur Rechtfertigung der Gesamtqualität der Lehre an der Universität und der Fächer dienen.

11.3 Rückmeldung der Ergebnisse an die Studierenden

In verschiedenen Konzepten der Lehrveranstaltungsevaluation wird es als sinnvoll und für die Qualität der Lehre förderlich betrachtet, dass die Evaluationsergebnisse den studentischen Teilnehmer(inne)n der betreffenden Lehrveranstaltung zurückgemeldet werden und mit ihnen über Interpretation und Folgerungen diskutiert wird (z.B. Webler, 1992).

Deshalb sollte den Dozent(inn)en empfohlen werden, die Evaluationsergebnisse noch im laufenden Semester in ihren Lehrveranstaltungen zu thematisieren. Dazu sollte ihnen eine geeignete Präsentationsvorlage mit den wichtigsten Ergebnissen zur Verfügung gestellt werden.

Die Vorstellung der Evaluationsergebnisse in der Lehrveranstaltung wird von der Arbeitsgruppe als ein wichtiger Schritt innerhalb der Lehrveranstaltungsevaluation angesehen. Sie dient dazu, die in Abbildung 3 dargestellte „Feedbackschleife“ auch für die Studierenden zu schließen und einer „Evaluationsmüdigkeit“ entgegenzuwirken (vgl. Abschnitt 6). Die Vorstellung der Evaluationsergebnisse macht den Studierenden deutlich, dass ihre

Meinung von den Lehrenden ernst genommen wird und dass die Lehrveranstaltungsevaluation ein sinnvolles Qualitätssicherungsinstrument ist.

Empfehlung 12: Ergebnismrückmeldung an die Studierenden

Die Arbeitsgruppe empfiehlt, dass die Dozent(inn)en die Evaluationsergebnisse in ihren Lehrveranstaltungen vorstellen und ggf. Unklarheiten sowie etwaige Veränderungswünsche mit den Studierenden diskutieren. Diese Ergebnispräsentation in den Lehrveranstaltungen sollte den Dozent(en) nachdrücklich empfohlen werden, aber dennoch freiwillig sein.

Dazu sollte die Ergebnismrückmeldung für die Dozent(inn)en eine Vorlage für die Ergebnispräsentation in der Lehrveranstaltung beinhalten.

11.4 Veröffentlichung der Ergebnisse

Evaluationsergebnisse einzelner Lehrveranstaltungen sollten zusammenfassend in Form von Berichten auf der Ebene von Fächern oder Studiengängen hochschulintern veröffentlicht werden.

Falls veranstaltungsbezogene Resultate veröffentlicht werden sollen (etwa in Form von Führern zur Veranstaltungswahl der Studierenden), müssen Datenschutzbestimmungen und Mitbestimmungsrechte beachtet werden (vgl. Abschnitt 15).

Empfehlung 13: Veröffentlichung der Ergebnisse

Es empfiehlt sich eine Veröffentlichung der Evaluationsergebnisse, bei der die Ergebnisse der einzelnen Veranstaltungen auf höherer Ebene (z.B. Fächer, Studiengänge) zusammengefasst werden (Aggregation der Daten).

Um eine Demotivierung der Dozent(inn)en bei der Verbesserung ihrer Lehrqualität zu vermeiden, empfiehlt die Arbeitsgruppe, vergleichende Bewertungen auf Ebene der einzelnen Lehrveranstaltungen (z.B. in Form von Rankings) *nicht* zu veröffentlichen.

12 Beratungsangebot für Dozent(inn)en

Für den Fall, dass die Lehrenden in den Evaluationsergebnissen Verbesserungsbedarf erkennen, ist eine beratende Unterstützung bei der Interpretation der Ergebnisse und der darauf aufbauenden Erarbeitung von Verbesserungsmöglichkeiten sinnvoll. Pädagogisch-psychologische Studien zeigten diesbezüglich, dass mit der Beratung von Dozent(inn)en

auf der Grundlage von studentischen Veranstaltungsbeurteilungen deutlich größere Verbesserungseffekte erzielt werden können als mit der ausschließlichen Rückmeldung der Evaluationsergebnisse (Dresel & Rindermann, 2006; Rindermann & Kohler, 2003; Penny & Coe, 2004). Eine wichtige Voraussetzung für ein Evaluationsverfahren, das effektiv zu Veränderungen bzw. Verbesserungen führen soll, ist also die Bereitstellung von Beratungs- und Weiterbildungsangeboten für Lehrende.

Empfehlung 14: Beratungsangebot für Lehrende

Für Lehrende, die Verbesserungsbedarf für ihre didaktischen Handlungsweisen erkennen, sollen Beratungsgespräche zur Interpretation der Evaluationsergebnisse und zur Verbesserung der individuellen Lehrkompetenz angeboten werden.

13 Form der Datenerhebung

Eine wichtige Voraussetzung, dass die studentische Lehrveranstaltungsevaluation als internes Instrument zur Optimierung der Lehrveranstaltungsqualität dienen kann, ist eine möglichst zeitnahe Ergebnisrückmeldung der studentischen Befragungen (vgl. Abschnitt 10). Dies setzt eine rasche und hocheffiziente Verarbeitung der in großer Zahl vorliegenden studentischen Veranstaltungsbeurteilungen voraus, die organisatorisch und finanziell durchführbar ist. Ziel ist es, verlässliche Ergebnisse bei möglichst geringer Ressourcenbelastung zu erhalten.

Um eine effiziente Datenerfassung zu ermöglichen, erschien zu Beginn der Tätigkeit der Arbeitsgruppe die auf einem Onlineevaluationssystem basierende Lehrveranstaltungsevaluation viel versprechend. Als zentraler Vorteil erschien der Arbeitsgruppe, dass alle Antworten, insbesondere auch jene auf offene Fragen, bereits maschinell gespeichert sind und damit unmittelbar zur Weiterverarbeitung zur Verfügung stehen.

Allerdings bestanden sowohl seitens des Senatsausschusses Lehre als auch innerhalb der Arbeitsgruppe Bedenken gegenüber der Lehrveranstaltungsevaluation mittels eines Onlinesystems. Befürchtet wurden vor allem niedrigere Beteiligungsquoten, eine Anfälligkeit für Manipulationen, Selbstselektionseffekte und damit eine Nichtrepräsentativität derjenigen Studierenden, die den Onlinefragebogen ausfüllen (für die Gesamtheit der Teilnehmer(innen) der betreffenden Lehrveranstaltung) sowie eine geringere Datenqualität.

Da die einschlägige Fachliteratur kaum Aufschluss zu diesen möglichen Verzerrungen liefert, beschloss die Arbeitsgruppe im Rahmen der Testevaluation im Sommersemester 2005 einen methodisch haltbaren Vergleich zwischen papiergebundener Evaluation und Onlineevaluation durchzuführen. Die Ergebnisse sind detailliert im Zwischenbericht der

Arbeitsgruppe (Dresel, 2005a) dargestellt. Ihre Veröffentlichung ist im Gange (Dresel & Tinsner, 2006).

Zusammenfassend lässt sich festhalten, dass starke Befürchtungen hinsichtlich der Manipulierbarkeit und geringen Datenqualität eher unberechtigt sind. Allerdings wird mit der Onlineevaluation derzeit nur ein geringer Teil der Studierenden erreicht. Mit der Onlineevaluation konnte nur gut ein Drittel der Studierenden erreicht werden, die mit der Papierevaluation erreicht werden. Dies alleine wäre noch nicht problematisch, wenn diese Teilstichprobe repräsentativ für die Gesamtheit der an einer Lehrveranstaltung teilnehmenden Studierenden wäre. Selbstselektionseffekte – vor allem in Bezug auf das Fachsemester, aber auch hinsichtlich der Beurteilung der Lehrqualität – deuten jedoch darauf hin, dass dies nicht der Fall ist. Ursache dafür dürfte vor allem die aus Sicht der Studierenden nicht geschlossene Feedbackschleife und eine daraus resultierende Evaluationsmüdigkeit mit zunehmendem Semester sein (vgl. Abbildung 3).

Von daher rät die Arbeitsgruppe – zumindest bis auf weiteres – davon ab, die Lehrveranstaltungsevaluation mit einem Onlinesystem durchzuführen. Stattdessen sollten maschinenlesbare Papierfragebögen Verwendung finden.

Nicht ausgeschlossen werden soll eine spätere Umstellung auf die onlinegestützte Lehrveranstaltungsevaluation, sofern der geringe Rücklauf und die dabei auftretende Selbstselektivität ausgeräumt sind.

Empfehlung 15: Form der Datenerhebung

Es wird empfohlen, die Evaluation der Lehrveranstaltungen bis auf weiteres mittels Papierfragebögen durchzuführen. Der mögliche Effizienzgewinn einer Datenerhebung mit einem reinen Onlinesystem wiegt momentan nicht deren Nachteile auf (geringe Beteiligungsquote und Selbstselektionseffekte).

Am Beispiel der Universität Stuttgart zeigte sich zudem, dass die Ausgabe von offenen Antworten in der Originalhandschrift der Befragten von der Zentralstelle für Datenschutz der baden-württembergischen Universitäten (ZENDAS) dann als datenschutzrechtlich unbedenklich eingestuft wird, wenn wenigstens sechs ausgefüllte Fragebögen der jeweiligen Lehrveranstaltung vorliegen (vgl. Abschnitt 15). Vor diesem Hintergrund fällt der Effizienzvorteil von Onlinesystemen ohnehin weit weniger gewichtig aus.

14 Evaluationsworkflow

14.1 Anforderungen an den Evaluationsworkflow

Basierend auf den Überlegungen der bisherigen Abschnitte ergeben sich die folgenden Anforderungen, die an einen effizienten Workflow bei der universitätsweiten Lehrveranstaltungsevaluation zu stellen sind:

1. Effiziente Datenerfassung, auch von handschriftlichen Antworten auf offene Fragen
2. Zeitnahe Ergebnismeldung noch im laufenden Semester
3. Unterstützung der Dozent(inn)en, bei der Rückmeldung der Evaluationsergebnisse an die Studierenden (automatische Erstellung von Folienvorlagen)
4. Automatisierte Erstellung von Berichten für Dozent(inn)en und Studiendekane/Studien-dekaninnen, Dekane/Dekaninnen und Prorektor für Lehre
5. Flexibilität hinsichtlich spezifischer Evaluationsinteressen (z.B. andere Veranstaltungsformen, neues Lehrkonzept, Fachspezifika)
6. Möglichkeit, englischsprachige Parallelversionen von Fragebögen einzubinden
7. Gewährleistung des Schutzes der Daten von Studierenden und Dozent(inn)en

Die Anforderungen 1 bis 4 sind zunächst auf das Ziel ausgerichtet, Rückmeldungen in einem sehr kurzen Zeitraum anzufertigen, um Dozent(inn)en Gelegenheit zu geben, noch in der laufenden Veranstaltung ihre Lehre anzupassen und gegenüber den Studierenden Stellung zu den Ergebnissen nehmen zu können (vgl. Abbildung 3).

Zwei weitere Anforderung (5 und 6) betreffen die Flexibilität des zu implementierenden Systems in Bezug auf unterschiedliche Evaluationsinteressen und die dabei zu verwendenden Fragebögen. Ein nützlicher Evaluationsworkflow sollte auch dann eine hocheffiziente Unterstützung bieten, wenn andere als die standardisierten Fragebögen zum Einsatz kommen (vgl. Abschnitt 8.4). So sollten in den Workflow schnell Fragebögen für beliebige Evaluationsinteressen integriert werden können. Eine zentrale Anlaufstelle sollte Unterstützung bei der Entwicklung der Fragebögen und der Implementation in den Evaluationsworkflow bieten. Der Integration von englischsprachigen Fragebögen dürfte dabei im Zuge der Internationalisierung der Lehre ebenfalls besonderer Stellenwert zukommen.

Schließlich ist bei einer zentral koordinierten Lehrveranstaltungsevaluation der Schutz der personenbezogenen Daten in stärkerem Ausmaß zu gewährleisten als in der bisherigen Evaluationspraxis (Anforderung 7). In diesem Zusammenhang besteht in der Arbeitsgruppe jedoch Konsens darüber, dass aus Gründen der Löschpolicy keinesfalls auf die Möglichkeit des chronologischen Vergleichs auf Ebene der Dozenten (Vergleich mit den Ergebnissen früherer Semester; vgl. Abschnitt 9.4) verzichtet werden darf.

14.2 Testweise Umsetzung eines Evaluationsworkflow

Nach Beschluss des Senatsausschuss Lehre in der Sitzung am 07.07.2005 wurden im Wintersemester 2005/06 alle Vorlesungen des ersten Semesters (Medizin zusätzlich noch diejenigen des dritten Semesters) testweise papiergebunden mit dem System *EvaSys* der Firma *Electric Paper* evaluiert. Für dieses System wurden eine Basislizenz sowie ein Hochleistungsscanner mit vollem Rückgaberecht erworben und vom Kommunikations- und Informationszentrum (kiz) implementiert und zur Verfügung gestellt. Die genannten Lehrveranstaltungen wurden aus dem Online-Veranstaltungsverzeichnis *LSF* extrahiert, nach *EvaSys* importiert und dort Teilbereichen in der Studienkommissionsstruktur zugeordnet. Für den Testlauf im Wintersemester 2005/06 wurden die organisatorische Abwicklung, inhaltliche Entwicklung, Systemeinführung und Erstellung von Ergebnismeldungen und Berichten von der Arbeitsgruppe betreut und koordiniert. Der Testlauf wurde in *EvaSys* nach dem Verfahren der „Zentralen Evaluation“ durchgeführt, unter Einbindung der Fachschaften bei der Datenerhebung, dem Scannen der Fragebögen und teilweise dem Versand der Ergebnismeldungen an die Dozenten.

Die zentral gesteuerte Evaluation hat sich bewährt und wird für den zukünftigen Evaluationsablauf empfohlen.

Empfehlung 16: Zentrale Evaluation unter Beteiligung der Fächer und Fachschaften

Es wird das Vorgehen einer zentralen Evaluation empfohlen. Dieses Vorgehen eignet sich besonders für die verpflichtende und in einigen Fächern flächendeckende Lehrveranstaltungsevaluation und bietet die Möglichkeit der Vergleichbarkeit der Ergebnisse. Die organisatorische, technische als auch die inhaltliche Betreuung sollte zentral koordiniert werden. Wichtig ist eine feste Einbindung der jeweiligen Fächern und Fachschaften in den Bereichen der Definition von Evaluationsschwerpunkten der Datensammlung, Datenerfassung und der konzeptionellen Weiterentwicklung von Erhebungsinstrumenten sowie der Rückmeldungen und Interpretation der Ergebnisse.

Die Arbeitsgruppe empfiehlt unabhängig davon, ob Lehrveranstaltungen obligatorisch oder freiwillig evaluiert werden, die Abwicklung über eine zentrale Stelle.

14.3 Empfohlener Evaluationsworkflow

Im Folgenden werden Empfehlungen zu dem zu realisierenden Evaluationsworkflow gegeben. Diese Empfehlungen stellen das Ergebnis des umfassenden Testlaufs im Wintersemester 2005/06 dar. Der empfohlene Evaluationsworkflow ist in Abbildung 4 dargestellt.

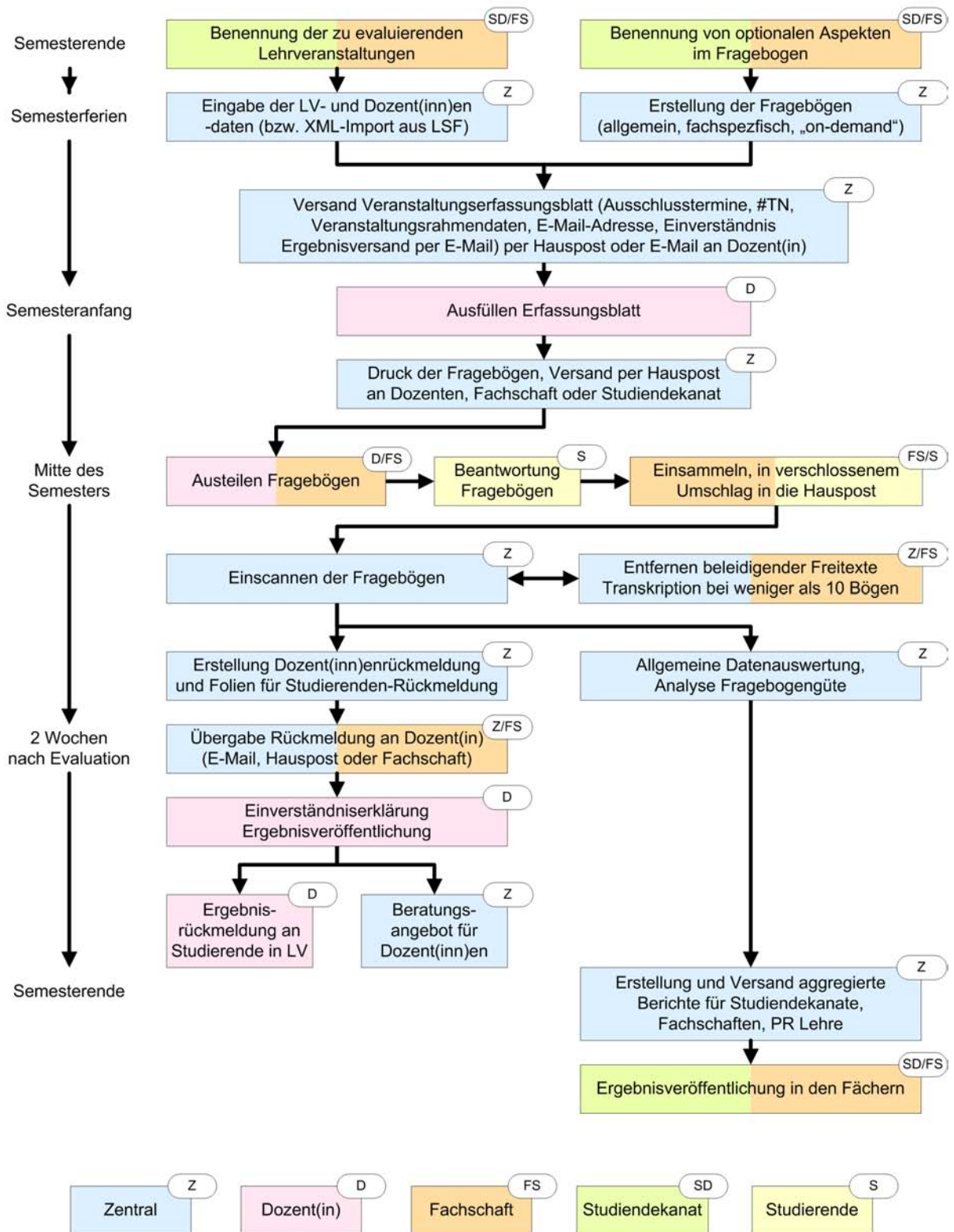


Abbildung 4. Empfohlener Evaluationsworkflow

Wie in Abschnitt 6 ausgeführt, empfiehlt die Arbeitsgruppe, in jedem Semester ein Viertel aller Lehrveranstaltungen jedes Fachs verpflichtend zu evaluieren und darüber hinaus die

Möglichkeit, beliebig viele Lehrveranstaltungen zu evaluieren. Die Festlegungen hinsichtlich der zu evaluierenden Lehrveranstaltungen werden von den Studiendekanaten bzw. den Studienkommissionen unter Einbeziehung der Fachschaften getroffen. Zur Koordination der Durchführung und Organisation der Lehrveranstaltungsevaluation ist dies jeweils bereits am Ende des vorangehenden Semesters erforderlich.

In diesem Rahmen können auch Anforderungen an das Erhebungsinstrument für die einzelnen Fächer festgelegt werden (z.B. optionale Fragen, fachspezifische Anforderungen). Dieser Schritt ermöglicht es neben der gezielten Planung auch, Schwerpunkte für ein Semester oder Studienjahr zu setzen (z.B. aufgrund von identifizierten Problem- oder Schwachstellen). So kann es beispielsweise für ein Fach sinnvoll sein, den Schwerpunkt auf Einführungsveranstaltungen, Vorlesungen, Praxisveranstaltungen o.ä. zu legen.

Auf der Grundlage der Benennungen der Studiendekanate können zum einen zu evaluierenden Lehrveranstaltungen korrekt in das Evaluationssystem eingepflegt werden (XML-Importschnittstelle zum *LSF*-System). Zum anderen können Fragebögen entsprechend der spezifischen Anforderungen erstellt werden. Die Fragebögen sollten zentral von fachkompetenter Seite erstellt werden. Dadurch wird der Einsatz inhaltlich wie psychometrisch haltbarer und nicht zuletzt einheitlich gestalteter Evaluationsbögen sichergestellt. Darüber hinaus ist bei der Fragebogenerstellung auch technische Kompetenz in der Bedienung des Fragebogeneditors *Teleform* erforderlich (vgl. Abschnitt 14.4), wobei dabei bereits auf gewünschte Darstellungen der Ergebnisse in den Dozent(inn)enrückmeldungen zu achten ist.

Zum Beginn des Semesters sollten die jeweiligen Dozenten bzw. Verantwortlichen über den Evaluationszeitraum (zwei Wochen in der Mitte des Semesters; vgl. Abschnitt 10) informiert und gleichzeitig um eine Rückmeldung hinsichtlich der Rahmendaten der zu evaluierenden Lehrveranstaltungen gebeten werden. Dies geschieht über den Versand eines Veranstaltungserfassungsblatts zu den folgenden Aspekten:

- Name des/der verantwortlichen Dozenten/Dozentin bei Lehrveranstaltungen mit mehreren Dozent(inn)en
- Etwaige Ausschlussstermine für die Studierendenbefragung
- Teilnehmer(innen)zahl (für Druck der Bögen und spätere Kontrolle des Rücklaufs)
- Veranstaltungsart (Wahlpflicht, Pflicht, freiwillig, Vorlesung mit/ohne Übung/Tutorium, Seminar, Praktikum)
- Vorgesehene Prüfung/Klausur
- Einverständniserklärung zum Ergebnisversand per Email (Angabe der Emailadresse)

Den Druck der Fragebögen sollte eine zentrale Stelle übernehmen (Druckauftrag an kiz/Medien). Der Versand der Bögen erfolgt per Hauspost an die Dozent(inn)en, die Fach-

schaften oder Studiendekanate (je nach vorheriger Absprache). Dadurch ist der rechtzeitige und korrekte Erhalt der einzusetzenden Fragebögen sichergestellt.

In den zu evaluierenden Lehrveranstaltungen werden die Fragebögen durch Fachschaftsvertreter(innen) oder durch die Dozent(inn)en am vorgesehenen Evaluationstermin ausgeteilt, von den Studierenden ausgefüllt, von den Fachschaftsvertreter(innen) oder einem vorab bestimmten Studierenden wieder eingesammelt und direkt nach der Erhebung in einem verschlossenen Hauspostumschlag an die zentrale Evaluationsstelle verschickt.

In Absprache mit den Fachschaften werden die Bearbeitung der offenen Fragen (Entfernen beleidigender Texte, manuelle Eingabe bei weniger als zehn Bögen) sowie Ergebnismeldungen an die Dozent(inn)en zentral oder von den Fachschaften übernommen. Die Rückmeldung der Ergebnisse sollte vorzugsweise per Email erfolgen, da ein Versand per Hauspost einen höheren administrativen Aufwand bedeutet.

Wie bereits ausgeführt, soll gewährleistet werden, dass den Dozent(inn)en unmittelbar eine Ergebnismeldung zur Verfügung gestellt wird (innerhalb von zwei Wochen) und sie somit die Möglichkeit haben, die Ergebnisse den Studierenden noch im laufenden Semester zurückzumelden und ggf. mit ihnen zu diskutieren (vgl. Abschnitt 11).

Nach Erhalt der Ergebnismeldung können die Dozent(inn)en ihr Einverständnis zur Veröffentlichung der Ergebnisse geben.

Eine Beratung der Dozent(inn)en hinsichtlich der Interpretation der Evaluationsergebnisse und möglichen Verbesserungen sollte von der zentralen Evaluationsstelle angeboten werden (vgl. Abschnitt 12).

Erforderliche allgemeine Datenauswertungen sowie Fragebogenanalysen sollten von zentraler, fachkompetenter Seite durchgeführt werden, um beispielsweise Daten aus weiterführenden Analysen zur Verfügung zu stellen oder eine konzeptionelle Weiterentwicklung von Erhebungsinstrumenten voranzutreiben.

Nach dem Abschluss der Erhebung werden von der zentralen Evaluationsstelle Berichte für Studiendekane, Dekane sowie den Prorektor für Lehre erstellt (vgl. Abschnitt 11.2). Dies erfordert neben methodischen auch kommunikative Kompetenzen (wie Übersetzungs-, Vermittlungs- und Serviceleistungen *zwischen den* und *für die* beteiligten Akteure). Es wird den verantwortlichen Studiendekanen und Dekanen empfohlen, mit Dozent(inn)en, bei denen Verbesserungsbedarf hinsichtlich der Lehrqualität erkennbar ist, ein klärendes Gespräch über mögliche Ursachen zu führen. In anschließenden Zielvereinbarungsgesprächen können spezifische Maßnahmen zur Verbesserung der Lehrqualität erörtert und zeitlich strukturiert werden.

Anschließend können die Fächer die sie betreffenden Evaluationsergebnisse veröffentlichen (vgl. Abschnitt 11.4).

14.4 Feldtest des Evaluationssystems *EvaSys*

Ein Schwerpunkt der Testevaluation im Wintersemester 2005/06 war der Feldtest des Evaluationssystems *EvaSys* bei der Umsetzung der in Abschnitt 14.1 dargestellten Anforderungen an den Evaluationsworkflow anhand eines großen Evaluationsvolumens.

EvaSys erwies sich prinzipiell geeignet zur effizienten Realisierung des empfohlenen Evaluationsworkflow. Mit dem System war es möglich, in kurzer Zeit eine große Anzahl an ausgefüllten Papierfragebögen per Scanner zu erfassen (7.866 ausgewertete Bögen) und für alle 102 am Testlauf beteiligten Vorlesungen eine zeitnahe Rückmeldung zu erstellen.⁸ Die von *EvaSys* automatisch erstellten Ergebnisrückmeldungen für Dozent(inn)en sowie Berichte für Studiendekane, Dekane und den Prorektor für Lehre genügten den daran gestellten Ansprüchen (vgl. Abschnitte 11.1 und 11.2). Vorteilhaft ist die durch das System unterstützte automatische Versendung der Rückmeldungen und Berichte per E-Mail.

Für die empfohlene zentrale Evaluation bietet *EvaSys* zahlreiche nützliche Funktionalitäten zur automatischen Verarbeitung von Umfragen. Notwendig für deren korrekten Ablauf ist eine geeignete Strukturierung des Systems (Teilbereiche entsprechend der Strukturierung des Online-Veranstaltungsverzeichnisses). Daneben müssen der Personenkreis, der mit der Durchführung der Lehrveranstaltungsevaluation betraut wird, und das resultierende Berechtigungskonzept festgelegt werden. Eine XML-Importschnittstelle zum *LSF*-System, mit dem das Online-Veranstaltungsverzeichnis realisiert ist, wird mittlerweile von der Firma Electric Paper angeboten und sollte künftig eingesetzt werden.

Für die Erstellung der Fragebögen wurde im Testlauf das *EvaSys*-Modul *VividForms* genutzt, das dies rein browserbasiert über eine intuitiv bedienbare Benutzeroberfläche ermöglicht. Die Fragebogenerstellung kann so von einem großen Anwenderkreis ohne Einsatz einer Clientsoftware durchgeführt werden. Allerdings sind mit *VividForms* auch einige Nachteile aufgrund des vorgegebenen eingeschränkten Layouts verbunden. Dadurch verlängerte sich im Testlauf der Fragebogen auf drei Seiten, was wiederum zu einer Erhöhung der Druckkosten und zur Notwendigkeit des Klammerns der Fragebogenseiten führte. Diese Klammern mussten dann vor dem Scannen der Bögen mit großer Vorsicht und somit aufwändig wieder entfernt werden. Dies ist als entscheidender Nachteil von *VividForms* anzusehen.

Neben *VividForms* bietet *Electric Paper* ein zusätzliches Modul, den *Teleform-Designer*, als professionelles Werkzeug zur Gestaltung von Fragebögen an, was allerdings mit zusätzlichen Kosten verbunden ist. Dieses Modul ermöglicht ein flexibles Design, die freie Anordnung sämtlicher gestalterischer Elemente sowie eine verbesserte Anpassung an die Corporate-Identity-Richtlinien der Universität Ulm. Es ist für einen kleineren Anwenderkreis

⁸ Tatsächlich ist es mit dem System ohne Weiteres möglich, den Dozent(inn)en noch am Tag der Studierendenbefragung ein „Sofort-Feedback“ zur Verfügung zu stellen, sofern die Fragebögen unmittelbar nach der Evaluation gescannt werden.

bestimmt und muss auf einem entsprechenden Arbeitsplatzrechner installiert werden, ist aber für die technische Umsetzung der eingesetzten Fragebögen weitaus geeigneter.

Im Rahmen des Testlaufs des Evaluationssystems *EvaSys* sollten auch Aussagen darüber getroffen werden, in welchem Umfang Fehler bei der scannergestützten Datenerfassung auftreten. Dazu wurden die 117 Fragebögen einer per Zufall ausgewählten Vorlesung sowohl scannergestützt, als auch per Hand erfasst. Die in Anhang A detailliert aufgeführten Ergebnisse zeigen, dass bei der scannergestützten Datenerfassung Fehler durch die Nichterkennung abgegebener Antworten auftraten, die in einer Erhöhung des Anteils fehlender Werte um etwa 2.5% resultierten. Weiterhin zeigte sich, dass diese Fehler vollkommen zufällig im Hinblick auf die abgegebenen Beurteilungen auftraten und die Evaluationsergebnisse nicht verzerrten. Insgesamt handelt es sich bei den Fehlern damit um eher geringe „Kosten“, die den Nutzen, der aufgrund des wesentlichen Effizienzgewinns als sehr groß einzuschätzen ist, kaum relativieren.

Zusammenfassend ist festzuhalten, dass sich das Evaluationssystem *EvaSys* im Rahmen der Testevaluation bei der Erfassung und Auswertung papiergebundener Fragebögen prinzipiell bewährt und als hocheffizient erwiesen hat. Sein Einsatz stellt nach der erforderlichen (und durchaus aufwändigen) Einarbeitung eine deutliche Arbeitserleichterung dar, insbesondere hinsichtlich der Datenerfassung und der Erstellung professioneller Auswertungen, die den definierten Anforderungen entsprechen. Das System erscheint auch in technischer Hinsicht ausgereift und läuft sehr stabil. Die Zusammenarbeit mit *Electric Paper* verlief sehr gut, Supportanfragen wurden sehr rasch bearbeitet. Allerdings zeigte sich, dass der in der Testlizenz enthaltene einfache *EvaSys*-Fragebogeneditor *VividForms* den Anforderungen der Universität Ulm nicht genügt und mit großen Einschränkungen bei der Fragebogenkonstruktion und Ergebnisrückmeldung einhergeht. Deshalb wird der Einsatz des umfangreicheren *EvaSys*-Fragebogendesigners *TeleForm* empfohlen.

Die Lizenzkosten für *EvaSys* und *TeleForm* sowie die jährlichen Supportgebühren werden durch den Rahmenvertrag zwischen der Fa. *Electric Paper* und dem Land Baden-Württemberg um 25% reduziert; auf den Scannerpreis wird ein Rabatt von 10 % gewährt.

Empfehlung 17: Evaluationssystem

Es wird der universitätsweite Einsatz des Evaluationssystems *EvaSys* empfohlen. Verwendet werden sollte der in der Testlizenz nicht enthaltene Fragebogendesigner *Teleform*, der alle erforderlichen Funktionalitäten zur Verfügung stellt.

Daneben ist der Einsatz der XML-Importschnittstelle von *LSF* nach *EvaSys* sowohl aufgrund der großen Zahl der zu verarbeitenden Veranstaltungsdaten wie auch zur Vermeidung fehlerhafter Daten durch Medienbrüche zwingend notwendig.

Es wird empfohlen, den technischen Support zentral an das Kommunikations- und Informationszentrum (kiz) anzubinden, das sich dazu bereits bereit erklärt hat.

14.5 Zusammenfassende und ergänzende Analysen

Die Lehrveranstaltungsevaluation an der Universität Ulm soll als internes Qualitätssicherungsinstrument dazu dienen, die Gesamtqualität der Lehre an der Universität Ulm und in den Fächern hinreichend valide zu bestimmen und darauf aufbauend Entscheidungen für Qualitätssicherungsmaßnahmen zu ermöglichen. Dies schließt die Identifikation von Lehrveranstaltungen ein, die nicht den Mindestanforderungen an Lehrqualität genügen (vgl. Abschnitt 5). Den verantwortlichen Akteuren (Studiendekane/Studiendekaninnen) sollten Informationen zur Verfügung gestellt werden, mittels derer spezifische Maßnahmen zur Verbesserung der Lehrqualität erörtert und festgelegt werden können. Die konkrete Ausgestaltung der Lehre ist nur mittelbar durch die Hochschul- oder Fakultätsleitung beeinflussbar (vgl. Wittenberg & Rothe, 1999). Es wird allerdings seitens der Fachbereiche und Hochschulen eine Selbstvergewisserung über die Qualität der Lehre als unerlässlich angesehen (Hochschulrektorenkonferenz, 1993). Hierfür wären am Ende jedes abgeschlossenen Semesters statistische Analysen notwendig, an denen eine bestmögliche Orientierung bzw. Einordnung stattfinden kann.

Es gilt zu bedenken, dass die zusammenfassende bzw. vergleichende Auswertung von studentischen Befragungsdaten auf der Ebene einzelner Lehrveranstaltungen, zur Lehrleistung einzelner Dozenten und Dozentinnen umstritten ist (Marsh, 1987; Marsh & Roche, 1987; Spiel, 2001a), wobei sich die Einwände vor allem auf die Validität studentischer Befragungsdaten beziehen (vgl. Kromrey, 2001; Rindermann, 2001b; siehe dazu auch Abschnitt 7.3).

Deshalb erfordern die notwendigen Analysen zur Bestimmung der Gesamtqualität der Universität und der Fächer sowie zur Korrektur möglicher Verzerrungen des studentischen Urteils zusätzliche komplexe Auswertungen mittels einer geeigneten Statistiksoftware. Sie können nicht mit dem Evaluationssystem durchgeführt werden.

Empfehlung 18: Ergänzende Auswertungen

Es empfiehlt sich, am Ende jedes evaluierten Semesters zusätzliche, komplexere statistische Auswertungen vorzunehmen, um die Definition der Qualitätskriterien, den Einfluss von Verzerrungsvariablen und die Güte neu entwickelter Fragebögen zu überprüfen sowie die längsschnittliche Entwicklung der Lehrqualität zu bestimmen.

15 Datenschutz

Bei der Evaluation der Lehrveranstaltungen der Universität Ulm müssen die damit gewonnenen personbezogenen Daten von Studierenden und Dozent(inn)en entsprechend der gesetzlichen Regelungen geschützt werden.

Laut Landeshochschulgesetzes (LHG) ist der Beschluss einer entsprechenden Satzung erforderlich. Die Vorschriften des LHG zum Datenschutz lassen die Durchführung von Evaluationsvorhaben nur zu, wenn eine Satzung der jeweiligen Hochschule die erforderlichen Regelungen trifft. Personenbezogene Daten dürfen im Zusammenhang mit einer Evaluation nur erhoben und weiterverarbeitet werden, wenn die Satzung dafür die notwendigen Festlegungen enthält (siehe § 5 Abs. 2 Satz 4 LHG). Die Satzung muss Aussagen dazu machen, welche Daten für ihren Zweck aus welchen Gründen erforderlich sind. Kann sie das nicht oder nicht überzeugend, bleibt die Verarbeitung personenbezogener Daten unzulässig.

Vor der Festlegung einer solchen Satzung, sollte deren Inhalt in einem gemeinschaftlichen Diskurs erörtert werden.

Im Rahmen der Einführung der universitätsweiten Lehrveranstaltungsevaluation mit *EvaSys* an der Universität Stuttgart, die durch die Zentrale Datenschutzstelle in Baden-Württemberg (ZENDAS) begleitet wurde, wurde sowohl der Ablauf der Evaluation, an dem sich die vorgeschlagene Vorgehensweise orientiert, wie auch das System *EvaSys* selbst kritisch begutachtet und für datenschutzrechtlich einwandfrei befunden.

Empfehlung 19: Datenschutz

Bei der Evaluation der Lehrveranstaltungen der Universität Ulm müssen die relevanten Bestimmungen zum Schutz der personbezogener Daten von Studierenden und Dozent(inn)en beachtet werden. Näheres dazu sollte eine Evaluationsatzung regeln. Bei der konkreten Durchführung der Evaluation sollten die Empfehlungen der Zentralen Datenschutzstelle in Baden-Württemberg (ZENDAS) berücksichtigt werden.

16 Erforderliche personelle und sachliche Ressourcen

Wie in den bisherigen Abschnitten deutlich wurde, empfiehlt die Arbeitsgruppe die Einrichtung einer zentralen Stelle für die Lehrveranstaltungsevaluation, in der Fachkompetenz im Hinblick auf theoretische Konzepte und die praktische Umsetzung von Lehrveranstaltungsevaluationen, die Entwicklung und empirische Überprüfung von Fragebögen, Hochschuldidaktische Konzepte, die Beratung von Lehrenden sowie die Bedienung des Evaluationssystems *EvaSys* (einschließlich des Fragebogeneditors *Teleform*) in hinreichendem Umfang vorhanden ist. Aufgrund dieses Anforderungsprofils soll die zentrale „Arbeitsstelle Lehrveranstaltungsevaluation“ dem Seminar für Pädagogik angegliedert werden. Die Arbeitsstelle soll die Lehrveranstaltungsevaluation an der Universität Ulm koordinieren, den empfohlenen Evaluationsworkflow (vgl. Abschnitt 13) verantwortlich durchführen und als Ansprechpartner für die Fächer fungieren.

In der folgenden, abschließenden Empfehlung der Arbeitsgruppe ist die personelle und sachliche Ausstattung der Arbeitsstelle aufgeführt.

Empfehlung 20: Einrichtung einer „Arbeitsstelle Lehrveranstaltungsevaluation“

Die Arbeitsgruppe empfiehlt die Einrichtung einer zentralen „Arbeitsstelle Lehrveranstaltungsevaluation“, die dem Seminar für Pädagogik angegliedert ist.

Diese Arbeitsstelle soll mit einer halben wissenschaftlichen Mitarbeiterstelle (BAT IIa/2) ausgestattet sein. Der/die wissenschaftliche Mitarbeiter(in) soll ein Diplom in Psychologie, Soziologie oder Pädagogik aufweisen und über sehr gute Kenntnisse in den Bereichen empirisch-sozialwissenschaftliche Methoden, Konstruktion und Validierung von Fragebögen sowie Evaluation verfügen. Spezifische Qualifikationen in den Bereichen der Lehrveranstaltungsevaluation, Hochschuldidaktik und Beratung sind wünschenswert. Auf der Mitarbeiterstelle soll die Möglichkeit zur Promotion gegeben sein.

Daneben soll die Arbeitsstelle mit 80 Hilfskraftstunden pro Monat zur organisatorischen und administrativen Unterstützung ausgestattet sein.

Im Hinblick auf die Sachmittel sollte die Arbeitsstelle mit zwei Notebooks mit Dockingstation, einem leistungsfähigen Laserdrucker und laufenden Sachmitteln (Papier, Druck, Softwarelizenzen) ausgestattet sein.

17 Literatur

- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Aufl.). Berlin: Springer.
- Cashin, W. E. (1996). Should student ratings be interpreted absolutely or relatively? Reaction to McKeachie. *Instructional Evaluation and Faculty Development*, 16, 14-19.
- Dresel, M. (2005a). *Universitätsweite Lehrveranstaltungsevaluation an der Universität Ulm. Zwischenbericht der Arbeitsgruppe Lehrveranstaltungsevaluation*. Ulm: Universität Ulm. Elektronisch verfügbar unter http://www.informatik.uni-ulm.de/t3-sfp/fileadmin/user/dresel/Zwischenbericht_AG_LVE_Juli_2005.zip (ca. 553 kB).
- Dresel, M. (2005b). *Empirische Überprüfung des einheitlichen Fragebogens zur Evaluation von Lehrveranstaltungen an der Universität Ulm (Fragebogenversion Sommersemester 2004)*. Ulm: Seminar für Pädagogik der Universität Ulm.
- Dresel, M. & Rindermann, H. (2006). *Beratung von Lehrenden auf der Grundlage von studentischen Beurteilungen der Lehrqualität: Mehrebenenanalytische Wirksamkeitsprüfung unter Berücksichtigung von Biasvariablen*. Manuskript eingereicht.
- Dresel, M. & Tinsner, K. (2006). *Ist die Onlinebefragung von Studierenden zur Evaluation von Lehrveranstaltungen ein faires, verzerrungsfreies, und ökonomisches Verfahren? Ergebnisse einer Messwiederholungsstudie unter Einbezug von papier- und online-basierten Lehrveranstaltungsbeurteilungen*. Manuskript in Vorbereitung.
- el Hage, N. (1999). *Lehrevaluation und studentische Veranstaltungskritik. Projekte, Instrumente, Grundlagen*. Bonn: Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. Introduction to the Current Issues section "student ratings of professors". *American Psychologist*, 52, 1182-1186.
- Hochschulrektorenkonferenz (HRK) (1993). *Umsetzung der Studienstrukturreform*. Hochschulrektorenkonferenz – 170. Plenum am 12. Juli 1993. URL: http://www.hrk.de/de/beschluesse/109_477.php?datum=170.+Plenum+am+12.+Juli+1993 (Stand: Januar 2006)
- Hochschulrektorenkonferenz (HRK) (Hrsg.) (2003): *Wegweiser 2003. Qualitätssicherung an Hochschulen. Sachstandsbericht und Ergebnisse einer Umfrage des Projektes Qualitätssicherung* (Beiträge zur Hochschulpolitik 7/2003). Bonn: HRK, Projekt Qualitätssicherung
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Beltz.
- Kromrey, H. (1995). Evaluation in der Lehre durch Umfragenforschung? Methodische Fallstricke bei der Messung der Lehrqualität durch Befragung von Vorlesungsteil-

- nehmern. In P. P. Mohler (Hrsg.), *Universität und Lehre. Ihre Evaluation als Herausforderung an die empirische Sozialforschung* (S: 105-127). Münster: Waxmann.
- Kromrey, H. (2001): Evaluation von Lehre und Studium. Anforderungen an Methodik und Design. In: C. Spiel (Hrsg.): *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S: 21-60). Münster: Waxmann,
- Locke, E. A. & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ: Prentice Hall.
- 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, 1187-1197.
- Marsh, H. W. (1987). Students' Evaluations of University Teaching: Research Findings, Methodological Issues and Directions for Future Research. *International Journal of Educational Research*. Vol. 11, 253-388.
- McKeachie, W. J. (1996). Do We Need Norms of Student Ratings to Evaluate Faculty? *Instructional Evaluation and Faculty Development*, 14, 14-17.
- Penny, A. R. & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research*, 74, 215-253.
- Pintrich, P. R. & Schunk, D. H. (2002). *Motivation in education: Theory, research and application* (2nd ed.). Englewood Cliffs, NJ: Merrill Prentice Hall.
- Rheinberg, F. (1998). Bezugsnormorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 39-43). Weinheim: Beltz.
- Rindermann, H. (2001a). *Lehrevaluation*. Landau: Verlag Empirische Pädagogik.
- Rindermann, H. (2001b). Die studentische Beurteilung von Lehrveranstaltungen - Forschungsstand und Implikationen. In C. Spiel (Hrsg.), *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck* (S. 61-88). Münster: Waxmann.
- Rindermann, H. & Kohler, J. (2003). Lässt sich Lehrqualität durch Evaluation und Beratung verbessern? Überprüfung eines Evaluations-Beratungs-Modells. *Psychologie in Erziehung und Unterricht*, 50, 71-85.
- Spiel, C. (Hrsg.) (2001). *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster: Waxmann.
- Spiel, C. (2004). Kommentar - Empfehlungen zur Ausgestaltung der internen Evaluation. In Hochschulrektorenkonferenz (Hrsg.), *Metaevaluation - Evaluation von Studium und Lehre auf dem Prüfstand* (Beiträge zur Hochschulpolitik 5/2004, S. 55-63). Bonn: Hochschulrektorenkonferenz.

- Spiel, C. & Gössler, P. M. (2000). Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende. *Zeitschrift für Pädagogische Psychologie*, 14, 38-47.
- Webler, W.-D. (1992). Evaluation der Lehre: Praxiserfahrungen und Methodenhinweise. In D. Grünh & H. Gattwinkel (Hrsg.), *Evaluation von Lehrveranstaltungen. Überfrachtung eines sinnvollen Instruments* (S. 143-161)? Berlin: FU-Dokumentationsreihe.
- Wittenberg, R. & Rothe, T. (1999). Studienabbruch sowie Studienfach- und Studienortwechsel an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Erlangen-Nürnberg. In M. Schröder-Gronostay & H.-D. Daniel (Hrsg.), *Studienerfolg und Studienabbruch. Beiträge aus Forschung und Praxis* (S. 105-131). Neuwied; Kriftel: Luchterhand (Hochschulwesen),
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation*. Bern: Huber.

Anhang A: Fehler bei der scannergestützten Datenerfassung mit EvaSys

Im Rahmen des Testlaufs des Evaluationssystems *EvaSys* sollten auch Aussagen darüber getroffen werden, in welchem Umfang Fehler bei der scannergestützten Datenerfassung auftreten.

Von Interesse sind insbesondere die *Nichtererkennung* von Antworten (wenn beispielsweise „knapp daneben“ gekreuzte Antworten als fehlende Antworten gewertet werden) und die *Fehlererkennung* von Antworten (wenn Antworten der falschen Antwortoption zugeordnet werden). Im Anschluss daran stellt sich die Frage, ob diese Fehler vollkommen zufällig oder in Abhängigkeit der Bewertungen der Lehrveranstaltungen auftreten (z.B. könnte es eine – gewagte – Hypothese sein, dass Studierende, die die Veranstaltung schlecht bewerten und negative Emotionen gegenüber dieser hegen, den Fragebogen weniger sorgfältig ausfüllen und häufiger neben die Antwortkästchen kreuzen, was dann zur Folge hätte, dass vor allem bei negativer Veranstaltungsbeurteilung eine erhöhte Anzahl an Nichterkennungen auftritt und damit die Ergebnisse systematisch verzerrt werden). Erwartet wurde, dass etwaige Fehler bei der scannergestützten Datenerfassung unabhängig von den abgegebenen Beurteilungen auftreten.

Die $N=117$ Fragebögen einer per Zufall ausgewählten Vorlesung wurden sowohl scannergestützt mit *EvaSys* als auch per Hand erfasst. Aus den Anteilen fehlender Werte und der Gesamtübereinstimmung der Daten lassen sich Schlüsse im Hinblick auf Nicht- und Fehlerkennungen ziehen.

Vergleich der Quoten fehlender Werte

Im ersten Schritt wurde für beide Datenerfassungsmethoden die Anzahl der fehlenden Werte im resultierenden Datensatz gezählt (nur echte fehlende Werte, d.h. ohne die Antwortoption „weiß nicht“). Die Ergebnisse sind in der folgenden Tabelle aufgeführt.

Anzahl fehlender Werte bei manueller Datenerfassung und scannergestützter Datenerfassung mit *EvaSys*

Art der Datenerfassung	Mittlere Anzahl fehlender Werte pro Fragebogen	Anteil an der Gesamtanzahl der 38 Fragen
Manuell	0.79 (s=1.88)	2.1%
scannergestützt mit <i>EvaSys</i>	1.74 (s=2.49)	4.6%

Anmerkung: $N=117$. s=Standardabweichung.

Da davon ausgegangen kann, dass bei der manuellen Dateneingabe nur solche Fragen als fehlend gewertet werden, bei denen die befragte Person tatsächlich kein Kreuz gesetzt hat, ist aus diesen Ergebnissen abzuleiten, dass bei der scannergestützten Datenerfassung etwa eine abgegebene Antwort pro Fragebogen nicht erkannt wird. Dies entspricht einer Erhöhung des Anteils fehlender Werte um etwa 2.5%. (Der Fragebogen umfasste 38 Items, wobei die sieben Besuchsgründe als separate Items gewertet wurden.)

Dieser Unterschied zwischen den beiden Methoden der Datenerfassung erwies sich als statistisch signifikant ($t(116)=6.609$; $p<.001$). Die Größe des Effekts beläuft sich auf $d=0.43$ Standardabweichungen. Bei der Nichterkennung abgegebener Antworten bei der scannergestützten Datenerfassung handelt es sich damit um einen nicht-vernachlässigbaren, aber dennoch eher kleinen Effekt.

Übereinstimmung der Daten

Anhand aller von den Studierenden abgegebenen Antworten wurde die Übereinstimmung der mit den beiden Methoden erfassten Daten überprüft. Bei 117 ausgewerteten Fragebögen und 38 im Fragebogen enthaltenen Items ergeben sich 4446 Einzelantworten. 4326 dieser Einzelantworten wurden mit manueller und scannergestützter Datenerfassung identisch erfasst. Dies entspricht einer prozentualen Übereinstimmung von 97.3%. Werden die oben genannten fehlenden Werte nicht berücksichtigt, erhöht sich die prozentuale Übereinstimmung auf 99.7%.

Da davon ausgegangen werden kann, dass auch bei der manuellen Dateneingabe Antworten fehlerhaft eingegeben werden, können beide Werte als sehr befriedigend gelten und deuten darauf hin, dass sich die Fehlererkennung bei der scannergestützten Datenerfassung in sehr engen Grenzen hält.

Potenzielle Ergebnisverzerrungen durch Erfassungsfehler

Im letzten Analyseschritt wurde die Annahme überprüft, dass Nicht- und Fehlererkennung bei der scannergestützten Datenerfassung völlig unabhängig von der Beurteilung der Lehrqualität (mithin: zufällig) auftreten.

Für die 28 intervallskalierten Fragen wurde dazu eine Reihe von ungepaarten t -Tests durchgeführt (Art der Datenerfassung als Gruppierungsvariable). Die Ergebnisse erbrachten wie erwartet, dass für keine der 28 Fragen Unterschiede zwischen den beiden Erfassungsmethoden bestanden ($t(s,dfs>104)<0.222$; $ps>.825$). Die standardisierten Mittelwertdifferenzen waren mit $ds<0.03$ praktisch Null. Für die beiden ordinalskalierten Fragen nach dem Fachsemester und der Nachbearbeitungszeit wurden analoge Mann-Whitney- U -Tests durchgeführt. Die Ergebnisse erbrachten auch hier, dass keine Unterschiede bestanden ($ps>0.985$). Schließlich wurden für die nominalskalierten Fragen nach dem Studiengang, den Besuchsgründen und dem Vorhandensein eines Skripts χ^2 -Tests angewendet. Auch hier zeigte sich, dass keine signifikanten Ergebnisverzerrungen eintraten ($\chi^2s(1)<0.456$; $ps>0.500$).

Zusammenfassende Bewertung

Die Analyseergebnisse zeigen, dass bei der scannergestützten Datenerfassung Fehler durch die Nichterkennung abgegebener Antworten auftraten, die in einer Erhöhung des Anteils fehlender Werte um etwa 2.5% resultieren. Dagegen traten keine nennenswerten Fehler durch Fehlererkennung auf. Weiterhin zeigte sich, dass beide Fehlerarten zusammen

genommen vollkommen zufällig im Hinblick auf die abgegebenen Bewertungen der Lehrveranstaltung auftraten und damit die Evaluationsergebnisse nicht verzerrten.

Insgesamt handelt es sich bei den Fehlern der scannergestützten Datenerfassung damit um eher geringe „Kosten“, die den Nutzen, der aufgrund des wesentlichen Effizienzgewinns als sehr groß einzuschätzen ist, kaum relativieren.