

Qualität in der Wissenschaft

Zeitschrift für Qualitätsentwicklung in
Forschung, Studium und Administration

Lehrveranstaltungsevaluation – zwischen methodischem Anspruch, Partizipation und Wirkung

- Studentische Evaluation von Lehre –
Eine themenorientierte Bestandsaufnahme
der wissenschaftlichen Literatur
- Die Evaluation von Lehrveranstaltungen an der
Medizinischen Hochschule Hannover
- Modellbasierte Lehrevaluation:
Konzept und empirische Ergebnisse
- Wie valide sind studentische Lehrveranstaltungsbewertungen?
Sachfremde Einflüsse, studentische Urteilerstandards, Selektionseffekte
- Elemente der Lehrveranstaltungsevaluation
an der Universität Koblenz-Landau:
Theoretische Einordnung und empirische Befunde
- Zur Wirkung hochschuldidaktisch fundierter
und in den Fachbereichen verankerter
Lehrveranstaltungsevaluation
- Evaluationsverfahren als Ausgangspunkt für
Diskussions- und Reflexionsprozesse.
Erfahrungen mit dem Aufbau eines hochschulweiten
Qualitätsmanagements an der Universität Hildesheim

2+3
2014

Herausgeberkreis

Hans-Dieter Daniel, Dr., Professor für Sozialpsychologie und Hochschulforschung, ETH Zürich (CH), Leiter der Evaluationsstelle der Universität Zürich

Susan Harris-Hümmert, Dr., Qualitätsbeauftragte, Referat Planung und Qualitätsmanagement, Universität Würzburg

Michael Heger, Prof. Dr., Geschäftsführer des Zentrums für Hochschuldidaktik und Qualitätsentwicklung in Studium und Lehre (ZHQ), Fachhochschule Aachen

Stefan Hornbostel, Dr., Professor für Soziologie (Wissenschaftsforschung), Institut für Sozialwissenschaften der Humboldt-Universität zu Berlin, Leiter des Instituts für Forschungsinformation und Qualitätssicherung (iFQ), Berlin

René Krempkow, Dr., Forschungsinstitut für Bildungs- und Sozialökonomie (FIBS), Berlin

Lukas Mitterauer, Dr., stellvertretender Leiter der besonderen Einrichtung für Qualitätssicherung, Universität Wien

Philipp Pohlenz, Dr., Professur für Hochschulforschung und Professionalisierung der akademischen Lehre an der Humanwissenschaftlichen Fakultät der Otto-von-Guericke-Universität Magdeburg

Uwe Schmidt, Dr., Leiter des Zentrums für Qualitätssicherung und -entwicklung der Universität Mainz, Geschäftsführer des Hochschulevaluationsverbundes Südwest

Wolff-Dietrich Webler, Dr., ehem. Professor of Higher Education, University of Bergen (Norway), Leiter des Instituts für Wissenschafts- und Bildungsforschung Bielefeld (IWBB)

Don Westerheijden, Dr., Center for Higher Education Policy Studies (CHEPS), University of Twente, Netherlands

Hinweise für die Autoren

In dieser Zeitschrift werden i.d.R. nur Originalbeiträge publiziert. Sie werden doppelt begutachtet. Die Autor/innen versichern, den Beitrag nicht zu gleicher Zeit an anderer Stelle zur Publikation angeboten zu haben. Beiträge werden nur dann angenommen, wenn die Autor/innen den Gegenstand nicht in vergleichbarer Weise in einem anderen Medium behandeln. Senden Sie bitte das Manuskript als Word-Datei und Abbildungen als JPG-Dateien per E-Mail an die Redaktion (Adresse siehe Impressum).

Wichtige Vorgaben zu Textformatierungen und beigefügten Fotos, Zeichnungen sowie Abbildungen finden Sie in den „Autorenhinweisen“ auf unserer Verlags-Homepage: „www.universitaetsverlagwebler.de“.

Ausführliche Informationen zu den in diesem Heft aufgeführten Verlagsprodukten erhalten Sie ebenfalls auf der zuvor genannten Verlags-Homepage.

Impressum

Anschrift Verlag, Redaktion, Abonnentenverwaltung:

UVW UniversitätsVerlagWebler
Der Fachverlag für Hochschulthemen
Bünder Straße 1-3 (Hofgebäude), 33613 Bielefeld
Tel.: 0521-92 36 10-12, Fax: 0521-92 36 10-22

Satz: UVW, info@universitaetsverlagwebler.de

Anzeigen: Die Zeitschrift „Qualität in der Wissenschaft“ veröffentlicht Verlagsanzeigen, Ausschreibungen und Stellenanzeigen. Aufträge sind an den Verlag zu richten. Die jeweils gültigen Anzeigenpreise sind der Homepage erhalten Sie auf Anfrage beim Verlag.

Redaktionsschluss dieser Ausgabe: 15.08.2014

Umschlaggestaltung: Wolff-Dietrich Webler, Bielefeld
Gesetzt in der Linotype Syntax Regular

Druck: Sievert Druck & Service GmbH,
Potsdamer Str. 190, 33719 Bielefeld

Abonnement/Bezugspreis ab 2014:

Jahresabonnement: 72 Euro zzgl. Versandkosten
Einzelpreis: 18,25 Euro zzgl. Versandkosten

Abo-Bestellungen und die Bestellungen von Einzelheften sind unterschrieben per Post oder Fax bzw. per E-Mail an den Verlag zu richten. Eine Abo-Bestellvorlage finden Sie unter www.universitaetsverlagwebler.de.

Das Jahresabonnement verlängert sich automatisch um ein Jahr, wenn es nicht 6 Wochen vor Jahresende gekündigt wird.

Erscheinungsweise: 4mal jährlich

Copyright: UVW UniversitätsVerlagWebler

Die mit Verfasseramen gekennzeichneten Beiträge geben nicht in jedem Falle die Auffassung der Herausgeber bzw. Redaktion wieder. Für unverlangt eingesandte Manuskripte/Rezenzionsexemplare wird keine Verpflichtung zur Veröffentlichung/Besprechung übernommen. Sie können nur zurückgegeben werden, wenn dies ausdrücklich gewünscht wird und ausreichendes Rückporto beigefügt ist. Die Urheberrechte der hier veröffentlichten Artikel, Fotos und Anzeigen bleiben bei der Redaktion. Der Nachdruck ist nur mit schriftlicher Genehmigung des Verlages gestattet.

Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Qualität in der Wissenschaft

**Zeitschrift für Qualitätsentwicklung in
Forschung, Studium und Administration**

Einführung des geschäftsführenden Herausgebers

II

*Susanne Weis, Christiane Karthaus
& Tanja Lischetzke*

Elemente der Lehrveranstaltungsevaluation
an der Universität Koblenz-Landau:
Theoretische Einordnung und empirische Befunde

61

Qualitätsforschung

Rüdiger Mutz & Hans-Dieter Daniel
Studentische Evaluation von Lehre –
Eine themenorientierte Bestandsaufnahme
der wissenschaftlichen Literatur

34

Volkhard Fischer
Die Evaluation von Lehrveranstaltungen an der
Medizinischen Hochschule Hannover

41

*Uwe Schmidt, Nicole Becker, Holger Lübke,
Marc Hümpfner, Franziska Schmidt, Natalia Yugay
& Cathrin Neßler*

Modellbasierte Lehrevaluation:
Konzept und empirische Ergebnisse

47

Tobias Wolbring
Wie valide sind studentische
Lehrveranstaltungsbewertungen?
Sachfremde Einflüsse, studentische Urteilerstandards,
Selektionseffekte

56

Qualitätsentwicklung/-politik

Jörg Jörissen & Michael Heger
Zur Wirkung hochschuldidaktisch fundierter
und in den Fachbereichen verankerter
Lehrveranstaltungsevaluation

70

Benjamin Ditzel
Evaluationsverfahren als Ausgangspunkt für
Diskussions- und Reflexionsprozesse.
Erfahrungen mit dem Aufbau eines hochschulweiten
Qualitätsmanagements an der Universität Hildesheim

81

Seitenblick auf die Schwesterzeitschriften

Hauptbeiträge der aktuellen Hefte
Fo, HSW, HM, P-OE und ZBS

IV

2+3 | 2014

Die Evaluation der Lehre umfasst unterschiedliche Blickrichtungen und Ansätze, die von der institutionellen Evaluation ganzer Hochschulen, Fachbereiche und Fächer, über die Betrachtung von Studiengängen und Modulen bis hin zur Ebene der Lehrveranstaltungen reichen. Betrachtet man die Häufigkeit der eingesetzten Instrumente, so kommt der Lehrveranstaltungsevaluation nach wie vor eine große Bedeutung zu. Nicht nur die Anzahl an durchgeführten Befragungen, die zuweilen eine gewisse Evaluationsmüdigkeit bei allen Beteiligten erzeugt, sondern auch die Vielzahl an unterschiedlichen Erhebungsinstrumenten erwecken zumindest den Eindruck, dass die Lehrveranstaltungsevaluation in den Hochschulen als wesentliches Element zur Messung der Lehrqualität angesehen wird. Diese – durchaus kritisch zu wertende – hohe Bedeutungszuschreibung korrespondiert häufig mit der Praxis, dass die Instrumente der Lehrveranstaltungsevaluation auf unterschiedlichen Ebenen der Hochschule im Sinne des Wortes verhandelt werden. Wenngleich die Befassung mit methodischen Aspekten in den vergangenen Jahren immer mal wieder an Bedeutung gewonnen hat, lässt sich mit Heiner Rindermann tendenziell auch heute noch schließen, dass „Evaluationsmaßnahmen wissenschaftlicher Institutionen [...] sich erstaunlicherweise oft gerade durch fehlende Rezeption wissenschaftlicher Forschung auszeichnen“; die Evaluation an Hochschulen geht praktizistisch vor, gesetzliche Aufgaben werden verwaltungskorrekt erfüllt [...]“ (Rindermann 2003, S. 234)¹. Dies impliziert unterschiedliche Praktiken im Hinblick auf die Entwicklung von Erhebungsinstrumenten als auch auf den Umgang mit Ergebnissen aus Lehrveranstaltungsevaluation.

Die vorliegende Ausgabe der QiW schließt hieran an und fasst Beiträge zusammen, in denen unterschiedliche theoretische Ansätze, methodische Prämissen und Formen der Rezeption und Ergebnisverwertung von Lehrveranstaltungsevaluationen behandelt werden.

Zu Beginn gibt der Beitrag von *Rüdiger Mutz* und *Hans-Dieter Daniel* einen komprimierten Überblick zur Entwicklung der Lehrveranstaltungsevaluation sowie zum aktuellen Forschungsstand. Grundlage hierzu ist eine umfangreiche Literaturanalyse, auf deren Grundlage unterschiedliche Funktionen und Modelle der studentischen Evaluation vorgestellt und methodische Fragen diskutiert werden. Hierbei wird sowohl auf Aspekte der Fragebogenkonstruktion als auch auf Fragen der Güte der studentischen Bewertung und die Relevanz der Ergebnisse für die Qualitätsentwicklung eingegangen. **Seite 34**

Volkhard Fischer resümiert die Erfahrungen im Bereich der Lehrveranstaltungsevaluation an der Medizinischen Hochschule Hannover der vergangenen Jahre. Von besonderem Interesse sind hierbei die Ergebnisse zum Einsatz eines zweistufigen Verfahrens, im Rahmen dessen zunächst ein modellgeleiteter Kurzfragebogen eingesetzt wird, dem bei unterdurchschnittlichen Resultaten eine vertiefende Analyse folgen kann. Zudem betrachtet Fischer die besonderen Anforderungen, die an ein Verfahren gestellt werden, wenn es sowohl als Feedback-Instrument als auch als ein Teil der leistungsorientierten Mittelverteilung eingesetzt wird. **Seite 41**



Uwe Schmidt

Uwe Schmidt, Nicole Becker, Holger Lübke, Marc Hümpfner, Franziska Schmidt, Natalia Yugay und Cathrin Neßler rekurren auf die Erfahrungen an der Universität Mainz und die dort eingesetzten Instrumente zur Lehrveranstaltungsevaluation. Ausgehend von einem systemtheoretischen Modell, das Qualität als Balance unterschiedlicher Anforderungen an die Ausgestaltung der Lehr-Lernbeziehung definiert, überprüfen sie die Tragfähigkeit des Modells mithilfe von Faktoren- und Regressionsanalysen. Als zentrale Dimensionen interpretieren sie in Anlehnung an das AGIL-Modell Talcott Parsons' die Ebenen der Kultur bzw. latenten Struktur, der Lehr- und Lernprozesse, des Zielerreichungsgrades und der vorhandenen Ressourcen bzw. der Voraussetzungen für die Gestaltung erfolgreichen Lehrens und Lernens. **Seite 47**

Tobias Wolbring widmet sich der Frage, ob und inwieweit Lehrveranstaltungsevaluationen eine valide Basis für ihre Nutzung im Rahmen von Anreizsystemen – wie die leistungsorientierte Mittelverteilung – bieten. Entlang drei „methodischer Fallstricke“, der sachfremden Urteile, der Urteilsstandards und der Selektionseffekte untersucht er die Validität studentischer Bewertungen von Lehrveranstaltungen und nimmt hierbei u.a. die Wirkung der Attraktivität des Lehrpersonals, geschlechtsspezifische Effekte, die durch die Studierenden genutzten Beurteilungsmaßstäbe, und Effekte aufgrund des Abbruchs der Veranstaltung in den Blick. **Seite 56**

Susanne Weis, Christiane Karthaus und Tanja Lischetzke gehen von Erfahrungen mit der Evaluation von Lehrveranstaltungen nach dem Trierer Inventar zur Lehrveranstaltungsevaluation (TRIL) an der Universität Koblenz-Landau aus und erweitern den Ansatz der Evaluation im engeren Sinne um Perspektiven des Workloads und der Lernzielerreichung. Hiervon ausgehend gelangen sie zu Ergebnissen, die nicht nur auf Fragen der Zufriedenheit mit Lehrveranstaltungen, sondern auch auf Aspekte des Studienerfolgs – wie den Zusammenhang zwischen Lernzeit oder objektiven Kompetenzen mit dem Lernerfolg – abstellen. **Seite 61**

Jörg Jörissen und Michael Heger fokussieren in ihrem Beitrag neben den für die Lehrveranstaltungsevaluation

¹ Rindermann, H. (2003): Lehrevaluation an Hochschulen – Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und Evaluation. Zeitschrift für Evaluation.

entwickelten Konstrukten und empirischen Erfahrungen an der Fachhochschule Aachen vor allem auf die Wirkungen und den Umgang mit Ergebnissen der Evaluation. Sie gelangen hierbei zu dem Ergebnis, dass die Entwicklung von methodisch fundierten Fragebogen alleine nicht ausreicht, um zu einer qualitativen Verbesserung der Lehre zu gelangen. Vielmehr – so die Autoren – sind unterschiedliche zusammenwirkende Erhebungsverfahren und Instrumente notwendig, um einen qualitativen Mehrwert aus Evaluationen zu erlangen bzw. um Evaluationen nicht zu einem Selbstzweck werden zu lassen, die keine Wirkung entfalten.

Seite 70

Benjamin Ditzel nimmt in seinem Beitrag „Evaluationsverfahren als Ausgangspunkt für Diskussions- und Reflexionsprozesse. Erfahrungen mit dem Aufbau eines hochschulweiten Qualitätsmanagements an der Universität Hildesheim“ ebenfalls die aus Qualitätssicherungsverfahren resultierenden Konsequenzen in den Blick und stellt in grundlegender Form die Frage nach Möglichkeiten und Grenzen der Hochschulsteuerung. Er fokussiert hierbei insbesondere auf die hohe Bedeutung der Selbstregulation in Organisationen und die Bedeutung von Feedbackmechanismen.

Seite 81

Uwe Schmidt

Leichter Zugang für Sie zur Expertise!

Bei 6 Zeitschriften im Themenfeld Wissenschaft und Hochschulen, die der UVW herausbringt, sammelt sich in kürzester Zeit eine erhebliche Expertise an.

Wir veröffentlichen 110 bis 120 Aufsätze pro Jahr. Da verlieren Leserinnen und Leser bei der Fülle schon mal leicht den Überblick. Wer weiß noch, was der Jahrgang 2010 in der Zeitschrift Hochschulmanagement für Themen bereit hielt? Seit Gründung hat die Zeitschrift „Qualität in der Wissenschaft (QiW)“ bisher rd. 120 Artikel publiziert – sorgfältig (i.d.R. doppelt) begutachtet. Ähnlich auch die anderen.

Daher bieten wir die Artikel aller unserer Zeitschriftenjahrgänge, die älter als zwei Jahre sind, *kostenlos* zum Herunterladen an.

Auf unserer Homepage finden Sie sie, wie unten angegeben.

Das Hochschulwesen (HSW)

<http://hochschulwesen.info/inhaltsverzeichnisse.html>

Forschung. Politik – Strategie – Management (FO)

<http://www.universitaetsverlagwebler.de/Forschung.html>

Zeitschrift für Beratung und Studium (ZBS)

<http://www.universitaetsverlagwebler.de/ZBS.html>

Qualität in der Wissenschaft (QiW)

<http://www.universitaetsverlagwebler.de/QiW.html>

Hochschulmanagement (HM)

<http://www.universitaetsverlagwebler.de/HM.html>

Personal- und Organisationsentwicklung in Einrichtungen der Lehre und Forschung (P-OE)

<http://www.universitaetsverlagwebler.de/P-OE.html>

Unser Gesamtangebot an Heften, Büchern und Zeitschriften finden Sie unter

<http://www.universitaetsverlagwebler.de>

Rüdiger Mutz & Hans-Dieter Daniel

Studentische Evaluation von Lehre – Eine themenorientierte Bestandsaufnahme der wissenschaftlichen Literatur



Rüdiger Mutz



Hans-Dieter Daniel

1. Problemstellung

Die studentische Evaluation der Lehre (SEL) gehört mittlerweile zum festen Bestand des Qualitätsmanagements vieler Hochschulen in Europa. Die große Verbreitung der SEL ist sicherlich auch auf den Bologna-Prozess zurückzuführen und den damit zusammenhängenden Bestrebungen, die Qualität von Bachelor- und Master-Studiengängen durch interne Evaluationen und Akkreditierungen zu sichern. Mit der Verbreitung der SEL nicht nur in Europa, sondern weltweit ist auch die Forschungsliteratur zu diesem Thema in den letzten 15 Jahren rasant angestiegen. So sind in der Datenbank ERIC insgesamt 1717 Publikationen in Fachzeitschriften allein dem Thesaurus-Begriff „course evaluation“ zugeordnet, mehr als die Hälfte dieser Publikationen stammen aus den Jahren 2000 bis 2014 (Abruf 17.3.2014). Diese Literatur und die damit verbundene empirische Evidenz werden von den Hochschulen meist nicht zur Entwicklung von Qualitätssicherungsinstrumenten für die Lehre herangezogen.

Vor diesem Hintergrund sind Überblicksartikel wichtig, die die Ergebnisse von Arbeiten auf dem jeweiligen Gebiet im Sinne eines State of the Art zusammenfassen. Solche Übersichten können von Hochschulen als Referenzrahmen zur Beurteilung ihrer eigenen Qualitätssicherungsinstrumente verwendet werden, z.B. ob die in einer Hochschule eingesetzten Fragebögen tatsächlich den in der Forschungsliteratur diskutierten Qualitätsstandards (z.B. Dimensionen, testtheoretische Gütekriterien) entsprechen. Während Reviews versuchen den Forschungsstand qualitativ zusammenzufassen, liefern Metaanalysen eine Zusammenfassung von quantitativen Forschungsergebnissen zu ausgewählten Fragestellungen (z.B. Validität der SEL).

2. Zielsetzung und methodischer Ansatz

Ziel dieses Beitrags ist es, die wissenschaftliche Literatur zu wichtigen Themenbereichen der SEL überblicksartig darzustellen. Die Literaturanalyse wurde im Rahmen eines Metaevaluationsprojektes an der ETH Zürich erstellt (Mittag/Mutz/Daniel, 2012) und für diese Publikation aktualisiert. Da die Anzahl von Zeitschriftenaufsätzen zum veröffentlichten Thema „Lehrveranstaltungsbeurteilung“ ein Ausmaß erreicht, das im Rahmen eines qualitativen Literaturreviews nicht mehr zu bewältigen

ist, behandelt die Literaturanalyse vor allem zentrale Übersichtsarbeiten, Review-Artikel und Monografien, die in den letzten fünfzehn Jahren veröffentlicht wurden. Im Rahmen einer Literaturrecherche wurde zum einen in den Datenbanken WISO, PSYCLIT, PSYINDEX, WEB OF SCIENCE und ERIC mittels verschiedener Schlagwörter (Lehrevaluation, course evaluation, student evaluation of teaching or instruction, student rating of instruction) nach einschlägigen Veröffentlichungen gesucht. Zum anderen wurden für wichtige hochzitierte Zeitschriftenartikel innerhalb von WEB OF SCIENCE neuere Arbeiten recherchiert, die diese hochzitierten Arbeiten in ihren Literaturlisten aufführen. Die recherchierte Literatur wurde im Hinblick auf international gültige fachliche Standards für Lehrveranstaltungsbeurteilungen ausgewertet.

3. Ergebnisse der Literaturanalyse

Die Literaturrecherche zeigt, dass in den letzten 15 Jahren eine Reihe von Übersichtsarbeiten publiziert wurden, die den aktuellen internationalen Forschungsstand im Bereich der studentischen Lehrveranstaltungsbeurteilung (Lehrevaluation, course evaluation, student evaluation of teaching or instruction, student rating of instruction) zusammenfassen und Hinweise geben, wie Lehrveranstaltungsbeurteilungen gestaltet werden sollten. Neben den klassischen Arbeiten von Daniel (1994, 1997, 1998), D'Appollonia und Abrami (1997), Greenwald (1997), Marsh (1983, 1984, 1987, 2007), Marsh and Roche (1997, 2000) und McKeachie (1997), deren Befunde immer noch Bestand haben und in nachfolgenden Arbeiten aufgenommen bzw. repliziert wurden, in den letzten Jahren eine Reihe von Übersichtsarbeiten zu Lehrveranstaltungsbeurteilungen erschienen, so von Apodaca und Grad (2006), Bargel und el Hage (2000), Berendt (2000), Chen und Hoshower (2003), Diehl (2003), Kromrey (2001), Lang und Kersting (2007), Marsh, Ginns, Morin, Nagengast und Martin (2011), McAlpine und Harris (2002), Mutz (2001, 2003), Mutz und Daniel (2006, 2008), Rindermann (2001a, 2001b, 2003), Rindermann und Kohler (2007), Perry und Smart (2007), Schmidt und Loßnitzer (2010), Schmidt und Tippelt (2005), Souvignier und Gold (2002, 2003), Spiel (2001a, 2001b), Spooren, Brockx und Mortelmans (2013), Webler (2008). In der internationalen Forschungsliteratur hat sich in den letzten Jahren der Be-

griff „student evaluation of teaching“ (SET) durchgesetzt, der im Deutschen mit „studentischer Lehrevaluation“ oder kurz „Lehrevaluation“ übersetzt werden kann.

Im Folgenden wird der Forschungsstand zu verschiedenen Themenbereichen der Lehrevaluation, die für die Qualitätssicherung an Hochschulen von besonderer Wichtigkeit sind, zusammengefasst:

3.1 Zweck und Funktion von studentischen Lehrevaluationen

Nach Marsh (1987, 2007) können Lehrevaluationen eine oder mehrere der folgenden Funktionen erfüllen:

- (1) „*Rückmeldung*“: diagnostisches Feedback zum Lehrangebot für die Lehrenden/Fakultäten/Departemente,
- (2) „*Steuerung*“: ein Maß für die Lehreffektivität als Grundlage für administrative Entscheidungen,
- (3) „*Kurswahl*“: Informationsgrundlage für Studierende für die Wahl von Lehrveranstaltungen,
- (4) „*Zufriedenheit mit der Lehrveranstaltung*“: ein Maß für die Zufriedenheit mit einer Lehrveranstaltung als Grundlage für die Curriculums-Entwicklung,
- (5) „*Forschung*“: für den Einsatz in der Lehrevaluationsforschung.

Für Souvignier und Gold (2002) ist es entscheidend für die Entwicklung eines Verfahrens zur studentischen Lehrveranstaltungsbeurteilung, welche Funktionen dieses Verfahren erfüllen soll. So müssen nach Souvignier und Gold (2002) höhere Anforderungen an das Instrumentarium gestellt werden, wenn es neben der Rückmeldung auch zur Steuerung eingesetzt werden soll. So müssten die forschungsmethodischen Standards empirischer Sozialforschung stärker eingehalten werden, das Bedingungsgefüge der Lehrqualität breiter abgebildet und die messtheoretischen Voraussetzungen erfüllt werden. In dem Sinne definieren Rossi und Freeman (1988, S. 5, zitiert nach Mutz 2001, S. 12) folgendermaßen: „Evaluation ist die systematische Anwendung von sozialwissenschaftlichen Methoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme“. Nach Schmidt und Timpelt (2005) erfolgt die Lehrevaluation auf drei Ebenen: a) der Mikroebene der einzelnen Lehrveranstaltung, b) der Mesoebene des Curriculums und des Fachbereichs und c) der Makroebene von Übergangs- und Verbleibsstudien. Die Evaluation einer Lehrveranstaltung erfasst zumeist nur Aspekte auf der Mikroebene.

Wenn Studierende danach befragt werden, was sie motiviert, an Lehrevaluationen teilzunehmen, so wird nach Chen und Hoshower (2003) die Verbesserung der Lehre als attraktivstes Ergebnis eines Evaluationssystems genannt sowie die Verbesserung der didaktischen Ausgestaltung und des Inhalts der Lehrveranstaltung. Dagegen ist die Möglichkeit, Informationsgrundlage für die Auswahl von Lehrveranstaltungen oder für die Beförderung bzw. das Gehalt der Dozierenden zu geben, für Studierende nicht entscheidend. Die Beteiligung an Lehrevaluationen wird darüber hinaus stark motiviert durch das Bewusstsein, ein bedeutsames Feedback geben zu können.

3.2 Lehrqualität und Modelle der Lehrevaluation

Mit der Durchführung von Lehrevaluationen verbinden sich unterschiedliche Vorstellungen über die Merkmale „guter Lehre“, die sich aus Theorien des Unterrichts, Befragungen der Lehrenden oder sehr häufig aus der Analyse bestehender Lehrevaluationsinventare ergeben (Berendt 2000; Rindermann 2003; Webler 1991). In Anlehnung an die Systematisierung von Evaluationskriterien für die Trainingsmaßnahmen von Kirkpatrick (Kirkpatrick 1959; 1960, zitiert nach Staufenberg 2000) lassen sich vier Kriterienarten für die Qualität der Lehre ableiten:

- (1) „*Akzeptanzkriterien*“: Wie positiv oder negativ bewerten Teilnehmerinnen und Teilnehmer die Lehrveranstaltung?
- (2) „*Lernkriterien*“: Haben sich die Studierenden durch die Veranstaltung im Sinne der Lehrziele geändert? Wissen sie mehr?
- (3) „*Transferkriterien*“: Sind die Studierenden in der Lage, das Wissen auch anzuwenden?
- (4) „*Ergebniskriterien*“: Zeigen sich Effekte in objektiven Indikatoren wie z.B. Studiendauer, Berufserfolg?

Innerhalb dieser Systematik werden durch Lehrevaluationen die Akzeptanzkriterien abgebildet. Von verschiedenen Autoren (Bargel/El Hage 2000; Rindermann 2003; Souvignier/Gold 2002) wird häufig ein Defizit an theoretischer Einbindung der Inventare beklagt, um „durch eine umfassende Modellierung der Determinanten des Studienerfolgs zugleich eine theoretische Basis für die Lehrevaluation zu Verfügung zu stellen“ (Souvignier/Gold 2002, p. 267). Vier Modelle können unterschieden werden:

- „*Angebots-Nutzungsmodell des universitären Lernprozesses*“ (Helmke 1996): In diesem Modell wird die Passung zwischen Merkmalen der Lehrveranstaltung (Angebot) und den Lernvoraussetzungen der Studierenden (Nutzung) untersucht.
- „*Multifaktorielles Modell der Lehrveranstaltungsqualität*“ von Rindermann (Rindermann 2001a, 2001b): Für den Erfolg einer Lehrveranstaltung, gemessen beispielsweise am subjektiven Lerngewinn, sind drei Determinanten maßgeblich: auf Seite der Dozierenden sind es u.a. die Strukturierung und der klare Aufbau einer Lehrveranstaltung, aber auch das Engagement der Lehrperson. Auf Seiten der Studierenden sind es Fleiß, Vorwissen und der Grad der Beteiligung und auf Seiten der Rahmenbedingungen sind es u.a. die Leistungsanforderungen, die Besucherzahl und die inhaltlichen und zeitlichen Überschneidungen mit anderen Lehrveranstaltungen.
- „*Person-Umwelt-Passungstheorie der Studienzufriedenheit*“ von Westermann, Spies, Heise und Wollburg-Claar (1998): Studienzufriedenheit wird als subjektive Passung zwischen den wahrgenommenen Anforderungen an die Person und die erlebten Fähigkeiten bzw. als Passung zwischen den persönlichen Bedürfnissen und den Angeboten der Umwelt verstanden.
- „*Kompetenzorientierte Lehrevaluation*“: Während die meisten eingesetzten Lehrevaluationsinstrumente eher Aspekte der Lehrveranstaltung und der Lehrperson erfassen, zielt die kompetenzorientierte Lehrevaluation

auf die Ergebnisse, d.h. auf die Kompetenzen. So erhebt das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp) den Zuwachs an Fach-, Methoden-, Präsentations-, Kommunikations-, Kooperations-, sowie Personal-kompetenzen (Braun 2008; Braun/Gusy/Leidner/Hannover 2008).

Souvignier und Gold (2002) bezeichnen diese theorieorientierte Entwicklung von Fragebögen als die dritte Phase in der Lehrevaluationsforschung, nach einer ersten Entwicklungsphase in den 1970er Jahren und einer kritischen Phase der Implementation in den 1990er Jahren. Insgesamt wird von allen Autoren die multiple Bedingtheit des Lehrerfolgs betont und die Tatsache, dass das Konstrukt „Lehrerfolg“ nicht durch ein einzelnes Kriterium erfasst werden kann (z.B. globale Bewertung einer Lehrveranstaltung).

3.3 Güte des studentischen Urteils

Von Kritikern studentischer Lehrevaluationen wird häufig die Urteilskompetenz der Studierenden in Frage gestellt (Kromrey, 1996, 2001). Die Kritikpunkte können an folgenden Gütekriterien psychologischer Tests in Anlehnung an Marsh (1984, 1987), Morley (2014) und Onwuegbuzie, Daniel und Collins (2009) festgemacht werden:

- **„Reliabilität“:** Mit Reliabilität wird die Messgenauigkeit der Lehrevaluationsbögen bzw. des studentischen Urteils bezeichnet. Diese drückt sich in der Interrater-reliabilität (z.B. Intraklassenkoeffizient ICC) aus, die aufzeigt, ob Studierende die gleiche Lehrveranstaltung sehr homogen, aber unterschiedliche Lehrveranstaltungen ganz unterschiedlich bewerten. Kromrey (2001) weist darauf hin, dass die Heterogenität des studentischen Urteils innerhalb einer Lehrveranstaltung sehr hoch sei. Empirische Studien zeigen, dass die Heterogenität der studentischen Urteile hoch ist, was sich in ICCs von .20 bis .30 (Feldman 1977; Rindermann 2001a, 2001b) ausdrückt. ICCs können Werte von minimal 0.0 und maximal 1.0 annehmen.
- **„Validität/Dimensionalität“:** Es wird häufig an studentischen Lehrevaluationen kritisiert, die erfassten Lehrdimensionen würden keine Zusammenhänge mit Kriterien des Lehrerfolgs aufweisen (Validität). Hier können angloamerikanische Überblicksarbeiten (z.B. Marsh/Roche 2000; McKeachie 1997; Spoor et al. 2013) und deutsche Übersichtsarbeiten (z.B. Rindermann 2003) allerdings Zusammenhänge aufzeigen. So berichtet Rindermann (2003) über Korrelationen von Skalen des HILVE-Lehrevaluationsinventars mit Leistungskriterien (mittlerer Prüfungserfolg) von .52. Der Zusammenhang von Fremdurteilen didaktisch geschulter Personen und Studierendenurteil liegt bei .54. Spoor et al. (2013) fasst im Hinblick auf die konvergente Validität den Stand der Literatur ab dem Jahr 2002 folgendermaßen zusammen: „In summary, the research literature revealed the existence of (small to strong) positive correlation between SET scores and student achievement, expert ratings of teaching behavior, self ratings, and alumni ratings“ (Spoor et al. 2013, p.

609). Das heißt, Lehrevaluationen weisen bedeutsame Zusammenhänge zu hochschuldidaktisch formulierter Lehrqualität auf. Behauptet wird auch, dass Studierende eindimensional urteilen würden und nicht die verschiedenen Facetten einer Lehrveranstaltung beurteilen könnten. Zwar wird in neuerer Zeit ein Generalfaktormodell diskutiert, in dem ein zentraler Bewertungsfaktor (z.B. gut-schlecht) die Urteile bestimmt (Souvignier/Gold 2002). Statistische Analysen (Faktorenanalyse) zeigen über eine Vielzahl von Studien hinweg jedoch immer wieder die Mehrdimensionalität des studentischen Urteils auf (Apodaca/Grad 2006; Rindermann 2001a, 2001b). Insgesamt wird beobachtet, dass Studierende eher zu positiv als zu negativ urteilen, was für die Konstruktion von Skalen und Antwortformaten von Bedeutung ist (Souvignier/Gold, 2002).

- **„Fairness“:** Die Ergebnisse einer Lehrevaluation sind fair, wenn Einflüsse von Variablen auf das studentische Urteil ausgeschlossen werden können, die nichts mit dem Lehrgeschehen selber zu tun haben. Wolbring (2013) behandelt Fragen der Validität und Fairness unter kausalanalytischer Perspektive. Ziel seiner Arbeit war es, methodische Fallstricke der studentischen Evaluation (z.B. Attraktivität, personenspezifische Beurteilungsstandards, Selbstselektion, studentischer Dropout) aufzuzeigen. Urteilsverzerrende Tendenzen lassen sich nach Wolf, Spiel und Pellert (2001) zu vier übergeordneten Bereichen zusammenfassen: Merkmale des Lehrenden (z.B. physische Attraktivität, Sympathie), Merkmale des Studierenden (z.B. Teilnehmermotivation, subjektive Theorien über gute Lehre, Fleiß, soziodemografische Variablen), Umgebungsscharakteristika (z.B. Raummerkmale) und Charakteristika der Studienrichtung (z.B. Größe des Studiengangs). Als zentrale Bias-Variablen werden in der Literatur die Größe der Lehrveranstaltung, das thematische Interesse an der Lehrveranstaltung unabhängig von der Darbietung, die Akquieszenz oder Zustimmungstendenz, die Arbeitsanforderungen, die Teilnehmermotivation und die erwartete Note in der Prüfung genannt. Die Ergebnisse zeigen (Daniel 1994; Klein/Rosar 2006; Marsh 1984, 1987; Rindermann 2001a, 2001b; Spiel 2001a; Spoor/Mortelmans/Thijssen 2007; Wolff et al. 2001), dass lediglich das persönliche Interesse am Thema der Veranstaltung über alle Studien hinweg einen Einfluss auf das studentische Urteil hat. So geben Studierende mit hohem Interesse am Thema häufig bessere Bewertungen in den Lehrveranstaltungsbeurteilungen ab als Studierende mit geringem Interesse am Thema – unabhängig davon, wie die Lehrveranstaltung didaktisch gestaltet ist. Damit eng verbunden ist auch der Besuchsgrund (Pflicht- vs. Wahlfach) als potenzielle Biasvariable zu nennen. Solche Einflüsse müssen gegebenenfalls kontrolliert werden, um die Fairness des Verfahrens zu garantieren.

Folgendes Resümee von Marsh (1987, p. 255) kann immer noch Geltung beanspruchen: „Based upon this overview, class-average student ratings are: (1) multidimensional; (2) reliable and stable; (3) primarily a function

of the instructor who teaches a course rather than the course that is taught; (4) relatively valid against a variety of indicators of effective teaching; (5) relatively unaffected by a variety of variables hypothesized as potential biases; and (6) seen to be useful by faculty as feedback about their teaching, by students for use in course selection, and by administrators for use in personnel decisions."

3.4 Inhaltlicher Aufbau der Fragebögen zur Lehrevaluation

Souvignier und Gold (2002) haben für den deutschsprachigen Bereich die eingesetzten psychometrisch überprüften Instrumente zur Lehrevaluation systematisiert. Entsprechend den Funktionen von Lehrevaluationen lassen sich die Verfahren gruppieren zum einen in Instrumente, bei denen eher die Rückmeldefunktion sowie die Akzeptanz- und Zufriedenheitsurteile im Vordergrund stehen, und zum anderen in Verfahren, die den Anspruch erheben, eine vergleichende Evaluation der Lehre zu ermöglichen (z.B. HILVE, Rindermann 2001b). Inhaltlich werden in den Bögen die Themen „Dozentenverhalten“, „Lerngegenstand“, „Lernverhalten“ und „Rahmenbedingungen“ in unterschiedlicher Weise thematisiert. Von den Dimensionen, die in den ersten Bögen in den 1970er Jahren enthalten waren, haben für Souvignier und Gold die Strukturierung der Veranstaltung, die Art der Stoffvermittlung und das Verhalten gegenüber den Studierenden als wichtige Gegenstände der Beurteilung durch Studierende immer noch Bestand. Wolf, Spiel und Pellert (2001) nennen in ihrem Literaturüberblick die Bereiche oder Dimensionen „Didaktik“, „Struktur und Organisation“, „Klima“ und „Engagement des Lehrenden“ als zentrale Bestandteile von Lehrwirksamkeit in den verschiedenen Fragebögen. Einen Theorie-orientierten Ansatz zur Erstellung eines Fragebogens wählen Souvignier und Gold (2003). Auf der Basis des Lernprozessmodells von Steiner (1996) begründen sie drei zentrale hochschuldidaktisch begründete Merkmalsbereiche guter Lehre:

- (1) „*Motivation und Interesse*“: Steuerung der Aufmerksamkeit, Wecken situationsbezogenen Interesses, Förderung gegenstandsbezogenen Interesses;
- (2) „*Lernen und Instruktion*“: Strukturierung der Lerninhalte, Hinweise zur Einbettung der Inhalte, Vermittlung von Lernstrategien, Durcharbeiten aus mehreren Perspektiven, Üben und Anwenden;
- (3) „*Unterstützende Faktoren*“: soziales Klima, Anforderungsniveau, soziales Lernen.

Weitere Übersichten und Beschreibungen deutschsprachiger Instrumente zur Lehrevaluation finden sich bei Diehl (2003) sowie Schmidt und Loßnitzer (2010). Ein Review der englischsprachigen Instrumente findet sich in Richardson (2005). Richtlinien zur Gestaltung von Fragebögen zur studentischen Evaluation der Lehre sind der Publikation von Kember und Ginns (2012) zu entnehmen.

3.5 Fragebogenkonstruktion

Bei der Fragebogenkonstruktion sind folgende Punkte zu beachten (Mutz 2001; Rindermann 2001a, 2001b;

Seldin 1999; Souvignier/Gold 2003; Wolff et al. 2001):

- „*Deckeneffekte*“: Da Studierende eine Lehrveranstaltung eher zu positiv bewerten, werden die negative Teile einer mehrstufigen Antwortskala z.B. „trifft überhaupt nicht zu“, weniger von den Studierenden verwendet, sodass die Antwortskala nur im positiven Bereich gut differenziert. Nach Wolf, Spiel und Pellert (Wolff et al. 2001) kann man die Antwortskala asymmetrisch gestalten oder die Itemformulierung modifizieren, indem die Aussagen extremer formuliert werden („Der Dozierende geht didaktisch sehr gut vor“). Diese Formulierung kann sprachlich zu Interpretationsproblemen führen, da diese extrem formulierten Aussagen mit dem negativen Ratingpol „trifft nicht zu“ keine eindeutigen Interpretationen erlauben.
- „*Bias-Variablen*“: Um den Einfluss von Bias-Variablen zu testen, sind diesbezügliche Variablen in den Fragebogen aufzunehmen (soziodemografische Variablen, Teilnahmegrund – z.B. Pflicht oder Wahlveranstaltung – Studienstufe, thematisches Interesse, Fleiß etc.).
- „*Antwortskala*“: Sedlmeier (2006) konnte zeigen, dass die spezifische Wahl der Antwortskala eines Fragebogens einen systematischen Einfluss auf das Antwortverhalten der Studierenden hat. Bipolare Skalen (-3 unzufrieden bis +3 zufrieden) erzeugen im Durchschnitt höhere Skalenwerte (z.B. eher zufrieden) als unipolare Skalen (1 unzufrieden 6 zufrieden), da der negative Teil bipolarer Skalen von den Befragten tendenziell weniger genutzt wird. Bei Zeitintervallangaben (z.B. 1-3 h, 4-5 h und mehr als 5 h) orientieren sich die Befragten an der mittleren Kategorie als Referenzwert. Je nach Zeitintervall können unterschiedliche Ergebnisse erzeugt werden (siehe auch Schaefer/Presser 2003).
- „*Modulevaluation oder Lehrveranstaltungsevaluation*“: Im Zuge der Bologna-Reform wird das Studium in der Regel modularisiert, was impliziert, dass neben der einzelnen Lehrveranstaltung auch das Modul als Zusammenfassung mehrerer Lehrveranstaltungen Gegenstand der Lehrevaluation sein kann. Entweder müssen dann modulbezogene Fragebögen entwickelt werden, oder die Evaluationsbögen verschiedener Lehrveranstaltungen über Identifikationscodes der Studierenden zusammengeführt werden.
- „*Offene Fragen*“: In den meisten Fragebögen wird dem Studierenden die Möglichkeit gegeben, Anmerkungen zur Lehrveranstaltung und zum Dozierenden zu machen durch Vorgabe von offenen Feldern (Leitfrage: „Was hat Ihnen an der Lehrveranstaltung besonders gut gefallen? Was hat Ihnen an der Lehrveranstaltung weniger gut gefallen?“). Antworten auf offene Fragen können dem Lehrenden wichtige Hinweise für die unmittelbare Verbesserung seiner Lehre geben.
- „*Länge des Fragebogens*“: Nach der Bestandsaufnahme deutschsprachiger Lehrevaluationsinventare von Souvignier und Gold (2002) lassen sich zwei Arten von Instrumenten der Lehrevaluation im Hinblick auf die Länge unterscheiden: a) kurze Fragebögen mit

16-20 Items, die unterschiedliche Aspekte des Lehrgeschehens abbilden und entweder auf eine umfassende Beschreibung des Lehrgeschehens verzichten oder diese nur mit jeweils einem Item repräsentieren oder b) lange Fragebögen mit 37 bis 55 Items, die beanspruchen, umfassend das Lehrgeschehen abzubilden (Dozentenverhalten, Aufarbeitung des Lehrstoffs, Rahmenbedingungen, Studierendenverhalten etc.), z.B. das Heidelberger Inventar zur Lehr-evaluation (Rindermann 2001a, 2001b). Kürzere Fragebögen haben eine größere Akzeptanz bei den Studierenden.

- „Dimensionen“: Der Fragebogen sollte so aufgebaut sein, dass er unterschiedliche Dimensionen des Lehrgeschehens möglichst mit jeweils 2-3 Items erfasst, die als Sub-skalen fungieren, um die Mehrdimensionalität des Unterrichtsgeschehens abzubilden. Wichtige Dimensionen von Lehr-evaluationsbögen sind: a) Dozentenverhalten und didaktische Kompetenz, b) Anforderungen, c) Strukturierung, d) Klima, e) Studierendenvariablen (Interesse am Thema, Fleiß etc.), f) Rahmenbedingungen und g) Materialien.
- „Globale Urteile“: Im Rahmen von Lehr-evaluationen ist es nicht nur wichtig, einzelne Aspekte einer Lehrveranstaltung bewerten zu lassen, sondern auch globale Urteile zu erheben, z.B. „Wie zufrieden waren Sie mit der Lehrveranstaltung insgesamt“? (Mutz 2001). Darüber hinaus gehen in das Gesamturteil relevante Aspekte des Lehrgeschehens ein, die aufgrund der Kürze des Fragebogens nicht berücksichtigt werden konnten.
- „Online-Befragung“: Ein an der Universität Zürich durchgeführter Pilotversuch zeigte, dass die Rücklaufquote bei den Online-Befragungen um rund 30 Prozentpunkte gegenüber der schriftlichen Befragung sank (von 67% auf 36%) (Mutz/Daniel 2006). Gamliel und Davidoviz (2005) berichteten ähnliche Evaluationsergebnisse.
- „Gütekriterien der Fragebögen“: Die Fragebögen zur Lehr-evaluation sollten die Gütekriterien klassischer psychologischer Tests erfüllen wie Objektivität, Reliabilität, Validität und Fairness (Marsh 1984, 1987; Mutz 2003).

3.6 Effektivität der Lehr-evaluation

Eine ältere immer wieder zitierte Metaanalyse (quantitative Zusammenfassung von wissenschaftlichen Arbeiten zu einem Thema) von Cohen (1980) zeigt, dass studentisches Veranstaltungsfeedback einen kleinen, aber doch bedeutsamen Beitrag zur Verbesserung der Lehre leistet. In einer Übersicht nennen Rindermann, Kohler und Meisenberg (Rindermann et al. 2007) vier Evaluationsmodelle im Hinblick auf die Effektivität der Lehr-evaluation:

- (1) „Aufmerksamkeitsmodell“: Die Implementation von Lehr-evaluation ist für sich genommen schon eine Form von Intervention und bewirkt eine Verbesserung der Lehre.
- (2) „Feedbackmodell“: Das Feedbackmodell geht davon aus, dass die Präsentation von Stärken und Schwächen einer Lehrveranstaltung auf der Basis stu-

dentischer Lehr-evaluationen den Lehrenden motiviert, seine Lehrveranstaltung zu verbessern.

- (3) „Diskussionsmodell“: Das Diskussionsmodell impliziert, dass die Lehrenden mit den Studierenden über die jeweilige Lehrveranstaltung sprechen. Im Rahmen solcher Diskussionen werden Ergebnisse der studentischen Lehr-evaluation berichtet, zusätzliche Informationen gesammelt und mögliche Veränderungen diskutiert und umgesetzt. Alean-Kirkpatrick, Hänni und Lutz (1997) empfehlen, dass solche Diskussionen von hochschuldidaktisch geschultem Personal moderiert werden sollten.
- (4) „Beratungsmodell“: Im Rahmen eines Beratungsgesprächs mit einer hochschuldidaktisch ausgebildeten Person werden (a) Lehrende über die Ergebnisse der Lehr-evaluation informiert, (b) ihre Wahrnehmungen des Lehrgeschehens erfragt, (c) modifizierbares Verhalten angesprochen. Es sollten (d) dysfunktionale Attributionsmuster unterbrochen, (e) konkrete Vorschläge für die Veränderung gemacht werden und (f) Lehrende zu Verhaltensänderungen angeregt werden.

Während die ersten zwei Modelle keine oder nur geringe Effekte im Bezug auf die Verbesserung zeigen, konnten Rindermann, Kohler und Meisenberg (2007) für das von den Autoren propagierte Beratungsmodell bedeutende Effekte zeigen, allerdings bisher nur für ein kleines Kollektiv von Lehrveranstaltungen an einer Sprachtherapieschule. Auch Hampton und Reiser (2004) konnten in einem Feldexperiment aufzeigen, dass ein Feedback an die Lehrenden auf der Basis eines theorieorientierten Fragebogens in Verbindung mit einer Beratung bedeutende Effekte auf die Verbesserung der Lehre hat. Lang und Kersting (2007) von der RWTH Aachen kommen aufgrund eigener umfangreicher Untersuchungen zu dem Schluss, dass Lehr-evaluation ohne Beratung in einer Langzeitperspektive weniger effektiv ist als in früheren Studien (z.B. Cohen 1980) angenommen.

Literaturverzeichnis

- Alean-Kirkpatrick, P./Hänni, H./Lutz, L. (1997): Internal quality monitoring of the teaching at the ETH Zurich: model design and initial impact. *Quality in Higher Education*, 3(1), pp. 63-71.
- Apodaca, P./Grad, H. (2006): The dimensionality of student ratings of teaching: integration of uni- and multidimensional models. *Studies in Higher Education*, 30(6), pp. 723-748.
- Bargel, T./El Hage, N. (2000): Evaluation der Hochschullehre – Modelle, Probleme und Perspektiven. *Zeitschrift für Pädagogik*, 11, S. 207-224.
- Berendt, B. (2000): Was ist gute Hochschullehre? *Zeitschrift für Pädagogik*, 11, S. 247-260.
- Braun, E. (2008): Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen (BEvaKomp). Göttingen.
- Braun, E./Gusy, B./Leidner, B./Hannover, B. (2008): Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54(1), S. 30-40.
- Chen, Y./Hoshower, L. B. (2003): Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), pp. 71-88.
- Cohen, P. A. (1980): Effectiveness of student rating feedback from improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 13(4), pp. 321-341.
- D'Appollonia, S./Abrami, P. C. (1997): Navigating student ratings of instruction. *American Psychologist*, 52, pp. 1198-1208.

- Daniel, H.-D. (1994): Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung. *Empirische Pädagogik*, 8, S. 109-129.
- Daniel, H.-D. (1997): Evaluierung der universitären Lehre durch Studenten und Absolventen. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 16, S. 149-164.
- Daniel, H.-D. (1998): Beiträge der empirischen Hochschulforschung zur Evaluierung von Forschung und Lehre. In: Teichler, U./Daniel, H.-D./Enders, J. (Eds.), *Brennpunkt Hochschule. Neuere Analyse zu Hochschule, Beruf und Gesellschaft* (S. 11-54). Frankfurt a. M.
- Diehl, J.M. (2003): Studentische Lehrevaluation in den Sozialwissenschaften: Fragebögen, Normen, Probleme. In: Keiner, S. (Ed.), *Evaluation (in) der Erziehungswissenschaft* (S. 63-90). Weinheim.
- Feldman, K. A. (1977): Consistency and variability among college students in rating their teachers and courses: a review and analysis. *Research in Higher Education*, 6, pp. 223-274.
- Gamliel, E./Davidovitz, L. (2005): Online versus traditional teaching evaluation: Mode can matter. *Assessment and Evaluation in Higher Education*, 30(6), pp. 581-592.
- Greenwald, A. G. (1997): Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), pp. 1182-1186.
- Hampton, S. E./Reiser, R. A. (2004): Effects of a theory based consultation process on instruction and learning in college classroom. *Research in Higher Education*, 45(5), pp. 497-525.
- Helmke, A. (1996): Studentische Evaluation der Lehre – Sackgasse und Perspektiven. *Zeitschrift für Pädagogische Psychologie*, 10, S. 181-186.
- Kember, D./Ginns, P. (2012): *Evaluating teaching and learning – A practical handbook for colleges, universities and the scholarship of teaching*. London: Routledge.
- Kirkpatrick, D. L. (1959): Techniques for evaluating training programs: Part 1 – Reactions, Part 2 – Learning. *Journal of the American Society of Training Directors*, 13 (3-9, 21-26).
- Kirkpatrick, D. L. (1960): Techniques for evaluating training programs: Part 3 – Behavior, Part 4 – Results. *Journal of the American Society of Training Directors*, 13(3-9), 13-18, 28-32.
- Klein, M./Rosar, U. (2006): Das Auge hört mit! Einfluss physischer Attraktivität des Lehrpersonals auf die studentische Evaluation von Lehrveranstaltungen – eine empirische Analyse am Beispiel der Wirtschaftswissenschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln. *Zeitschrift für Soziologie*, 35(4), S. 305-316.
- Kromrey, H. (1996): Qualitätsverbesserung in Lehre und Studium statt sogenannter Lehrevaluation. *Zeitschrift für Pädagogische Psychologie*, 10(3/4), S. 153-166.
- Kromrey, H. (2001): Evaluation von Lehre und Studium – Anforderungen an Methodik und Design. In Spiel, C. (Ed.), *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S. 21-60). Münster.
- Lang, J.W.B./Kersting, M. (2007): Regular feedback from student rating of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, 35, pp. 187-205.
- Marsh, H.W. (1983): Multidimensional ratings of teaching effectiveness. *Journal of Educational Psychology*, 75(1), pp. 150-166.
- Marsh, H.W. (1984): Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76(5), pp. 707-754.
- Marsh, H.W. (1987): Students' evaluation of university teaching: research issues, and directions for future research. *International Journal of Educational Research*, 11, pp. 243-388.
- Marsh, H.W. (2007): Students' evaluation of university teaching: A multidimensional perspective. In Perry, R. P./Smart, J. C. (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-384). New York.
- Marsh, H.W./Ginns, P./Morin, A.J.S./Nagengast, B./Martin, A. J. (2011): Use of student ratings to benchmark universities: Multilevel modeling of responses to the Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology*, 103(3), pp. 733-748.
- Marsh, H.W./Roche, L. A. (1997): Making student evaluation of teaching effective. The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), pp. 1187-1197.
- Marsh, H.W./Roche, L. A. (2000): Effects of grading leniency and low workload on students' evaluation of teaching: popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92(1), pp. 202-228.
- McAlpine, L./Harris, R. (2002): Evaluating teaching effectiveness and teaching improvement: A language for institutional policies and academic practices. *The International Journal of Academic Development*, 7(1), pp. 7-19.
- McKeachie, W. J. (1997): Student ratings. The validity of use. *American Psychologist*, 52(11), pp. 1218-1225.
- Mittag, S./Mutz, R./Daniel, H.-D. (2012): Institutionelle Qualitätssicherung der Lehre auf dem Prüfstand: Eine Fallstudie an der ETH Zürich. *Bielefeld*.
- Morley, D. (2014): Assessing the reliability of student evaluation of teaching: choosing the right coefficient. *Assessment & Evaluation in Higher Education*, 3(2), pp. 127-139.
- Mutz, R. (2001): Studienreform als Programm. *Programmevaluation zur Akzeptanz des reformierten Studiengangs „Forstwissenschaft“ bei Lehrenden und Studierenden*. Landau.
- Mutz, R. (2003): Multivariate Reliabilitäts- und Generalisierbarkeitstheorie in der Lehrevaluationsforschung. *Zeitschrift für Pädagogische Psychologie*, 17(3/4), S. 245-254.
- Mutz, R./Daniel, H.-D. (2006): Entwicklung eines Verfahrens für die studentische Beurteilung von Lehrveranstaltungen an der Universität Zürich. Universität Zürich: Unveröffentlichter Forschungsbericht.
- Mutz, R./Daniel, H.-D. (2008): Nutzung von Lehrevaluationsdaten für die Qualitätssicherung der Evaluationsinstrumente am Beispiel der Universität Zürich. *Beiträge zur Hochschulforschung*, 30(2), S. 34-55.
- Onwuegbuzie, A. J./Daniel, L. G./Collins, K. M. T. (2009): A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity*, 43, pp. 197-209.
- Perry, R. P./Smart, J. C. (2007): *The scholarship of teaching in higher education: An evidence-based perspective*. Dordrecht.
- Richardson, J. T. E. (2005): Instruments for obtaining student feedback: A review of literature. *Assessment & Evaluation in Higher Education*, 30(4), pp. 387-415.
- Rindermann, H. (2001a): Die studentische Beurteilung von Lehrveranstaltungen – Forschungsstand und Implikationen. In Spiel, C. (Ed.), *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S. 61-88). Münster.
- Rindermann, H. (2001b): *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierter Unterrichts*. Landau.
- Rindermann, H. (2003): *Lehrevaluation an Hochschulen – Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und Evaluation*. *Zeitschrift für Evaluation*, 2, S. 233-256.
- Rindermann, H./Kohler, J./Meisenberg, G. (2007): Quality of instruction improved by evaluation and consultation of instructors. *International Journal of Academic Development*, 12(2), pp. 73-85.
- Rossi, P. H./Freeman, H. E./Hofmann, G. (1988): *Programm-Evaluation. Einführung in die Methoden der angewandten Sozialforschung*. Stuttgart.
- Schmidt, B./Loßnitzer, T. (2010): *Lehrveranstaltungsevaluation: State of the Art, ein Definitionsvorschlag und Entwicklungslinien*. *Zeitschrift für Evaluation*, 9(1), S. 49-72.
- Schmidt, B./Tippelt, R. (2005): *Lehrevaluation*. *Zeitschrift für Erziehungswissenschaften*, 8(4), S. 227-241.
- Sedlmeier, P. (2006): The role of scales in student ratings. *Learning and Instruction*, 16(5), pp. 401-415.
- Seldin, P. (1999): *Changing Practice in Evaluating Teaching. A practical guide to improve faculty performance and promotion/tenure decisions*. Ballville.
- Sougnier, E./Gold, A. (2002): Fragebögen zur Lehrevaluation: Was können sie leisten? *Zeitschrift für Evaluation*, 2, S. 265-280.
- Sougnier, E./Gold, A. (2003): *Lehrevaluation als Feedback für Lehrende: Entwicklung eines Fragebogens unter hochschuldidaktischer Perspektive*. In G. Krampen & H. Zayer (Eds.), *Psychologiedidaktik und Evaluation IV – Neue Medien, Konzepte, Untersuchungenbefunde und Erfahrungen zur psychologischen Aus-, Fort- und Weiterbildung* (S. 129-144). Bonn.
- Spiel, C. (2001a): Der differentielle Einfluss von Biasvariablen auf die studentische Lehrveranstaltungsbewertung. In Engel, U. (Ed.), *Hochschul-Ranking – zur Qualitätsbewertung von Lehre und Studium* (S. 61-82). Frankfurt a. M.
- Spiel, C. (2001b): *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck*. Münster.
- Spooren, P./Broock, B./Mortelmans, D. (2013): On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), pp. 598-642.
- Sporen, P./Mortelmans, D./Thijssen, P. (2007): Content vs. style. Acquiescence in student evaluation of teaching. *British Journal of Educational Research Journal*, 38(3), pp. 3-21.
- Steiner, G. (1996): *Lernverhalten und Instruktionsmethoden*. In W. F. E. (Ed.), *Psychologie des Lernens und der Instruktion, D/I/2, Enzyklopädie der Psychologie* (S. 279-317). Göttingen.
- Webler, W.-D. (1991): Kriterien für gute akademische Lehre? *Das Hochschulwesen*, 6, S. 243-249.
- Webler, W.-D. (2008): Was taugen Fragebögen zur Lehrveranstaltungsbeurteilung – und was ihre Ergebnisse? *Praxishinweise aus einem Vergleich üblicher Fragebögen zur Evaluation von Lehre und Studium. Qualität in der Wissenschaft – Zeitschrift für Qualitätsentwicklung in Forschung, Studium und Administration*, 3(4), S. 96-105.
- Westermann, R./Spies, K./Heise, E./Wolburg-Claar, S. (1998): *Bewertung von Lehrveranstaltungen und Studienbedingungen durch Studierende: Theorieorientierte Entwicklung von Fragebögen*. *Empirische Pädagogik*, 12(2), 133-166.

Wolbring, T. (2013): Fallstricke der Lehrevaluation. Frankfurt a. M.

Wolff, P./Spiel, C./Pellert, A. (2001): Entwicklung eines Fragebogens zu globalen Lehrveranstaltungsevaluation – ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz. In Spiel, C. (Ed.), Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck (S. 89-109). Münster.

■ **Dr. Rüdiger Mutz**, Mitarbeiter an der Professur für Sozialpsychologie und Hochschulforschung, D-GESS, ETH Zürich, E-Mail: mutz@gess.ethz.ch

■ **Dr. Hans-Dieter Daniel**, Professur für Sozialpsychologie und Hochschulforschung, D-GESS, ETH Zürich, wiss. Leiter Evaluationsstelle, Universität Zürich, E-Mail: daniel@gess.ethz.ch

Wolff-Dietrich Webler (Hg.): Studieneingangsphase?

Das Bachelor-Studium braucht eine neue Studieneingangsphase!

Band I: Studierfähigkeit für ein frei(er)es Studium

Die Qualität der Studieneingangsphase hat wesentlichen - u.U. sogar entscheidenden - Einfluss auf den Studienerfolg. Trotzdem ist sie in der Vergangenheit in ihrer Bedeutung oft unterschätzt und vernachlässigt worden. Die Relevanz dieses Studienabschnitts wird jedoch inzwischen hoch eingestuft, eine Korrektur aufgrund z.T. falscher Signale aus der Bologna-Reform immer dringender. Für die Bewältigung des Übergangs in ein wissenschaftliches Studium aus der Schule oder aus dem Beruf ist wesentlich mehr nötig als ein Orientierungswochenende, (fach-)einführende Veranstaltungen und fachliche Begleittutorien. Mit der Feststellung „Das Bachelor-Studium braucht eine neue Studieneingangsphase - Studierfähigkeit für ein frei(er)es Studium!“ hatte das IWBB eine Initiative gestartet (u.a. eine Tagung), die den Erkenntnisstand zur Gestaltung der Studieneingangsphase bilanzieren, an solchen Fragen arbeitende Kolleg/innen zusammenführen und die Verbreitung besonders fortgeschrittener Modelle befördern sollte. Die Ergebnisse liegen nun in einem Doppelband vor.

Als zentrales Ziel wollen die Texte dazu beitragen, in den ersten beiden Semestern konzentriert die Studierenden zu befähigen, für sich anschließend ein motivierendes, durch Wahlmöglichkeiten möglichst selbst organisiertes, selbst verantwortetes und lerneffektives Studium zu organisieren.

Da das Themenfeld sehr groß ist, werden Fragen des Hochschulzugangs und von Auswahlverfahren nur am Rande angeschnitten. Die empirisch gewonnen Erkenntnisse zur Situation der Studierenden am Studienbeginn liegen bereits öffentlich vor, können also vorausgesetzt werden.

Im Mittelpunkt dieses Doppelbandes stehen daher in Band I die Ableitung und Begründung der Struktur des Problems und des Handlungsbedarfs (einschließlich des Beratungsbedarfs) sowie Gestaltungsprinzipien der Studieneingangsphase auf curricularer Ebene und der Ebene der Lehr-/ Lernprozesse. Behandelt wird die Frage, warum die Studieneingangsphase die ersten beiden Semester umfasst und wie sich fachliche und die Befriedigung des Orientierungsbedarfs über weite Strecken in den gleichen Veranstaltungen verschränken bzw. integrieren lassen, statt sie in Sonderveranstaltungen auszulagern.

Band II: Lösungsmodelle

Im Band II dieses Doppelbandes werden vielfältige Lösungen für Transfer und Gestaltungspraxis vorgelegt.

Diese Lösungen sind z.T. strategisch als umfassende Handlungsprogramme konzipiert, z.T. setzen sie bei einzelnen Lehrveranstaltungen an, z.T. bestehen sie aus einzelnen Lernmaterialien, die in besonderem Maße geeignet sind, einen Lernprozess in Richtung der Studienziele und der Ziele der Studieneingangsphase zu provozieren und zu unterstützen. Das Spektrum der Beispiele wird durch besonders eindrucksvolle Modelle aus der Zeit vor der Bologna-Reform ergänzt, an denen auch heute noch viel gelernt werden kann.



ISBN 10: 3-937026-76-2, Bielefeld 2012, 477 S., Doppelband im Schuber, 69.50 Euro zzgl. Versandkosten

Beziehbar im Fachbuchhandel und direkt beim Verlag – selten im Versandbuchhandel (z.B. nicht bei Amazon).

Bestellung - E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22

Volkhard Fischer

Die Evaluation von Lehrveranstaltungen an der Medizinischen Hochschule Hannover



Volkhard Fischer

Die Medizinische Hochschule Hannover (MHH) wurde im Jahre 1965 als autonome medizinische Universität mit dem Ziel gegründet, einen Campus der kurzen Wege zur Integration von medizinischer Lehre, Forschung und universitärer Krankenversorgung zu schaffen. Auch wenn für die MHH innovative Instrumente teilweise zuerst in den kleineren Studiengängen erprobt werden, richten sich auch in den naturwissenschaftlichen Masterstudiengängen die Bewertungskriterien an den für die Humanmedizin gewählten aus. Dies hat einerseits viel mit dem Selbstverständnis der MHH als medizinische Spartenuniversität zu tun und andererseits werden für diesen Studiengang mit Abstand die meisten Ressourcen benötigt. Verglichen mit den großen Volluniversitäten, die auch eine medizinische Fakultät haben, fällt dieser Anteil an der MHH relativ noch stärker ins Gewicht und ist durch die Organisation der MHH nach dem Integrationsmodell auch politisch wichtiger.

Am 2006 konzipierten Evaluationskonzept mit seinen verschiedenen Ebenen (Fischer 2008) hat auch die 2011 grundlegend überarbeitete Evaluationsordnung nichts geändert („Ordnung über die Evaluation von Studium und Lehre an der Medizinischen Hochschule Hannover“ 2011). Die MHH unterscheidet weiterhin zwischen einer Evaluation ganzer Studiengänge, der Evaluation spezifischer Aspekte, wie der Studienmotivation bzw. den individuellen Studienbedingungen, und der Evaluation einzelner Module. Insbesondere im Bereich der Lehrveranstaltungsevaluation wurden die eingesetzten Instrumente weiterentwickelt. Deshalb liegt der Fokus in diesem Artikel auf diesem dritten Bereich.

Die mit der Lehrveranstaltungsevaluation gewonnenen Daten werden nicht nur für den kritischen Dialog zwischen Lehrenden und Lernenden genutzt, sondern dienen auch den Studienkommissionen als Grundlage für die Qualitätssicherung in der Lehre. So hat die MHH vor drei Jahren zur finanziellen Stärkung der Lehre eine leistungsorientierte Mittelvergabe in der Lehre eingeführt. Der wichtigste Parameter hierbei ist die globale Modulbewertung aus der studentischen Lehrveranstaltungsevaluation.

1. Die Evaluation der Lehrveranstaltungen

Für die medizinische Lehre wurde im Jahre 2000 durch eine Arbeitsgruppe des Rektors ein Fragebogen zur Evaluation von Lehrveranstaltungen entwickelt, um so die

Akzeptanz eines zentralen Fragebogens zur Lehrveranstaltungsevaluation auf Seiten der Lehrenden zu sichern (Fischer 2000). Dieser Fragebogen zur Evaluation von Lehrveranstaltungen durch Studierende (FELS) orientierte sich an der Faktorenstruktur des Fragebogens „Student Evaluation of Educational Quality“ (Marsh 1982, 1984). Besonderes Augenmerk galt dabei dem Versuch, einen Fragebogen zu entwickeln, der den Lehrenden eine differenzierte Rückmeldung über ihre Lehrveranstaltung gibt und auf Itemebene eindeutige Empfehlungen für Konsequenzen zur Verbesserung des Lehrangebotes ermöglicht. Gleichzeitig sollte der FELS einen Vergleich zwischen verschiedenen Lehrveranstaltungen über die Skalenwerte ermöglichen (Fischer/Beyer 2001, 2004). Eine besondere Schwierigkeit war dabei von Anfang an, dass medizinische Lehrveranstaltungen nicht von ein oder zwei Lehrenden durchgeführt werden. Dies liegt zum einen daran, dass das Medizinstudium den Studierenden einen Überblick über die gesamte Medizin geben soll. Zum anderen ist für eine optimale Krankenversorgung eine zunehmende Spezialisierung des einzelnen Dozenten vorteilhaft. Wegen der Jahrgangsstärke normaler deutscher medizinischer Fakultäten sind dann Lehrveranstaltungen mit einer einstelligen Dozentenzahl eher die Ausnahme. In der Konsequenz erfolgt die Evaluation also veranstaltungsbezogen, obwohl auch an der MHH davon ausgegangen wird, dass der wesentliche Faktor zur Erklärung guter Lehre die Unterrichtenden sind und deshalb eine personenscharfe Evaluation sinnvoller wäre.

Bis 2005 wurde der Fragebogen als vierseitiger Papierfragebogen in allen Lehrveranstaltungen des Medizinstudiengangs und einigen des Zahnmedizinstudiengangs eingesetzt. Wurden einzelne Fragen überarbeitet, so geschah dies, um drei Aspekten mehr Beachtung zu schenken:

- 1) Die Fragen sollten sich auf von den Studenten direkt beobachtbares Verhalten beziehen,
- 2) die Fragen sollten so formuliert sein, dass sie eine Nutzung der gesamten Antwortskala provozieren,
- 3) die Antwortskala sollte eine gerade Anzahl an Ankreuzmöglichkeiten haben, um eine Wahl der Mittelkategorie aus Unentschiedenheit zu erschweren.

Eine Umrechnung der Rohdaten in einen bereinigten Lehreffekt wurde dabei nicht angestrebt, weil die von den Lehrenden nicht zu beeinflussenden Faktoren in der

medizinischen Lehre als zu vielfältig eingestuft wurden. So gibt es in fast keinem Modul der Studiengänge Humanmedizin und Zahnmedizin nur eine Veranstaltungsart, die Anzahl der Wahlmöglichkeiten zwischen verschiedenen Modulen ist extrem eingeschränkt und die Module haben nicht nur kognitive Lernziele, sondern auch psychomotorische. Außerdem ist das Verhältnis theoretischer Lehrveranstaltungen zu praktischen Lehrveranstaltungen durch Approbationsordnungen stärker zementiert als in anderen Studiengängen.

1.1 Entwicklung eines Kurzfragebogens

Da die flächendeckende Evaluation mit Papierfragebögen von allen Beteiligten als zu kostenintensiv und zu schwerfällig erlebt wurde, erfolgte die Entwicklung eines verkürzten Fragebogens und sukzessive die Umstellung auf eine elektronische Befragung. Insbesondere die als lang erlebten Zeiträume bis zur Rückmeldung der Ergebnisse an die Beteiligten hatten dazu geführt, dass sich die Akzeptanz für die papierbasierte Evaluation auf Seiten der Dozenten zunehmend verschlechterte. Diese Umstellung verlief parallel zur Einführung von schriftlichen Prüfungen mit elektronischen Eingabegeräten, zunächst auf Tablets, später auf Laptops.

Der entscheidende Antrieb für einen raschen Umbau ergab sich aus dem Aufbau eines Modellstudiengangs Medizin: Seit 2005 ist das Medizinstudium an der MHH modular aufgebaut, alle Module sind mit mindestens einer benoteten Prüfung am Modulende versehen (Haller/Fischer 2007). Als Besonderheit der Module in den ersten beiden Studienjahren sind diese mit anderen Modulen zeitlich verschränkt, während die Studierenden der späteren Studienjahre ganztags ausschließlich in einem Modul unterrichtet werden. Die minimale Länge dieser Module beträgt eine Woche, die maximale zehn Wochen. Gleichzeitig sind die Prüfungen in den Modulen der ersten beiden Studienjahre nicht nur Prüfungen der Universität, sondern auch Teil eines über zwei Jahre verteilten Staatsexamens. Sie werden ein Mal im Jahr angeboten, während die Module der folgenden drei Studienjahre mindestens drei Mal im Jahr angeboten werden.

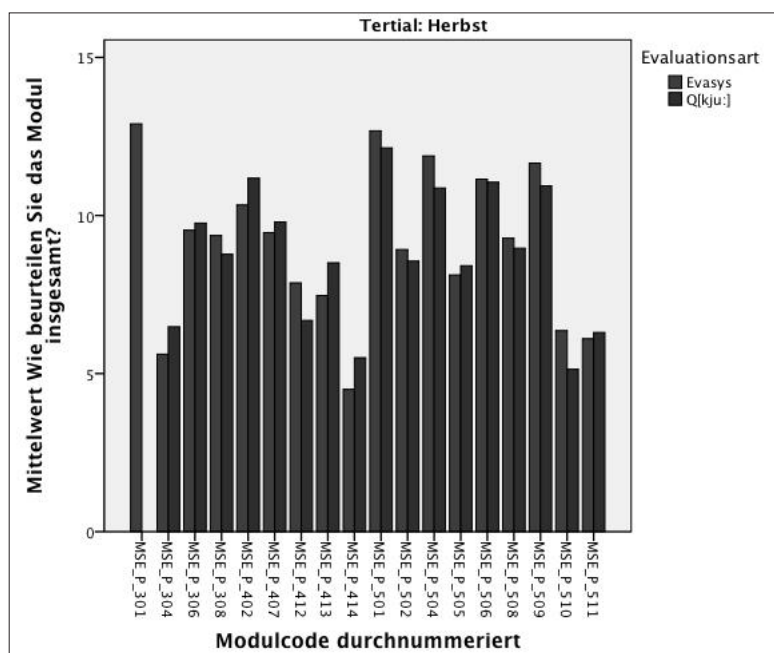
Der Aufwand für die Entwicklung neuer Module bzw. die Anpassung bestehender Lehrveranstaltungen sollte aber nicht zu Lasten der Qualität gehen. Um dies zu verhindern, wurde vom Senat der MHH beschlossen, neben der schon etablierten leistungsorientierten Mittelvergabe für Forschungsleistungen auch eine leistungsbezogene Mittelvergabe für gute Lehre einzuführen. D.h. die Qualitätsverbesserung in der Lehre sollte über monetäre Anreize gefördert werden, wobei die Mittelvergabe im Wesentlichen an die Lehrveranstaltungsevaluation gekoppelt werden sollte. Im Vorfeld der geplanten leistungsbezogenen Mittelvergabe für die Lehrqualität wurde eine alle medizinischen Fächer einbeziehende einjährige Studie geplant, die sich mit zwei Vorwürfen beschäftigen sollte: 1) Die Erhebungen mit dem von den Studierenden akzeptierten FELS seien nicht repräsentativ, weil der Rücklauf deutlich unter 80% läge. 2) Die

Evaluationen könnten durch hohe Anforderungen in der Lehrveranstaltung, insbesondere durch schwierige Klausuren, negativ beeinflusst werden.

1.2 Validierungsstudie

Im Studienjahr 2007/08 wurde im Anschluss an die meisten Module des Studiengangs Medizin mit schriftlichen Prüfungen sowohl eine Online-Evaluation (EvaSys) durchgeführt, als auch parallel – im Anschluss an die elektronischen Prüfungen – eine „Vor-Ort-Evaluation“ (Q[kju:]). 26 Module wurden in den drei Tertialen des Studienjahres auf diese Weise doppelt evaluiert (Fischer/Paulmann/Seibicke/Haller 2008). Damit gingen 57,8% der Module des Studiengangs in die Studie ein. Die Fragensätze waren bei beiden Vorgehensweisen identisch, wenn man davon absieht, dass in Q[kju:] zum Testzeitpunkt keine Eingabe von Freitextkommentaren möglich war (Möbs 2014). Bei der Online-Befragung mit EvaSys wurden alle Teilnehmer/innen per E-Mail zur Teilnahme an der Evaluation aufgefordert und konnten sich dann per TAN am Server anmelden. Die Eingabe erfolgte ortsungebunden an einem beliebigen PC. Die Aufforderung wurde am Tag vor der Prüfung verschickt, der Zugang blieb für vier Tage geöffnet. Da die überwiegende Mehrheit der Befragten diese Evaluation nach der Prüfung vornahm, wurde der Einfluss der Prüfung auf das Evaluationsergebnis in dieser Studie nicht weiter betrachtet. Die Vor-Ort-Befragung wurde mit Q[kju:] durchgeführt. Die Eingabe erfolgte an den normalen in den Prüfungen eingesetzten Tablett-PCs per Stift. Für die Evaluation wurden im Anschluss an die Klausur die Evaluationsfragen auf den Bildschirm dargeboten. Durch einen automatisierten Neustart zwischen Prüfung und Evaluation blieb die Anonymität der Befragten gewahrt. Der Rücklauf in den mit EvaSys durchgeführten Befragungen schwankte zwischen 10 und 51%, der Rücklauf

Abbildung 1a: Vergleich der mittleren Evaluationsergebnisse nach Art der Evaluation bei ausgewählten Modulen im Herbstterial 2007.



bei den Q[kju:]-Befragungen zwischen 60 und 97% der Modulteilnehmer. Damit entsprach dieser Aspekt des Methodenvergleichs den Erwartungen. Erfreulich war die große Bereitschaft der Studierenden, überhaupt an dieser doppelten Evaluation teilzunehmen und damit ihr Interesse an einer möglichst effektiven und unverzerrten Lehrrevaluation zu dokumentieren. Die zentrale Fragestellung der beiden Hypothesen betraf aber die mittlere Bewertung der Lehrveranstaltungen. Diese ist ge-

Abbildung 1b: Vergleich der mittleren Evaluationsergebnisse nach Art der Evaluation bei ausgewählten Modulen im Wintertertiar 2008.

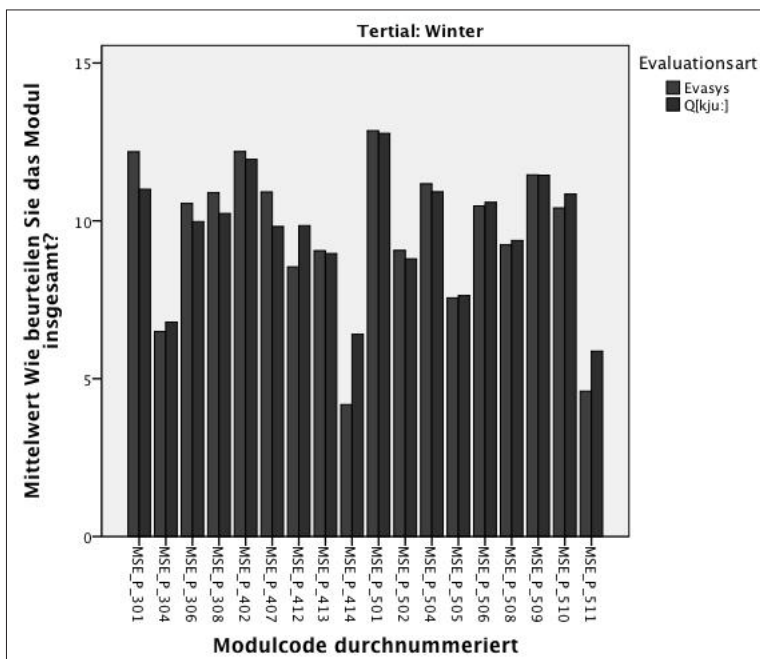
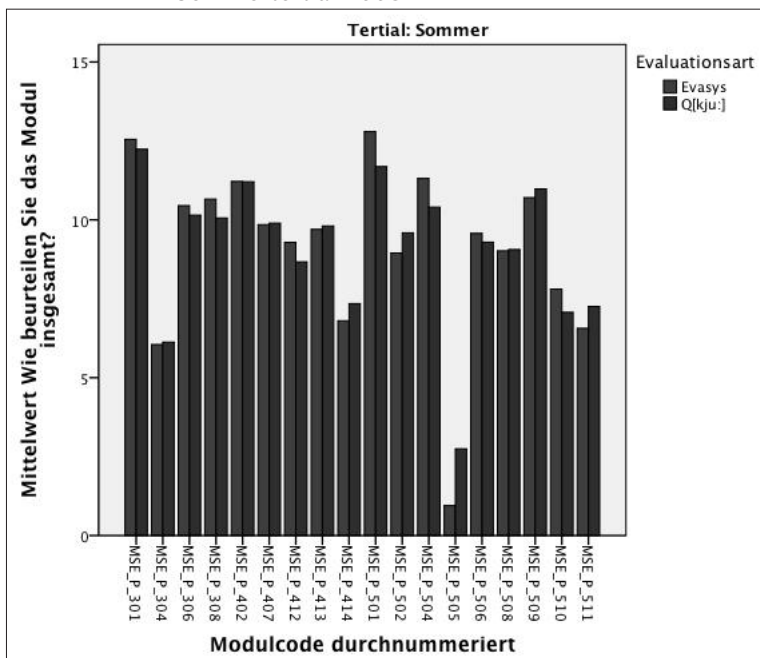


Abbildung 1c: Vergleich der mittleren Evaluationsergebnisse nach Art der Evaluation bei ausgewählten Modulen im Sommertertiar 2008.



trennt für die drei Tertiale des Studienjahres 2007/08 in der Abbildung 1a, b, c dargestellt. Den Einfluss der Variablen Geschlecht der Befragten, Tertial, in dem das Modul evaluiert wurde und Evaluationsmethode wurde in je einer ANCOVA pro Modul untersucht (Fischer, 2011), da sich die Stichproben in den einzelnen Modulen teilweise überlappten, aber nicht identisch waren. Hätte man statt der separaten ANCOVAs personidentische Stichproben über alle Module eines Studienjahres

gebildet, wäre die resultierende Schnittmenge so klein ausgefallen, dass eine ernsthafte Auswertung nicht mehr sinnvoll gewesen wäre. Den größten Einfluss haben danach saisonale Effekte, die man als den Kontext, in den das Modul jeweils im Studienjahr eingebettet ist, betrachten muss.

Ungeklärt blieb, ob Online- und Vor-Ort-Evaluationen auch in Modulen mit mündlichen Prüfungen identische Ergebnisse erbringen würden, wenn die Studierenden für die Vor-Ort-Evaluation einen Extratermin im Hörsaal bekommen würden. Die geplante analoge Doppelevaluation für das Studienjahr 2008/09 für alle Module mit mündlich/praktischen Prüfungen unterblieb wegen des erheblich höheren Aufwandes und der in Abbildung 1a, b, c dargestellten Ergebnisse bei den schriftlichen Prüfungen.

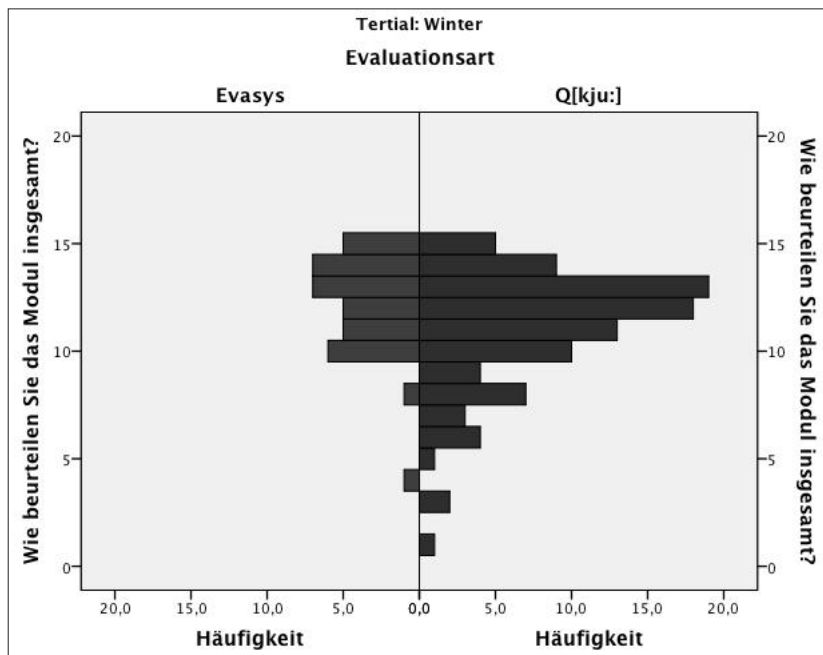
Das Beispiel der Verteilung der Gesamtbewertung des Moduls 301 im Wintertertiar ist in Abbildung 2 dargestellt. Es fällt auf, dass der geringere Rücklauf offensichtlich daher rührte, dass sich weniger Studierende an der Evaluation mit EvaSys beteiligt hatten, die eine mittlere Bewertung des Moduls abgeben wollten. Abgesehen von der Rücklaufquote lassen sich nur bei einzelnen Modulen Unterschiede in der mittleren Bewertung zwischen den beiden Befragungsmethoden nachweisen. Da die Anzahl der aufgetretenen Unterschiede aber statistisch nicht signifikant ist, lässt sich ein Methodeneffekt nicht belegen.

Die durchgeführten Online-Evaluationen waren also nicht verzerrt, bezogen auf die unmittelbar nach den Prüfungen erhobenen, elektronischen „Vor-Ort-Evaluationen“, und deshalb repräsentativ, auch wenn der Rücklauf deutlich geringer ausfiel. Es fanden sich lediglich Unterschiede in der Steilheit der Verteilung der Bewertungen. Auf der anderen Seite konnte auch die Befürchtung, schwierige Prüfungen könnten unmittelbar anschließend durchgeführte Evaluationen negativ verzerren, nicht bestätigt werden. Summa summarum zeichnete sich das Vor-Ort-Evaluationsmodell als das vorteilhaftere bei schriftlichen Prüfungen ab, da der höhere Rücklauf auch Fragen an Subgruppen erfolgversprechend erscheinen ließ.

1.3 Aktuelle Umsetzung der Lehrveranstaltungsevaluation

Inzwischen werden alle Module der Studiengänge Humanmedizin, bzw. Public Health

Abbildung 2: Vergleich der globalen Modulbewertung im Wintertertia 2008 für das Modul 301 für zwei Evaluationsarten.



und die scheinpflichtigen Module des Studiengangs Zahnmedizin mit einem nach der Validierungsstudie angepassten Kurzfragebogen (Kurzes Inventar zur Evaluation durch Studierende: KIES) evaluiert, der nach der Umstellung auf elektronisch dargebotene Fragebögen die wichtigsten Dimensionen des FELS abbildet. Wichtig war dabei, dass die Pflichtlehrveranstaltungen eine oder mehrere Prüfungen mit einschließen. Deshalb sollte die Evaluation möglichst zeitnah nach jeder dieser Prüfungen erfolgen, um die Einheit von Lehren, Lernen und Prüfen zu betonen. Eine Evaluation ungefähr in der Mitte einer Lehrveranstaltung, würde dieses Prinzip brechen und die mit der Wahl eines solchen Termins verbundene Auflösung des Anspruchs, dass Lehren, Lernen und Prüfen aufeinander abgestimmt erfolgen müssen, für alle Beteiligten klar dokumentieren.

Wegen der Vielfalt an zu beurteilenden Module mussten es Fragen sein, die auf jede Art von Präsenzveranstaltung anwendbar sind, bzw. durch sprachliche Umformulierungen angepasst werden können. In der gegenwärtigen Fassung sind dies: Je eine Frage zu den Stärken bzw. Schwächen des Moduls. Die Studierenden können dabei für sechs Dimensionen angeben, ob sie diese für eine besondere Stärke und/oder für einen verbesserungswürdigen Aspekt des Moduls halten. Die Dimensionen sind dem FELS entnommen und werden durch die Labels Lehrpersonal, Lehr- und Lernmaterialien, Organisation der Lehrveranstaltung, Inhalte der Lehrveranstaltung, Patientenbezug und Prüfung dargestellt. Weiterhin sollen die Studierenden auf sechsstufigen Skalen angeben,

- wie hoch der Vor- und Nachbereitungsaufwand für sie war,
- ob sie das Verhältnis von Lernaufwand und Zugewinn an Kenntnissen und Fertigkeiten für angemessen ansehen,
- ob die wichtigsten theoretischen Themen praktisch vertieft wurden,

- ob die Bezüge zu anderen Modulen verdeutlicht wurden
- und ob die Positionierung des Moduls im Studienablauf zu früh, angemessen oder zu spät ist.

Je nach der Häufigkeit, mit der bestimmte Stärken oder Schwächen von den Studierenden genannt werden, und dem Punktwert in der globalen Veranstaltungsbewertung kann dann innerhalb kürzester Zeit eine vertiefende Evaluation mit auf die fünf Aspekte zugeschnittenen Skalen nachgeschoben werden. Die Entscheidung über die Ansetzung einer Nachbefragung trifft für die Module im Modellstudiengang HannibaL das Evaluationsbüro des Studiendekanates, im Falle der Module des Studiengangs Zahnmedizin ist die Freischaltung an das Erreichen eines Quorums in der Evaluation gekoppelt. Für diese ad hoc Befragung ist zwar mit einem niedrigeren Rücklauf zu rechnen. Weil die Antworten der Nachbefragung aber über den persönlichen Code mit der Basisevaluation verknüpft werden können, ist eine differenzierte Analyse einschließlich der Prüfung auf mögliche Stichprobenverzerrungen möglich. Außerdem wird durch diesen manuell vom Evaluationsbüro ausgelösten Qualitätssicherungsprozess den Studierenden und den Dozenten gezeigt, dass Evaluationen nicht ohne Resonanz bleiben, zumal die Nachbefragungsergebnisse ebenfalls im Lernmanagementsystem der MHH veröffentlicht werden.

An diese sieben Fragen schließen sich noch zwei weitere Fragen an, die auf alle an der MHH durchgeführten Module anwendbar wären:

- Die allgemeine Veranstaltungszufriedenheit
- und ein Freitextfeld für weitere Rückmeldungen.

Der Frage nach der allgemeinen Veranstaltungszufriedenheit kommt dabei eine zentrale Bedeutung zu. So stellt sie den zentralen Parameter im Rahmen der Diskussion über notwendige Modulverbesserungen dar. Die Antworten auf die Freitextfrage werden insbesondere von den Lehrenden, aber auch von der Studienkommission als direkte Rückmeldung zur Interpretation der Bewertungen in den geschlossenen Fragen genutzt. Damit stellt die Freitextfrage für die Lehrenden die wichtigste Datenquelle für die weitere Verbesserung ihrer Module dar. Denn die prinzipielle Möglichkeit, bis zu drei modulspezifische Fragen an diesen allgemeinen Teil anzuhängen, wird von den wenigsten Lehrverantwortlichen genutzt. Und weil bei größeren Änderungen am Modulaufbau schon im Vorfeld der Kontakt zum Evaluationsbüro gesucht wird, um diese Änderungen durch spezifische Fragen in ihrer Wirksamkeit besser beurteilen zu können, kann man davon ausgehen, dass der Kurzfragebogen KIES von Lehrenden akzeptiert ist.

Generell bereitet aber der Rücklauf in den Online-Evaluationen Probleme. Vor allem in den kleineren Studiengängen führt die Freiwilligkeit der Evaluation bei, verglichen mit der Vor-Ort-Evaluation, höherem Aufwand

für die Studierenden dazu, dass die Ergebnisse einzelner Evaluationen nicht hochschulöffentlich gemacht werden können, weil sonst die Gefahr der Deanonymisierung der wenigen Befragten bestünde. Häufig genannte Gründe für die Nichtteilnahme sind Probleme mit dem offiziellen eMail-System der MHH und die Vermutung, dass sich in den Lehrveranstaltungen ja doch nichts ändern würde.

2. Leistungsbezogene Mittelvergabe für die Lehre

Die globale Lehrveranstaltungsbewertung geht darüber hinaus als ein Messwert in die qualitative leistungsbezogene Mittelvergabe für die Lehre (LOM-L) im Studiengang Humanmedizin ein. Der zweite Parameter für LOM-L ist die Prüfungsqualität, ermittelt über die Notenverteilung, den dritten Parameter bildet die Modulgröße, ermittelt über die curricularen Stunden. Da die Gesamtbewertung des Moduls in der Lehrveranstaltungsevaluation mit 0 bis 15 Punkten erfolgt, für die Prüfungsqualität bis zu 8 Punkte und die Modulgröße bis zu 3 Punkte vergeben werden, kommt der Lehrveranstaltungsevaluation das größte Gewicht zu. Für die Mittelvergabe berücksichtigt werden die 15 Module mit dem höchsten Punktwert, sofern sie in der Evaluation nicht im letzten Drittel platziert waren. Die ausgewählten Module werden dann abschließend entsprechend ihrer Größe mit einem Faktor bewertet, der festlegt, welcher Anteil an der zur Verfügung stehenden Gesamtsumme von 700.000,- € auf das jeweilige Modul entfällt.

Dieses Ranking hat die Aufmerksamkeit der Abteilungen für ihr Abschneiden in der schon länger existierenden hochschulöffentlichen Darstellung der Lehrqualität erhöht. Erste explorative Auswertungen zeigen zwar keine Veränderungen in der mittleren Bewertung aller Module oder der mittleren Bewertung aller Prüfungen.

Betrachtet man dagegen einzelne Module, so lassen sich mit Hilfe der Lehrveranstaltungsevaluation aber sehr wohl die Effekte von Interventionen in die Moduldurchführung nachweisen. So konnten Bintaro et al. zeigen, dass die simple erneute gegenseitige Abstimmung der Dozenten in einem Modul, welches an 15 Tagen die wichtigsten internistischen Erkrankungen aus grundlagenwissenschaftlicher, klinisch-praktischer und klinisch-theoretischer Perspektive jeweils anhand eines konkreten Patienten behandelt, deutliche Effekte bei der globalen Veranstaltungsbewertung, aber auch bei der Bewertung der Dozenten, des Lehrmaterials und des Patientenbezuges auslöste (Bintaro/Haller/Just/Fischer 2012). Auch für andere Module gibt es vergleichbare, meist weniger spektakuläre Daten (Gutenbrunner et al. 2010). Da die Bewertungen der Studierenden aber die Veränderungen der Module im Laufe der Jahre nicht berücksichtigen (sollten), bleibt das Problem, wie man generelle Qualitätsverbesserungen in der Lehre abbilden kann, zumindest für die MHH eine noch zu lösende Frage.

3. Diskussion

Mehr als 10 Jahre nach der Einführung eines Systems zur Evaluation von Studium und Lehre an der Medizini-

schen Hochschule Hannover kann man feststellen, dass das System zwar permanent an die Bedürfnisse der Beteiligten angepasst worden ist, sich im Kern aber nicht verändert hat. Technische Veränderungen in der Art, wie die Lehrveranstaltungsevaluation umgesetzt wird, wurden zwar nicht so systematisch untersucht, wie dies z.B. von Dresel und Tinsner (Dresel/Tinsner 2008) beschrieben wurde. Die Ergebnisse weisen aber auf vergleichbare Methodeneinflüsse hin. Obwohl die Grundlage der Lehrveranstaltungsevaluation an der MHH ein personenzentrierter Fragebogen ist (Marsh 1982; Marsh/Roche 2000), ist die aktuelle flächendeckende Realisierung eine veranstaltungsbezogene Kurzform. Die Überprüfung, ob die Kurzform wirklich das Gleiche erfasst wie eine ausführliche Fassung, steht allerdings aus. Trotz dieser damit nur eingeschränkten wissenschaftlichen Herleitung des eingesetzten Instrumentes hat es sich in der Praxis einer medizinischen Hochschule bewährt, die mehr als zwei Drittel ihres Umsatzes in der Krankenversorgung erwirtschaftet und deren Drittmittelaufkommen auf dem zweiten Platz in Deutschland liegt. Die zentrale Durchführung der Lehrveranstaltungsevaluation sichert einen gleichbleibenden Standard. Die wahrgenommene Neutralität der Ergebnisinterpretation durch eine zentrale Einrichtung hat im Laufe der Jahre die Akzeptanz zusätzlich gefördert. Die Validität der Ergebnisse wird zwar in Einzelfällen, insbesondere bei extrem niedrigen Rückläufen angezweifelt. Alle durchgeführten empirischen Überprüfungen konnten aber keine systematischen Verzerrungen nachweisen. Unsere Daten bestätigen auch, dass Maßnahmen zur spezifischen Verbesserung der Lehrqualität sich auf die studentische Bewertung dieser Module auswirken, obwohl sich in der globalen Betrachtung keine nennenswerten Veränderungen in der Gesamtbewertung der Lehrveranstaltungen ergeben.

Mit der Einführung einer leistungsbezogenen Mittelvergabe für die Lehre (Just/Beinhoff/Haller/Fischer 2010), die als Gegengewicht zur schon länger existierenden leistungsbezogenen Mittelvergabe für die Forschung intendiert war, konnte die Beachtung, die grundständige Lehre in den akademischen Gremien erfährt, gesteigert werden. Auffällig ist aber, dass sich dadurch der Fokus in der Lehrveranstaltungsevaluation, zumindest in der Diskussion der Ergebnisse in den Gremien, weg von einem kritischen Dialog mehr und mehr hin zu einer Diskussion über die zusätzliche Finanzierung der Abteilungen verschiebt. Ob dieser Eindruck mit dem extremen Kostendruck an einer medizinischen Hochschule zusammenhängt, der auf einmal alles auf diesen Aspekt reduziert, oder ob sich hier eine Verschiebung von einer normativ propagierten intrinsischen Motivation hin zu einer extrinsischen abzeichnet, sei dahingestellt. An der Notwendigkeit, mit Hilfe der Lehrveranstaltungsevaluation den Lehrenden eine konkrete Rückmeldung über Stärken und Schwächen ihrer Lehre zu geben, ändert sich dadurch nichts.

Literaturverzeichnis

Bintaro, P./Haller, H./Just, I./Fischer, V. (2012): „The times are changing“ – Bewertung eines Moduls im Modellstudiengang durch fünf aufeinander

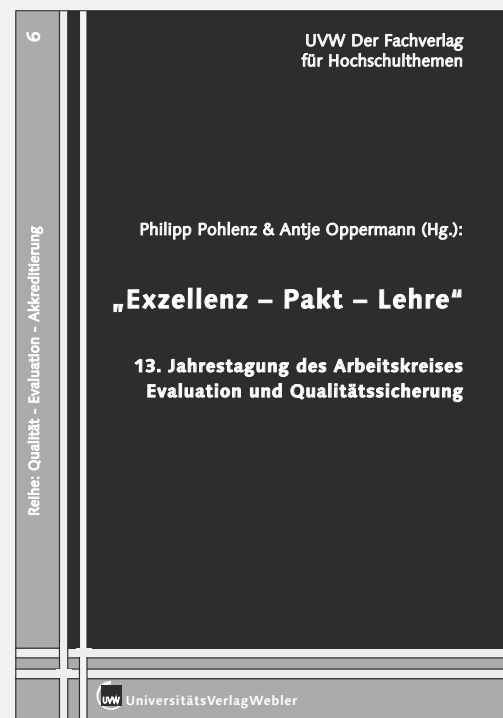
- folgende Jahrgänge. Paper presented at the Jahrestagung der Gesellschaft für Medizinische Ausbildung, Aachen. Poster retrieved from
- Dresel, M./Tinsner, K. (2008): Onlineevaluation von Lehrveranstaltungen: Methodeneffekte bei der Onlineevaluation von Lehrveranstaltungen. Zeitschrift für Evaluation, 7, S. 183-211.
- Fischer, V. (2000): Prädiktoren der Beurteilung von Vorlesungen durch Studierende der Humanmedizin. Paper presented at the Jahrestagung der GMA: Lehre 2000, Köln, 12.-13.5.2000.
- Fischer, V. (2008): Ebenen der Evaluation eines modularisierten Studienganges, in: Rudinger, G./Krahn, B./Rietz, C. (Eds.): Evaluation und Qualitätssicherung von Forschung und Lehre im Bologna-Prozess. Göttingen. S. 163-176.
- Fischer, V. (2011): Do different evaluation modes produce different results? Paper presented at the International Meeting der Association for Medical Education in Europe. Wien.
- Fischer, V./Beyer, O. (2001): Zur faktoranalytischen Struktur des Fragebogens für die Evaluation von Lehrveranstaltungen durch Studierende, in: von der Hardt, H./Knapp, W. (Eds.): Lehrbericht 2000. Hannover: Medizinische Hochschule Hannover.
- Fischer, V./Beyer, O. (2004): Factor structure of a multidimensional questionnaire for the evaluation of medical courses. Paper presented at the International Meeting of the Association for Medical Education in Europe, Edinburgh, 7.9.2004.
- Fischer, V./Paulmann, V./Seibicke, N./Haller, H. (2008): Vergleich eines Onlineevaluationssystems und eines ortsgebundenen Systems hinsichtlich möglicher Ergebnisverzerrungen. Paper presented at the Jahrestagung der Gesellschaft für Medizinische Ausbildung. Greifswald.
- Gutenbrunner, C./Schiller, J./Schwarze, M./Fischer, V./Paulmann, V./Haller, H./Küther, G. (2010): The Hannover Model for the Implementation of Physical and Rehabilitation Medicine into the Undergraduate Medical Training. Journal of Rehabilitation Medicine, 42, pp. 206-213.
- Haller, H./Fischer, V. (2007): Bericht über den Stand der Realisierung des Modellstudienganges Hannibal. Hannover: Medizinische Hochschule Hannover. S. 45.
- Just, I./Beinhoff, S./Haller, H./Fischer, V. (2010): Kriterien für eine leistungsorientierte Mittelvergabe in der Lehre. Paper presented at the Jahrestagung der Gesellschaft für Medizinische Ausbildung, Bochum, 23.09.-25.09.2010.
- Marsh, H. W. (1982): SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. British Journal of Educational Psychology, 52, pp. 77-95.
- Marsh, H. W. (1984): Student evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 76, pp. 707-754.
- Marsh, H. W./Roche, L. A. (2000): Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? Journal of Educational Psychology, 92, pp. 202-228.
- Möbs, D. (2014): IQ[kju:]L GmbH > e-Prüfungen. from <http://www.iqul.de/ePr%C3%BCfungen.aspx>
- Ordnung über die Evaluation von Studium und Lehre an der Medizinischen Hochschule Hannover (2011).

■ **Volkhard Fischer**, Leiter des Evaluationsbüros im Studiendekanat der Medizinischen Hochschule Hannover,
E-Mail: fischer.volkhard@mh-hannover.de

Philipp Pohlenz & Antje Oppermann „Exzellenz – Pakt – Lehre“

Reihe: Qualität - Evaluation - Akkreditierung

Unter dem Titel „Exzellenz – Pakt – Lehre“ veranstaltete der Arbeitskreis Qualitätsmanagement und Evaluation der Berliner und Brandenburger Hochschulen seine 13. Jahrestagung in Berlin (24./25. Mai 2012). Im Mittelpunkt der Veranstaltung standen Projekte, die in den jüngsten Förderprogrammen zur Steigerung der Qualität in Lehre und Studium initiiert wurden. Neben dem Bund-Länder Programm „Qualitätspakt Lehre“ ist hier insbesondere der Wettbewerb „Exzellente Lehre“ von Stifterverband und Kultusministerkonferenz zu nennen. Die Tagung stellte die Projektideen einer breiten Öffentlichkeit aus dem deutschsprachigen Hochschulwesen und der Hochschulpolitik vor. Der Tagungsband gibt darauf aufbauend Einblick in ausgewählte Vorhaben. Die thematische Vielfalt der dargestellten Projekte reicht von der Gestaltung der Studieneingangsphase über E-Learning Initiativen bis zu Projekten im Bereich des „Bologna-sensiblen“ Curriculumentwicklung. Durch die Diskussion erster Erfahrungen sollen die Projektideen bewusst zur Nachahmung in anderen Hochschulen anregen.



ISBN 10: 3-937026-84-3, ISBN 13: 978-3-937026-84-8, Bielefeld 2013, 200 Seiten, 34.90 Euro

Erhältlich im Fachbuchhandel und direkt beim Verlag – selten im Versandbuchhandel (z.B. nicht bei Amazon).
Bestellung - E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22

*Uwe Schmidt, Nicole Becker, Holger Lübke, Marc Hümpfner,
Franziska Schmidt, Natalia Yugay & Cathrin Neßler*

Modellbasierte Lehrevaluation: Konzept und empirische Ergebnisse

Evaluation an Hochschulen kann im Bereich der Forschung auf eine lange Tradition zurückblicken, wogegen Lehrevaluation im engeren Sinne sich erst im Verlaufe des 20. Jahrhunderts nennenswert ausbildete. Während in den USA die Evaluation der Lehre mit ersten Erhebungsinstrumenten auf das Ende der 1920er Jahre datiert, spielt sie im deutschen Sprachraum lange Zeit keine bzw. eine nachgeordnete Rolle. Erst im Verlaufe der 1990er Jahre gewinnen Evaluationsverfahren in der Lehre in Form von Peer Reviews oder der Bewertung von Lehrveranstaltungen im deutschen Sprachraum allmählich an Bedeutung, die durch die Einführung des Akkreditierungssystems nochmals verstärkt wurde. Wenn auch insgesamt eine quantitative Zunahme an Evaluationsverfahren und Erhebungen unterschiedlicher Art festzustellen ist, folgt hieraus nicht zwingend auch eine Zunahme an Akzeptanz für Evaluation. Die nach wie vor vorhandene Skepsis rekurriert in der Regel auf eine nicht immer transparente und nachvollziehbare Verwendung der Ergebnisse, das enorme Ausmaß an Evaluationen, welches allen Beteiligten zu einer gewissen Evaluationsmüdigkeit führt und zunehmend auch auf den zum Teil geringen Erkenntnisgewinn, den manche Evaluationsverfahren mit sich bringen. Dies wiederum ist unter anderem auf die zum Teil starke politische Überformung der Verfahren und ein damit verbundenes Defizit an theoretischer Fundierung und Verfahrensvalidität zurückzuführen (vgl. u.a. Schmidt 2005).

Der folgende Beitrag greift die Frage der theoretischen bzw. Modellorientierung der Hochschulevaluation am Beispiel der Bewertung von Lehrveranstaltungen auf. Hierbei wird ausgehend vom strukturfunktionalistischen Modell Talcott Parsons' ein Erklärungsmodell vorgeschlagen, in dessen Mittelpunkt ein ausbalanciertes Verhältnis unterschiedlicher Qualitätsdimensionen steht. Im Anschluss an die Darlegung der Operationalisierung des Modells werden Erklärungswert und Modellstabilität empirisch überprüft.

1. Ansätze der Entwicklung von Instrumenten der Lehrveranstaltungsevaluation

Die Entwicklung von Instrumenten zur Bewertung der Lehre lässt sich in Anlehnung an Marsh (1987) und Rindermann (2001) unterschiedlichen Vorgehensweisen bzw. Modi zuordnen. Danach lassen sich sechs Zugänge benennen, die in der Regel nicht trennscharf sind, sondern in der Praxis parallel Anwendung finden. Der (1)

synkretische Ansatz rekurriert auf bereits vorhandene Instrumente und adaptiert diese auf die jeweilige Situation bzw. Fragestellung. Der (2) Befragungs-Ansatz geht davon aus, dass die befragten Studierenden oder Lehrenden als Experten zumindest mit Blick auf ihre praktischen Erfahrungen zu werten sind und ein Eigenverständnis guter Lehre entwickeln, das im Rahmen der Befragung als Bewertungsmaßstab dienen kann. Der (3) theoretische Ansatz leitet relevante Bewertungsindikatoren aus (hochschuldidaktischen) Theorien bzw. Modellen der Lehr- und Lernforschung ab. Der (4) Experten-Ansatz basiert auf der Einschätzung anderer Forscherinnen und Forscher im Evaluationsfeld und macht sich deren Einschätzungen bspw. in Form von Literaturanalysen zum Forschungsstand zu eigen. Von einem (5) Autoren-Ansatz lässt sich sprechen, wenn die Entwicklung der Instrumente in Eigenregie erfolgt, ohne dass auf Modelle oder bereits vorhandene Instrumente Bezug genommen wird. Schließlich ist der (6) Konsens-Ansatz zu nennen, bei dem die Entwicklung der Bewertungskriterien und -instrumente im Sinne eines partizipativen Prozesses in Gremien oder Arbeitsgruppen erfolgt (vgl. zusammenfassend Hiltmann 2002, S. 28).

Betrachtet man die Praxis der Lehrveranstaltungsevaluation, so lassen sich – wenngleich die Ansätze selten in dieser idealtypischen Form alleine zu beobachten sind – drei dominierende Ansätze identifizieren. Zum einen ist dies der Befragungs-Ansatz, der insbesondere auf die Einschätzungen Studierender abstellt und bspw. in Form von Zufriedenheitsmaßen Lehrqualität zu messen versucht. Kontrovers diskutiert wird in diesem Zusammenhang, inwieweit Studierende als Experten zu werten sind, wenn sie zugleich Betroffene sind und ob die Wertungen valide sind, wenn unklar ist, welcher Maßstab sich hinter diesen verbirgt (vgl. u.a. Kromrey 2003).

Weit verbreitet und häufig überschneidend zum Befragungs-Ansatz ist der Konsens-Ansatz. Hier kommen in der Regel nicht in erster Linie Standards empirischer Sozialforschung zum Tragen, sondern Instrumente werden in Gremien entwickelt, deren Akteure unterschiedlich stark mit Fragen der Evaluations- oder Hochschulforschung befasst sind. "In many institutions, the instruments used to obtain student feedback have been constructed and developed in-house and may never have been subjected to any kind of external scrutiny." (Richardson 2005, S. 388) Wesentlich ist hierbei auch weniger die Validität der Instrumente als viel mehr deren Statusgruppen übergreifende Akzeptanz innerhalb der

Hochschule. Dies geht häufig mit dem zuvor genannten Autoren-Ansatz einher, beim dem die erfahrungsbasierte Plausibilität einzelner Items als Grundlage für die Entwicklung der Fragebogen im Vordergrund steht.

Schließlich ist der synkretische Ansatz weit verbreitet, indem auf bereits vorhandene Instrumente zurückgegriffen wird. Für den deutschen Sprachraum ist dies insbesondere das durch Rindermann und Amelang (1994) entwickelte Erhebungsinstrument, das sogenannte Heidelberger Inventar. Aber auch das von Gollwitzer und Schlotz (2002) entwickelte Trierer Inventar zur Lehrveranstaltungsevaluation (TRIL) findet vereinzelt Anwendung.

Betrachtet man die Ansätze der Lehrveranstaltungsevaluation insgesamt, so fällt auf, dass der theoretische Ansatz vergleichsweise selten vertreten ist. So folgt bspw. auch die Entwicklung des Heidelberger Inventars in erster Linie einem induktiven Ansatz, der auf die Einschätzungen von Lehrenden und Studierenden, nicht aber auf ein übergreifendes didaktisches oder theoretisches Konzept Bezug nimmt. Dies hat unterschiedliche Gründe. Wesentlich scheinen uns hierbei drei Aspekte, die zum Teil mit den vorherigen Ausführungen korrespondieren: Zum einen ist zu beobachten, dass die Lehrveranstaltungsbewertung aufgrund ihrer individuellen Ausrichtung nach wie vor einen sensiblen Bereich adressiert. Zwar sind die Akteure an Hochschulen aufgrund der permanenten Bewertung von Forschungsleistungen – sei es im Rahmen von Anträgen oder sei es im Kontext von Publikationen, Vorträgen und Qualifikationsarbeiten – gewohnt evaluiert zu werden, doch mit Blick auf die individuelle Lehrleistung zeigen sich insofern Unterschiede, als sich eine entsprechende Tradition erst allmählich herausbildet und die asymmetrische ‚Schüler-Lehrer-Beziehung‘ und die damit verbundenen Rollen in Frage gestellt werden. Hieraus resultieren vielschichtige Diskussionen innerhalb der Hochschulen und Fächer, die im Sinne des Konsens-Ansatzes nicht zuletzt der Intention entspringen, die „Spielregeln“ der Evaluation mit definieren zu können.

Zum zweiten folgt – wie zu Beginn angeführt – Hochschulevaluation häufig politischen Vorgaben, die vor allem auf die Funktion der Rechenschaftslegung abstellen. Evaluation wird in diesem Sinne als notwendiges Übel, nicht aber als Erkenntnis generierendes Instrument verstanden: „Es wird evaluiert, um eine an die Universität gestellte Forderung nach Evaluation zu erfüllen. Dass in einem solchen Fall möglichst wenig Aufwand getrieben wird und lediglich leicht zugängliche Informationen gesammelt werden, dass zudem vor allem „unproblematische“ Daten präsentiert werden [...], darf nicht verwundern“ (Kromrey 2003). Wenn aber die Verfahren nicht dazu dienen, Wissen über Ursachen und Wirkungen bspw. unterschiedlicher Lehrformate und Lernszenarien zu gewinnen, spielt letztlich auch die Qualität des Erhebungsinstrumentes eine nachgeordnete Rolle.

Schließlich zeichnet für die geringe Modellorientierung im Kontext der Lehrveranstaltungsevaluation zumindest im deutschen Hochschulraum die insgesamt geringe Bezugnahme von Evaluation und Hochschuldidaktik verantwortlich. Dies lässt sich zunächst auf die lange Zeit geringe empirische Orientierung der Hochschuldidaktik zurückführen, die erst allmählich an Gewicht gewinnt.

Zudem ist eine gewisse institutionelle Differenzierung zwischen Qualitätsmanagement, das primär auf Instrumente der Evaluation und Akkreditierung rekurriert, und Hochschuldidaktik festzustellen. Dies hat mit Blick auf die hier behandelte Fragestellung insbesondere zwei Effekte: Auf der einen Seite nimmt die Hochschuldidaktik Ergebnisse der Lehrveranstaltungsevaluation kaum wahr bzw. sieht diese als nur in geringem Maße dazu geeignet, um hochschuldidaktische Interventionen zu implementieren. Dies mag an dem partiell fehlenden empirischen Hintergrund derer, die im Bereich der Hochschuldidaktik agieren, aber ggf. auch an der Form der Aufbereitung der Befragungsergebnisse liegen, die häufig nicht über ein deskriptives Niveau in Form von pdf-Reports hinausgeht. Auf der anderen Seite erfolgt die Entwicklung der Befragungsinstrumente häufig ohne Beteiligung der Hochschuldidaktik, was zu dem diagnostizierten Defizit an theorie- oder modellorientierter Entwicklung der Erhebungsinstrumente beiträgt.

2. Modell der Lehrevaluation

Das Modell zur Erfassung guter Lehre orientiert sich an systemtheoretischen Erwägungen und insbesondere am strukturfunktionalistischen Paradigma Talcott Parsons'. Im Sinne eines handlungstheoretischen Ansatzes steht bei Parsons die Annahme im Vordergrund, dass sich Systeme und entsprechend auch Handlungssysteme an unterschiedlichen Anforderungen ausrichten müssen, um neuen Anforderungen genügen zu können. Als wesentliche Anforderungen lassen sich hierbei vier Systemebenen benennen, die als sogenanntes AGIL-Schema in die Literatur eingegangen sind: Adaption, Goal Attainment, Integration und Latent Pattern Maintenance (vgl. zusammenfassend u.a. Parsons 1996; Münch 1982; Parsons/Platt 1990).

Adaption lässt sich als Aufgabe fassen, das System zu öffnen und sich externen Bedingungen anzupassen, um zusätzliche bzw. neue Ressourcen für ein System zu gewinnen. Für den Bereich der Lehre handelt es sich hierbei um strukturelle Voraussetzungen im Sinne von angemessener sächlicher Ausstattung und ausreichenden personellen Ressourcen, sowie intellektuelle Potenziale, Wissensbestände und Fähigkeiten bspw. im Hinblick auf didaktische Kompetenzen. Als Indikatoren auf Ebene der Lehrveranstaltung sind mithin die Qualität der Räumlichkeiten und deren Ausstattung, die Ausbildung der Lehrenden und deren quantitative Relation zur Anzahl der Studierenden wie auch das Leistungsvermögen der Studierenden zu nennen.

Goal Attainment umschreibt den Grad der Zielerreichung. Die Definition bzw. das Vorhandensein erreichbarer Ziele wird bei Parsons als wesentlicher Motivationsfaktor gesehen, um überhaupt zu handeln. Ziele, die als nicht realisierbar erachtet werden, erzeugen auch keinen Impuls zu agieren. Für die Betrachtung von Studium und Lehre lässt sich diese Ebene entsprechend als Ergebnisebene fassen und operationalisieren in Form des Lernfortschrittes. Dieser wiederum betrifft im Sinne des Gewinns bzw. des Zuwachses von Kompetenzen alle vier Kompetenzfelder nach Weinert, d.h. Fachkompetenzen, Methodenkompetenzen, Sozialkompetenzen

und Selbstkompetenzen. Fachkompetenzen werden hierbei differenziert nach theoretischen, methodischen und praktischen fachlichen sowie fachübergreifenden Kompetenzen. Methodenkompetenzen umschreiben solche Fähigkeiten und Fertigkeiten, die auf fachübergreifende Aspekte abstellen, wie analytische Kompetenzen, die Fähigkeit wissenschaftlich zu schreiben oder Informations- und Wissensmanagement sowie Präsentationskompetenzen. Unter Sozialkompetenzen werden Teamfähigkeit, Konfliktfähigkeit und kommunikative Fähigkeiten zusammengefasst. Schließlich werden als Selbstkompetenzen die Fähigkeit selbstständig zu arbeiten, Belastungsfähigkeit und Selbstdisziplin gemessen. Idealerweise wären diese Kompetenzen in Form externer, vergleichender Überprüfung abzubilden. Im Rahmen der Lehrveranstaltungsevaluation wird aufgrund der zum Zeitpunkt der Befragung fehlenden Messung des Kompetenzzuwachses bzw. aufgrund fehlender Messung spezifischer Kompetenzbereiche, wie bspw. von Sozial- und Selbstkompetenzen, auf die subjektive Einschätzung der Befragten zurückgegriffen, die jeweils den Grad des wahrgenommenen Kompetenzgewinns angibt.

Integration lässt sich nach Parsons als Anforderung beschreiben, die im System vorhandenen Elemente sinnhaft miteinander zu verbinden. Auf der Ebene von Gesellschaften ist hiermit der Prozess der Institutionalisierung von Werten und gleichzeitig der Zusammenführung der Gesellschaftsmitglieder in sozialen Systemen gemeint. Überträgt man dies auf Lehren und Lernen, so sind alle Prozesse der Interaktion zwischen Lehrenden und Studierenden, d.h. Lehr-Lernprozesse adressiert. Als relevante Items zur Operationalisierung dieser Ebene eignen sich solche, die sich auf die didaktische Qualität wie auch auf Kommunikations- und Betreuungsprozesse beziehen.

Schließlich ist die Ebene der Latent Pattern Maintenance, der Bewahrung latenter Strukturen, jene, die übergrei-

Abbildung 2: Operationalisierung des AGIL-Schemas auf Ebene von Lehrveranstaltungen

<u>L=KULTUR-ZIELE</u> - Nachvollziehbarkeit der Ziele der Lehrveranstaltung - Leistungsanforderungen (Angemessenheit von Tempo, Menge der Lehrinhalte und Anspruch)	<u>I=INTEGRATION-PROZESSE</u> - Strukturierung der LV - Vergewisserung über Lerneffekte - Arbeitsmaterialien - Leistungsrückmeldung - Anregen zur kritischen Auseinandersetzung
<u>A=RESSOURCEN-STRUKTUREN</u> - Motivation des Lehrenden - Motivation und Interesse der Studierenden - Leistungsvermögen (ausreichendes Vorwissen, der Veranstaltung folgen können) - Räumlichkeiten, Ausstattung - Gruppengröße	<u>G=ZIELERREICHUNG-ERGEBNISSE</u> - Fachkompetenzen (fachspezifisch theoretisch, methodisch, praktisch, fachübergreifendes Wissen) - Methodenkompetenzen (analytisch, Präsentation, Informations-, Wissensmanagement, wiss. Schreiben) - Soziale Kompetenzen (Kommunikation, Konflikt-, Teamfähigkeit) - Selbstkompetenzen (Selbstdisziplin, Belastungsfähigkeit, selbst. Arbeiten)

fende Zielsetzungen eines Systems umschließt. Diese lassen sich weniger im Sinne konkreter Handlungsziele, sondern vielmehr als Leitideen und Werte benennen. Bezogen auf Studium und Lehre, ließen sich hier bspw. grundlegende Einstellungen zur Wissenschaft oder zum Studium und dessen Ausrichtung sowohl auf Seiten der Studierenden als auch der Lehrenden subsumieren. Diese lassen sich in Form von Einstellungselementen abbilden. Im Folgenden wird in einem ersten Schritt die Operationalisierung auf Ebene von Studium und Lehre insgesamt beispielhaft abgebildet (Abb. 1), um den Charakter der unterschiedlichen Systemebenen zu verdeutlichen. Anschließend wird die im Rahmen von Lehrveranstaltungsevaluationen genutzte Operationalisierung betrachtet (Abb. 2).

Betrachtet man die folgende Abbildung, in der die Items auf den vier Systemebenen zusammengefasst sind, so zeigt sich mit Blick auf die Operationalisierung auf Ebene von Lehrveranstaltungen insgesamt ein Übergewicht an Items auf den Ebenen der Lehr- und Lernprozesse (Integration), der Kompetenzerfassung (Goal Attainment) sowie der vorhandenen Strukturen und Ressourcen (Adaption), wogegen bislang die Ebene der Einstellungen bzw. der Kultur noch nicht hinreichend abgebildet ist. Dies – so viel sei vorweggenommen – wird auch durch die nachfolgende Analyse bestätigt.

Wie lässt sich nun nach Parsons die erfolgreiche Anpassung von Systemen an sich verändernde Umweltbedingungen erklären? Der zentrale Gedanke Parsons ist, dass dies am ehesten gelingt, wenn eine Balance im Hinblick auf die Anforderungen auf allen vier Systemebenen gegeben ist. Alleine die Verbesserung der Ressourcen ist demnach nur zielführend, wenn dies mit guten Lehr- und Lernprozessen, systemadäquaten Einstellungen und guten Lernergebnissen korrespondiert. Umgekehrt ist eine gute Lehr- und Lernkultur und sind elaborierte Prozesse nur dann erfolgversprechend, wenn sie mit ausrei-

Abbildung 1: Beispielhafte Operationalisierung des AGIL-Schemas in Studium und Lehre

<u>L=KULTUR-ZIELE</u> - Selbstverständnis - Standards (Definition guter Lehre und Leistungsstandards) - Programme und Profile (Schwerpunktsetzung) - Einstellungen zur Wissenschaft und zu Studium und Lehre	<u>I=INTEGRATION-PROZESSE</u> - Lehrorganisation - Didaktik - Betreuung und Beratung - Interne und externe Kooperation - Abstimmungsprozesse von Lehrinhalten und Leistungsanforderungen
<u>A=RESSOURCEN-STRUKTUREN</u> - Materielle Ressourcen (Räume, Raum- und Medienausstattung) - Personelle Ressourcen (Betreuungsrelation, Qualifikation etc.) - Wissensbestände, Leistungsvermögen auf Seiten von Studierenden	<u>G=ZIELERREICHUNG-ERGEBNISSE</u> - Kompetenzgewinn - Prüfungsergebnisse - Abbruchquoten - Persönlichkeitsentwicklung - Erfolgreiche Berufsfindung

chenden Ressourcen hinterlegt sind. Schließlich sind auch positive Lernergebnisse nach diesem Modell nicht per se positiv zu werten, wenn nicht zugleich auch bspw. ein angemessenes Grundverständnis universitärer Bildung vorhanden ist. So lässt sich bspw. die konsequente Output-Orientierung in der Forschung in Form möglichst vieler Publikationen oder der Kreierung neuer Forschungsergebnisse systemdestabilisierend interpretieren, wenn damit ein Mehr an Plagiaten bzw. wissenschaftlichem Fehlverhalten insgesamt verbunden ist und somit die Wissenschaftskultur und die durch die Scientific Community vereinbarten Standards gefährdet werden. Qualität von Forschung und Lehre manifestiert sich entsprechend in der gleichzeitigen oder ggf. auch sukzessiven Erfüllung unterschiedlicher Systemanforderungen.

Ausgehend von dieser Prämisse spiegelt die Gesamtbewertung einer Lehrveranstaltung die Balance unterschiedlicher Perspektiven, die durch die vier Systemebenen abgebildet sind. Unabhängig hiervon lassen sich selbstverständlich im Hinblick auf die Erklärung von Zusammenhängen und Wirkmechanismen auch einzelne Variablen insbesondere auf der Ebene der erlangten Kompetenzen als abhängige Variablen definieren, um den Einfluss und die Interdependenzen zwischen den unterschiedlichen Systemebenen und einzelnen Indikatoren abbilden zu können. Dies gilt insbesondere für Variablen auf der Ebene der Zielerreichung, wie den subjektiv wahrgenommenen Erwerb von Kompetenzen oder objektive Leistungen in Form von Noten.

3. Empirische Analyse

3.1 Stichprobe

Die Daten für die nachfolgende Analyse entstammen Lehrveranstaltungsevaluationen, die an der Johannes Gutenberg-Universität Mainz durchgeführt werden. Die Universität Mainz verfügt über ein zentrales Qualitätssicherungssystem, das Grundlage für die deutschlandweit erste erfolgreiche Systemakkreditierung war. Die Befragungen werden vom Zentrum für Qualitätssicherung und -entwicklung (ZQ) zentral für alle Fachbereiche in der Regel in einem dreisemestrigen Zyklus im Paper & Pencil-Verfahren durchgeführt. Hierbei werden nicht alle Veranstaltungen in die Untersuchung einbezogen, sondern die Lehrveranstaltungen werden nach bestimmten „Quotenmerkmalen“ ausgewählt. Quotenmerkmale für die hier vorliegende Stichprobe sind der Studiengang, die Lehrveranstaltungsform (Seminar, Übung) sowie die Studienphase in Form des Semesters, in dem die Veranstaltung im Curriculum verankert ist. Die Stichprobe umfasst insgesamt N=6532 Studierende aus dem Bereich der Geistes-, Sprach- und Kulturwissenschaften, die im Sommersemester 2012 an 254 verschiedenen Seminaren und Übungen teilgenommen haben. Von der gesamten Fallzahl entfallen N=4726 auf seminareilnehmende und N=1806 auf übungsteilnehmende Studierende.

3.2 Faktorenanalyse

Um die Stabilität des hier vorgeschlagenen Erklärungsmodells zu prüfen, wird in 2 Schritten verfahren: Zunächst wird die Anzahl der oben aufgeführten Varia-

blen reduziert, indem auf Grundlage von Faktorenanalysen¹ innerhalb der Subsysteme Indizes gebildet werden. Danach wird zur Überprüfung des Modells eine globale Faktorenanalyse über alle Indizes berechnet. Alle vorgenommenen Faktorenanalysen werden als Hauptkomponentenanalysen (Rotationsmethode: Varimax) berechnet, d.h. es wird angenommen, dass die eingesetzten Items erschöpfend die Varianz des jeweiligen Faktors erklären. Zudem ergibt sich vor dem Hintergrund der getroffenen Annahme ein exploratives Vorgehen: die Faktoren werden auf Grundlage ihrer Eigenwerte > 1 extrahiert.

3.2.1 Ressourcen- bzw. Strukturebene

Die Anwendung der Faktorenanalyse auf die oben benannten Ressourcen- bzw. Strukturitems (Adaption) ergibt eine 3-faktorielle Lösung (vgl. Tab. 1) mit einer Varianzaufklärung von insgesamt 64,6%. Davon verteilen sich 24,0% der Varianz auf den ersten Faktor, 21,5% der Varianz auf den zweiten Faktor und 19,1% der Varianz auf den dritten Faktor. Bei schwachen bis mittleren Korrelationen zwischen den eingesetzten Variablen ($0,202 < r < 0,732$) sind die Ladungen auf den Faktoren insgesamt als sehr gut zu betrachten. Allein das Item „Ich beteilige mich, soweit möglich, aktiv an der Veranstaltung (z.B. an Diskussionen; bei Fragen der Lehrkraft)“ korreliert nur schwach mit dem entsprechenden Faktor (,413). Gemäß dem Sphärentest nach Bartlett² ($p=,000$) und dem KMO-Kriterium (KMO-Wert = ,629) sind die Voraussetzungen der Faktorenanalyse akzeptabel³. Inhaltlich lassen sich die Komponenten „studentisches Vorwissen/Motivation“, „Lehrendengagement“ und „Rahmenbedingungen“ unterscheiden, die zu entsprechenden Indizes zusammengefasst werden.

Der Cronbach-Alpha-Wert⁴ liegt bei $\alpha=,641$ und ist als nicht ganz zufriedenstellend einzustufen. Generell sind Werte ab 0,8 erst akzeptabel, doch in der Praxis sind

¹ Die Faktorenanalyse ist ein statistisches Verfahren zur Datenreduktion, mit dessen Hilfe eine Vielzahl von Variablen auf eine geringere Anzahl von gemeinsamen „Faktoren“ zurückgeführt werden kann (Backhaus et al. 2003, S. 260).

² Bartlett-Sphärentest: Wenngleich in der Korrelationsmatrix der Tabelle 2 für zahlreiche Variablenpaare relativ starke Korrelationen ausgewiesen werden, so kann daraus nicht gefolgert werden, dass diese Zusammenhänge gleichfalls in der Grundgesamtheit vorliegen. Bei Vorliegen von Stichprobendaten ist es durchaus möglich, dass die ausgewiesenen Korrelationen lediglich ein zufälliges Stichprobenergebnis darstellen, obwohl die Variablen in der Grundgesamtheit unkorreliert sind. Ist dies der Fall, so werden mit einer Faktorenanalyse nur solche Faktoren extrahiert, die auf zufällige Gemeinsamkeiten der Variablen zurückzuführen sind und demzufolge einen geringen Erklärungsgehalt besitzen. Es erscheint daher angeraten, zunächst der Frage nachzugehen, ob die Stichprobenwerte aus einer Grundgesamtheit entstammen, in der die Variablen unkorreliert sind (Backhaus et al. 2003, S. 274).

³ Kaiser-Meyer-Olkin-Kriterium (KMO): Das KMO-Kriterium bildet ein zusammenfassendes Beurteilungsmaß für die faktorenanalytische Angemessenheit der Untersuchungsvariablen und kann sowohl für die Gesamtheit aller Variablen als auch für einzelne Variablen berechnet werden (Backhaus et al. 2003, S. 276). Die vorstehenden Überlegungen zusammenfassend, sind Variablen dann für eine Faktorenanalyse geeignet, wenn diese einerseits hoch mit anderen Variablen korrelieren und andererseits weitgehend durch die anderen Variablen erklärt werden können. Dies ist gleichbedeutend damit, dass die einfachen Korrelationskoeffizienten möglichst groß, die partiellen Korrelationskoeffizienten hingegen möglichst gering sind.

⁴ Mittels Reliabilitätsanalyse wurde die interne Konsistenz aller Itemskalen berechnet. Die interne Konsistenz einer Skala beschreibt das Ausmaß, in dem die Items der Skala miteinander in Beziehung stehen. Dabei ist der Cronbach-Alpha-Wert das am häufigsten verwendete objektive Maß zur Reliabilitätsbestimmung, der zwischen minus unendlich und eins liegt (Tavakol/Dennick 2011, S. 53). Es ist anzumerken, dass nur positive Cronbach-Alpha-Werte sinnvoll interpretierbar sind.

Tabelle 1: Diagramm der rotierten Faktorladungen (Adaption)

Ressourcen/Struktur (Adaption)	Komponente		
	1	2	3
1. Dem Dozenten/der Dozentin liegt etwas daran, sein/ihr Wissen zu vermitteln und den Teilnehmern etwas beizubringen	,178	,032	,908
2. Der Dozent/die Dozentin wirkt engagiert.	,096	,040	,922
3. Das Thema/die Themen der Veranstaltung hat/haben mich schon vor Beginn der Veranstaltung interessiert.	,689	,013	,101
4. Es fällt mir in der Regel leicht, die Inhalte dieser Lehrveranstaltung zu verstehen.	,865	,045	,053
5. Um der Veranstaltung folgen zu können, ist mein Vorwissen ausreichend.	,848	,076	-,006
6. Ich beteilige mich, soweit möglich, aktiv an der Veranstaltung (z.B. an Diskussionen; bei Fragen der Lehrkraft)	,413	-,025	,188
7. Die Räumlichkeiten (Größe, Lage, etc.) sind für die Veranstaltung angemessen.	,021	,877	-,009
8. Die Ausstattung des Raumes (Medien, Technik etc.) ist für die Veranstaltung angemessen.	,047	,832	,005
9. Die Gruppengröße ist meines Erachtens der Art und Anforderung der Veranstaltung angemessen.	,011	,678	,062

auch etwas niedrigere Cronbach-Alpha-Werte (um 0,7) zulässig (Schnell et al. 2005, S. 153). Bei der Betrachtung der korrigierten Item-Skala-Statistik zeigt sich, dass alle Items offensichtlich die Ressourcen- bzw. Strukturebene messen. Alle Items weisen eine Trennschärfe zur Gesamtskala $>0,2$ auf, d.h. die Items korrelieren gut mit der Gesamtskala und können daher beibehalten werden

3.2.2 Zielerreichungs- bzw. Ergebnisebene

Auf der Zielerreichungs- bzw. Ergebnisebene (Goal Attainment) werden über die Faktorenanalyse 4 Komponenten extrahiert (vgl. Tab. 2), welche 64,0% der Gesamtvarianz erklären. Davon verteilen sich 16,8% der Varianz auf den ersten Faktor, 16,7% der Varianz auf den zweiten Faktor, 15,7% der Varianz auf den dritten Faktor und 14,7% der Varianz auf den vierten Faktor. Bei schwachen bis mittleren Korrelationen zwischen den eingesetzten Variablen ($,260 < r < ,660$) sind die Ladungen auf den 4 Faktoren mit Werten über 0,5 als gut zu betrachten. Entsprechend dem Sphärentest nach Bartlett ($p=,000$) und dem KMO-Kriterium (KMO-Wert = ,879) sind die Voraussetzungen der Faktorenanalyse sehr gut erfüllt. Inhaltlich lassen sich die oben bereits beschriebenen vier Kompetenzbereiche: Fach-, Methoden-, Sozial- und Selbstkompetenz differenzieren, die im Folgenden zu entsprechenden Indizes zusammengefasst werden.

Der Cronbach-Alpha-Wert der Ergebnisebene liegt bei $\alpha=,878$ und ist beinahe als exzellenter Wert anzusehen. Die korrigierte Item-Skala-Statistik zeigt hier, dass die genutzten Items die Zielerreichungs- bzw. Ergebnisebene messen, da alle Items gut mit der Gesamtskala korrelieren ($>0,2$).

3.2.3 Integrations- bzw. Prozessebene

Die Anwendung der Faktorenanalyse auf die oben benannten Integrations- bzw. Prozessitems (Integration) ergibt eine Lösung mit nur einem Faktor (vgl. Tab. 3) bei einer Varianzaufklärung von insgesamt 55,1%. Bei schwachen bis mittleren Korrelationen zwischen den eingesetzten Variablen ($,372 < r < ,605$) sind die Ladungen auf dem Faktor mit Werten über 0,6 als sehr gut zu betrachten. Entsprechend dem Sphärentest nach Bartlett ($p=,000$) und dem KMO-Kriterium (KMO-Wert = ,893) sind die Voraussetzungen der Faktorenanalyse sehr gut erfüllt. Da nur eine Komponente extrahiert wurde, konnten alle

Items zu einem Index zusammengefasst werden.

Die Reliabilitätsprüfung ermittelte einen Cronbach-Alpha-Wert von $\alpha=,860$, der ebenfalls als nahezu exzellenter Wert einzustufen ist. Auch hier verweist die korrigierte Item-Skala-Statistik darauf, dass die Items mit der Gesamtskala gut korrelieren ($>0,2$) und die Integrations- bzw. Prozessebene gut messen.

3.2.4 Kultur- bzw. Zielebene

Für die Dimension Kultur bzw. Ziele (Latent Pattern Maintenance) werden über die Faktorenanalyse 2 Komponenten extrahiert (vgl. Tab. 4), welche 57,2% der Gesamtvarianz erklären. Dabei erklärt der erste Faktor 30,1% der Varianz und der zweite Faktor 27,1% der Varianz. Bei schwachen bis mittleren Korrelationen zwischen

Tabelle 2: Diagramm der rotierten Faktorladungen (Goal Attainment)

Zielerreichung/Ergebnis (Goal Attainment)	Faktoren			
	1	2	3	4
1. Fachspezifische theoretische Kenntnisse	-,174	,354	,207	,624
2. Fachspezifische methodische Kenntnisse	,255	,116	,123	,811
3. Praktische fachspezifische Fähigkeiten	,346	,068	,116	,721
4. Fächerübergreifendes Wissen	,225	,506	,111	,337
5. Wissenschaftliches Schreiben : Fähigkeit, Berichte, Protokolle oder wissenschaftliche Aufsätze zu verfassen	,119	,595	,225	,223
6. Analytische Fähigkeiten : Fähigkeit, ein komplexes System in Teile zu zerlegen und die Beziehungen zwischen den einzelnen Teilen zu analysieren	,049	,606	,226	,355
7. Präsentationsfähigkeit : Fähigkeit, Produkte, Ideen oder Berichte einem Publikum zu präsentieren	,527	,636	,033	-,117
8. Informations- und Wissensmanagement : Fähigkeit, Informationen aus verschiedenen Quellen zu erschließen und aufzubereiten	,270	,750	,230	,011
9. Teamfähigkeit : Fähigkeit, mit anderen produktiv und kooperativ zusammenzuarbeiten	,793	,155	,148	,110
10. Kommunikative Fähigkeiten : Kenntnisse und Fertigkeiten der wissenschaftlichen Gesprächsführung	,674	,282	,223	,254
11. Konfliktfähigkeit : Fähigkeit und Bereitschaft, eine Auseinandersetzung aufzunehmen, auszuhalten und konstruktiv zu bewältigen	,720	,176	,267	,204
12. Selbstdisziplin : Fähigkeit, konsequent seine Ziele zu verfolgen und sich nicht von anderen Dingen ablenken zu lassen	,230	,166	,829	,157
13. Belastungsfähigkeit : Fähigkeit, unter physischen und psychischen Belastungen arbeiten zu können	,252	,156	,798	,076
14. Selbständiges Arbeiten	,075	,270	,705	,191

Tabelle 3: Diagramm der rotierten Faktorladungen (Integration)

Integration/Prozess (Integration)	Komponente	
	1	
1. Der Dozent/die Dozentin strukturiert die Veranstaltung; es ist ein roter Faden erkennbar.		,724
2. Der Dozent/die Dozentin gibt Beispiele, die zum Verständnis der Lehrinhalte beitragen.		,791
3. Der Dozent/die Dozentin vergewissert sich in der Veranstaltung, ob die behandelten Themen verstanden wurden.		,778
4. Der Dozent/die Dozentin gab mir eine Rückmeldung zu meinen Leistungen, die mir weitergeholfen hat.		,679
5. Ich finde, Arbeitsmaterialien (z.B. Handouts, Skripte, Literaturhinweise), die der/die Dozent/-in bereitstellt, hilfreich.		,706
6. Die Zufriedenheit der Studierenden mit der Lehrveranstaltung scheint dem Dozenten/der Dozentin wichtig zu sein.		,808
7. Der Dozent/die Dozentin regt zur kritischen Auseinandersetzung mit den behandelten Themen an.		,702

den eingesetzten Variablen ($300 < r < 500$) sind die Ladungen auf beiden Faktoren mit Werten über ,6 als gut zu betrachten. Entsprechend dem Sphärentest nach Bartlett ($p=,000$) und dem KMO-Kriterium (KMO-Wert = ,669) sind die Voraussetzungen der Faktorenanalyse mittelmäßig erfüllt. Inhaltlich lässt sich die Zielkomponente „Kreditierung und Arbeitsaufwand“ von der Zielkomponente „Passung von Modul- und Veranstaltung/Transparenz“ unterscheiden, die schließlich zu Indizes zusammengefasst werden.

Der Cronbach-Alpha-Wert liegt bei $\alpha=,589$ und gibt an, dass die interne Konsistenz der Skala der Items als schlecht zu bewerten ist. Dennoch zeigt die Betrachtung der korrigierten Item-Skala-Statistik, dass die Items mit der Gesamtskala gut korrelieren ($>0,2$). Das bedeutet, dass die verwendeten Items zwar die Kultur- bzw. Zielebene messen, aber die Beziehung der Items zueinander eher schlecht zu bewerten ist. Daraus lässt sich ableiten, dass das zu Anfang beschriebene Problem die Kultur- bzw. Zieldimension adäquat zu erfassen durch die vorgeschlagene Operationalisierung nicht umfassend gelöst werden kann.

3.2.5 Fazit

Die Reliabilitätsanalysen zeigen, dass die Items der Prozess-/Integrations- sowie Zielerreichungs-/Ergebnisebene gut geeignet sind, um Indizes zu bilden. Die Cronbach-Alpha-Werte liegen jeweils bei nahezu 0,9 und sind als beinahe exzellente Werte einzustufen. Die jeweiligen Ebenen werden durch die vorgeschlagenen Items gut erfasst.

Tabelle 4: Diagramm der rotierten Faktorladungen (Kultur/Ziele)

Kultur/Ziel (Latent Pattern Maintenance)	Komponenten	
	1	2
1. Der Dozent/die Dozentin stellt die Ziele der Veranstaltung nachvollziehbar dar.	,154	,645
2. Die Menge an behandeltem Lehrstoff/die Breite an Themen ist ...	,834	,092
3. Das Tempo der Lehrveranstaltung ist ...	,780	,107
4. Ich finde, die Lerninhalte dieser Veranstaltung sind auf andere Veranstaltungen innerhalb dieses Moduls abgestimmt.	,015	,772
5. Einmal alle Lehrveranstaltungen dieses Moduls zusammengekommen, finden Sie die Anforderungen sind ...	,687	,073
6. Inwieweit trifft auf Sie die Aussage zu, dass die Veranstaltungen des Moduls gut auf die Modulabschlussprüfung vorbereiten?	,088	,765

Ein anderes Bild ergibt sich auf der Ressourcen-/Struktur- bzw. Ziel-/Kulturebene. Hier liegen die Cronbach-Alpha-Werte jeweils unter 0,7 und sind als weniger zufriedenstellend zu bewerten. Insbesondere auf der Ziel-/Kulturebene weist der Wert darauf, dass die verwendeten Items diese Dimension nicht adäquat erfassen. Dieser empirische Beleg korrespondiert mit der zuvor angedeuteten Vermutung, dass die Ziel-/Kulturebene bislang noch nicht erschöpfend operationalisiert ist.

Die Faktorenanalysen belegen auf allen vier Modellebenen jeweils eine Gesamtvarianzaufklärung zwischen 55 – 64%. Das KMO-Kriterium verweist in allen Analysen darauf, dass die Voraussetzungen zur Berechnung einer Faktorenanalyse entweder akzeptabel (Ressourcen-/Strukturebene; Ziel-/Kulturebene) oder sehr gut (Integrations-/Prozessebene; Zielerreichungs-/Ergebnisebene) sind. Die Analysen zeigen, dass hinter den extrahierten Faktoren Variablen/Items stehen, die ähnliche Informationen erfassen, wenngleich der KMO-Wert der Analyse auf der Ressourcen-/Strukturebene und der Ziel-/Kulturebene nur akzeptabel ist. Korrespondierend mit den Reliabilitätsanalysen auf diesen Ebenen wird deutlich, dass wiederum die Ziel-/Kulturebene bislang nicht erschöpfend operationalisiert ist. Insgesamt betrachtet, ist es dennoch möglich auf dieser Grundlage innerhalb der Modellebenen Indizes zu bilden und diese empirisch zu überprüfen.

3.3 Modellpassung

Die an die vorherige Bildung der Indizes anschließende Faktorenanalyse über alle Subsysteme hinweg ergibt eine 3-faktorielle Lösung (vgl. Tab. 5) mit einer Varianzaufklärung von insgesamt 61,1%. Davon verteilen sich 23,9% der Varianz auf den ersten Faktor, 23,6% der Varianz auf den zweiten Faktor und 13,5 % der Varianz auf den dritten Faktor. Bei teilweise sehr schwachen bis mittleren Korrelationen zwischen den eingesetzten Variablen ($0,078 < r < ,756$) sind die Ladungen auf den Faktoren mit Werten über ,546 als gut zu betrachten. Entsprechend dem Sphärentest nach Bartlett ($p=,000$) und dem KMO-Kriterium (KMO-Wert = ,820) sind die Voraussetzungen der Faktorenanalyse auch hier sehr gut erfüllt.

Inhaltlich lassen sich mehrere Schlussfolgerungen ziehen: Zunächst wird deutlich, dass entgegen der durch das Modell getroffenen Annahme nur 3 statt 4 Dimensionen empirisch differenzierbar sind. Die Indizes auf der Kultur-/Zielebene verteilen sich sowohl auf die Ressourcen-/Struktur- als auch auf die Integrations-/Prozessebene. Auf Ressourcen-/Strukturebene werden allein die materiellen

Tabelle 5: Diagramm der rotierten Faktorladungen (Modell)

Indizes	Komponente		
	Prozess	Ergebnis	Struktur
Kreditierung und Arbeitsaufwand	,236	-,043	,717
Passung von Modul- und Veranstaltung	,684	,229	,220
Integration/Prozess	,860	,293	,105
Fachkompetenz	,430	,546	,116
Methodenkompetenz	,207	,781	,133
Sozialkompetenz	,135	,786	,169
Selbstkompetenz	,157	,792	-,143
Lehrengagement	,863	,121	,001
studentische/s Vorwissen/Motivation	,321	,237	,500
Rahmenbedingungen	-,103	,053	,671

(Rahmenbedingungen z.B. Räumlichkeiten, Ausstattung) und die studentischen Ressourcen (Vorkenntnisse, Motivation) als ein Faktor ausgegeben, wohingegen die Voraussetzungen/Ressourcen auf Seiten der Lehrenden der Integrations-/Prozessebene zugeordnet werden müssen. Vollständig wird allein die Zielerreichungs-/Ergebnisebene – wie durch das Modell vorhergesagt – auch empirisch repräsentiert. Die Vermischung von Integration/Prozess und Ressourcen/Struktur lässt sich zum Teil über den Modus der Lehrevaluation erklären: Die im Sinne einer Veranstaltungsqualität interpretierte Lehrqualität könnte dem entgegen aus Perspektive der Studierenden in erster Linie als Dozierenden-Evaluation wahrgenommen werden. Dozierende wiederum erscheinen als eine „Einheit“, die im Rahmen der Bewertung durch Studierende keine analytische Trennung erfährt und lediglich von den Rahmenbedingungen und den eigenen Voraussetzungen, die als relativ unabhängig und faktisch gegeben wahrgenommen werden, abgegrenzt werden. Auf der anderen Seite lässt sich auf Grundlage des Modells argumentieren, dass eine trennscharfe Abgrenzung zwischen den Struktur- und Prozessvariablen im Bezug auf die Dozierenden gerade da nicht möglich ist, wo gute Prozesse (Didaktik, Veranstaltungsdurchführung) Folge der strukturellen Voraussetzungen des Dozierenden sind, wie bspw. die Motivation und das Engagement des Dozierenden. Dies wäre auch im Sinne Parsons nachvollziehbar, der nicht von einer überschneidungsfreien Differenzierung von Subsystemen, sondern von deren gegenseitiger Durchdringung und Anschlussfähigkeit ausgeht.

Die fehlende Passung der Kultur-/Zieldimension als eigenständiger Faktor kann mehrere Gründe haben. Einerseits bezog sich die Interpretation dieser Dimension bislang in erster Linie auf den Aspekt der Ziele, weniger auf jenen der Kultur, der im Sinne Parsons als wesentlich relevanter zu erachten ist. Die Ebene der Ziele im engeren Sinne weist gemäß der konkreten Ziele bei Parsons starke Überschneidungen mit der Ebene des Goal Attainment auf, da hier die Definition erreichbarer Ziele zugleich zentraler Aspekt von Handlungsmotivation ist. Zum zweiten bezieht sich die bisherige Definition der Items auf dieser Ebene vor allem auf indirekte Faktoren der Zielqualität, wie die Transparenz von Zielen oder die Nachvollziehbarkeit und Angemessenheit von inhaltlichen Strukturen und Anforderungen. Diese wiederum können nachvollziehbarerweise als Dozierenden- oder auch Ressourcen-Strukturvariablen interpretiert werden.

Eine mögliche Lösung könnte sein, die konkreten inhaltlichen Ziele jeder Veranstaltung zu explizieren und deren Grad der Zielerreichung zu messen, was allerdings im Rahmen einer umfassenden Veranstaltungsevaluation kaum leistbar ist. Zudem stellt sich die Frage, ob nicht im Sinne Parsons die Ebene der Latent Pattern Maintenance weniger konkrete Zieldefinitionen als viel mehr Aspekte eines kulturellen Verständnisses von Lehre und Studium im Fokus stehen sollten.

Entsprechend wären verstärkt Items einzubinden, die subjektive Einstellungen, Ziele und Werte auf Seiten der Studierenden zu Studium, Wissenschaft und Bildung erfassen.

3.4 Multivariate lineare Regression

Ausgehend von den Faktorenanalysen und den daraus gebildeten Indizes, sowie der Prämisse, dass die Gesamtbewertung einer Lehrveranstaltung das Zusammenspiel und die Balance der vier Systemebenen widerspiegelt, wurde im folgenden Abschnitt mit Hilfe linearer Regressionsverfahren geprüft, in welcher Form und welcher Stärke die einzelnen Ebenen des theoretischen Modells die Gesamtzufriedenheit erklären.

Die abhängige Variable Gesamtzufriedenheit wurde hierbei mittels des Items „Insgesamt bin ich mit der Veranstaltung zufrieden“ sowie einer 7-stufigen Likert-Skala mit den Ausprägungen 1 „trifft gar nicht zu“ bis 7 „trifft voll und ganz zu“ operationalisiert. In der vorliegenden Stichprobe liegt der Mittelwert der Zufriedenheit bei 5,53 (SD = 1,55).

Neben der Zufriedenheit als abhängiger Variable und den oben gebildeten Indizes (Ziel-/Kulturebene: Kreditierung und Arbeitsaufwand, Passung von Modul und Veranstaltung, Integrations-/Prozessebene: Integration/Prozess, Zielerreichungs-/Ergebnisebene: Fachkompetenzen, Methodenkompetenzen, Sozialkompetenzen, Selbstkompetenzen, Ressourcen-/Strukturebene: Lehrerengagement, studentische/s Vorwissen/Motivation, Rahmenbedingungen) gehen Soziodemographika (Geschlecht, Fachsemester, Master vs. Bachelorstudiengang, Hauptfach vs. Nebenfach) als Kontrollvariablen in das Regressionsmodell ein. Die Annahmen des Regressionsmodells wurden geprüft.⁵

Nach listenweisem Fallausschluss ergab sich für die Regression eine Fallzahl von N = 3380. Das Gesamtmodell sagt die Zufriedenheit der Studierenden signifikant voraus und kann insgesamt 60% der Varianz der Gesamtzufriedenheit erklären. Die im Rahmen der Lehrveranstaltungsevaluation gemessenen Dimensionen können somit einen Großteil der abhängigen Variable vorhersagen. Den stärksten Einfluss mit einem standardisierten Beta-Koeffizienten von 0,365 hat hierbei der Index *Integration/Prozess*. Erhöht sich dieser Index um eine Einheit,

⁵ Störgrößen sind normalverteilt (optische Prüfung), keine Heteroskedastizität (optische Prüfung), keine Autokorrelation (Durbin-Watson-Koeffizient = 1,8), keine Multikollinearität (alle VIF-Indizes < 4)

Tabelle 6: Regressionsanalyse: Gesamtzufriedenheit mit der Lehrveranstaltung

Abhängige Variable: Gesamtzufriedenheit mit der Lehrveranstaltung	Nicht standardisierte Koeffizienten		Standard- disierte Koeffi- zienten	t
	B	Standard- fehler	Beta	
(Konstante)	-3,231***	,152		-21,28
Kreditierung/Arbeitsaufwand	,374***	,029	,154***	13,073
Passung von Modul und Veranstaltung	,123***	,018	,093***	6,738
Integration/Prozess	,566***	,030	,365***	19,001
Fachkompetenzen	,140***	,017	,112***	8,404
Methodenkompetenzen	,016	,017	,014	,953
Sozialkompetenzen	-,016	,014	-,016	-1,144
Selbstkompetenzen	,082***	,014	,077***	5,721
Lehrerengagement	,206***	,025	,138***	8,388
studentische/s Vorwissen/Motivation	,259***	,016	,196***	16,148
Rahmenbedingungen	,015	,012	,013	1,200
Geschlecht	-,012	,041	-,003	-,302
Fachsemester	-,011	,008	-,018	-1,365
Hauptfach	-,080*	,036	-,024*	-2,197
Master	,038	,054	,009	,711
* p<0,05, ** p<0,01, ***p<0,001	N=3380 R ² =0,60			

steigert sich der Wert der Gesamtzufriedenheit um 0,566 Punkte. Hinter dem Prozessindex verbergen sich, wie oben sowohl theoretisch als auch statistisch anhand der Faktoranalysen ermittelt, einerseits Variablen, die die didaktische Qualität, andererseits aber auch Variablen, die Kommunikations- und Betreuungsprozesse abbilden.

Beide Indizes der *Kultur-/Zielebene* weisen einen signifikanten Einfluss auf die Gesamtzufriedenheit auf. Mit einem standardisierten Beta von 0,154 stellt der Index „Kreditierung/Arbeitsaufwand“, welcher die Angemessenheit von Breite und Menge des Lehrstoffs sowie des Tempos der Lehrveranstaltung misst, den stärkeren Einflussfaktor dar.

Auf *Ressourcen-/Strukturebene* wirken sich vor allem studentenbezogene Ressourcen (Motivation, Vorkenntnisse und Interesse), aber auch dozentenbezogene Variablen (Engagement und Motivation des Lehrenden) signifikant auf die abhängige Variable aus. Keinen bedeutsamen Einfluss zeigen hingegen strukturelle Rah-

menbedingungen wie Gruppengröße, Ausstattung des Raumes oder die Angemessenheit der Räumlichkeiten. Auch dies deutet darauf hin, dass die Lehrveranstaltung von Studierenden stärker als Dozierenden-Evaluation betrachtet wird und die Rahmenbedingungen einen vergleichsweise geringeren Einfluss auf die Gesamtzufriedenheit mit der Lehrveranstaltung haben.

Auch auf *Zielerreichungs-/Ergebnisebene* können signifikante Zusammenhänge zur Zufriedenheit festgestellt werden. Den stärksten Effekt hat hierbei die subjektiv wahrgenommene Verbesserung der Fachkompetenz. Zusätzlich nachweisbar ist ein, wenn auch geringer, dennoch signifikanter Einfluss des wahrgenommenen Zuwachses an Selbstkompetenz. Als nicht relevant erweisen sich hingegen der wahrgenommene Gewinn an Methoden- und Sozialkompetenzen.

Die Kontrollvariablen Geschlecht, Fachsemester und Art des Studienabschlusses zeigen keinen signifikanten Einfluss auf die Zufriedenheit. Nur die Variable *Hauptfach* zeigt zwar einen signifikanten, jedoch nur sehr niedrigen Effekt. Studierende im Hauptfach sind tendenziell etwas unzufriedener mit der Lehrveranstaltung als Studierende im Nebenfach.

Insgesamt zeigt sich, dass die im Befragungsinstrument eingesetz-

ten Items und die aus ihnen gebildeten Indizes die Zufriedenheit der Studierenden mit der Lehrveranstaltung zu einem großen Teil erklären können ($R^2 = 0,60$). Geht man davon aus, dass die Gesamtzufriedenheit die perfekte Balance der unterschiedlichen Perspektiven des AGIL-Schemas im Sinne einer „guten“ Lehre widerspiegelt, konnten die theoretischen Vorüberlegungen des „Mainzer“-Modells zu großen Teilen bestätigt werden, insofern sich auf allen vier Ebenen signifikante Effekte auf die Gesamtzufriedenheit zeigen.

4. Zusammenfassung und Ausblick

Unseren Ausgangsüberlegungen lag zugrunde, dass das Modell zur Erfassung guter Lehre sich besonders am strukturfunktionalistischen Paradigma Talcott Parsons orientiert. Der Fokus des Modells liegt darauf, dass sich (Handlungs-)Systeme stets an (externen) Anforderungen orientieren müssen, um dadurch neuen Anforderungen gerecht zu werden. Die zentrale Anforderung dabei ist in

vier Dimensionen differenzierbar: Adaption, Goal Attainment, Integration und Latent Pattern Maintenance. Bezogen auf die Lehrveranstaltungsevaluation lassen sich die Dimensionen spezifizieren: Ressourcen-/Struktur (Adaption), Zielerreichung/Ergebnis (Goal Attainment), Integration/Prozess (Integration), Ziel/Kultur (Latent Pattern Maintenance). Rückblickend auf die Operationalisierung der vier Modellebenen zeigt sich, dass insbesondere die Ziel-/Kulturebene bislang unterrepräsentiert ist.

Die empirische Überprüfung des theoretischen Ausgangsmodells zeigt, dass eine Balance im Hinblick auf die Anforderungen der vier Ebenen bisher noch nicht vollständig gegeben ist. Gerade die Ziel-/Kulturdimension erweist sich insbesondere im Hinblick auf die Kulturkomponente als nicht hinreichend erfasst.

Den zentralen Ergebnissen der Regressionsanalyse kann entnommen werden, dass sich der Zufriedenheitsgrad mit der Lehrveranstaltung, wenngleich unterschiedlich stark, durch Merkmale auf allen vier Systemebenen erklären lässt. Den größten signifikanten Einfluss auf die Gesamtzufriedenheit mit der Lehrveranstaltung hat die Prozessebene ($\beta=,365$), gefolgt von strukturellen Bedingungen seitens der Studierenden ($\beta=,196$), der Kreditierung und dem Arbeitsaufwand ($\beta=,154$) sowie den strukturellen Bedingungen seitens der Dozierenden ($\beta=,138$). Hingegen konnte bei den Indizes zum subjektiv empfundenen Kompetenzzuwachs (Fach- und Selbstkompetenzen) auf der Zielerreichung-/Ergebnisebene sowie zur Passung von Modul und Veranstaltung auf der Ziel-/Kulturebene ein geringer, aber dennoch signifikanter Einfluss auf die Gesamtzufriedenheit mit der Lehrveranstaltung nachgewiesen werden. Externe Rahmenbedingungen wie bspw. die Gruppengröße, die Ausstattung oder die Räumlichkeiten haben nach den vorliegenden Ergebnissen keinen Einfluss auf die Gesamtzufriedenheit der Studierenden mit der Lehrveranstaltung. Dieses Ergebnis deutet darauf hin, dass die Lehrveranstaltungsevaluation von den Studierenden stärker als Dozierenden-Evaluation wahrgenommen wird. Insgesamt ist zu konstatieren, dass die gebildeten Indizes die Gesamtzufriedenheit mit der Lehrveranstaltung mehrheitlich erklären können. Jede Ebene des theoretischen Ausgangsmodells hat einen mehr oder weniger starken Einfluss auf die Zufriedenheit der Studierenden mit der Lehrveranstaltung. Eine Weiterentwicklung des Erhebungsinstrumentes, insbesondere auf der Ziel-/Kulturebene ist allerdings erforderlich. Zur Erfassung dieser Ebene sollte der Fokus auf Items und Fragen liegen, die u.a. auf Einstellungen der Studierenden zu Wissenschaft, Studium und zur

Universität abstellen und somit in deutlich stärkerem Maß als bisher den Bereich der Kultur- und Wertvorstellungen abdecken.

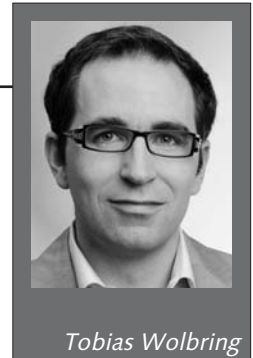
Literaturverzeichnis

- Backhaus, B./et al. (2003): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Berlin/Heidelberg/New York.
- Gollwitzer, M./Schlotz, W. (2002): Das „Trierer Inventar zur Lehrveranstaltungsevaluation“ (TRIL): Entwicklung und erste testtheoretische Erprobungen. In G. Krampen; H. Zayer (Hg.): Psychologiedidaktik und Evaluation IV, Bonn, S. 114-128.
- Hiltmann, M. (2002): Die Erfassung des Konstruktes „guter Lehre“ in Fragebogen-Verfahren (Mainzer Beiträge zur Hochschulentwicklung Bd. 6). Mainz.
- Kromrey, H. (2003): Qualität und Evaluation im System Hochschule. In: Stockmann, Reinhard (Hg.): Evaluationsforschung. 2. Aufl., Opladen, S. 233-258.
- Marsh, H. W. (1987): Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. In: Journal of Educational Research 11, S. 253-388.
- Münch, R. (1982): Theorie des Handelns. Zur Rekonstruktion der Beiträge von Talcott Parsons, Emile Durkheim und Max Weber. Frankfurt/M.
- Parsons, T. (1996): Das System moderner Gesellschaften, 4. Aufl. München
- Parsons, T./Platt, G. M. (1990): Die amerikanische Universität. Frankfurt a.M.
- Richardson, J. T. E. (2005): Instruments for obtaining student feedback: a review of the literature. In: Assessment & Evaluation in Higher Education 30, S. 387-415.
- Rindermann, H. (2001): Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Landau.
- Rindermann, H./Amelang, M. (1994): Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE). Handanweisung. Heidelberg.
- Schmidt, U. (2005): Zwischen Messen und Verstehen Anmerkungen zum Theoriedefizit in der deutschen Hochschulevaluation (<http://www.forschungsnetzwerk.at/downloadpub/messen%20und%20verstehen.pdf>)
- Schmidt, U. (2008): Aufbau, Funktionsweisen, Effekte und Wirkungsgrenzen einer systematischen hochschuleigenen Qualitätssicherung. In: Handbuch Qualität in Studium und Lehre, 19. Ergänzungslieferung, E 9.5., S. 1-22. Stuttgart.
- Schnell, R. et al. (2005): Methoden der empirischen Sozialforschung. München/Wien.
- Tavakol, M./Dennick, R. (2011): Making sense of Cronbach's alpha. International Journal of Medical Education, Number 2: S. 53-55.

■ **Dr. Uwe Schmidt**, Leiter Zentrum für Qualitätssicherung und -entwicklung, Johannes Gutenberg-Universität Mainz, E-Mail: Uwe.Schmidt@zq.uni-mainz.de

■ **Nicole Becker, Holger Lübke, Marc Hümpfner, Franziska Schmidt, Natalia Yugay und Cathrin Neßler**, wissenschaftliche Mitarbeiter/innen, Zentrum für Qualitätssicherung und -entwicklung, Johannes Gutenberg-Universität Mainz

Tobias Wolbring



Tobias Wolbring

Wie valide sind studentische Lehrveranstaltungsbewertungen?¹

Sachfremde Einflüsse, studentische Urteilerstandards, Selektionseffekte

In jüngster Zeit wird vermehrt diskutiert, studentische Lehrveranstaltungsevaluationen zur Setzung von Anreizen für ‚gute Lehre‘ einzusetzen. Grundvoraussetzung für die Implementierung eines solchen Anreizsystems ist, dass sich Lehrqualität mit diesem Instrument valide und fair messen lässt. Die Gültigkeit dieser Bedingung ist jedoch in der umfangreichen Fachliteratur zu dem Thema durchaus umstritten. Im vorliegenden Beitrag wird anhand eigener empirischer Untersuchungen ein Beitrag zu dieser Diskussion geleistet und auf bisher vernachlässigte Punkte aufmerksam gemacht. Die Analysen deuten auf Einflüsse von Dozierendenattraktivität und Notengebung, auf individuelle Unterschiede in der studentischen Urteilsstrenge und -differenziertheit sowie auf Selektionseffekte hin. Studentische Lehrveranstaltungsevaluationen bilden damit die Lehrleistung nicht unverzerrt ab, sondern spiegeln auch den Veranstaltungskontext und -verlauf wider. Dies legt einen auf Leitungsebene zurückhaltenden Umgang mit den Ergebnissen von Lehrveranstaltungsevaluationen nahe. Im Gegensatz dazu wird dafür plädiert, Lehrveranstaltungsbeurteilungen verstärkt als Feedback-Instrument zu nutzen.

1. Einleitung

Lehrveranstaltungsevaluationen (LVE) sind nicht nur im angloamerikanischen Hochschulwesen fest etabliert, sondern haben auch an europäischen Hochschulen mittlerweile weite Verbreitung gefunden. Ziel des Einsatzes dieses Instruments ist dabei einerseits, Informationen für die Lehrenden bereitzustellen und somit einen Impuls für die Verbesserung der Lehre zu geben. Andererseits wird mit der Befragung der Studierenden aber auch das hochschulpolitische Ziel verfolgt, entscheidungsrelevante Informationen für Bewerbungsverfahren und die Vergabe von Prämien und Lehrpreisen zu sammeln. Insbesondere in den letzten Jahren sind in der deutschen Wissenschaftslandschaft verstärkt Forderungen zu vernehmen, LVE für diesen zweiten Zweck zu nutzen. Eine entsprechende Implementierung leistungsabhängiger Anreizmechanismen setzt natürlich voraus, dass LVE die Lehrqualität valide messen und ein kausaler Wirkungsnachweis des Treatments ‚Lehre‘ mittels LVE gelingt. Wenn finanzielle Entlohnungen und weitere Handlungsanreize von derartigen Messungen abhängen, sollte Klarheit darüber bestehen, ob studentische Qualitätseinschätzungen belastbare Aussagen liefern oder

nicht. Schließlich sollten Einzelpersonen anhand meritokratischer Prinzipien entlohnt werden und nicht anhand von Kriterien, welche von zufälligen Schwankungen und systematischen Verzerrungen dominiert werden oder leicht manipulierbar sind.

Die Frage nach der Datenqualität ist somit mit Pohlentz (2009) als Schlüsselfrage evidenzbasierter Qualitätssicherung von Studium und Lehre zu betrachten, die daher viele Forschergenerationen beschäftigt hat (siehe überblicksweise Marsh 2007; Spooren et al. 2013). Neben der Validität der Messungen werden hierbei auch Aspekte der Fairness derartiger Bewertungsverfahren berührt. Würde sich etwa herausstellen, dass LVE von lehrunabhängigen Faktoren beeinflusst werden, die von den Dozierenden nur schwer zu beeinflussen sind, so wäre das ein Fairness-Problem (siehe Rindermann 2009).

Im vorliegenden Beitrag wird daher nach möglichen methodischen Fallstricken bei der studentischen Evaluation universitärer Lehre durch sachfremde Einflüsse, Unterschiede in den studentischen Urteilerstandards und Selektionseffekte gefragt (siehe Abbildung 1). Zunächst werden mit der Attraktivität des Lehrpersonals, dem Studierenden- und Dozierendengeschlecht sowie der Notengebung lehrunabhängige Einflüsse bei studentischen LVE betrachtet. Im dritten Abschnitt werden individuelle Unterschiede im studentischen Anspruchsniveau und Urteilsverhalten und ihre Folgen für LVE in den Blick genommen. Schließlich wird in Abschnitt 4 die Möglichkeit selektionsbedingter Verzerrungen geprüft, da in vielen Veranstaltungen zum Evaluationszeitpunkt nicht mehr die gesamte ursprüngliche Teilnehmerschaft anwesend und der Ausfall vermutlich systematisch ist.

2. Lehrunabhängige Einflüsse: physische Attraktivität und Notengebung

Physisch attraktive Menschen werden (im Vergleich zu unattraktiven und durchschnittlich attraktiven Menschen) u.a. als sozial kompetenter, geselliger, moralischer, verlässlicher, intelligenter, kreativer, anpassungsfähiger, fleißiger, erfolgreicher und (physisch und psychisch) gesünder eingeschätzt (zum Beispiel Eagly et al.

¹ Es handelt sich bei dem vorliegenden Beitrag um eine Zusammenfassung des Buches *Fallstricke der Lehrevaluation. Möglichkeiten und Grenzen der Messbarkeit von Lehrqualität* (Wolbring 2013).

Abbildung 1: Drei methodische Fallstricke bei der Evaluation universitärer Lehre

Abschnitt 2 Sachfremde Einflüsse	Abschnitt 3 Urteilerstandards	Abschnitt 4 Selektionseffekte
• Attraktivität des Lehrpersonals • Geschlechtseffekte • Benotung, insbesondere Reziprozität • Wechselwirkung dieser drei Einflussfaktoren	• Werden unterschiedliche Maßstäbe angelegt? • Unterscheiden sich Studierende in der Urteilsstrenge und -differenziertheit? • Wodurch lassen sich personenspezifische Standards erklären?	• Ausmaß der Selektion und des Selektionsbias • Erklärungen für den studentischen Absentismus • Wirkung des Selektionsprozesses auf Veranstaltungsratings und -rankings

1991; Jackson et al. 1995). Dieses Attraktivitätsstereotyp führt dazu, dass sie im Leben zahlreiche Vorteile genießen: Sie erfahren mehr Aufmerksamkeit und Zuwendung, haben erhöhte Chancen mit Anderen zu interagieren, sind populärer, weisen mehr sexuelle Erfahrungen auf, können Andere eher zur Kooperation bewegen, werden in simulierten Gerichtsverfahren seltener verurteilt, erhalten bei politischen Wahlen mehr Stimmen, schneiden in der Schule besser ab und haben im Arbeitsmarkt bessere Einstellungs- und Aufstiegschancen sowie höhere Gehälter (für eine Zusammenschau siehe Hamermesh 2011; Patzer 2007).

Es würde daher nicht überraschen, wenn attraktive Lehrende von den Studierenden auch bessere Lehrveranstaltungsbewertungen erhalten. Ein solcher Befund wäre jedoch nur dann problematisch, wenn das physische Erscheinungsbild einer Person unabhängig von ihrer Lehrleistung ist, also der Attraktivitätsbonus aufgrund von Diskriminierung und nicht aufgrund von Unterschieden in der Produktivität zustande kommt. Zur empirischen Überprüfung dieser Überlegungen wurden zunächst prozessproduzierte Daten zur LVE an der Sozialwissenschaftlichen Fakultät der LMU München genutzt und auf Grundlage von im Internet verfügbarer Potraitphotos um Angaben zur physischen Attraktivität der Dozierenden ergänzt. Empirisch können, auch bei Kontrolle zentraler Drittvariablen wie Kurstyp, Kursgröße und Vorinteresse der Studierenden am Kursthema, schwache, aber systematische Einflüsse der physischen Attraktivität der Lehrenden auf die Gesamtbewertung der Veranstaltung (1 ‚sehr gut‘ – 5 ‚mangelhaft‘) nachgewiesen werden. Der maximale Effekt des physischen Erscheinungsbildes beträgt dabei 0,3 Notenpunkte. Der Attraktivitätseffekt variiert zudem mit dem Geschlecht der Dozierenden und der Studierenden. Durchgehend positive Effekte zeigen sich für weibliches Lehrpersonal, allerdings können entsprechende Effekte in den meisten Analysen auch für Dozenten nachgewiesen werden.

Die Frage, ob diese Befunde von Produktivitätsunterschieden oder Diskriminierung herrühren, lässt sich auf Grundlage dieser prozessproduzierten Daten jedoch nicht mit Sicherheit beantworten – produktivitätsrelevante Faktoren können in den Modellen immer nur unvollständig kontrolliert werden. Im Anschluss an diese Studie wurde daher ein Laborexperiment durchgeführt, in dem Studierende einen Lebenslauf mit Photo erhiel-

ten, eine Audioaufzeichnung eines Vortrags der abgebildeten Lehrperson hörten und eine LVE dazu ausfüllten. Da das Erscheinungsbild der auf dem Photo abgebildeten Person systematisch variiert wurde, während die Lehrqualität in Form der Audioaufzeichnung konstant gehalten wurde, können Produktivitätsunterschiede bei diesem Design ausgeschlossen werden. Im Ergebnis zeigt sich nichtsdestotrotz erneut ein schwacher, aber statistisch signifikanter Effekt der physischen Attraktivität auf die Lehrbewertung, der in seiner Stärke den vorherigen Befunden sehr ähnelt. Insbesondere Dozentinnen profitieren gemäß der Befunde des Laborexperiments von ihrem Erscheinungsbild. Überdies absolvierten die Studierenden im Experiment einen sehr schwierigen Test zum Inhalt des Vortrags, wobei ein zufällig ausgewählter Teil der Probanden vor, ein anderer Teil nach dem Test evaluierte. Wird die Prüfung vor der LVE gestellt, so hat das Abschneiden im Test einen sehr starken Effekt auf die Bewertung des Vortrags und der Lehrperson. Studierende folgen also offensichtlich dem Prinzip der Reziprozität („Wie Du mir, so ich Dir!“) und strafen Lehrende für eine Verletzung der Fairnessnorm bei der Testgestaltung systematisch ab.² Interessanterweise werden überdurchschnittlich attraktive Dozierende und v.a. attraktive männliche Lehrende für Verfehlungen in Form eines zu schwierigen Tests von den Studierenden signifikant stärker abgestraft als andere Lehrende. Die Erwartungen an diese Personen sind aufgrund positiver Attraktivitätsstereotype grundsätzlich höher, sodass die Enttäuschung über eine unfaire Behandlung durch diese ursprünglich als sympathisch wahrgenommenen Personen ebenfalls stärker ausfällt.

3. Personenspezifische Ansprüche und Bewertungsstandards

Während im letzten Abschnitt mit der physischen Attraktivität, dem Geschlecht und dem Notengeungsverhalten Merkmale der Dozierenden im Vordergrund standen, werden nun individuelle Unterschiede im studentischen Urteilsverhalten behandelt. Grundthese ist dabei, dass sich Studierende in ihren Anspruchshaltungen gegenüber der Lehre unterscheiden und unterschiedlich großen kognitiven Aufwand bei der Beurteilung einzelner Aspekte der Lehre betreiben. Erste empirische Evidenz zu entsprechenden interindividuellen Unterschieden findet sich bereits in den Arbeiten von Kromrey (1994, S. 157f.):

² Weitere Befunde von Krempkow (2007, S. 173) deuten an, dass vermutlich nicht nur ein (zu) hohes Anspruchsniveau, sondern auch ein (zu) niedriges Anspruchsniveau zu negativeren LVE-Urteilen führt. Demnach lassen sich LVE zwar durch Vermeidung besonders anspruchsvoller Prüfungen positiv beeinflussen, aber nicht unbedingt – wie vielfach von Kritikern angeführt wird – durch Senkung der Anforderungen immer weiter verbessern.

„Für einen Teil der Befragten (nahezu die Hälfte) gilt: [...] Die Antworten auf die Detailfragen sind Indikatoren für ein dahinterstehendes ‚Urteilssyndrom‘ [...]. Ist man generell eher zufrieden (bzw. unzufrieden), werden Qualitätsabweichungen bei dem einen oder anderen Teilaspekt weder kritisiert noch honoriert. [...] Etwas mehr als die Hälfte zeichnet sich [demgegenüber, T.W.] durch Urteilmuster aus, deren Verläufe sich überschneiden. Zum Beispiel: positive Einschätzungen für Aspekte des sprachlichen Vortrags [...] und Stoffaufbereitung [...], aber negative Urteile über die Strukturierung der Veranstaltung [...], das persönliche Engagement des/der Lehrenden und die Aufgeschlossenheit für Belange der Studierenden.“

Würden verschiedene Personen unterschiedliche Standards bei der Bewertung anlegen, so wäre dies für die Güte von LVE aus mehreren Gründen problematisch. Erstens würde die alleinige Rückmeldung der durchschnittlichen Bewertungen an die Dozierenden eine unsachgemäße Informationsreduktion darstellen. Die Berechnung des arithmetischen Mittelwerts setzt nämlich voraus, „dass die zu beschreibende Reihe eine gewisse Konzentration innerhalb der beobachteten Elemente aufweis(t)“ und eine „homogene statistische Masse“ vorliegt (Kromrey 2002, S. 425). Beide Annahmen wären bei Vorliegen variierender Urteilerstandards verletzt. Zweitens würden auch die gemittelten Angaben zu einzelnen Dimensionen der Lehre verwässert und damit an Informationsgehalt verlieren, wenn zumindest ein Teil der Urteilenden bei der Bewertung schlicht seinem Gesamteindruck folgt. Schließlich würden sich beim Vergleich der LVE verschiedener Veranstaltungen Probleme ergeben, wenn deren Besucherinnen und Besucher unterschiedliche Standards anlegen.

Im Gegensatz zu Kromrey, der seine Befunde auf Latent-Class- und Clusteranalysen von Querschnittsdaten stützt, werden im Folgenden zum Nachweis des Vorliegens interindividueller Unterschiede in der Strenge und Differenziertheit des studentischen Urteils Paneldaten genutzt. Die Intuition hierfür ist, dass Personen möglicherweise in einer Reihe von Kursen die Lehrqualität systematisch besser oder schlechter (Urteilsstrenge) bzw. auf einer Itematterie zu verschiedenen Einzeldimensionen differenzierter oder undifferenzierter (Urteilsdifferenzierung) beurteilen als ihre Kommilitoninnen und Kommilitonen. Grundlage stellen erneut Daten der LVE an der Sozialwissenschaftlichen Fakultät der LMU München dar, wobei der Fragebogen seit dem Wintersemester 2008/09 einen selbstgenerierenden Panelschlüssel enthält. Die Erhebung dieser zeitkonstanten, personenspezifischen Merkmale (erste beide Buchstaben des Vornamens der Mutter und des Vaters, erste beide Ziffern des Geburtstags der Mutter, Zahl älterer Schwestern und Brüder) erlaubt dabei die Unterscheidung verschiedener Urteilender, ohne deren Anonymität zu gefährden. Somit lassen sich systematische Abweichungen einzelner Personen von der durchschnittlichen Kursbewertung über die Zeit beobachten.

Mit Blick auf Standards bei der Gesamtbeurteilung, d.h. die Urteilsstrenge, zeigt sich ein beträchtliches Maß an interindividueller Variation. Es gibt demnach Personen, die

regelmäßig deutlich besser oder schlechter als die restlichen Kursteilnehmer urteilen. Dies gilt selbst dann, wenn man nur Personen betrachtet, für die zehn oder mehr ausgefüllte Evaluationsbögen vorliegen. Weitere Analysen zeigen, dass die studentischen Anspruchshaltungen und damit die Urteilstrenge mit der sozialen Herkunft, dem Fachsemester und der vorherigen Absolvierung eines Studiums variieren. Daneben tragen auch die Abiturnote, die allgemeine Vorbildung und spezifisches Vorwissen zur Erklärung der Urteilstrenge bei. In diesem Zusammenhang wird auch die Bedeutung der Passung zwischen Vorwissen und Kursanforderungen beispielhaft anhand von Veranstaltungen aus dem Bereich der quantitativen Methoden deutlich: Personen mit einem mathematisch-naturwissenschaftlichen Schwerpunktfach in der Schule (z.B. Informatik, Mathematik, Physik) beurteilen formal anspruchsvollere Lehre wohlwollender.

Als bedeutsamste Einflussgröße ist schließlich die studentische Motivation hervorzuheben. Intrinsische Motivation wirkt sich positiv auf die Bewertungen aus, während extrinsische Teilnahmeanreize einen negativen Effekt haben. Zudem moderiert die intrinsische Motivation letzteren Effekt: Der negative Effekt einer Pflichtveranstaltung auf die Bewertung entfaltet sich nur im Zusammenspiel mit geringem Vorinteresse. Das Vorinteresse am Veranstaltungsthema ist somit – insbesondere bei Veranstaltungen, in denen ein Leistungsnachweis erworben werden muss – für eine positive Bewertung entscheidend. Studienkosten erwiesen sich dagegen als wenig bewertungsrelevant, obwohl sie aus theoretischer Perspektive die studentische Anspruchshaltung steigern sollten. Jedoch zeigt sich, dass Studierende mit zunehmendem Studienverlauf wohlwollender urteilen, was einerseits mit dem Erwerb des notwendigen Vorwissens und Sozialisationseffekten, andererseits aber auch mit den zunehmenden Wahlmöglichkeiten im Zeitverlauf, also mit Selektionseffekten, zusammenhängen kann. Eine Korrektur der LVE um entsprechende Kompositionseffekte bewirkt starke Veränderungen von Veranstaltungsratings und -rankings, was eine begrenzte Vergleichbarkeit von LVE für Veranstaltungen mit unterschiedlichem Zielpublikum andeutet.

Die interpersonelle Heterogenität hinsichtlich der Urteilsdifferenzierung scheint mit Blick auf die LVE-Daten stärker ausgeprägt zu sein, als die Effekte genereller Standards bei der Gesamtbewertung ausfallen. Jedoch wird der Grad der studentischen Urteilsdifferenzierung zum Teil von ähnlichen Faktoren wie die Urteilsstrenge beeinflusst. Dies gilt etwa für die kognitiven Fähigkeiten der Studierenden, das Fachsemester und die vorherige Absolvierung eines Studiums. Darüber hinaus ergeben sich interessante Befunde zur Wirkung der Befragtenmotivation und zu Ermüdungseffekten, die auch zum generellen Verständnis menschlichen Verhaltens in Befragungssituationen beitragen. Demnach urteilten teilnahmemotivierte Studierende, d.h. Personen die vormals freiwillig an einer allgemeinen Umfrage zur Verbesserung von Studium und Lehre teilgenommen haben, bei der LVE differenzierter auf die Fragen einer Itematterie, während zu häufiges Ausfüllen der LVE-Fragebögen Ermüdung hervorruft und eine Reduktion des Informationsgehalts der Angaben bewirkt. In der Evaluationspra-

xis an deutschen Hochschulen wird dieser Aspekt der Belastung der Studierenden durch die Vielzahl an Befragungen bisher meist kaum berücksichtigt. Um reliable und valide Bewertungen der Lehre zu gewährleisten, sollten diese Befunde jedoch bei der Gestaltung von Evaluationsverfahren in Zukunft stärker beachtet werden. Insbesondere ist etwa davon abzuraten, jede Lehrveranstaltung eines Studiengangs jedes Semester von allen Studierenden zu evaluieren lassen.

4. Selbstselektion, studentischer Dropout und Antwortverweigerung

Verzerrungen aufgrund von Selektionseffekten sind eines der Hauptprobleme jeder empirischen Studie. Dies gilt gleichermaßen für studentische LVE und könnte von erheblicher Relevanz für die Reliabilität, Validität und Fairness der Messung sein. Bei Online-Evaluationen außerhalb des Seminarraums wird bekanntlich meist nur ein geringer Rücklauf erzielt. Zu befürchten ist daher, dass besonders teilnahmemotivierte Studierende, d.h. Personen mit ausgeprägter Zufriedenheit oder Unzufriedenheit, überrepräsentiert sind (Dresel/Tinsner 2008; Wolbring 2010). Bei papier-basierten LVE sind dagegen in vielen Lehrveranstaltungen, insbesondere in Vorlesungen, zum Zeitpunkt der Evaluation keineswegs alle ursprünglichen Besucher noch anwesend. Auch hier ist zu vermuten, dass sich abwesende Studierende bezüglich ihrer Zufriedenheit mit der Lehre, ihrem Interesse an den Veranstaltungsinhalten und ihrer Motivation zur Teilnahme systematisch von der evaluierenden Hörerschaft unterscheiden. Bei beiden Designs sind daher verzerrte Schätzungen der Lehrqualität sehr wahrscheinlich. Da zudem die Selektionseffekte in verschiedenen Veranstaltungstypen (z.B. Wahl vs. Pflicht; Seminar vs. Vorlesung) nicht gleich ausfallen, ergeben sich vermutlich überdies Probleme beim Vergleich von LVE in Form von Rankings.

Die Wissenschaft ist jedoch bisher eine systematische empirische Untersuchung zu möglichen selektionsbedingten Verzerrungen bei papier-basierten LVE schuldig geblieben (siehe aber Esser 1995). Dieses Defizit wird mit den nun vorgestellten Untersuchungen behoben. Die Determinanten des Besuchs und deren Konsequenzen für Ratings und Rankings wurden einerseits auf Grundlage regulärer LVE-Daten, andererseits anhand zusätzlich zu zwei Messzeitpunkten erhobener Daten für 26 ausgewählte Kurse untersucht. Die empirischen Analysen zeigen starke Einflüsse des Veranstaltungstermins (Wochentag, Uhrzeit) und der Veranstaltungsqualität auf die Häufigkeit des studentischen Besuchs auf. Darüber hinaus tragen auch Anwesenheitsregelungen, der Grad der sozialen Kontrolle (u.a. operationalisiert mittels der Kursgröße) und Studierendenmerkmale (u.a. Fähigkeiten, Motivation, Umfang an Erwerbsarbeit) zur Erklärung von Unterschieden im Besuchsverhalten bei. Überdies hat auch die Dozierendenattraktivität eine besuchssteigernde Wirkung, was auf den nutzenstiftenden Charakter sozialer Interaktionen mit attraktiven Menschen hindeutet.

Da einige dieser Einflussfaktoren zudem mit der Veranstaltungsbewertung zusammenhängen und auch die

Veranstaltungsqualität selbst die Besuchsentscheidung beeinflusst, wird anschließend nach dem Ausmaß der vorliegenden Selektionseffekte gefragt. Anhand der regulären LVE-Daten wird hierzu zunächst mit einer besuchsbedingten Gewichtung versucht, der selektionsbedingten Unsicherheit bei der Erstellung von Lehrveranstaltungsratings und -rankings Rechnung zu tragen. Die resultierenden Veränderungen deuten die Möglichkeit des Vorliegens erheblicher Verzerrungen für einzelne Veranstaltungen und damit auch des Rankings als Ganzem an. Anschließend wird auch anhand der Daten zu zwei Messzeitpunkten für 26 ausgewählte Veranstaltungen gezeigt, dass sich bei Berücksichtigung der fehlenden Werte für abwesende Studierende zwar nur geringe Veränderungen der Veranstaltungsratings, aber – trotz einer Rangkorrelation von $r = 0,9$ – erhebliche Veränderungen der resultierenden Ränge für einzelne Veranstaltungen ergeben. Verschiedene Versuche zur Korrektur der Verzerrungen werden erprobt, erweisen sich aber als unzureichend bzw. unausgereift. Selektionsbias liegt also vor und ist selbst mit zusätzlichen Informationen aus einer zweiten Erhebung und mit fortgeschrittenen statistischen Verfahren nicht behebbar.

5. Implikationen für die Validität von LVE und ihre hochschulpolitische Nutzung

Anhand des Nachweises von lehrunabhängigen Einflüssen, Urteilerstandards und Selektionseffekten wird deutlich, dass das Instrument ‚studentische Lehrveranstaltungsevaluation‘ nicht nur die Lehrqualität erfasst. Entsprechende Verzerrungen sind dabei zum Teil nur bedingt von den Lehrenden beeinflussbar, weshalb die Messung nicht nur nicht vollkommen valide, sondern auch unfair ist. Es ist daher an entsprechende Korrekturmaßnahmen der Rohdaten zu denken. Dadurch lässt sich jedoch nur imperfekt für unerwünschte Einflüsse korrigieren, sodass man eine Form statistischer Diskriminierung institutionalisieren würde, wenn man etwa die Evaluationen schöner Lehrender standardmäßig mit einem Attraktivitätsmalus versehen würde. Dies impliziert, dass Vergleiche zwischen den Bewertungen verschiedener Veranstaltungen selbst nach entsprechenden Adjustierungsmaßnahmen keine belastbaren Aussagen zu Unterschieden in der Lehrqualität liefern.

Es ist daher von der Implementierung lehrleistungsabhängiger Anreizmechanismen auf Grundlage studentischer LVE abzuraten. Eine solche Anreizsetzung würde auf verzerrten und unfairen Messungen basieren und somit einige Dozierende ungerechtfertigt benachteiligen. Dies könnte un intendierte Folgewirkungen haben. Denkbare und teilweise bereits dokumentierte Reaktionen sind beispielsweise die Demotivation der betroffenen Lehrenden, Verzicht auf schwierigen Stoff, Senkung des Anspruchsniveaus bei Prüfungen, Noteninflation und Versuche der Manipulation der Messung (etwa durch starke Auslese der Hörerschaft, Wahl günstiger Zeitpunkte bzw. beliebter Themen oder das Verteilen kleiner Geschenke).

Da LVE mehrere positive Funktionen zukommen, wird jedoch keineswegs für deren Abschaffung, sondern nur für einen sachgemäßen, auf Leitungsebene zurückhal-

tenden Umgang plädiert. LVE sollten als das angesehen werden, was sie sind: Akzeptanz- und Zufriedenheitsmessungen, die vom Veranstaltungskontext, Veranstaltungsverlauf und der Zusammensetzung der Teilnehmer-schaft beeinflusst werden. Am besten können diese Aspekte und ihre Folgen für die Lehrbewertung die Dozierenden selbst beurteilen. Man sollte LVE daher weniger als summatives, sondern als formatives Instrument zur Qualitätssicherung begreifen, das den Lehrenden einen Anstoß zur systematischen Auseinandersetzung mit der eigenen Lehre auf Grundlage des Feedbacks der Hörerschaft gibt.

Demnach liegt es nahe, Studierende bereits zu Beginn oder zur Mitte des Semesters nach ihrer Zufriedenheit mit der Lehre zu befragen, sodass Dozierende noch im laufenden Semester Verbesserungsvorschläge umsetzen können. Zudem dokumentieren neuere Befunde positive Wirkungen von LVE in Kombination mit Weiterbildungsangeboten (Dresel/Rindermann 2011; Penny/Coe 2004). Es ist daher auch empfehlenswert, das Instrument mit der Möglichkeit freiwilliger Fortbildungen durch Peers und Experten zu koppeln.

Literaturverzeichnis

- Dresel, M./Rindermann, H. (2011): „Consultation of University Instructors Based on Student Evaluations of Teaching Effectiveness: Multilevel Test of its Effectiveness under Consideration of Bias and Unfairness Variables“. In: Research in Higher Education, Vol. 52, No. 7, pp. 717-737.
- Dresel, M./Tinsner, K. (2008): Onlineevaluation von Lehrveranstaltungen: Methodeneffekte bei der Onlineevaluation von Lehrveranstaltungen. In: Zeitschrift für Evaluation, 7. Jg., H.2, S. 183-211.
- Eagly, A.H./Ashmore, R.D./Makhijani, M.G./Longo, L.C. (1991): „What is Beautiful is Good, but ...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype“. In: Psychological Bulletin, Vol. 110, No. 1, pp. 109-128.
- Esser, H. (1995): No Shows, Karteileichen, Schleifendreher. In: Deutsche Universitätszeitung, 18. Jg., S. 22-25.
- Hamermesh, D.S. (2011): Beauty Pays. Princeton.
- Jackson, L.A./Hunter, J.E./Hodge, C.N. (1995): „Physical Attractiveness and Intellectual Competence: A Meta-analytic Review.“ In: Social Psychology Quarterly Vol. 58, No. 2, pp. 108-122.
- Krempkow, R. (2007): Leistungsbewertung, Leistungsanreize und die Qualität der Hochschullehre. Konzepte, Kriterien und ihre Akzeptanz. Bielefeld.
- Kromrey, H. (1994): Wie erkennt man gute ‚Lehre‘? Was studentische Vorlesungsbefragungen (nicht) aussagen. In: Empirische Pädagogik, 8. Jg., H.2, S. 153-168.
- Kromrey, H. (2002): Empirische Sozialforschung (10. Aufl.). Opladen.
- Marsh, H.W. (2007): Students' Evaluations of University Teaching: A Multidimensional Perspective. In: Perry, R.P./Smart, J.C. (Ed.): The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective. New York, pp. 319-384.
- Patzner, G.L. (2007): Why Physically Attractive People are More Successful. The Scientific Explanation, Social Consequences, and Ethical Problems. Lewiston.
- Penny, A.R./Coe, R. (2004): „Effectiveness of Consultation on Student Ratings Feedback: A Meta-analysis.“ In: Review of Educational Research, Vol. 74, No. 2, pp. 215-253.
- Pohlenz, P. (2009): Datenqualität als Schlüsselfrage der Qualitätssicherung von Lehre und Studium. Bielefeld.
- Rindermann, H. (2009): Lehrrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts. Landau.
- Spooren P./Brocx, B./Mortelmans, D. (2013): „On the Validity of Student Evaluation of Teaching: The State of the Art.“ In: Review of Educational Research, Vol. 83, No. 4, pp. 598-642.
- Wolbring, T. (2010): Weshalb die Separierung von Produktivitätseffekten und Diskriminierung bei der studentischen Lehrveranstaltungsbewertung misslingt. Selektive Stichproben, fehlende Drittvariablenkontrolle und die Konfundierung von Effekten. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 62. Jg., H.2, S. 317-326.
- Wolbring, T. (2013): Fallstricke der Lehrrevaluation. Möglichkeiten und Grenzen der Messbarkeit von Lehrqualität. Frankfurt am Main und New York.

■ Dr. Tobias Wolbring, Postdoctoral Research Fellow an der Professur für Sozialpsychologie und Hochschulforschung, ETH Zürich, E-Mail: tobias.wolbring@gess.ethz.ch

Sandra Mittag, Rüdiger Mutz & Hans-Dieter Daniel: Institutionelle Qualitätssicherung der Lehre auf dem Prüfstand: Eine Fallstudie an der ETH Zürich



Im Rahmen der vorliegenden Studie wurde das Qualitätssicherungssystem der ETH Zürich im Bereich Lehre einer umfassenden Meta-Evaluation unterzogen.

Das Qualitätssicherungssystem stützt sich auf die vier Instrumente Lehrveranstaltungsbeurteilung, Absolventenbefragung, Selbstevaluation und Peer Review.

Die Ergebnisse zeigen unter anderem, dass die ETH Zürich über etablierte Qualitätssicherungsinstrumente verfügt, die weitestgehend akzeptiert sind.

Allerdings bestehen bei allen vier Instrumenten Optimierungspotentiale.

ISBN 3-937026-74-6, Bielefeld 2012, 115 S.,
19.50 Euro zzgl. Versandkosten

Erhältlich im Fachbuchhandel und direkt beim Verlag – selten im Versandbuchhandel (z.B. nicht bei Amazon).
Bestellung – E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22

Reihe: „Qualität – Evaluation – Akkreditierung“

Susanne Weis, Christiane Karthaus & Tanja Lischetzke

Elemente der Lehrveranstaltungsevaluation an der Universität Koblenz-Landau: Theoretische Einordnung und empirische Befunde

„Teachers are the single most important learning resource available to most students. It is important that those who teach have a full knowledge and understanding of the subject they are teaching, have the necessary skills and experience to transmit their knowledge and understanding effectively to students in a range of teaching contexts, and can access feedback on their own performance.“ European Association of Quality Assurance in Higher Education (2009; S. 18)

Die „Standards and Guidelines for Quality Assurance in the European Higher Education Area“ der *European Association of Quality Assurance in Higher Education* (ENQA 2009) weisen in dieser Leitlinie knapp auf Inhalt und Bedeutung von Lehrveranstaltungsevaluationen hin. Diese stellen typischerweise die Bewertung einer Lehrveranstaltung durch Studierende mit Hilfe von Fragebogeninstrumenten dar (Marsh 1997; Rindermann 2009). Gegenstand der Bewertung ist der Prozess der Vermittlung von Lehrinhalten durch die Lehrkräfte. Durch die Rückmeldung der aggregierten Studierendenbewertungen an die Lehrkräfte soll diesen ermöglicht werden, ihre Lehre kontinuierlich zu verbessern (ENQA 2009; Gollwitzer/Schlotz 2003; Spiel/Wolf/Popper 2002).

Bereits lange vor der Einführung des europäischen Hochschulraums war das Thema Lehrveranstaltungsevaluation fest in der Forschungsliteratur verankert. Wiederkehrende Forschungsfragen betreffen in erster Linie die Gültigkeit der Studierendenbewertungen als einen Indikator für Lehrqualität (Validität), aber auch die Entwicklung von reliablen Fragebogeninstrumenten oder die Gestaltung der Rückmeldung an die Lehrkräfte (für einen umfassenden Überblick über viele der genannten Forschungsfragen und noch weitere s. Rindermann 2009). Ein weiterer, durch das einheitliche Leistungspunktesystem notwendig gewordener Bestandteil von Lehrveranstaltungsevaluation im weiteren Sinne ist die Erfassung des studentischen Workload. Außerdem existieren diverse Ansätze zur Erfassung von Ergebnisvariablen universitärer Lehre (z.B. Lernerfolg oder Kompetenzzuwachs). Bislang wenig untersucht ist der Zusammenhang der unterschiedlichen Ansätze (z.B.: Fällt der Lernerfolg in besser bewerteten Lehrveranstaltungen größer aus?). Dabei existieren diverse theoretische Modelle aus der schulischen Lehr-Lern-Forschung (u.a. Bloom 1976; Carroll 1963; Helmke 2012), die auf die Lehrveranstaltungsevaluation an Hochschulen übertragbar sind und testbare Annahmen formulieren.



Susanne Weis



Christiane Karthaus



Tanja Lischetzke

Der vorliegende Beitrag stellt die Elemente der Lehrveranstaltungsevaluation der Universität Koblenz-Landau vor, ordnet sie in theoretische Vorhersagemodelle für akademischen Lernerfolg ein und bringt die somit erfassten Konstrukte in Relation zueinander. Jeweils ergänzend werden Daten aus verschiedenen empirischen Studien berichtet, welche u.a. die psychometrische Qualität der einzelnen Elemente und deren Zusammenhang miteinander beleuchten und deren Beziehung zu weiteren Lernerfolgsindikatoren untersuchen.

1. Lehrveranstaltungsevaluation an der Universität Koblenz-Landau

Das Konzept zur Lehrveranstaltungsevaluation an der Universität Koblenz-Landau verfolgt verschiedene Ziele: a) Bereitstellung von empirischen Informationen zur kontinuierlichen Verbesserung der Lehre auf Veranstaltungs- und Studiengangebene und für Akkreditierungsverfahren, b) die Sicherstellung der wissenschaftlichen Güte der eingesetzten Instrumente und Methoden sowie deren kontinuierliche, bedarfsorientierte Weiterentwicklung und c) die Berücksichtigung von Machbarkeits- und Ökonomiekriterien.

1.1 Allgemeine Lehrveranstaltungsbewertungen

1.1.1 Fragebogen

Zur allgemeinen Bewertung einer Lehrveranstaltung wird eine adaptierte Form des Trierer Inventars zur Lehrveranstaltungsevaluation (TRIL; Gollwitzer/Schlotz 2003) eingesetzt. Das TRIL versteht sich als ein Instrument, welches sowohl die diagnostische Güte als auch die Nützlichkeit zur konkreten Lehrveranstaltungsverbesserung auf Basis der Rückmeldungen zum Ziel hat (Fondel/Lischetzke/Weis/Gollwitzer 2014; Gollwitzer/Schlotz 2003). Die ursprüngliche Form umfasste 42

Fragen auf sechs Skalen mit drei veranstaltungsartspezifischen Fragebogenformen (für Vorlesungen, Seminare, Übungen). Seit Beginn der zentral durchgeführten Lehrveranstaltungsevaluation wurde das TRIL sowohl auf Basis von testtheoretischen Erkenntnissen (Trennschärfen, Interkorrelationen der Fragen und Skalen etc.), als auch auf Basis spezifischer Bedarfe in Kooperation mit den Studiengangvertreterinnen und -vertretern angepasst, wobei sowohl inhaltliche als auch ökonomische Gesichtspunkte ausschlaggebend waren. Im Kern beinhaltet das adaptierte TRIL nun zwischen 15 und 34 Fragen auf vier Skalen (Beispielfrage s. Tabelle 1 oberste Zeilen): 1) Struktur und Didaktik, 2) Anregungsgehalt, 3) Veranstaltungsklima und 4) Anwendungs- und Praxisbezug (Fondel et al. 2014). Studierende geben ihre Antworten auf einer sechsstufigen Ratingskala von 1 (trifft gar nicht zu) bis 6 (trifft voll zu). Die in Tabelle 1 (Punkt 1) genannten Beispielfragen sind solche, die sich in dieser oder sehr ähnlicher Formulierung als fachübergreifend angemessen herausgestellt haben. Fachübergreifend wird außerdem die Gesamtnote für die Lehrveranstaltung auf einer Skala von 1 (sehr gut) bis 6 (ungenügend) erfasst und es werden zwei offene Fragen dazu gestellt, was besonders gut und was verbesserungswürdig war. Die fächer- oder fachbereichsspezifischen Adaptationen und Ergänzungen des TRIL unterscheiden sich sowohl in Länge und Inhalt (s. Tab. 1, Punkt 3), als auch darin, ob sie die Differenzierung in veranstaltungsartspezifische Fragebogenformen aufrechterhalten (s. Tab 1, Punkt 2). Ein Verzicht auf veranstaltungsartspezifische Fragebogenformen resultiert in aller Regel in der Aufnahme einer Enthaltungskategorie (nicht relevant). Fondel et al. (2014) untersuchten verschiedene Aspekte der Messeigenschaften der fächerübergreifend eingesetzten Fragen des TRIL und konnten insgesamt gute Reliabilitäten und Hinweise zur Validität des Instruments zeigen.

1.1.2 Verwertung der Evaluationsergebnisse

Das adaptierte TRIL wird üblicherweise online nach etwa zwei Dritteln der Veranstaltungszeit eingesetzt. Lehrkräfte erhalten im Anschluss einen Bericht, der die Antworten quantitativ zusammenfasst (Häufigkeitsverteilungen und Profillinien) bzw. die offenen Antworten auflistet. Lehrkräfte werden ermuntert, die Ergebnisse der Befragung in der Lehrveranstaltung mit den Studierenden zu besprechen. Empirische Studien (Spiel et al. 2002) und auch die Rücksprache mit den Beteiligten konnten zeigen, dass dieser Austausch mittel- und langfristig zu einer Verbesserung der Lehre und zu einer erhöhten Bereitschaft der Studierenden führt, an den Befragungen teilzunehmen.

Nach Ende der Veranstaltungszeit werden die Ergebnisse in aggregierter und anonymisierter Form den für Qualitätssicherung zuständigen Kommissionen in den Fachbereichen zur Verfügung gestellt. Die Kommissio-

Tabelle 1: Übereinstimmung und Variabilität der Fragen des adaptierten und ergänzten TRIL

1. Unterschiede zwischen Skalen über Fächer und Veranstaltungsarten hinweg			
Struktur und Didaktik	Anregungsgehalt	Veranstaltungsklima	Anwendungs- und Praxisbezug
„Die Veranstaltungsmaterialien (z.B. Folien, Skripte) waren hilfreich für das Verständnis der Veranstaltungsinhalte.“	„Der/Die Lehrende schuf eine anregende Atmosphäre.“	„Der/Die Lehrende war im Umgang mit den Studierenden freundlich und aufgeschlossen.“	„Die Inhalte der Veranstaltung wurden in angemessenem Umfang durch praxisnahe Beispiele (z.B. Anschauungsmaterial, Fallberichte etc.) veranschaulicht.“
2. Unterschiede zwischen Veranstaltungsarten (prototypische Fragen)			
Vorlesung	Seminar	Übung	
„Der/Die Lehrende ließ inhaltliche Fragen zu.“	„Der / die Lehrende gab auf Diskussionsbeiträge der TeilnehmerInnen ein hilfreiches Feedback.“	„Die Erläuterungen und Anweisungen für die praktische Arbeit durch die Lehrende/den Lehrenden waren hilfreich.“	
3. Unterschiede zwischen Fächern (prototypische Fragen)			
Fach 1 (Naturwissenschaften)		Fach 2 (Geisteswissenschaften)	
„Die Inhalte der Veranstaltung wurden anhand von Experimenten/Versuchen veranschaulicht?“		„In den einzelnen Lehrereinheiten erkennt man einen Bezug zum Gesamtzusammenhang der Veranstaltung.“	

nen können die Identität einzelner, in der Gesamtzahl aber beschränkter Datenpunkte einfordern, sobald dies für qualitätssichernde Maßnahmen notwendig ist. Die weitere Nutzung der Informationen zur Qualitätssicherung ist in den Fachbereichen sehr unterschiedlich. Die Ergebnisse werden u.a. für die Vergabe von Lehrpreisen für sehr gute Evaluationsergebnisse und für Peer-Gespräche zur Identifikation von notwendigen Verbesserungen bei schlechten Evaluationsergebnissen genutzt.

1.2 Workload

Zur Überprüfung des studentischen Workload sind an deutschen Hochschulen unterschiedliche Ansätze entwickelt worden (s. Blüthmann/Ficzko/Thiel 2006 und Burck/Heil/Böhres 2011 für einen Überblick). Sie unterscheiden sich sowohl in den Methoden (Fragebogen vs. Tagebuchaufzeichnungen vs. Interviews), den Forschungsdesigns (quer- vs. längsschnittlich mit einer variierenden Zahl von Messzeitpunkten) als auch den Inhalten (Zeiteinschätzungen in Stunden oder Minuten vs. pauschale, verbal verankerte Beurteilungen, z.B. „Haben Sie weniger/genauso viel/mehr Zeit investiert als im Modulhandbuch vorgesehen?“). Verschiedene Aspekte des Designs einer Workload-Erhebung (z.B. einmaliges retrospektives Urteil vs. wiederholte Befragungen, modulspezifische vs. modulübergreifende Angaben) haben typischerweise gegenteilige Effekte auf die Datenqualität auf der einen Seite sowie auf Rücklauf und Akzeptanz auf der anderen Seite (Blüthmann/Thiel 2011): Ein Instrument mit hohem Aufwand für die Studierenden (z.B. mit vielen Befragungszeitpunkten), welches sich damit durch einen geringeren Grad der Retrospektivität und validen Angaben auszeichnet, führt im Evaluationsalltag zu einer geringeren Rücklaufquote und damit wiederum zu einer verminderten Datenqualität und Akzeptanz.

1.2.1 Fragebögen

Zwei Befragungsvarianten zur Erfassung des studentischen Workload stehen den Fächern und Fachbereichen zur Wahl:

Ansatz 1: Dreifach-Erhebung. Angelehnt an das Modell der Studentischen Online Workload Erfassung Aachener Hochschulen (StOEHN; RWTH Aachen 2014) wird der studentische Workload zu drei Befragungszeitpunkten online erfasst. Dabei schätzen Studierende die Zeit für einen jeweils 4-5-wöchigen Bezugszeitraum pro Lehrveranstaltung („Wie viel Zeit [in Zeitstunden] im Bezugszeitraum haben Sie für folgende Tätigkeiten im Rahmen der angegebenen Veranstaltung aufgewendet? a) Besuch der Lehrveranstaltung, b) Tutorium/Repetitorium, c) Vor- und Nachbereitung der Veranstaltung und d) sonstiges“). Zum letzten Befragungszeitpunkt wird zusätzlich die Frage nach der prospektiv zu investierenden Zeit gestellt („Wie viel Zeit [in Zeitstunden] werden Sie voraussichtlich in der veranstaltungsfreien Zeit für die vorliegende Lehrveranstaltung aufwenden [inkl. Lernen für die Modulprüfung zu dieser Veranstaltung]?“). Die Angaben werden über die Zeit aufsummiert und den Lehrkräften und Kommissionen zurückgemeldet.

Ansatz 2: Einfach-Erhebung. Es wird eine über die Zeit hinweg aggregierte Zeiteinschätzung der Studierenden erfragt („Zusätzlich zu Ihren Anwesenheitszeiten in der Veranstaltung: Wie viel Zeit [in Zeitstunden] haben Sie für die vorliegende Veranstaltung im Schnitt pro Woche aufgewendet?“ und „Zusätzlich zu der oben angegebenen Zeit: Wie viel Zeit [in Zeitstunden] haben und werden Sie bezogen auf die vorliegende Veranstaltung für die Prüfungsvorbereitung insgesamt aufwenden?“). Die Präsenzzeit wird über die angegebenen Fehlzeiten und Semesterwochenstunden errechnet. Die weitere Verwertung der Informationen erfolgt, indem der Ist- mit dem Soll-Workload der Lehrveranstaltung absolut und prozentual in Beziehung gesetzt wird. So können Abweichungen vom Soll zusätzlich gewichtet werden: eine Abweichung von 30 bei 120 Stunden Soll-Workload ist damit weniger bedeutend als bei 60 Stunden Soll-Workload.

Beide Ansätze werden aktuell an der Universität Koblenz-Landau eingesetzt, wobei der zweitgenannte häufiger verfolgt wird. Beide bergen gleichsam Vor- und Nachteile. Während der zweite Ansatz mit weniger Aufwand verbunden ist und damit auf eine hohe Akzeptanz stößt, ist grundsätzlich davon auszugehen, dass die Datenqualität im erstgenannten Ansatz besser ist, wenn vertretbare Rücklaufquoten eingehalten werden. Dies begründen Blüthmann et al. (2006) damit, dass „die Ergebnisse punktueller Befragungen, bei denen längere Zeitabschnitte von den Befragten in ihrem Zeitaufwand eingeschätzt werden sollen, nicht zufriedenstellend wären“ (S. 4) und verweisen auf eine Studie von Jahnke (1971, zitiert nach Blüthmann et al. 2006). Nichtsdestotrotz bleibt auch die Datenqualität des Ansatzes der dreifachen Erhebung zunächst ungeklärt.

Tabelle 2: Stichprobengrößen (N), Mittelwerte (M) und Standardabweichungen (SD) der Workload-Einschätzung und der Kontrollvariablen in den drei Erhebungsbedingungen und der Gesamtstichprobe für beide evaluierten Lehrveranstaltungen

		Vergleichs- bedingung	Ansatz 1	Ansatz 2	
		Wöchentliche Befragung nach FELZ (Blüthmann et al., 2006)	Dreifach- Erhebung nach StOEHN (RWTH Aachen University, 2014)	Einfach-Erhebung zur aggregierten Zeiteinschätzung („im Schnitt pro Woche“)	Gesamt- stichprobe
Seminar B.Sc. Psychologie	N (vollst. Daten)	31	29	12	72
	M (SD) Workload	35.73 (15.93)	34.25 (15.67)	28.42 (8.37)	33.94 (14.81)
	M (SD) Gewissen- haftigkeit	3.54 (.62)	3.45 (.67)	3.32 (.53)	3.44 (.62)
	M (SD) Akademisches Selbstkonzept	4.67 (.55)	4.62 (.53)	4.55 (.69)	4.62 (.58)
Vorlesung B.Ed. Mathematik	N (vollst. Daten)	16	22	4	42
	M (SD) Workload	40.14 (19.03)	35.61 (25.78)	-	40.14 (23.64)
	M (SD) Gewissen- haftigkeit	2.73 (.63)	2.73 (.45)	-	2.72 (.52)
	M (SD) Akademisches Selbstkonzept	3.71 (.41)	3.49 (.50)	-	3.55 (.46)

1.2.2 Empirische Befunde zur Vergleichbarkeit und weiterführende Fragen

Begleitend zu der Einführung der Workload-Erhebung wurde die Vergleichbarkeit der mit den beiden Ansätzen gewonnenen Informationen empirisch überprüft. Dazu wurden Studierende zweier Lehrveranstaltungen (Seminar im Bachelor-Studiengang Psychologie, Vorlesung im Fach Mathematik des lehramtsbezogenen Bachelor-Studiengangs) randomisiert drei Bedingungen zugewiesen. Tabelle 2 zeigt die drei Bedingungen (Einfach- und Dreifach-Erhebung und eine Vergleichsbedingung) und die Stichprobengrößen. Als Vergleichsbedingung gelten die Workload-Einschätzungen mit einer an FELZ (Blüthmann et al. 2006) angelehnten wöchentlichen Befragung, weil davon auszugehen ist, dass damit die validesten Einschätzungen gewonnen werden können. Zu Beginn der Lehrveranstaltung wurde bei allen Studierenden zusätzlich Gewissenhaftigkeit (Goldberg et al. 2006) und akademisches Selbstkonzept (Dickhäuser/Spinath/Schöne/Stiensmeier-Pelster 2002) erhoben, um die Äquivalenz der Versuchsgruppen zu überprüfen.

Tabelle 2 zeigt Mittelwerte und Standardabweichungen der Workload-Einschätzungen als auch der Kontrollvariablen in den drei Bedingungen und der Gesamtstichprobe. Es zeigten sich in den beiden Lehrveranstaltungen keine statistisch signifikanten Mittelwertunterschiede zwischen den Erhebungsbedingungen, wobei die einmalige Erhebung in der Vorlesung aufgrund einer zu geringen Fallzahl (N=4) von den Analysen ausgeschlossen wurde.

Auch wenn die vorliegenden Befunde beide Vorgehensweisen an der Universität Koblenz-Landau stützen und keine Bedenken nahelegen, solange nur die Mittelwerte

der studentischen Workload-Einschätzungen im Rahmen qualitätssichernder Maßnahmen genutzt werden, bleiben offene Fragen zu diskutieren. Die Zeit, die Studierende in der veranstaltungsfreien Zeit für Prüfungen oder ggf. Scheinerwerb investieren, wird nur prospektiv geschätzt. Zuverlässigkeit und Validität dieser Schätzung sind bislang nicht geprüft. Zusätzlich zu diskutieren ist die Lehrveranstaltung als Evaluationsgegenstand. Zwar ist diese die naheliegendste Einheit zur Veränderung der Lehre (z.B. Reduktion der Anforderungen), allerdings wird dieser Fokus der Studienstruktur nur bedingt gerecht. So ist z.B. bei zeitgleich mehreren Veranstaltungen eines Moduls die Zeit für die Prüfungsvorbereitung nicht den unterschiedlichen Veranstaltungen zuordenbar. Darüber hinaus stellt sich manche Über- oder Unterschreitung des Soll-Workload als ein Problem des Moduls heraus, wenn bspw. der Soll-Workload von vornherein in einem nicht erreichbaren Bereich angesetzt war.

Eine weitere wichtige Frage betrifft die teils große Variabilität der Workload-Einschätzungen. Auf der einen Seite steht dabei die Idee, Zeiteinheiten in Leistungspunkte zu übersetzen, um damit die studentische Arbeitsbelastung zu regulieren und Studierbarkeit für den/die durchschnittlichen Studierenden (KMK 2010) zu gewährleisten. Auf der anderen Seite steht die empirisch gut belegte Erkenntnis, dass Individuen für einfache und hoch komplexe Aufgaben unterschiedlich viel Zeit benötigen. Sowohl die Differentielle als auch die Kognitions-Psychologie hat sich mit dem Phänomen auseinandergesetzt (s. das Konzept der *mental speed or processing speed*, exemplarisch dazu Ackerman/Beier/Boyle 2002; Rindermann/Neubauer 2004). Somit birgt die Variabilität Erklärungspotential für Erfolgs- oder Misserfolgsgründe im Studium und kann somit wiederum für die Studiengangentwicklung interessant sein.

1.3 Lernzielerreichung

Während mit dem adaptierten TRIL und dem Workload Prozessvariablen der universitären Lehre erfasst werden, liegt der Fokus der Erfassung von Lernzielerreichung auf dem Ergebnis von Lehre. Die Umsetzung einer ergebnisorientierten Evaluation, wie sie bspw. in englischsprachigen Ländern bereits etabliert ist, steckt an deutschen Hochschulen noch in den Kin-

derschuhen (Ausnahme z.B. das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen; BEvaKomp; Braun/Gusy/Leidner/Hannover 2008). Beim BEvaKomp handelt es sich um ein standardisiertes Fragebogeninstrument, welches auf sechs fach- und lehreveranstaltungsübergreifenden Kompetenzskalen (Methoden-, Fach-, Personal-, Präsentations-, Kommunikations- und Kooperationskompetenz) den subjektiven Kompetenzzuwachs nach einer Lehrveranstaltung durch die Studierenden einschätzen lässt. Das Erhebungskonzept der Lernzielerreichung an der Universität Koblenz-Landau dagegen ist zwar im Ablauf standardisiert, aber in den Inhalten an die konkrete Lehrveranstaltung angepasst (s. Beispiel in Abb. 1). Es umfasst ein mehrstufiges Vorgehen, welches sich zu Nutzen macht, dass die Modulbeschreibungen Qualifikationsziele benennen. Zunächst erhalten die Lehrkräfte einen teilstandardisierten Fragebogen zu den von ihnen durchgeführ-

Abbildung 1: Mehrstufiges Vorgehen bei der Evaluation der Lernzielerreichung

Schritt 1:

Kreuzen Sie bitte zunächst in nachfolgender Tabelle an, welche Qualifikationsziele aus der für Sie relevanten Modulbeschreibung für Ihre Lehrveranstaltung relevant sind.

Qualifikationsziele aus der Modulbeschreibung Die Studierenden ...	Relevanz	Lfd. Nr.
... besitzen Kenntnis univariater und bivariater deskriptiv-statistischer Kennwerte, deren Aussagekraft und Anwendbarkeit (etwa auf Daten unterschiedlichen Skalenniveaus).	☐	A
... besitzen Kenntnis unterschiedlicher Wahrscheinlichkeitsverteilungen.	☐	B
...	☐	C

Schritt 2:

Gehen Sie nun bitte noch etwas genauer auf die Qualifikationsziele ein, die Sie in der vorangegangenen Tabelle als relevant angekreuzt haben.

Listen Sie zu jedem **relevanten** Qualifikationsziel zugehörige, **möglichst konkrete** Lehrziele auf, indem Sie den Buchstaben des relevanten Ziels voranstellen (a-m) und dann mit 1., 2. ... die Teillehrziele nummerieren. Zusätzlich können Sie in der gleichen Tabelle weitere, über die vorgegebenen Kategorien hinausgehende Lehrziele benennen. Bitte beschränken Sie sich auf max. 3 Lehrziele für jedes ausgewählte Qualifikationsziel und auch max. 3 zusätzliche Lehrziele. Nutzen Sie für die Beschreibung der Lehrziele möglichst Formulierungen der folgenden Art: Studierende „beherrschen“, „sind in der Lage“, „können“, „haben Fertigkeiten in“ o.ä. und beschreiben Sie das angestrebte Lernziel so konkret wie möglich. Wir haben versucht, Ihnen ein Beispiel vorzugeben.

Lfd. Nr.	Indikatoren für die Qualifikationsziele für diese Lehrveranstaltung „Studierende am Ende der Lehrveranstaltung ...“
Bsp. zu B	1. ... können stetige von diskreten Verteilungen unterscheiden. 2. ... können empirische und theoretische Verteilungen unterscheiden
...	...
...	...

Schritt 3 (Übertragung der Lernziele in einen Studierendenfragebogen)

Bitte geben Sie an, wie stark Sie den folgenden Aussagen zustimmen. Die Aussagen beziehen sich auf Lernziele dieser Lehrveranstaltung. Kreuzen Sie ganz links an, wenn Sie der Aussage überhaupt nicht zustimmen und ganz rechts, wenn Sie der Aussage voll und ganz zustimmen. Falls Sie sich nicht zwischen zwei Antwortalternativen entscheiden können, kreuzen Sie bitte diejenige an, die aus Ihrer Sicht am ehesten zutrifft.

Durch den Besuch dieser Lehrveranstaltung...

... kann ich stetige von diskreten Verteilungen unterscheiden.	trifft überhaupt nicht zu	☐	☐	☐	☐	☐	☐	trifft voll und ganz zu
... kann ich empirische und theoretische Verteilungen unterscheiden.	trifft überhaupt nicht zu	☐	☐	☐	☐	☐	☐	trifft voll und ganz zu
... bin ich in der Lage, wissenschaftliche Hypothesen in Null- und Alternativhypothesen zu übertragen.	trifft überhaupt nicht zu	☐	☐	☐	☐	☐	☐	trifft voll und ganz zu

ten Lehrveranstaltungen, die sie hinsichtlich der Lernzielerreichung evaluieren lassen wollen. Der Fragebogen gibt die Qualifikationsziele der Modulbeschreibung vor. Lehrkräfte sollen ankreuzen, welche für ihre Lehrveranstaltung relevant sind. Da die Qualifikationsziele meist sehr allgemein formuliert sind, werden Lehrkräfte danach gebeten, zu den relevanten Qualifikationszielen konkrete, auf ihre Lehrveranstaltung inhaltlich zugeschnittene Indikatoren zu formulieren. Dabei kann eine unterschiedliche Anzahl an Indikatoren (max. drei) pro Qualifikationsziel genannt werden. Zusätzlich können Lehrkräfte weitere Indikatoren formulieren, die keinem Qualifikationsziel zuordenbar sind (max. drei). Alle von den Lehrkräften formulierten Lernziele werden in einen Fragebogen umformuliert, durch den Studierende einschätzen sollen, wie stark sie durch den Besuch der Lehrveranstaltung das Lernziel als erfüllt betrachten. Da das Vorgehen nicht in einem einzelnen Fragebogen resultiert, können keine globalen Befunde zur psychometrischen Güte oder zur Validität geliefert werden. Bislang wurden 142 Lehrveranstaltungen hinsichtlich der Lernzielerreichung evaluiert. Lehrkräfte nannten zwischen 3 und 22 Indikatoren, die Reliabilität (Cronbach's alpha) lag zwischen .73 und .94 u.a. abhängig von der Anzahl der genannten Indikatoren. Neben den guten Befunden zur Reliabilität empfinden viele Lehrkräfte den Prozess der Identifikation relevanter Qualifikationsziele und dazu gehöriger Indikatoren als eine gute Strukturierungs- und Planungshilfe für Lehrveranstaltungen und Prüfungen.

2. Modelle zur Vorhersage von Lernzeit und Lernerfolg

Vorhersagemodelle von Lernzeit und akademischem Lernerfolg (u.a. Bloom 1976; Carroll 1963; Helmke 2012) dienen der theoretischen Einordnung der im Rahmen der Lehrveranstaltungsevaluation eingesetzten Instrumente. Somit bieten sie prüfbare Hypothesen zur Rolle von Prozessvariablen der Lehrveranstaltung (allg. Lehrveranstaltungsbewertung und Workload) bei der Vorhersage von Ergebnisvariablen bzw. Lernerfolgsindikatoren (z.B. Lernzielerreichung, Klausurnoten, etc.). Die Modelle differenzieren zwischen individuellen Lernvoraussetzungen (z.B. kognitive und motivationale Merkmale der Lernenden), Unterrichtsmerkmalen (z.B. Qualität und Anforderungen) und Kontextmerkmalen (z.B. soziales Umfeld, Lernbedingungen). Speziell für den Hochschulkontext haben sich Helmke und Schrader (1996) mit der Rolle der Lernzeit im Selbststudium und ihren Determinanten auseinandergesetzt. Das Modell schulischen Lernens (Carroll 1963; s. Abbildung 2a). Nach dem Modell schulischen Lernens nach Carroll (1963) ist der Lernerfolg eine Funktion aus aufgewendeter und benötigter Lernzeit. Die benötigte Lernzeit hängt von der aufgabenspezifischen Begabung und

der Fähigkeit, den Unterricht zu verstehen, ab. Letztere wird durch allgemeine kognitive und verbale Fähigkeit determiniert und steht in Interaktion mit der Unterrichtsqualität. Die tatsächlich aufgewendete Lernzeit ist abhängig von Ausdauer und der verfügbaren Lernzeit. Ausdauer ist in diesem Kontext als die Zeit konzipiert, die ein Schüler bereit ist, aktiv zu lernen. Als zentralen Punkt kann man bezeichnen, dass das Unterrichtsgeschehen für die Vorhersage von Lernerfolg zunächst ein distaler Prädiktor ist. Nur durch die Tatsache, dass ein Schüler das Unterrichtsangebot aktiv nutzt (d.h. aktive Lernzeit investiert), kann sich Lernerfolg einstellen. Der Lernerfolg ist dabei umso größer, je positiver das Verhältnis von aufgewendeter zu benötigter Lernzeit ausfällt (Carroll 1963).

Determinanten der Lernzeit (Helmke/Schrader 1996; s. Abbildung 2b). Helmke und Schrader (1996) gehen davon aus, dass die aktive Lernzeit – und durch sie vermittelt der Lernerfolg – ebenfalls von unterschiedlichen Bedingungsfaktoren determiniert werden: a) psychologische Personenvariablen (objektive und subjektiv empfundene Kompetenzen, Ziele wie das Anspruchsniveau und die Valenz wie Interesse und intrinsische Motivation für das eigene Studium), b) Merkmale des persönlichen Hintergrunds (z.B. familiäre und finanzielle Situation) und c) institutionelle und ökologische Bedingungsfaktoren (z.B. Rahmenbedingungen im Studium, Qualität der Lehre). Abbildung 2b zeigt, dass Helmke und Schrader (1996) eine negative Beziehung zwischen objektiver/subjektiver Kompetenz und der investierten Lernzeit annehmen. Ihre Annahme ist, dass „je leistungsstärker man ist und je mehr man sich zutraut, desto weniger Zeit wird man [...] letztendlich auch investieren“ (S. 42). Dahingegen wirken sich Interesse, Motivation und Ziele positiv auf die investierte Lernzeit aus.

Synopse. Bezüglich des Wirkgefüges zwischen objektiven (und subjektiven) Kompetenzen, der Lernzeit und dem Lernerfolg sind das Modell des schulischen Lernens (Carroll 1966) und das Modell der Determinanten der Lernzeit (Helmke/Schrader 1996) ineinander überführbar. Die objektiven Kompetenzen entsprechen in der Terminologie Carrolls der Fähigkeit, den Unterricht zu verstehen, der allgemeinen und aufgabenspezifischen Begabung und dem fachspezifischen Vorwissen und bestimmen so die benötigte Lernzeit. Die tatsächlich aufgewendete Lernzeit wird in beiden Modellen durch die

Abbildung 2a: Modell des schulischen Lernens adaptiert und vereinfacht nach Carroll (1963)

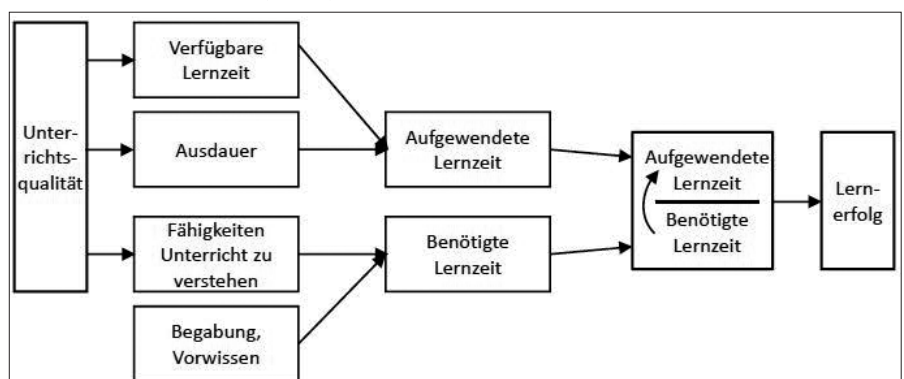
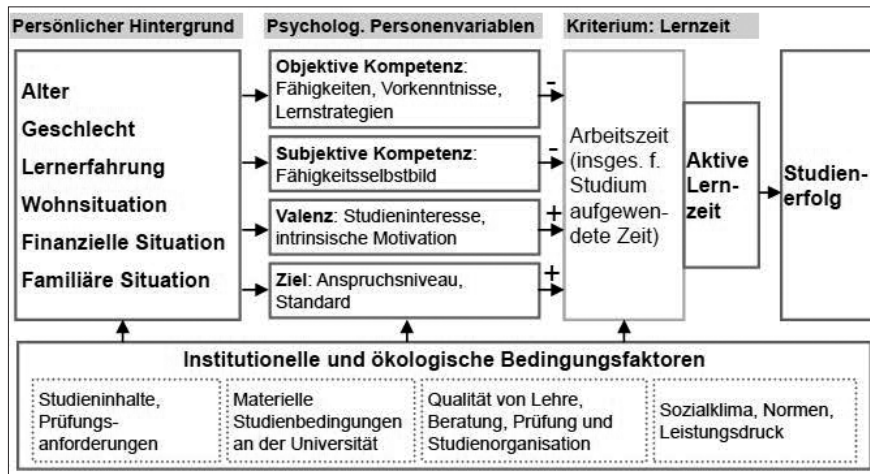


Abbildung 2b: Determinanten der Lernzeit nach Helmke & Schrader (1996)



benötigte Lernzeit beeinflusst, aber auch durch motivationale Faktoren (Ausdauer im Modell von Carroll, Interesse und Anspruchsniveau im Modell von Helmke und Schrader). Der Einfluss der subjektiven Kompetenz wird nur bei Helmke und Schrader spezifiziert. Auch wenn beide Modelle die Art des Zusammenhangs zwischen den Prädiktoren nicht explizieren, ist zum einen davon auszugehen, dass sie miteinander korreliert sind, und zum anderen, dass sie nicht nur additiv, sondern auch in Interaktion miteinander den Lernerfolg determinieren können (so genannte Moderatoreffekte). So kann theoretisch z.B. mangelndes Vorwissen durch mehr tatsächlich aufgewendete Lernzeit ausgeglichen werden. Bisherige Befunde zur Vorhersage der Lernzeit und des Lernerfolgs sind uneinheitlich (Helmke/Schrader 1996; Süß/Oberauer/Gelbert 2005) und scheinen vom zugrundeliegenden Studiengang (hier: Psychologie vs. Lehramt) und den gewählten Indikatoren für die einzelnen Merkmalsbereiche abhängig zu sein (z.B. $\beta = -.16$ bei der Vorhersage der Lernzeit durch die Mathe-Note bei Psychologie-Studierenden vs. $\beta = .25$ bei der Vorhersage der Lernzeit durch die Fremdsprache-Note bei Lehramts-Studierenden; Helmke/Schrader 1996). Moderatoreffekte wurden in keiner der beiden Studien geprüft.

3. Empirische Studien zum Zusammenhang der Elemente der Lehrveranstaltungsevaluation und zu ihrer Kriteriumsvalidität für Lernerfolg

In bisher drei Studien wurden die Elemente der Lehrveranstaltungsevaluation an der Universität Koblenz-Landau auf ihre Zusammenhänge miteinander und ihre Kriteriumsvalidität für Lernerfolg hin überprüft. Entsprechend der theoretischen Vorannahmen stellen die Prozessmerkmale (die Veranstaltungsbewertung mit dem adaptierten TRIL und die Lernzeit) Prädiktoren für die Erreichung von Lernzielen und Klausurergebnissen dar. Dabei ist davon auszugehen, dass die Lernzielerreichung ein proximales Kriterium zum Lehrprozess darstellt im Vergleich zur Klausurleistung als distales Kriterium, da die Lernzielerreichung explizit das Ausmaß erfragt, in dem der Lernerfolg durch die Lehrveranstaltung vermittelt

wurde („Durch den Besuch der Lehrveranstaltung kann ich ...“, s. Abbildung 1). Wenn möglich, wurden zusätzlich psychologische Personenvariablen erfasst (u.a. als Indikatoren für die objektive Kompetenz).

3.1 Studien 1 und 2

Stichproben, Instrumente und Prozedere. In zwei Vorlesungen (*Statistik II: Inferenzstatistik* im Studiengang Diplom Psychologie ($N=47$) und *Lehr-Lernmedien* im lehramtsbezogenen Bachelor-Studiengangs ($N=43$)) wurden folgende Instrumente eingesetzt: a) adaptierter TRIL (über alle vier Skalen aggregierter Gesamtwert) und Gesamtnote für die Lehrveranstaltung (Notenskala 1 bis 6),

- b) Vor- und Nachbereitung mit verbaler Verankerung („Ich habe mich auf die Veranstaltung regelmäßig vorbereitet.“, „Ich habe die einzelnen Sitzungen regelmäßig nachbereitet.“ von 1 „trifft gar nicht zu“ bis 6 „trifft voll zu“) als Indikator für den betriebenen Lernaufwand,
- c) veranstaltungsspezifische Lernziel-Skala mit 13 bzw. 10 Lernzielen,
- d) Punktschumme in der Abschlussklausur.

Die Erhebung des adaptierten TRIL, der Vor- und Nachbereitung und der Erreichung der Lernziele erfolgte gegen Ende der Lehrveranstaltungszeit, die Klausuren wurden 2-3 Wochen danach geschrieben.

Korrelative Ergebnisse für Studie 1 und 2 (beide Lehrveranstaltungen). Tabelle 3 zeigt die Mittelwerte, Standardabweichungen, Reliabilitätskoeffizienten und Interkorrelationen für beide Lehrveranstaltungen.

Das Muster der Interkorrelationen war für beide Lehrveranstaltungen vergleichbar. TRIL-Gesamtwert und -Gesamtnote waren hoch miteinander korreliert ($r = .77/- .81$). Der Grad der Vor- und Nachbereitung hing kaum mit der Veranstaltungsbewertung durch den TRIL zusammen, zeigte aber positive Zusammenhänge mit der Lernzielerreichung ($r = .46/.28$) und der Punktschumme in den Klausuren ($r = .21/.33$). Die größten Zusammenhänge mit der Klausurleistung zeigte die Lernzielerreichung ($r = .36/.23$).

Regressionsanalysen Vorlesung Statistik II (Studie 1). Bei simultaner Aufnahme der Prozessvariablen zur Vorhersage der Lernzielerreichung zeigten der TRIL-Gesamtwert ($\beta = .52$; $t = 3.15$; $p < .01$) und die Vor- und Nachbereitung ($\beta = .34$; $t = 3.08$; $p < .01$) signifikante Vorhersagebeiträge, während die TRIL-Gesamtnote ($\beta = -.01$; $t = -.06$; n.s.) keine eigenständige Varianz an der Lernzielerreichung erklären konnte (bei einem $R^2 = .46$; $F = 13.87$; $p < .01$). Bei der Vorhersage der Klausur-Punktschumme ($R^2 = .13$; $F = 8.60$; $p < .01$); zeigte der Grad der Lernzielerreichung als einzige Variable einen substantiellen Einfluss ($\beta = .31$; $t = 1.86$; $p = .07$). Der TRIL-Gesamtwert ($\beta = .01$; $t = .35$; n.s.) und die Vor- und Nachbereitung ($\beta = .05$; $t = .06$; n.s.) zeigten keinen eigenständigen Anteil mehr, obwohl beide mit

Tabelle 3: Mittelwerte, Standardabweichungen, Reliabilität und Interkorrelationen des adaptierten TRIL, des Grades der Vor- und Nachbereitung, der Lernziel-Skala und der Klausur-Punktsommen in Studien 1 und 2

Variable	M	SD	1	2	3	4	5
1 TRIL Gesamtwert*	4.14 4.56	.79 .51	.84 .79				
2 TRIL Gesamtnote ^(c)	2.93 2.25	.92 .71	-.77** -.81**	-			
3 Vor-/Nachbereitung	3.74 2.60	1.36 1.13	.10 -.08	-.29* -.16	.87 .57		
4 Lernziel-Skala*	4.00 3.92	.90 .75	.50** .25*	-.51** -.36**	.46** .28*	.89 .87	
5 Klausur-Punktsomme	84.03 25.75	14.94 4.13	.20 -.09	-.16 -.04	.21 .33*	.36** .23	-.80

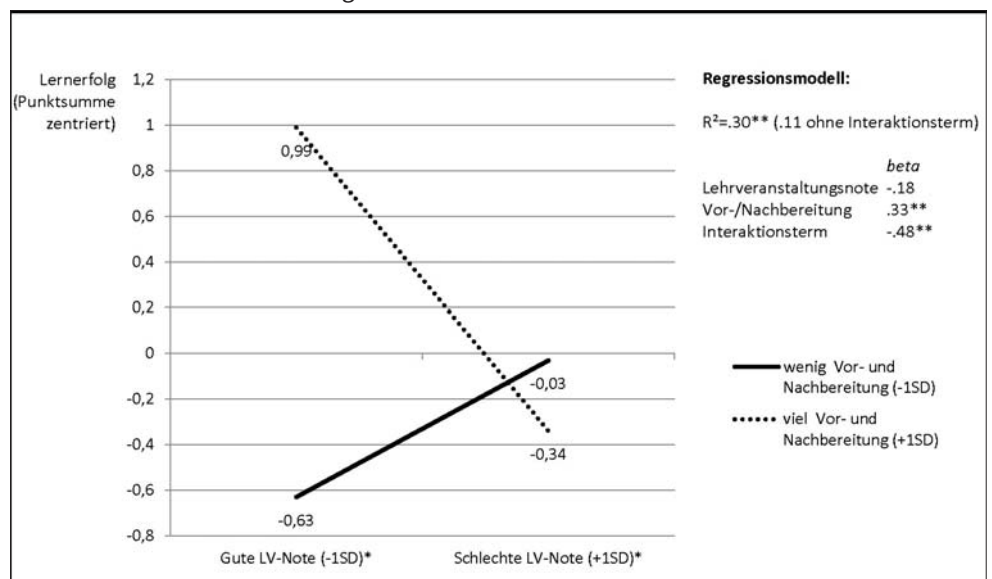
Die jeweils obere Zeile pro Zeile zeigt die Kennwerte der Vorlesung Statistik II (Studiengang Psychologie, die jeweils untere Zeile die der Vorlesung Lehr-Lernmedien (Studiengang Bildungswissenschaften). Die Diagonale stellt die Reliabilitätskoeffizienten dar (Cronbach's alpha) mit Ausnahme der TRIL Gesamtnote, die mit nur einer Frage erfasst wurde.

* alle adaptierten TRIL-Fragen bzw. Fragen zur Lernzielerreichung aggregiert

$r=.20/.21$ mit der Punktsomme korreliert waren. Statistisch deutet dieser Befund darauf hin, dass der Grad der Lernzielerreichung den Einfluss der Prozessvariablen vollständig vermittelt (Baron/Kenny 1986). Dieses Ergebnis legt eine nicht-kompensatorische Verknüpfung der Prädiktoren untereinander nahe, da eine schlechte Lehrveranstaltung bzw. mangelnder Aufwand für die Vor- und Nachbereitung eher mit einer schlechten Lernzielerreichung und damit einer schlechten Klausurleistung einhergehen.

Regressionsanalysen Vorlesung Lehr-Lernmedien (Studie 2). Die Regressionsanalysen wiesen zunächst vergleichbare Befunde zu Studie 1 auf. Der TRIL-Gesamtwert ($\beta=.28$; $t=2.45$; $p<.05$) und der Grad der Vor- und Nachbereitung ($\beta=.30$; $t=2.66$; $p<.05$) zeigten signifikant positive Vorhersagebeiträge für den Grad der Lernzielerreichung ($R^2=.15$; $F=5.98$; $p<.01$), während die TRIL-Gesamtnote keinen eigenständigen Vorhersagebeitrag mehr leisten konnte. Bei der Vorhersage der Klausur-Punktsomme zeigte sich dagegen ein deutlich komplexeres Variablengefüge. Ein Mediatormodell wie in Studie 1 konnte nicht bestätigt werden, da die eigenständigen Vorhersagebeiträge bei der Hinzunahme der Lernziel-Skala stabil blieben und der Einfluss der Note sich sogar im Vorzeichen umkehrte. Da dies auf eine Interaktion der Prädiktoren untereinander hindeutete, wurde ein Moderatormodell berechnet mit einer Interaktion zwischen der TRIL-Gesamtnote und der Vor- und Nachbereitung (mehr Prädiktoren konnten aufgrund der geringen Fallzahlen nicht ins Modell aufgenommen werden).

Abbildung 3: Ergebnis des Regressionsmodells mit Moderatoreffekt zwischen der Lehrveranstaltungsnote



* Abi-Schnitt negativ gepolt, kleinere Werte bedeuten eine bessere Note

Abbildung 3 stellt das Ergebnis graphisch dar und zeigt die beta-Koeffizienten und die Gesamtvarianzaufklärung.

Der Interaktionseffekt zwischen der TRIL-Note und der Vor- und Nachbereitung wird signifikant ebenso wie der eigenständige Vorhersagebeitrag der Vor- und Nachbereitung. Die Gesamtvarianzaufklärung liegt mit $R^2=.30$ weit über der ohne den Interaktionsterm ($R^2=.11$). Inhaltlich bedeutet der Moderatoreffekt, dass Studierende mit viel Vor- und Nachbereitungsaufwand und einer gleichzeitig guten Bewertung der Lehrveranstaltung ganz besonders profitieren (gestrichelte Linie in Abbildung 3). Umgekehrt hat eine wahrgenommene gute Qualität der Lehrveranstaltung bei wenig Vor- und Nachbereitungsaufwand noch zusätzlich einen negativen Effekt auf das Klausurergebnis. Ggf. könnte hier die Erwartung der Studierenden zu Grunde liegen, dass sie fälschlicherweise auf eine gute Klausur hoffen (durchgezogene Linie in Abbildung 3) auch und gerade wenn sie sich nicht viel vorbereiten. Über den Moderationseffekt hinaus zeigt sich der uneingeschränkte Einfluss der Vor- und Nachbereitung, was darauf schließen lässt, dass mangelnder Lernaufwand kaum durch eine bessere Lehrveranstaltung kompensiert werden kann.

3.2 Studie 3

Stichproben, Instrumente und Prozedere. Die Stichprobe bestand aus Studierenden in der Vorlesung Deskriptive Statistik im Studiengang Bachelor Psychologie ($N=55$). Es wurden folgende Variablen erhoben:

- Abitur-Durchschnittsnote als Indikator für die objektive Kompetenz im Sinne des Helmke/Schrader-Modells (1996),
- Investierte aktive Lernzeit,

c) Punktschnee in der Abschlussklausur der jeweiligen Lehrveranstaltung.

In einem längsschnittlichen Design wurden die Abiturnote und das akademische Selbstkonzept zu Beginn der Lehrveranstaltung erhoben, die aktive Lernzeit wurde zu drei Befragungszeitpunkten erfasst, die Punktschnee in der Klausur nach Ende der Lehrveranstaltung.

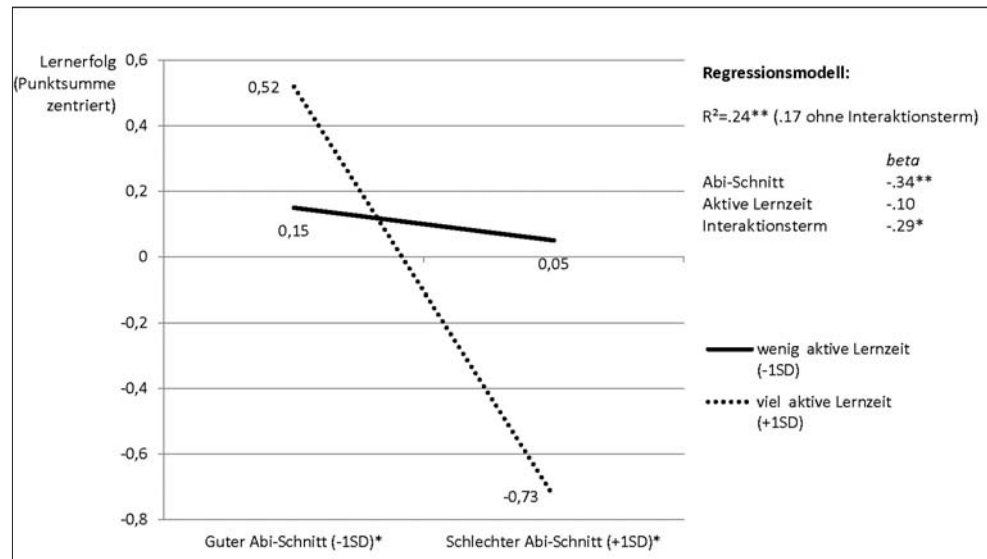
Ergebnisse. Korrelationsanalysen zeigten erwartungskonform hohe Zusammenhänge zwischen dem Abi-Schnitt (negativ gepolt) und der Punktschnee in der Klausur ($r = -.38$). Dagegen war die Korrelation zwischen aktiver Lernzeit und Klausurleistung mit einem negativen Vorzeichen versehen ($r = -.22$). Da diese negative Korrelation aus theoretischer Sicht nicht erklärbar ist, wurde wiederum ein Moderatormodell mit einer Interaktion zwischen dem Abi-Schnitt und der investierten Zeit gerechnet.

Abbildung 4 stellt das Ergebnis dar und zeigt einen Interaktionseffekt, der neben dem Haupteffekt des Abi-Schnitts maßgeblich zur Gesamtvarianzaufklärung der Klausurleistung beiträgt ($R^2 = .24$; $F = 5.498$; $p < .01$). Studierende, die einen guten Abi-Schnitt mitbringen, profitieren am meisten, wenn sie viel Zeit investieren (linker Pol der gestrichelten Linie). Dagegen scheinen sich Studierende mit schlechtem Abi-Schnitt zu benachteiligen, wenn sie viel Zeit investieren, was ggf. auf ungünstige Lernstrategien hindeutet (rechter Pol der gestrichelten Linie).

3.3 Zusammenfassung und methodenkritische Bewertung der dargestellten Befunde

Die bislang vorliegenden Daten (s. auch Fondel et al., 2014 zur Validität des adaptierten TRIL) sind vielversprechend. Die Lernzielerreichung erweist sich als wichtiger Indikator, der sowohl das Kriterium Lernerfolg abbildet, als auch auf den Prozess der Lehrveranstaltung bezogen ist und in beide Richtungen erwartungskonforme Zusammenhangsmuster zeigt. Die Ergebnisse der randomisierten Workload-Befragung stützen beide Erhebungsvarianten (Einfach- und Dreifach-Erhebung) zumindest für die Nutzung der mittleren Workload-Einschätzungen für qualitätssichernde Maßnahmen, wobei die Datenlage zur Einfach-Erhebung relativ schwach ist. Insgesamt zeichnen die Daten der Studien 1 bis 3 aber ein komplexes Bild. Die investierte Lernzeit zeigt sich als relevanter Einflussfaktor für Lernerfolg in Übereinstimmung mit den theoretischen Modellen von Helmke und Schrader (1996) und Carroll (1963), wobei die Lernzeit mit weiteren Prädiktoren auf unterschiedliche Art und Weise interagiert. Die objektive Kompetenz weist erwartungskonform einen hohen Vorhersagebeitrag für

Abbildung 4: Ergebnis des Regressionsmodells mit Moderatoreffekt zwischen dem Abi-Schnitt und der investierten, aktiven Lernzeit zur Vorhersage der Klausur-Punktschnee in Studie 3



* Abi-Schnitt negativ gepolt, kleinere Werte bedeuten eine bessere Note

den Lernerfolg (Studie 3) auf. Alle Befunde sprechen wiederum dafür, den Zusammenhang zwischen Variablen nicht nur auf der Ebene der bivariaten Zusammenhänge zu betrachten, sondern Modelle mit komplexeren Interaktionen zu testen.

Aufgrund der geringen Fallzahlen und der Tatsache, dass alle Befunde immer nur auf eine Lehrveranstaltung bezogen sind, bedürfen sie allerdings der Replikation auf einer breiteren Datenbasis über mehrere Lehrveranstaltungen hinweg. So könnten weitere Einflussfaktoren identifiziert werden, die nicht bei den Studierenden, sondern in den Merkmalen des Fachs oder denen der Lehrkräfte zu finden sind. Des Weiteren sind – mit Ausnahme von Studie 3 – alle Prädiktoren zu einem Erhebungszeitpunkt erfasst worden, so dass Rückschlüsse auf kausale Zusammenhänge kritisch zu sehen sind (z.B.: Wird die Lehrveranstaltung besser beurteilt, weil die Studierende sich mehr vor- und nachbereiten und so den Unterricht besser verstehen können? Oder bereiten sie sich mehr vor und nach, weil sie die Lehrveranstaltung als besser wahrnehmen?).

3.4 Fazit und Ausblick

Abgesehen von den Einschränkungen der Generalisierbarkeit der Befunde über die untersuchten Lehrveranstaltungen hinaus, können bei vorsichtiger Interpretation auch weiterführende Schlussfolgerungen für die Qualitätssicherung in der universitären Lehre gezogen werden. So erscheint eine Passung des Fähigkeits- und Kompetenzprofils der Studierenden zentral für den Lernerfolg. Dies spricht u.a. für eine umfassende Information der Studieninteressierten über die Anforderungen und Erfolgsfaktoren eines Studiums, wie es seit einigen Jahren im Rahmen von Online Self-Assessments erfolgreich geschieht (s. u.a. Hasenberg/Schmidt-Atzert 2014). Ebenfalls für qualitätssichernde Maßnahmen in einem Studiengang interessant erscheint eine nähere Betrachtung der differentiellen Aspekte des Workload. Auch

wenn per definitionem nur der Workload eines/einer mittleren Studierenden von Interesse ist, ist doch auch relevant, a) aus welchen Gründen Studierende mehr oder weniger Zeit investieren, b) ob sie sich damit im Vergleich zu ihren Kommilitonen und Kommilitoninnen benachteiligen oder benachteiligt sind und c) ob diese Benachteiligung potentiell sogar zu einem Studienabbruch oder einer verlängerten Studiendauer führen kann.

Literaturverzeichnis

- Ackerman, P. L./Beier, M. E./Boyle, M. B. (2002): Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. In: *Journal of Experimental Psychology: General*, No. 131, pp. 567-605.
- Baron, R. M./Kenny, D. A. (1986): The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. In: *Journal of Personality and Social Psychology*, No. 51, pp. 1173-1182.
- Bloom, B. S. (1976): *Human Characteristics and School Learning*. New York.
- Blüthmann, I./Ficzko, M./Thiel, F. (2006): FELZ – ein Instrument zur Erfassung der studienbezogenen Arbeitsbelastung. In: Berendt, B./Wild, J./Voss, H.-P. (Hg.), *Neues Handbuch Hochschullehre*, 2. Aufl., S. 1-30. Stuttgart, Berlin.
- Blüthmann, I./Thiel, F. (2011): Sind pauschale Workload-Einschätzungen aussagekräftig? In: Steinhardt, I. (Hg.), *Studierbarkeit nach Bologna*, Mainzer Beiträge zur Hochschulforschung, Bd. 17, S. 82-95.
- Braun, E./Gusy, B./Leidner, B./Hannover, B. (2008): Kompetenzorientierte Lehrrevaluation – Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). In: *Diagnostica*, H. 54, S. 30-42.
- Burck, K./Heil, K./Böhres, M. (2011): Quantitative Workload Erhebungen an der Johannes Gutenberg-Universität Mainz. In: Steinhardt, I. (Hg.): *Studierbarkeit nach Bologna*. Mainzer Beiträge zur Hochschulentwicklung, Bd. 17, S.49-66.
- Carroll, J. B. (1963): A model of school learning. *Teachers College Record*, No. 64, pp. 723-733.
- Dickhäuser, O./Schöne, C./Spinath, B./Stiensmeier-Pelster, J. (2002): Die Skalen zum akademischen Selbstkonzept. Konstruktion und Überprüfung eines neuen Instrumentes. In: *Zeitschrift für Differentielle und Diagnostische Psychologie*, H. 23, S. 393-405.
- European Association of Quality Assurance in Higher Education (2009): *The Standards and Guidelines for Quality Assurance in the European Higher Education Area* (3rd Edition). Helsinki.
- Fondel, E./Lischetzke, T./Weis, S./Gollwitzer, M. (2014, in *Begutachtung*): Zur Validität von studentischen Lehrveranstaltungsevaluationen: Messinvarianz über Veranstaltungsarten, Konsistenz von Urteilen und Gründe für ihre Heterogenität. In: *Diagnostica*, Sonderheft.
- Goldberg, L. R./Johnson, J. A./Eber, H. W./Hogan, R./Ashton, M. C./Cloninger, C. R./Gough, H. C. (2006): The International Personality Item Pool and the future of public-domain personality measures. In: *Journal of Research in Personality*, No. 40, pp. 84-96.
- Gollwitzer, M./Schlotz, W. (2003): Das „Trierer Inventar zur Lehrveranstaltungsevaluation“ (TRIL): Entwicklung und erste testtheoretische Erprobungen. In: Krampen, G./Zayer, H. (Hg.): *Psychologiedidaktik und Evaluation IV*, S. 114-128. Bonn.

- Hasenberg, S./Schmidt-Atzert, L. (2014): Internetbasierte Selbsttests zur Studienorientierung. In: *Beiträge zur Hochschulforschung*, 36. Jg, H. 1., S. 8-28.
- Helmke, A. (2012): Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts. 4. überarb. Auflage. Seelze.
- Helmke, A./Schrader, F.-W. (1996): Kognitive und motivationale Bedingungen des Studierverhaltens: Zur Rolle der Lernzeit. In: Lompscher, J./Mandl, H. (Hg.): *Lehr-Lernprobleme im Studium – Bedingungen und Veränderungsmöglichkeiten*. Bern.
- KMK Kultusministerkonferenz (2010): *Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen*. Beschluss der Kultusministerkonferenz vom 10.10.2003 i.d.F. vom 04.02.2010.
- Marsh, H. W. (1987): Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. In: *International Journal of Educational Research*, No. 11, pp. 253-388.
- Rindermann, H. (2009): Lehrrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierter Unterrichts. In: *Psychologie*, Bd. 42, 2. leicht korrigierte Auflage. Landau.
- Rindermann, H./Neubauer, A. C. (2004): Processing speed, intelligence, creativity, and school performance: testing of causal hypotheses using structural equation models. In: *Intelligence*, No. 32, pp. 573-589.
- RWTH Aachen (2014): Studentische Online Workload Erfassung Aachener Hochschulen (StOEHN). http://www.rwth-aachen.de/cms/root/Studium/Im_Studium/Lehrveranstaltungen/~bmfu/StOEHN/. (Zugriff: 05.06.2014)
- Spiel, C./Wolf, P./Popper, V. (2002): Lehre und Lehrrevaluation – (un)geliebt? Die Perspektive der Universitätslehrenden. In: *Zeitschrift für Psychologie/Journal of Psychology*, H. 210, S. 27-39.
- Süß, H.-M./Oberauer, K./Gelbert, T. (2005): The determinants of successful statistics learning: A longitudinal study based on John B. Carroll's model of school learning. In: Beauducel, A./Biehl, B./Bosniak, M./Conrad, W./Schönberger, G./Wagener, D. (Hg.), *Multivariate Research Strategies – Festschrift in Honor of Werner W. Wittmann*, pp. 283-301. Aachen.

■ **Dr. Susanne Weis**, wissenschaftliche Mitarbeiterin, Zentrum für Methoden, Diagnostik und Evaluation, Universität Koblenz-Landau (Campus Landau), Schwerpunkt Lehrrevaluation, E-Mail: weis@uni-landau.de

■ **Christiane Karthaus**, wissenschaftliche Mitarbeiterin Zentrum für Methoden, Diagnostik und Evaluation, Universität Koblenz-Landau (Campus Koblenz) mit dem Schwerpunkt Lehrrevaluation, E-Mail: karthaus@uni-koblenz.de

■ **Dr. Tanja Lischetzke**, Professorin für Methodenlehre und Evaluation, Universität Koblenz-Landau (Campus Landau), E-Mail: lischetzke@uni-landau.de

im Verlagsprogramm erhältlich:

**Ludwig Huber, Margot Kröger & Heidi Schelhowe (Hg.):
Forschendes Lernen als Profilvermerkmal einer Universität.
Beispiele aus der Universität Bremen**

ISBN-13: 978-3-937026-83-1, Bielefeld 2013, 266 Seiten, 38.60 Euro

Bestellung: E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22

Jörg Jörissen & Michael Heger

Zur Wirkung hochschuldidaktisch fundierter und in den Fachbereichen verankerter Lehrveranstaltungsevaluation



Jörg Jörissen



Michael Heger

Wird über Lehrveranstaltungsevaluation geschrieben, steht häufig eine Methodendiskussion über die Befragungsinstrumente im Vordergrund¹. Insbesondere mit Blick auf die Wirkung der Lehrveranstaltungsevaluation können methodische Qualität und Differenzierungsgrad der eingesetzten Fragebögen aber durchaus als zweitrangig (Kromey 2003, S. 14) und Fragen nach der konzeptionellen Einbettung und nach verallgemeinerbaren Standards für die Verfahren der Lehrveranstaltungsevaluation als vorrangig (Schmidt 2008, S. 30) betrachtet werden.

Diesem Artikel liegen über 16 Jahre Erfahrungen aus der konzeptionellen und institutionellen Kopplung von Hochschuldidaktik und Evaluation an der FH Aachen zugrunde. Dies sind im Einzelnen Erfahrungen in der Entwicklung entsprechender Erhebungsinstrumente, in der Einführung und Umsetzung einer Evaluationsordnung, im Aufbau von Strukturen und Evaluationskreisläufen zu Studium und Lehre unter Koordination des Zentrums für Hochschuldidaktik und Qualitätsentwicklung in Studium und Lehre (ZHQ)².

Diese Anstrengungen sind auf ihre Wirksamkeit hin zu untersuchen. Am Beispiel der Lehrveranstaltungsevaluation ist zu fragen: Inwieweit ist es gelungen, die Evaluation mit hochschuldidaktischem Grundlagen- und Erfahrungswissen konzeptionell zu fundieren sowie wirkungsvolle Strukturen, Verfahren und Instrumente zur Durchführung aufzubauen und den Evaluationskreislauf mit hochschuldidaktischer Weiterentwicklung (Beratung, Weiterbildung, Curriculumsentwicklung) zu schließen? Auf den folgenden Seiten wird anhand von Zahlen, Fakten und Eindrücken diesen Fragen zur Wirksamkeit hochschuldidaktisch fundierter Lehrveranstaltungsevaluation nachgegangen und insbesondere auf die wichtige Rolle der fachbereichsinternen Strukturen und Verfahren eingegangen.

1. Konzept: Ausgangspunkte, Ziele, Wirkungsplan

Grundlage der folgenden Ausführungen sind theoretische Ansätze, die anwendungsorientiert erschlossen und in der fortwährenden alltäglichen hochschuldidaktischen und evaluativen Arbeit angewandt sowie weiter entwickelt wurden (vgl. Heger 1984 bis Heger 2009).

1.1 Lehrveranstaltungsevaluation als angewandte Hochschuldidaktik

Unter Hochschuldidaktik verstehen die Autoren die Anwendung und Erweiterung wissenschaftlicher Erkenntnisse über Studium und Lehre an Hochschulen

- im Alltag von Studium und Lehre,
- in der Aus- und Weiterbildung sowie Beratung zu Studium und Lehre,
- in Studienreform und Wissenschaftsentwicklung und
- in der dazugehörigen fortwährenden anwendungsorientierten Hochschulforschung (insbesondere in Form der Evaluation).

Sind die Adressaten/innen und Hauptkooperationspartner/innen so verstandener Hochschuldidaktik vordergründig meist Lehrende, grundsätzlich und letztendlich aber Studierende, so gilt dasselbe für die Evaluation. Durch sie sollen Entwicklungspotentiale im Alltag von Studium und Lehre identifiziert und Innovationsprozesse angestoßen werden, die vorrangig in der Lehrpraxis und durch Lehrende realisiert werden, aber letzten Endes das Studieren verbessern sollen.

Mit dieser Orientierung auf Innovation und Weiterentwicklung folgt die Evaluation an der FH Aachen einem 'Entwicklungsparadigma' (vgl. Kromey 2003, S. 8 ff.). Sie wird zur Vorbereitung rational fundierter Entscheidungen zum einen als formative Selbstevaluation und zum anderen als anwendungsorientierte und methodisch anspruchsvolle Wissenschaft (vgl. Standards für Evaluation, DeGEval 2008) verstanden, die in Abgrenzung zur strengen empirischen Grundlagenforschung zumindest bei

¹ Genügen sie methodischen Anforderungen? Sind sie im Stande alle relevanten Aspekte 'guter Lehre' zu erfassen und verlässliche Hinweise auf die Lehrveranstaltungsqualität zu liefern? Wo können Messfehler auftreten?

² 2013 wurde die lokale hochschuldidaktische Einrichtung der FH Aachen als 'Zentrum für Hochschuldidaktik und Qualitätsentwicklung in Studium und Lehre' (ZHQ) neu aufgestellt und verankert. Ihre Vorgänger reichen bis ins Jahr 1998 zurück. Die lokale hochschuldidaktische Einrichtung war von Beginn an – neben dem 2005 abgeschlossenen Aufbau eines landesweiten hochschuldidaktischen Weiterbildungs- und Beratungsangebotes der Fachhochschulen (s. www.hdw-nrw.de), der bis 2006 ebenfalls integrierten Studienberatung und der seit 2007 integrierten Unterstützung bei Akkreditierung und Reakkreditierung – auch mit dem Aufbau und der Weiterentwicklung der Evaluation von Studium und Lehre betraut. So werden auch in der neuen Verwaltungsordnung 'Evaluation und Qualitätsentwicklung in Studium und Lehre' als Anwendungsgebiete hochschuldidaktischer Expertise explizit aufgeführt (Verwaltungsordnung des ZHQ 2013, S. 2 ff.).

Einzelprojekten als Aktionsforschung (vgl. Heger 2004a) praktiziert wird.

Eine so konzipierte Evaluation orientiert sich im Gegensatz zur summativen Evaluation von Studienprogrammen (z.B. Akkreditierung und Reakkreditierung) nicht an fremdgesetzten Standards, sondern misst das Erreichte an den eigenen Zielen und Konzepten (Stake 1972)³. Gemäß der Devise ‚Tu Gutes und sprich drüber!‘ verfolgt sie auf diese Weise das Doppelziel der Qualitätsentwicklung (Innovation/Improvement, kontinuierlicher Verbesserungsprozess) einerseits und der Rechenschaftslegung (Legitimation/Accountability) andererseits. Der hierin potentiell angelegte Zielkonflikt soll durch die Wirkungskonzeption ‚Legitimation durch Innovation‘ aufgelöst werden. Im Zentrum einer so verstandenen Rechenschaftslegung steht der Nachweis über fortgesetzte Bemühungen um Weiterentwicklung bzw. Verbesserung.

Die hierbei zugrunde gelegte Wirkungskonzeption der Autoren setzt als Selbstevaluation einzelner Lehrender bzw. Fachbereiche bei der Eigenverantwortung an und differenziert grundsätzlich in allen Belangen zwischen individuellen (einzelne Lehrende bzw. Studierende) und strukturellen (Curriculum, Rahmenbedingungen) Faktoren.

Auch das hochschuldidaktische und evaluative Alltagswissen ist ein wichtiger Ausgangspunkt für die Wirkungsplanung der Lehrveranstaltungsevaluation an der FH Aachen. Als Erfahrungsquellen gerade auch für die Konzipierung und Umsetzung der Lehrveranstaltungsevaluation an der FH Aachen zu nennen sind dabei insbesondere ca. 70 sogenannte lehrbegleitende Beratungen (Lehrveranstaltungsbesuche mit Vor- und Nachgesprächen, meist auch Videoaufzeichnung) bei Professorinnen und Professoren (vgl. Heger 1997), davon über 20 an der FH Aachen sowie Forschungsarbeiten und Lehrtätigkeiten zum ‚Studieren lernen‘ (vgl. Heger/Sell 1986 und Heger/Mundyanapurath 2004). Die Entwicklungen zwischen 2008 und 2014, die in diesem Artikel resümiert werden, wurden gemeinsam durch die beiden Autoren initiiert und koordiniert. So wissen die Autoren aus vielfältigen Praxiserfahrungen (eigene Hochschullehre, Mitarbeit in Hochschulgremien, Einsatz der Lehrveranstaltungsevaluation im Rahmen der lehrbegleitenden Beratungen) um die oft aufgezwungene und wenig wirksame Handhabung der Lehrveranstaltungsevaluation: Interpretation der studentischen Rückmeldungen als zu heterogene, nicht wirklich kompetente Einzelmeinungen, Lochen, Abheften bzw. heute Speichern ohne weitere Verständigung bzw. Schlussfolgerungen. Wenn Lehrende ihr Lehrveranstaltungsfeedback nicht als Instrument zur Qualitätsverbesserung auffassen und für die Studierenden spürbar nutzen, ist erfahrungsgemäß auch auf Seiten der Studierenden trotz einer grundsätzlichen Offenheit für die Lehrveranstaltungsevaluation (vgl. Schmidt 2010, S. 183) eine Evaluationsmüdigkeit zu erwarten (vgl. Bargel/Hage 2000, S. 221). Um drohenden Ermüdungs- bzw. Verpuffungstendenzen vorzubeugen gilt es, von vornherein ein ‚Diskurs- und Beratungsmodell‘ mit konsequentem Follow-Up und wahrnehmbaren Verbesserungen in Folge der Lehrveranstaltungsevaluation zu entwickeln⁴.

1.2 Kompetenzorientierung und Lernpartnerschaft

Wird die Lehrveranstaltungsevaluation als angewandte Hochschuldidaktik verstanden, bilden ‚Kompetenzorientierung und -entwicklung‘ einen ganz zentralen theoretischen Rahmen. Im Vordergrund steht dabei die Lehrkompetenz. Sie ist freilich auch zu messen an und zu beziehen auf die Berücksichtigung und Verbesserung der Studierkompetenz der Lehrveranstaltungsteilnehmer und -teilnehmerinnen. Die wesentliche Evaluation der Studienkompetenz selbst sollte aber nicht über die veranstaltungsbegleitende Lehrevaluation, sondern in den Prüfungen erfolgen (im besten Fall formativ als studienbegleitende Prüfung zur Selbstvergewisserung anstatt nur summativ zur modulbezogenen Abschlussprüfung). Kompetenz wird hier verstanden als das Zusammenspielen von Person und Umwelt in der jeweiligen Situation (vgl. Hof 2001). Die Person, hier zentral der/die Lehrende, wird bzgl. ihres Wollens, Könnens und Wissens betrachtet. Trautwein und Merkt (2013) sprechen von Lehr-/Lernphilosophie, Handlungsstrategien und Kontextwissen. Die Umwelt, hier im Wesentlichen der Fachbereich, wird durch die Handlungserwartungen und die bereit gestellten Handlungsressourcen kompetenzrelevant. Die Situation ist die Lehrveranstaltung bzw. das fortwährende Geschehen in Hörsaal, Seminar- oder Laborraum.

Mit Bezug auf das Modell des Didaktischen Dreiecks (vgl. Heger 2004b) ergeben sich folgende Orientierungen:

- Inwieweit wird mit Blick auf die inhaltlichen und interaktionellen Tätigkeiten der Lehrenden ein *Ansatz/Selbstverständnis der Stoffdarstellung* und/oder *der Lernorganisation* praktiziert (darbietende Lehre vs. angeleitetes Lernen)?
- Inwieweit werden die verschiedenen Lernzielebenen – Faktenwissen, Zusammenhangsverständnis, Anwendungskönnen und kreatives Problemlösen (vgl. Deutscher Bildungsrat 1971) – berührt?
- Inwieweit wird die studentische Auseinandersetzung mit den Lehrinhalten (in der Veranstaltung, im Selbststudium und in der Prüfung) gefördert?

Neben den Lernzielebenen stehen also die Lehr-, Lern- und Prüfungsformen im Zentrum der Betrachtung. Die dahinterliegenden theoretischen Bezüge (vgl. Heger 1985) lassen sich als Handlungs-, Tätigkeits- und Prozessorientierung (vgl. jeweils Leontjew 1977; Hacker 1978; Volpert 1980) fokussieren. Deutlich different zum

³ Ein hervorragendes Evaluationsmodell wurde 1972 von Robert E. Stake (1972) vorgestellt: Voraussetzungen, Prozesse und Ergebnisse werden nicht nur unter einander abgeglichen, sondern bezüglich Konzept (intendierte Voraussetzungen, Prozesse und Ergebnisse) und Realisierung verglichen (vgl. Heger 1984).

⁴ Inzwischen ist empirisch unstrittig, dass die reine Evaluation zu gar keinen und erst mündliche Erläuterungen und persönliche Beratungen in Folge des Lehrveranstaltungsfeedbacks zu wirklich substantiellen Verbesserungen führen (zusammenfassend Rindermann 2009, S. 227 ff.). Anbieter und Nutzer der Lehrveranstaltungsevaluation sind sich auch darüber einig, dass im Wesentlichen „das Fehlen von Folgemaßnahmen wie Beratung und Weiterbildung sowie die mangelnde Einbindung der LVE [Lehrveranstaltungsevaluation, Anm. der Verfasser] in ein übergreifendes Konzept zur Kompetenzentwicklung“ zum Scheitern derselben führt (Schmidt 2008, S. 22). Schmidt fordert entsprechend, „zukünftig mehr Augenmerk darauf zu richten, welche Verfahren als Standards etabliert werden können“ (ebd. S. 32), einer Aufforderung der wir mit diesem Artikel gerne nachkommen wollen.

aktuell oft benutzten Kundenbegriff wird zumindest idealistisch das Konzept einer Lernpartnerschaft zwischen Lehrenden und Studierenden zugrunde gelegt, ohne dabei die Prüfungsabhängigkeit und -orientierung der Studierenden und die resultierenden Hierarchien zu ignorieren.

Der Ansatz einer formativen Selbstevaluation in Kombination mit dem Konzept einer Lernpartnerschaft entspricht einer sehr ursprünglichen Auffassung von Lehrveranstaltungsevaluation (vgl. Souvignier/Gold 2002) und legt als Mindeststandard für sie das ‚Diskursmodell‘ (vgl. Schmidt 2010) nahe, in dem Lehrende und Studierende als Akteure über Verlauf und Erfolge ihrer Veranstaltung diskutieren und gemeinsam Verbesserungsvorschläge erörtern.

2. Realisierung: Umsetzung, Ergebnisse, Entwicklungen

2.1 Grundsätze einer hochschuldidaktisch fundierten Evaluationsordnung

Als wichtiger Schritt hin zu einer hochschuldidaktisch fundierten Evaluation wurde Anfang 2004 eine bis heute gültige Ordnung zur Evaluation von Studium und Lehre⁵ im Sinne der in Abschnitt 1 erläuterten Konzeption an der FH Aachen eingeführt. Laut dieser Evaluationsordnung leistet die regelmäßig mit „standardisierten Instrumenten und Verfahrensweisen“ (§3 Abs.3) durchgeführte Evaluation einen Beitrag zur „systematischen Selbstanalyse, Qualitätssicherung und -verbesserung [...] sowie der Rechenschaftslegung“ (§2 Abs.1) auf Lehrveranstaltungs-, Fachbereichs- und Hochschulebene und geht im Sinne der Selbstevaluation „von den Zielvorstellungen der evaluierten Einheiten aus“ (§2 Abs.2).

Verantwortlich für die Qualität von Studium und Lehre und demnach auch für die Durchführung der Evaluation ist die jeweilige Fachbereichsleitung, die in dieser Aufgabe von einer Evaluationskommission mit Vertretern/innen aller Statusgruppen und dem/der aus ihrer Mitte gewähltem Evaluationsbeauftragten unterstützt wird. Allen Akteuren/innen steht das ZHQ bei der Durchführung der Evaluationsverfahren zur Seite. Entsprechend der Evaluationsordnung ist es „für die wissenschaftliche Beratung und Begleitung bei der Konzeption, Durchführung und Umsetzung der Evaluationsaktivitäten verantwortlich [...] und unterstützt] bei der Bewertung der Evaluationsergebnisse bzw. bei der Umsetzung dieser Ergebnisse in Maßnahmen zur Qualitätsverbesserung.“ (§10, Abs.2). Diese Maßnahmen sollen in den Ziel- und Leistungsvereinbarungen (ZLV) und der Mittelvergabe berücksichtigt werden. Im Sinne der Wirkungskonzeption ist der Erfolg ihrer Umsetzung – und weder die schlichte, regelmäßige Durchführung noch etwa das Erreichen bestimmter Kennzahlen – Gegenstand der ZLV-Gespräche mit dem Rektorat (§8, Abs.2) und einer Zwischenevaluation nach vier Semestern (§3, Abs.6).

Neben der Selbstevaluation von Fachbereich und Studiengängen im Vierjahresrhythmus (mit Befragungen der Studierenden, Lehrenden und Absolventen) gehört das Instrument der Lehrveranstaltungsevaluation zu den Kernelementen der Qualitätsentwicklung.

In ihrem Zentrum steht entsprechend dem Verständnis einer formativen Evaluation die „Optimierung des Lehr- und Lernprozesses innerhalb der einzelnen Lehrveranstaltung“ (§5 Abs.1). Hierzu erhalten dem Ansatz der Selbstevaluation folgend ausschließlich die Lehrenden selbst die sie betreffenden kompletten (bzgl. der urteilenden Studierenden anonymen) Evaluationsergebnisse einschließlich der Kommentare. Um ein ‚Diskursmodell‘ als Standard zu gewährleisten, „haben [Lehrende] die Ergebnisse mit den Studierenden der entsprechenden Lehrveranstaltungen zu diskutieren“ (§5, Abs.5). Gleichzeitig wird darüber hinaus geregelt, dass die Evaluationskommission z.B. im Rahmen einer flächendeckenden Lehrveranstaltungsevaluation eine Zusammenfassung der Ergebnisse aller Lehrveranstaltungsbefragungen des Fachbereichs erhält, damit sie ihren Aufgaben nachkommen kann (§7, Abs.2). Hierbei sind auch personenbezogene Rückschlüsse zulässig.

Kontinuierlich evaluieren müssen nur Neuberufene in den ersten vier Semestern insbesondere während eines einjährigen, an nordrhein-westfälischen Fachhochschulen vorgesehenen Verfahrens zur Feststellung ihrer pädagogischen Eignung (§5, Abs.3). Während die 2004 geschriebene Evaluationsordnung für den Regelfall eine Lehrveranstaltungsbefragung alle acht Semester verlangt (§3, Abs. 6), einigten sich Hochschul- und Fachbereichsleitungen bereits 2005 auf eine Evaluation alle vier Semester (vgl. FH Aachen 2005). In allen ZLV seit 2007 wiederum wurde eine Evaluation je Studienjahr vereinbart. Letztlich haben die Fachbereiche seit 2008 im Schnitt in acht von zehn Semestern und in insgesamt über 6.000 Lehrveranstaltungen mit insgesamt über 150.000 ausgefüllten Fragebögen eine Evaluation durchgeführt.

Auch in anderen Feldern hat es infolge der Evaluationsordnung ergänzende Vereinbarungen gegeben. So wurde etwa in einem 2005 zwischen Rektorat und Fachbereichen vereinbarten ‚Konzept zur Verbesserung des Qualitätsmanagements in Studium und Lehre an der Fachhochschule Aachen‘ (FH Aachen 2005) festgehalten, dass Lehrende in Folge mehrfach kritisch evaluierter Veranstaltungen zur Teilnahme an hochschuldidaktischen Beratungen und Weiterbildungen verpflichtet werden können. Während für Neuberufene die Teilnahme an entsprechenden Workshops bis 2005 dringend empfohlen wurde, wird auch mit ihnen eine Teilnahme im Umfang von sechs Tagen seither im Rahmen der Berufungsgespräche verbindlich vereinbart. Im Gegenzug erhalten sie eine Deputatsreduktion über sechs Semesterwochenstunden und erhöhte Erstausrüstungsmittel. Für alle Lehrenden der FH Aachen gilt seit dem Aufbau des landesweiten hochschuldidaktischen Netzwerks hdn-nrw generell, dass sie sich auch auf Eigeninitiative hin an die lokale Hochschuldidaktik (ZHQ) wenden und eine Beratung oder Lehrhospitation in Anspruch nehmen oder an einer für Lehrende unentgeltlichen Weiterbildung des hdn-nrw teilnehmen können.

⁵ Die Evaluationsordnung ist zu finden unter http://www.fh-aachen.de/fileadmin/org/org_zqe/ZQE_Evaluation/2004_02_EvaOrd_Teil_A.pdf, Zugriff 07.07.2014.

2.2 Hochschuldidaktisch gestützte Fragebogenentwicklung und -revision

Eine Lehrveranstaltungsevaluation nach dem ‚Diskursmodell‘ geschieht nicht durch die Umfrage, wohl aber mit Hilfe eines Feedbackinstruments: dem Lehrveranstaltungsfragebogen. Seine Ergebnisse fließen als Anlass und Ausgangspunkt zur Diskussion in den Evaluationsprozess ein.

Mit Bezug auf den Dienstleistungscharakter von Lehrveranstaltungen verweist Kromey (2003, S. 16) darauf, dass es nicht möglich ist, die Qualität von Lehre kontextunabhängig verbindlich zu definieren. *„Wo dennoch der Versuch unternommen wird, Merkmale „guter Lehre“ aufzulisten, setzt dieser – unabhängig vom Lehr-Inhalt – an der didaktischen Oberfläche an (Webler 1991, S. 246); und selbst da fällt es schwer, Einigkeit über einen Kriterienkatalog für „gute Didaktik“ zu erzielen.“* (Kromey 2003, S. 16).

Damit die Ergebnisse eines Lehrveranstaltungsfeedbacks auch konstruktiv aufgegriffen werden, müssen die erhobenen Items und die ihnen zugrundeliegenden Kriterien ‚guter Didaktik‘ oder ‚guter Lehre‘ in erster Linie von den betroffenen ‚Lernpartnern‘ selber, insbesondere den Lehrenden als relevant erachtet werden. In einer Umfrage (n=102) unter Dozenten der Universität Paderborn (Fendler 2004) waren 80% der Befragten motiviert, ihre eigenen Lehrveranstaltungen zu verbessern. Die von ihnen genannten Optimierungsziele (n=161) bezogen sich weitgehend auf die Aktivierung der Studierenden und die Interaktion in der Veranstaltung, auf die Rhetorik bzw. das Auftreten des Dozenten sowie auf die Strukturierung der Inhalte und ihre Vermittlung unter Einsatz verschiedener Methoden und Neuer Medien (vgl. ebd.). Das Interesse der Lehrenden konzentriert sich demnach auf die Aspekte, die durch ihr eigenes Handeln direkt verändert und als wesentliche Aspekte der Lehrkompetenz verstanden werden können. Das sind zugleich auch genau die Bereiche, in denen verschiedene Studien zur Wirksamkeit der Lehrveranstaltungsevaluation substantielle Verbesserungen feststellen konnten, wenn das Lehrveranstaltungsfeedback mit persönlicher Beratung verbunden (Rindermann 2003, S. 241) und Kompetenzentwicklung angeregt wurde.

Ein Fragebogen, in dessen Konstruktion entsprechende Aspekte berücksichtigt werden, würde das Interesse der Lehrenden wecken, aber ohne offene Felder für handschriftliche Kommentare auf kurz oder lang an Bedeutung verlieren. Es käme wenig Neues. Für offene Felder bieten sich positive-, negative- und unspezifische Anmerkungen sowie insbesondere Verbesserungsvorschläge an, die von Lehrenden als besonders wertvoll erachtet werden und eher zur Auseinandersetzung mit den Ergebnissen anregen (vgl. Schmidt/Loßnitzer 2010).

Ausgehend von den bisherigen Annahmen, sollte ein Lehrveranstaltungsfragebogen auf dem Verständnis einer ‚Lernpartnerschaft von Lehrenden und Studierenden‘ fußen und demnach unter den Prozessvariablen neben den genannten Aspekten des Lehrendenhandelns und der Interaktion auch solche des Studierendenverhaltens erfassen. Darüber hinaus sollte er Items zu den Voraussetzungen (z.B. Rahmenbedingungen, Vorinteresse der Teilnehmer/innen) und den Ergebnissen der Ver-

anstaltungen beinhalten. Um diese Aspekte ergänzt, sollte er ausreichend Ansatzpunkte für die Diskussion zwischen Lehrenden und Studierenden, zur Beschreibung und Bewertung der Veranstaltung und zur Identifikation von Zusammenhängen und Verbesserungspotentialen bieten.

Das Itemgenerierungsverfahren im Jahre 2004 entsprach dem von Rindermann skizzierten Beispiel eines Mischverfahrens. Unter der Regie des in der Evaluation von Lehrveranstaltungen erfahrenen Leiters der lokalen Hochschuldidaktik wurden verschiedene von Lehrenden der FH Aachen entwickelte Fragebögen sowie Items aus insgesamt über 20 Lehrveranstaltungsfragebögen aus dem deutschsprachigen Raum (u.a. Sturm/Csanyi 1991; Koch 2004) herangezogen und mit bereits validierten Fragebögen und ihren Dimensionen guter Lehre abgeglichen (u.a. Rindermann 2009 in seiner ersten Fassung). Der fertige Entwurf eines vierseitigen Fragebogens wurde letztlich der Senatskommission für Studium und Lehre (K1) zur Diskussion und Verabschiedung vorgelegt. Im Ergebnis ist ein typischer Fragebogen entsprechend des ‚State of the Art‘ bzw. der Definition von Schmidt und Loßnitzer (2010) entstanden.

Der Fragebogen wurde noch 2004 eingesetzt. In den Jahren darauf wurde er um fachbereichsspezifische Fragen ergänzt und intensiv genutzt. Je häufiger der flächendeckende Einsatz in den Fachbereichen (siehe 2.1), desto größer der Wunsch nach einem Standardfragebogen, der möglichst kurz ist, möglichst viele Anregungen zur Qualitätsverbesserung und zugleich gute Indikatoren zu wesentlichen Aspekten der Lehrveranstaltung für die Qualitätssicherung auf Fachbereichsebene liefert.

Im Jahre 2007 ergab eine Umfrage in den Fachbereichen zwei wesentliche und nicht unerwartete Kritikpunkte bezüglich der Lehrveranstaltungsevaluation: Der Fragebogen sei mit vier Seiten und 42 Fragen sowie mittlerweile zum Teil zahlreichen fachbereichsspezifischen Ergänzungen – bei immer häufigeren Befragungen – zu umfangreich und die Folgenhaftigkeit des Verfahrens insgesamt zu gering.

Wesentlich für die Akzeptanz studentischer Rückmeldungen ist erfahrungsgemäß ein angemessener Umfang, der bei Standardverfahren 25 Items nicht überschreiten sollte und unumgänglich einen *„Kompromiss zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz“* bedeutet (Souvignier/Gold 2002, S. 271). Ein kurzer und weniger differenzierter Fragebogen ist zum Zwecke einer Lehrveranstaltungsevaluation nach dem ‚Diskursmodell‘ aber völlig ausreichend, soweit Lehrende ihre Ergebnisse im Diskurs mit den Studierenden reflektieren und weiter konkretisieren (vgl. Kromey 2003, S. 14).

Damit Lehrende unmittelbar Schlussfolgerungen zur Verbesserung ihrer Lehrtätigkeit ziehen können, sollte der Fragebogen vor allem Aspekte des Lehrendenverhaltens erfassen sowie ausreichend Platz für freie Kommentare zur Qualitätsverbesserung geben. Der Bogen sollte folglich auf maximal zwei Seiten gekürzt und hierzu insbesondere auf nur wenige Fragen zum Studierendenverhalten und zu den Rahmenbedingungen reduziert werden. Eine aufwendige Analyse des Fragebogens fand noch 2007 im Rahmen einer Diplomarbeit statt, die von

einem der Autoren betreut wurde. Mit dem Ziel der Itemreduktion wurde der Fragebogen anhand einer Stichprobe von 81 evaluierten Lehrveranstaltungen (n=3412) auf inhaltliche Redundanzen untersucht. Das Ergebnis ist ein multidimensionaler Kurzfragebogen mit drei offenen Fragen und 26 geschlossenen Items⁶, von denen 14 zuverlässig über alle Fachbereiche hinweg fünf akzeptable und inhaltlich eigenständige Dimensionen von Lehrveranstaltungsqualität⁷ erfassen:

1. Soziale Interaktion & Motivation ($\alpha=0,804$ bei vier Items)
2. Didaktischer Aufbau & Methodik ($\alpha=0,852$ bei drei Items)
3. Angebote für Selbststudium & Prüfungsvorbereitung ($\alpha=0,688$ bei drei Items)
4. Verdeutlichung der Relevanz für Studium & Beruf ($\alpha=0,682$ bei zwei Items)
5. Studentisches Vorinteresse & Lernengagement ($\alpha=0,509$ bei zwei Items)

Die ersten vier Indikatoren beziehen sich auf die Beiträge des Lehrenden zur Prozessqualität der Lehrveranstaltung. Sie ähneln erwartungsgemäß den Dimensionen anderer Instrumente der Lehrveranstaltungsevaluation (vgl. u.a. Gollwitzer/Schlottz 2003, zusammenfassend Schmidt/Lößnitzer 2010).

Der Indikator ‚Studentisches Vorinteresse & Lernengagement‘ soll als Kontrollvariable herangezogen werden können. Er setzt sich aus zwei Studierendenbezogenen Items zu Vorinteresse und eigenem Lernverhalten in der Veranstaltung zusammen. Unter den Voraussetzungen der Lehrveranstaltungsteilnehmer (z.B. schulische Leistungen und Schwerpunkte) ist insbesondere das durchschnittlich mitgebrachte Vorinteresse entscheidend für eine positive Bewertung der Veranstaltung insgesamt (vgl. zusammenfassend Bargel/Hage 2000, Rindermann 2009 und Wolbring 2013). Unter den studentischen Tätigkeiten ist wiederum das eigene Lernverhalten während der Veranstaltung für ihren Erfolg besonders bedeutend (vgl. Rindermann 2009, S. 72 ff.).

In Tabelle 1 sind die Indikatoren, ihre Items und psychometrische Kennwerte aufgeführt.

Diese Indikatoren bilden gemeinsam mit den Ergebnisvariablen ‚Gesamtbeurteilung‘ und ‚Lernerfolg‘ die Globalwerte, die für alle evaluierten Lehrveranstaltungen den Evaluationsbeauftragten des jeweiligen Fachbereichs zu Verfügung gestellt werden, um kritische Lehrveranstaltungen und erste Erklärungsansätze hierzu im Rahmen eines Monitorings zu identifizieren.

2.3 Praxiserfahrungen aus der evaluationsbezogenen Beratung zum Fragebogen

Bei aller inzwischen erreichten grundsätzlichen Akzeptanz des nunmehr langjährig eingesetzten Fragebogens werden in Einzelmeldungen oder in hochschuldidaktischen Beratungsgesprächen vor allem folgende einzelne Items/Aspekte immer wieder, wie nachstehend dargestellt, kritisch kommentiert bzw. gehandhabt:

Prüfungstransparenz: „Kann man jetzt nicht erfragen/ beantworten.“

Lehrende lehnen diese Frage häufig ab, da die aktuelle Prüfung zum Erhebungszeitpunkt noch nicht konzipiert

oder zumindest den Studierenden noch nicht bekannt sei. Die ja wirklich in nahezu jeder Studienhandlung all-täglich greifende Prüfungsorientierung der Studierenden wird durch die Lehrenden nicht zur konkret inhaltlichen Lernsteuerung schon während der Vorlesungsperiode genutzt. Die hochschuldidaktische Leitlinie ‚angstfreies und gezieltes Studieren schon während der Vorlesungsperiode durch rechtzeitige Prüfungstransparenz‘ wird nicht unterstützt bzw. nicht realisiert.

Prüfungsformulierung und Lernen für die Prüfung finden dementsprechend erst gegen Lehrveranstaltungsende bzw. in der vorlesungsfreien Periode statt. Eine geheim gehaltene Prüfungsformulierung wird von einigen Lehrenden als (einziges) Macht- und Druckmittel verstanden, um Studierende zum beliebig umfassenden, nur ja nicht exemplarischen, Lernen zu bringen, das dann eher zum Pauken bzw. „Bulleimie-Lernen“ entartet. Prüfung wird weniger als Honorierung von Lernerfolgen und inhaltsbezogenem Kompetenznachweis gesehen, sondern vor allem als Selektionsakt und hierzu zwingend mit Überraschungscharakter.

Ein Vorschlag wäre, in der Lehrveranstaltungsevaluation nur zu erfassen, inwieweit Prüfungsinhalte und -formen angekündigt und in Verbindung zum behandelten Stoff gesetzt werden. Auf diese Weise würde es zumindest thematisiert und kann ggf. in Einzelberatungen aufgegriffen werden. Die Verläufe und Ergebnisse der Prüfungen wären ohnehin in eine über den Hörsaal hinausführende formative Modul- bzw. Curriculumsevaluation mit einzubeziehen.

Lernerfolg: „Den kann man doch nicht mitten im Semester erfragen/ beantworten.“

Der Lernerfolg stellt allerdings den zentralen Bewertungsmaßstab für Lehr- bzw. Lernveranstaltungen dar, der viel ausführlicher erfasst und nach Kompetenzbereichen differenziert (z.B. BEvaKomp, mehr bei Braun 2007) erfragt werden könnte oder müsste, was im Rahmen kurzer Fragebögen mit den oben aufgeführten Zielen nur begrenzt möglich ist.

Der Lernerfolg ist regelmäßig das am schlechtesten bewertete Item. Ausführliche Detailanalysen machen deutlich, dass es weniger eine studentische Kritik an den Lehrenden, sondern vor allem als realitätsgerechte stu-

⁶ Die geschlossenen Items werden mit einer Fünferskala von 1 „trifft völlig zu“ bis 5 „trifft gar nicht zu“ angeboten. Ausnahmen finden sich in den beiden Einzelitems zum Veranstaltungserfolg (Lernerfolg von 1 „sehr hoch“ bis 5 „sehr niedrig“; Gesamturteil von 1 „sehr gut“ bis 5 „mangelhaft“), in den Einzelitems zum Vorgehen im Stoff (von 1 „sehr schnell“ bis 5 „sehr langsam“), zum fachlichen Niveau (von 1 „sehr hoch“ bis 5 „sehr niedrig“), zum Anteil der Übungen (von 1 „sehr umfangreich“ bis 5 „sehr gering“) und in verschiedenen Items zur studentischen Motivation und Mitarbeit (Teilnahmegewohnheit in Prozent und wöchentlicher Zeitaufwand in Stunden).

⁷ Anhand der Daten aus fünf Jahren Lehrveranstaltungsevaluation mit dem gekürzten Fragebogen (n=114.782) wurde 2013 eine explorative Faktorenanalyse mit durchweg erstaunlichen Maßen für die Stichprobeneignung durchgeführt (Hauptachsenanalyse mit Equamax-Rotation und 11 Iterationen). Unter Einbezug ausschließlich Lehrendenbezogener Items legt die Faktorenlösung vergleichbar mit anderen Kurzskalen zur Lehrveranstaltungsevaluation zunächst Eindimensionalität und damit einen generellen Faktor nahe (vgl. Zumbach et al. 2007, STUD-FEL in Tillmann et al. 2011). Unter Einbezug der studentischen Variablen und unter Berücksichtigung theoretischer Ansätze lässt sich allerdings eine inhaltlich erwogene Fünf-Faktoren-Lösung mit einer Varianzaufklärung von 70% ohne Unterschiede in den fachbereichsspezifischen Stichproben replizieren. Das Ergebnis sind fünf eigenständige Dimensionen, die als inhaltlich selbstständige Indikatoren für die Lehrveranstaltungsqualität genutzt werden können.

Tabelle 1: Indikatoren für eine gute Lehrveranstaltung, ihre Items und psychometrischen Kennwerte

Indikatoren (α)	Aspekte einer Lehrveranstaltung	Item im Fragebogen	Rotierte Faktorenmatrix (Faktorladungen ab ,300)				
			1	2	3	4	5
Soziale Interaktion & Motivation ($\alpha=,804$)	Freundlichkeit	Der/die Lehrende ist im Umgang mit uns Studierenden freundlich und aufgeschlossen:	,767				
	Kritikfähigkeit	Der/die Lehrende greift studentische Beiträge und Kritik angemessen auf:	,667				
	Interesse am Lernerfolg	Der/die Lehrende zeigt Interesse am Lernerfolg der Studierenden:	,579	,307			
	Schaffung einer Lernatmosphäre	Der/die Lehrende schafft eine zur Mitarbeit anregende Lernatmosphäre:	,517	,406			
Didaktischer Aufbau & Methodik ($\alpha=,852$)	Rhetorik	Diese LV ist für mich rhetorisch gut gestaltet (Vortragsstil, Beispiele usw.):		,743			
	logische Struktur	Diese LV ist für mich klar strukturiert (Roter Faden, Beziehungen zwischen Einzelfakten usw.):		,571			
	Medieneinsatz	Der Einsatz von Präsentationsmedien bietet in dieser LV eine gute Lernhilfe:		,497			
Angebote für Selbststudium & Prüfungsvorbereitung ($\alpha=,688$)	Material zum Selbststudium	Die Materialien bzw. Hinweise zum Selbststudium erachte ich als ausreichend und lernfördernd:			,615		
	Begleitende Übungen	Die Übungen sind gut auf die LV abgestimmt:			,551		
	Prüfungstransparenz	Der/die Lehrende macht Prüfungs- und Leistungsanforderungen rechtzeitig transparent:			,525		
Verdeutlichung der Relevanz für Studium & Beruf ($\alpha=,682$)	Relevanz für Ausbildungsziel	Die Relevanz der LV-Inhalte für das Ausbildungsziel wird gut verdeutlicht:				,665	
	Beziehungen zu anderen Fächern	Beziehungen zu anderen Fächern werden gut gestaltet:				,537	
Studentisches Vorinteresse & Lernengagement ($\alpha=,509$)	Lernverhalten	Mein eigenes Lernverhalten in der LV beurteile ich als (Mitarbeiten, Mitdenken, Mitschreiben usw.):					,635
	Vorinteresse	Mein Interesse am Thema war vor Beginn dieser LV:					,509

dentische Selbstkritik verstanden werden muss. Nicht nur, aber auch um diese studentische Selbstkritik zu unterstützen, wäre ein Vorschlag der Autoren, die Frage nach dem Kompetenzerwerb unter der Überschrift ‚Lernergebnisse & -prozesse‘ in Form offener Kommentarfelder an das Ende des Fragebogens zu stellen, z.B. wie folgt:

- „Was ich in dieser Veranstaltung bisher insbesondere gelernt bzw. verstanden habe und in Zukunft kann bzw. weiß ist, ...“
- „Gelernt habe ich dies vor allem dadurch, dass ...“
- „Folgende Kenntnisse/Fähigkeiten/Kompetenzen würde ich gerne noch lernen bzw. vertiefen oder hätte ich gerne noch gelernt bzw. vertieft...“

Solche freien Felder orientieren beide Seiten auf das Wesentliche, nämlich den Lernerfolg und sie geben reichlich Anlass zur individuellen oder gemeinsamen Reflexion.

Workload (durchschnittliche wöchentliche Vor- und Nachbereitungszeit):

Entsprechende Angaben werden seitens der Lehrenden nicht geglaubt bzw. haarklein nachgerechnet und als Bologna-Untererfüllung eines Vollzeitstudiums nachgewiesen und von Studierenden z.T. spontan oder strategisch hochgepuscht. Studierende, Lehrende und hochschuldidaktische Berater sollten auch mit Blick auf den vorigen Aspekt die Frage in den Mittelpunkt stellen: Wieviel Zeit investieren Studierende für welche Art von studienbezo-

gener Tätigkeit, um sich mit den vorgesehen Inhalten auseinanderzusetzen? Wieviel Zeit und welche Tätigkeiten sind kompetenzfördernd gut investiert?

Medien (nicht nur Präsentationsmedien, sondern solche zum Selbststudium):

Hin und wieder herrscht die Verwechslung von Teilnahme/Lernen mit Anwesenheit (Zuhören, wach bleiben und Material sammeln, lochen, abheften). Die hochschuldidaktische Option medienunterstützter Vor- und Nachbereitung der Präsenzhäsen und die Möglichkeit, hierdurch Lerninhalte auszulagern, wird selten realisiert. Dies vor allem weil Erstellungsaufwand und Nutzungsertrag kaum in angemessenem Verhältnis zueinander zu stehen scheinen, auch in Bezug auf e-learning-Angebote. Insbesondere neuberufene Professoren und Professorinnen könnten hier – ggf. mit entsprechenden Angeboten innerhalb und außerhalb der Hochschule – eine Wende in Gang setzen.

Offene Kommentare:

Sie werden von Lehrenden nach wie vor ganz häufig als „die wichtigste Rückmeldung“ bezeichnet. In der hochschuldidaktischen Beratung erlebt man bisweilen noch die selbstschützende und entlastende selektive Wahrnehmung lediglich der erwünschten Kommentare. In der Regel werden aber alle, fast durchweg konstruktiven Kommentare verständigungsorientiert im Zusammenhang mit den quantitativen Items aufgegriffen. Die zum Glück wenigen, z.T. absolut unverantwortlichen und ‚politisch wirklich nicht korrekten‘ Kommentare werden zu-

nehmend in den Evaluationskommissionen thematisiert und in den Lehrveranstaltungen zum Zwecke einer gemeinsamen ‚Kulturverbesserung‘ angesprochen.

2.4 Praxiserfahrungen aus der evaluationsbezogenen Beratung zum Verfahren

Die ausschließlich durch den fachbereichsexternen Leiter der lokalen Hochschuldidaktik realisierten Follow-Up-Beratungen wurden zum einen gelegentlich als evaluationsverursachte Zwangsberatungen empfunden. Sie stoßen zum anderen entsprechend nunmehr langjähriger Erfahrungen an kapazitative und strukturelle Grenzen (Analysen und Fallbeispiele in Heger 2009). Denn insgesamt entsprechen nur wenige Fälle der Kombination: hochmotivierte/r Lehrende/r, der/die einerseits individuelle Lehrprobleme hat, andererseits aber hochschuldidaktische Beratung von sich aus hartnäckig immer wieder nachfragt und kontinuierlich aufgreift. Um in der Mehrzahl der hochschuldidaktisch nicht so hartnäckig intrinsisch motivierten Problemfälle Verbesserungen auf lange Sicht zu erreichen, müssten die hochschuldidaktischen Berater/innen mehr als deren Kapazität erlaubt ihrerseits immer wieder nachsetzen und dran bleiben. Ansonsten bleiben die Beratungsergebnisse doch irgendwie unverbindlich und selbst sehr intensive und anfänglich auch deutlich wirksame Beratungen verpuffen auf Dauer wieder, zum Teil sogar recht schnell.

Selbst bei weitgehend individuell begründeten Lehrveranstaltungsproblemen ist demnach bzgl. Kapazität und Verbindlichkeit eine Entwicklung sehr zu begrüßen, nach der evaluationsbezogene Beratungsgespräche jetzt vielfach von den Fachbereichen moderiert und (mit-)gestaltet werden. Die Fachbereiche zeigen ihren Lehrenden damit, dass sie es inzwischen ernst meinen mit einem in sich geschlossenen Evaluationskreislauf, dass sie sich auch selbst verantwortlich fühlen, die Prozesse nicht mehr komplett nach außen delegieren und sich auch zutrauen, solche sensiblen Prozesse konstruktiv zu gestalten.

Dann trägt vor allem bei strukturell begründeten Lehrveranstaltungsproblemen das seitens der Fachbereichsvertreter/innen vorhandene Hintergrundwissen zu Inhalten, Formen und zeitlichen Positionierungen der jeweiligen Module im Curriculum ganz wesentlich zur Wirksamkeit bei, sowohl in der Analyse der problemverursachenden Zusammenhänge, als auch in der problemlösenden weiteren Handhabung innerhalb der Fachbereichsgremien.

Verschiedene Personen und Fachbereiche setzen diese Aufgabe, manchmal auch individuell schwankend, ganz unterschiedlich um. Das reicht von enger Zusammenarbeit mit dem/der Dekan/in bis hin zu wirklich weitgehender, vertrauensvoller Entlastung des/der Dekans/in. Es gibt Fachbereiche in denen aktuell (noch bzw. je nach spezieller Entwicklung immer mal wieder) die Aufwandsminimierung bzgl. Evaluation im Vordergrund steht: nur das absolut Unumgängliche tun. Selbst ausgehend von solchen Grundpositionen haben Fachbereiche aber im Kontext der hochschulweit vorgegebenen Strukturen und Beratungs- bzw. Unterstützungsangebote inzwischen die in Abschnitt 2.5 erläuterten Verfahrensstandards weitgehend entwickelt.

Die Kooperation der Fachbereiche mit dem ZHQ existiert, z.T. auch zeitlich wechselnd, in Varianten von weitgehendem Auslagerungsversuch an das ZHQ über mehr oder weniger intensive Kooperation bis hin zu fast völliger, durchaus sehr erfolg- und folgenreicher Eigenständigkeit in Evaluation bzw. hochschuldidaktischer Qualitätsentwicklung, z.T. mit Nutzung bundesweiter Kooperationen und Beratungszusammenhänge.

Engagement und Einbezug der Studierenden in den Evaluationskommissionen fallen auch unterschiedlich aus. Werden hier die Prozesse und Ergebnisse der Lehrveranstaltungsevaluation ausreichend transparent gehandhabt, ergibt sich z.T. eine starke studentische Identifikation dieser Kommissionsmitglieder, eine engagierte Mitarbeit im Gremium und im Idealfall externes ‚Marketing‘ unter Studierenden für die hier entwickelten Evaluationsmaßnahmen.

Eine weitere erwähnenswerte Ambivalenz liegt in der Verwendung der Evaluationsergebnisse, die z.B. auch der Evaluationskommission vorliegen und von ihr ggf. für die Vergabe von Lehrprämien herangezogen werden. Instrumente der Lehrveranstaltungsevaluation messen gerade nicht (nur) die Lehrleistung der Dozenten. Sie sind hingegen multifaktoriell und unterliegen selbst unter Berücksichtigung aller Faktoren noch etwaigen Messfehlern. Im Einverständnis mit einer Mehrheit von Anbietern und Nutzern der Lehrveranstaltungsevaluation (vgl. Schmidt 2008, S. 28) muss daher aus Sicht der Autoren eine kritische Position zur Vergabe von Lehrprämien oder Besoldungszulagen in Abhängigkeit von Lehrleistungsbewertung eingenommen werden, soweit sich diese vorrangig auf das Lehrveranstaltungsfeedback stützen soll. Damit Follow-Up-Maßnahmen an der FH Aachen allerdings nicht ausschließlich kritisch evaluierte Veranstaltungen betreffen, sondern auch herausragende Veranstaltungen honoriert werden, vertreten die Autoren – insbesondere auch gegenüber teils wesentlich weniger kritischen Lehrenden und Studierenden – die Auffassung, auf anderen Wegen ermittelte Vorschläge anhand der Globalwerte der Lehrveranstaltungsbefragung zu prüfen. Globalwerte sollten korrigierend nur dann zum Tragen kommen, wenn sie aufgrund unterdurchschnittlicher Bewertungen der Vergabe einer Lehrprämie widersprechen.

2.5 Denkbare Standards für Strukturen und Verfahren

Insgesamt haben sich einige Verfahren und Strukturen als bewährt herausgestellt. Zunächst hat sich ein anerkanntes Verfahren für das Follow-Up bei kritischen Lehrveranstaltungen etabliert. Zur Definition kritischer Veranstaltungen – sogenannter ‚Problemmodule‘ – haben sich die Fachbereiche einen Vorschlag der lokalen Hochschuldidaktik zu eigen gemacht, der in vergleichbarer Form auch an anderen Hochschulen Einzug gehalten hat (vgl. z.B. Craanen 2011). Demnach gelten Lehrveranstaltungen als unauffällig, wenn Globalindikator und Gesamtnote besser als 2,5 liegen. Liegen einer der beiden Werte oder gar beide Werte bei mindestens 3, so wird die Veranstaltung der Problemstufe 1 oder 2 zugeordnet. Die Evaluationskommissionen kontaktieren in erster Linie Lehrende dieser beiden Problemstufen. Nur wenn Lehrveranstaltungen wiederholt ein Ergebnis zwischen 2,5 und 3,0 aufweisen, werden in einigen Fachbe-

reichen auch diese Lehrenden kontaktiert (vgl. Heger 2009). Nicht mehr wie in der Evaluationsordnung eigentlich vorgesehen in allen Fällen, sondern nur in diesen kritischen Fällen wird zunächst um eine meist schriftliche Stellungnahme bzgl. der jeweiligen Veranstaltung gebeten. Sieht die Kommission in Folge der Stellungnahme noch Klärungs- und Handlungsbedarf, wird ein Gespräch initiiert, an dem im Idealfall neben einem Vertreter der lokalen Hochschuldidaktik auch ein professorales Mitglied der Kommission (in der Regel Evaluationsbeauftragte und/oder Dekane) als Fachvertreter/in teilnimmt. Gemeinsam werden die Ergebnisse aus Befragung und Diskussion interpretiert und Ursachen für das schlechte Feedback sowie Verbesserungsmaßnahmen eruiert.

Im Rahmen der aktuell geplanten Überarbeitung der Evaluationsordnung Anfang 2014 hat es eine intensive Diskussion über die Arbeit der Evaluationskommissionen und eine erfolgreiche Evaluationskultur in den Fachbereichen gegeben. Aus einem Experteninterview mit langjährigen Evaluationsbeauftragten zweier Fachbereiche und einer anschließenden Diskussion der Ergebnisse unter anderen mit Mitgliedern der Hochschulleitung, der Studierendenvertretung und der lokalen Hochschuldidaktik haben sich wesentliche, in einigen Fachbereichen bereits realisierte Empfehlungen und Standards zu Aufbau und Arbeit der Evaluationskommissionen herauskristallisiert, die z.T. erst in einer zukünftigen Evaluationsordnung berücksichtigt werden sollen:

- **Eigenständige Kommission aus Mitgliedern aller Statusgruppen unter Moderation eines professoralen Evaluationsbeauftragten:**

Die Evaluationskommission ist – in Abgrenzung zu einigen Versuchen, diesem Aufgabenbereich als Tagesordnungspunkt einer Dekanatssitzung ausreichend nachzukommen – ein eigenständiges, themenspezifisch funktionales Gremium für das operative Geschäft der Evaluation von Studium und Lehre und besteht entsprechend der Evaluationsordnung aus Mitgliedern aller Statusgruppen. Diese Kommissionsmitglieder werden von ihrer jeweiligen Statusgruppe aus dem Fachbereichsrat gewählt. Aus ihrer Mitte wiederum wird der/die Evaluationsbeauftragte gewählt, der/die den Vorsitz der Kommission übernimmt und Professorenenstatus haben muss.

- **Regelmäßig durchgeführte Kommissionssitzungen zur Evaluationsplanung und -durchführung sowie zur Ergebnisdiskussion:**

Zur Planung und Diskussion aller Evaluationsbemühungen und zur kontinuierlichen Nachverfolgung der Evaluationsergebnisse finden regelmäßige Sitzungen (min. 1x im Semester) statt, die in einer abschließenden in der Evaluationskommission öffentlich diskutierten Bewertung münden (gegebenenfalls inklusive Maßnahmenplanung) und hierzu mindestens einen schriftlichen Austausch mit den Lehrenden der als kritisch eingestuften Veranstaltungen voraussetzen. Als Grundlage finden kontinuierliche Lehrveranstaltungsbefragungen statt, die in der Regel flächendeckend sind, mindestens aber die Veranstaltungen von Neuberufenen sowie zuletzt kritisch evaluierte Lehrveranstaltungen betreffen.

- **Kooperation mit der zentralen Hochschuldidaktik durch gemeinsame Sitzungen und/oder Beratungsgespräche mit Lehrenden:**

Es findet eine enge Kooperation mit dem ZHQ statt, etwa über die Möglichkeit Vertreter des ZHQ als Gast beratend zu den Sitzungen einzuladen und Follow-Up-Gespräche mit Lehrenden kritisch evaluierter Veranstaltungen gemeinsam, also in Anwesenheit von Fachvertreter und Hochschuldidaktik durchzuführen.

Die Existenz einer Evaluationskommission entsprechend dieser Kriterien gewährleistet zunächst – so die Annahme – eine Bündelung von Erfahrungswissen aus Hochschul- und Fachdidaktik sowie der Evaluation von Studium und Lehre und führt zu Konsistenz und Dauerhaftigkeit in seiner Anwendung.

Die Wahl durch den Fachbereichsrat sowie der im Laufe einer regelmäßigen Arbeit wachsende Bekanntheitsgrad von Kommission, Kommissionsmitgliedern und ihren Aufgaben scheinen Verankerung und Akzeptanz der Evaluationsverfahren insgesamt zu steigern. So kann – wie es an der FH Aachen heute häufiger geschieht – ein solches Gremium zunehmend auch Adressat von Hinweisen werden, die nicht unmittelbar über die standardisierten Evaluationsverfahren erfasst werden, aber zur Identifizierung von kritischen Veranstaltungen, problematischem Lehrverhalten oder zur Interpretation bestimmter Ergebnisse relevant sind.

Werden die genannten Standards zu Aufbau und Arbeit der Evaluationskommission erfüllt und der Evaluationskreislauf im Sinne einer kontinuierlichen Qualitätsverbesserung geschlossen, ist zu erwarten, dass sich die Ergebnisse der Lehrveranstaltungsevaluation steigern und Studierende diese Verbesserungen in Folge der Evaluation bewusst wahrnehmen, was wiederum die erwähnten Ermüdungserscheinungen verhindern dürfte.

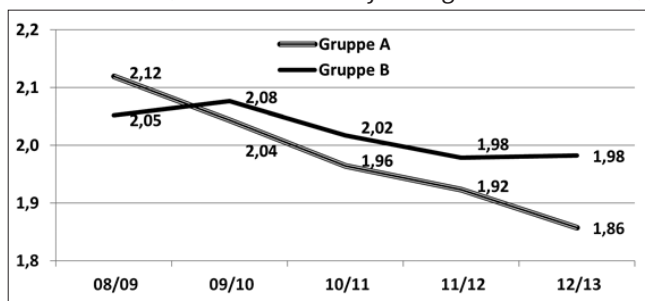
2.6 Empirische Ergebnisse zur Qualitätsentwicklung unter diesen Standards

Trotz derselben Evaluationsordnung und der gleichen standardisierten Instrumente und Verfahren ist generell anzunehmen, dass verschiedene Fachbereiche unterschiedliche ‚Evaluationskulturen‘ herausbilden (vgl. Böttcher/Grewe 2010). Auch an der FH Aachen kann von einer Entwicklung solcher fachspezifischer Evaluationskulturen mit fachspezifischen Evaluationsstrukturen und -verfahren zwischen 2004 und 2008 ausgegangen werden (vgl. zum damaligen Stand der Evaluationsverfahren der FH Aachen Heger 2009).

Nach weiteren sechs Jahren stellt sich nun die Frage, welche Evaluationsstrukturen und -verfahren etabliert sind und inwieweit diese Wirkungen nach sich ziehen. Die zuvor beschriebenen Standards sind in nahezu allen Fachbereichen der FH Aachen weitgehend umgesetzt, in immerhin schon einem Drittel der Fachbereiche vollständig. Untersucht wurde nun, ob sich die evaluierten Veranstaltungen der Fachbereiche, in denen die zuvor genannten Standards vollständig etabliert sind ($n_A=2244$) über die Jahre hinweg anders entwickelt haben als die Veranstaltungen der übrigen Fachbereiche ($n_B=2728$). Ein nennenswerter Unterschied zwischen diesen beiden Gruppen A und B liegt in der Verwendung

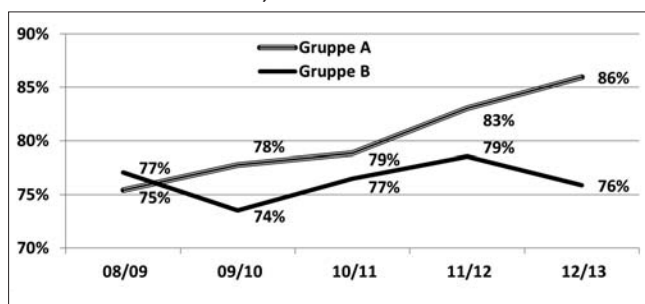
von Onlinebefragungen, die in Gruppe B ausgeprägter ist, was auch entsprechend eigener empirischer Analysen zu leicht schlechteren Ergebnissen in der Lehrveranstaltungsbewertung führt (vgl. auch Dresel/Tinsner 2008, Wolbring 2013). Dieser Einflussfaktor bleibt aber über den betrachteten Zeitraum hinweg konstant. Fasst man je Gruppe die Indikatoren 1 bis 4 aller Lehrveranstaltungen je Studienjahr zusammen und vergleicht die Entwicklung dieses ‚Globalindikators‘ über die Jahre hinweg, so zeigt sich eine stetige Verbesserung der Veranstaltungen aus Gruppe A und eine weniger starke Verbesserung und weniger kontinuierliche Entwicklung der Veranstaltungen aus Gruppe B.

Abbildung 1: Durchschnitt Globalindikator aller Veranstaltungen je Studienjahr und Gruppe (Durchschnitt der jeweiligen Fachbereiche)



Betrachtet man die Entwicklung des Anteils gänzlich unkritischer Lehrveranstaltungen über die Studienjahre hinweg, so zeigt sich parallel auch hier eine stetige Steigerung in Gruppe A um insgesamt 11 auf zuletzt 86 Prozent, während in Gruppe B der Anteil unkritischer Lehrveranstaltungen bei +/- 2 um die 77 Prozent stagniert.

Abbildung 2: Anteil unkritischer Lehrveranstaltungen (Globalindikator <2,5) je Studienjahr und Gruppe (Durchschnitt der jeweiligen Fachbereiche)



Ist die Ausgangslage im Studienjahr 2008/2009 nahezu die gleiche, so zeigen die Veranstaltungen der Gruppe A eine moderat bessere Entwicklung als die der Gruppe B, auch wenn erst die Betrachtung der Folgejahre hier abschließende Kausalattributionen möglich machen würde.

Neben der Betrachtung der Evaluationsergebnisse über einen Zeitraum von fünf Jahren haben wir eine Befragung unter den Absolventen und Absolventinnen durchgeführt, die in diesen fünf Jahren Studierende der FH Aachen gewesen sind⁸. Gefragt wurde, ob sie ausreichend Gelegenheit hatten, ihre Rückmeldung zu Lehr-

veranstaltungen im Rahmen von Lehrveranstaltungsbeurteilungen zu geben und ob Lehrende die Ergebnisse in der Folge mit den Studierenden besprochen haben. Hieran lässt sich vermuten, inwieweit das ‚Diskursmodell‘ als etabliert angesehen werden kann. Nicht zuletzt wurden sie gefragt, ob durch diese Lehrveranstaltungsevaluation aus ihrer Sicht Verbesserungen erreicht wurden. Denn auch wenn eine positive Entwicklung der Lehrveranstaltungen anhand der Ergebnisse aus Gruppe A festgestellt werden konnte, bedeutet das nicht, dass diese Verbesserungen auch als solche von Studierenden wahrgenommen wurden.

Absolventen/innen der Fachbereiche, die die genannten Standards vollständig erfüllen ($n_A = 173$), sind zunächst eher der Meinung, dass sie ausreichend Rückmeldung geben konnten (2,25 zu 2,39), geben darüber hinaus deutlich häufiger an, dass Lehrende die Ergebnisse mit ihnen besprochen haben (2,52 zu 3,03) und sind zuletzt etwas seltener als Absolventen/innen der übrigen Fachbereiche ($n_B = 409$) der Meinung, dass durch die Lehrveranstaltungsevaluation keine Verbesserungen erreicht wurden (3,31 zu 3,55).

Diese Ergebnisse legen insgesamt nahe, dass es in den als vorbildlich zu verstehenden Fachbereichen eher gelungen ist, den Evaluationskreislauf durch ein etabliertes Diskurs- und Beratungsmodell zu schließen und Verbesserungen in der Lehrveranstaltungsqualität, die als solche auch von Studierenden wahrgenommen werden, zu erreichen.

3. Schlussfolgerungen und Perspektiven

Bezugnehmend auf Ausgangspunkte, Ziele und Wirkungsplan wurden in diesem Beitrag die Realisierung und die Wirksamkeit der im Verlauf von inzwischen 16 Jahren entwickelten Lehrveranstaltungsevaluation der FH Aachen untersucht.

So wichtig methodisch fundierte Fragebögen sind, so sehr sind die Autoren dieses Beitrags der Auffassung, dass es für die Wirksamkeit der Lehrveranstaltungsevaluation vorrangiger ist, die Fachbereiche darin zu unterstützen, eigenständige und tragfähige Strukturen und Verfahren der Evaluation zu etablieren. Die Befragungsinstrumente müssen anerkannt und akzeptiert werden, wozu methodisch geprüfte Fragebögen ganz sicher beitragen. Aber ein nach allen Regeln der Messtechnik konstruierter und mehrfach validierter Fragebogen ist kein Selbstzweck. Erst das Zusammenwirken von Lehrveranstaltungsfragebogen und anderen vor allem qualitativen Instrumenten und Verfahren aus Evaluation und Qualitätsentwicklung kann offenbar dazu beitragen, Verbesserungspotentiale zu identifizieren und wirksame Maßnahmen abzuleiten, um dauerhaft und in der Fläche Verbesserungen in der Qualität von Studium und Lehre zu erreichen.

Dabei waren einzelne hochschulpolitische Entwicklungen der letzten Jahrzehnte hilfreich, so etwa die Pro-

⁸ Die Absolventenbefragung wurde im Februar 2014 mit einer Brutto-Rücklaufquote von 51 Prozent aller Absolventen des Prüfungsjahrgangs 2012 bzw. einer Netto-Rücklaufquote von 60 Prozent aller postalisch erreichten Absolventen abgeschlossen. Zu den genannten Fragen wurde eine Skala von 1 „trifft völlig zu“ bis 5 „trifft gar nicht zu“ angeboten. Dargestellt ist jeweils der arithmetische Mittelwert.

gramme zur Verbesserung der Lehre und – initiiert durch verschiedene Prozesse – auch die Einführung der Akkreditierung, die landesweite Etablierung des hochschuldidaktischen Weiterbildungsnetzwerks hdw-nrw, ebenso die ursprünglichen Ansätze der Bologna-Reform sowie die aktuelle Förderung durch den sogenannten ‚Qualitätspakt Lehre‘. Andere hochschulpolitische Entwicklungen und Vorgaben haben die Entfaltung bzw. das Wachstum einer formativ wirksamen Evaluation an der FH Aachen eher gefährdet, vor allem die überladenen und ausschließlich quantitativen Vorgaben einer falsch verstandenen Output- und Effizienzorientierung (z.B. Absolventenquoten in der Regelstudienzeit). Die Legitimation erfolgt dann allenfalls durch den Nachweis geforderter Prozesse der Evaluation und nicht durch deren Wirksamkeit. Bei solchen Entwicklungen sehen dann auch die Akkreditierungsagenturen die Lehrveranstaltungsevaluation nicht mehr als das Herzstück der innovationsorientierten Evaluation, sondern eher als ‚nice to have‘ aber insgesamt recht harm- und folgenlosen Standard. Eine Systemakkreditierung, so sie gewollt ist, kann und muss solchen Entwicklungen entgegen wirken.

An der FH Aachen sind aus Sicht der Autoren – unabhängig von der zukünftigen Akkreditierungsform – schon im Kontext der anstehenden Überarbeitung der Evaluationsordnung auch folgende nur stichwortartig benannten Weiterentwicklungen der Lehrveranstaltungsevaluation geboten:

- Stärkung der Konzepte der ‚Kompetenzorientierung‘ und ‚Lernpartnerschaft‘ einschließlich der Berücksichtigung von Rahmenbedingungen und Niederschlag all dessen in einzelnen Items des Fragebogens.
- Stärkere Nutzung und Langzeitauswertung der Ergebnissrückkopplungen innerhalb der Lehrveranstaltungen.
- Offensivere, gleichwohl weiterhin vertrauliche Veröffentlichung der Evaluationsergebnisse.
- Systematische Nutzung und Langzeitauswertung der strukturellen und curriculumsbezogenen studentischen Hinweise u.a. durch die Kombination mit oft entsprechenden Rückmeldungen aus der ebenfalls vom ZHQ betreuten Maßnahmen StAB (Stelle für Anregungen und Beschwerden).
- Ergänzung der vorlesungsbegleitenden Lehrveranstaltungsevaluation um eine Prüfung einbeziehende Modulevaluation, dabei Sicherstellung der konstruktiven kooperativen Datennutzung aller am Modul lehrend und unterstützend Beteiligten, auch der Mitarbeiter/innen.
- (Weiter-)Entwicklung geeigneter Erhebungsbögen für verschiedene Lehr- und Lernformen: Laborpraktika, Übungen, Seminare.
- Fachkultur- und personenspezifische Optimierung des Verhältnisses von Häufigkeit vs. Nachhaltigkeit und Tiefgang der Erhebungen.
- Weiterentwicklung des Einbezugs der Lehrveranstaltungsevaluation in den Prozess der pädagogischen Eigenprüfung.

Zu denken wäre im Kontext der ‚pädagogischen Eignung‘ an eine reguläre Nachbesprechung der Evaluationsergebnisse im Kreise des/der Neuberufenen und jeweils einem/r Vertreter/in aus Fach- und Hochschuldi-

daktik. Das Ergebnis dieser Diskussion könnte schriftlich festgehalten Teil eines Lehrportfolios werden und als weiterführende Information über reine Kennzahlen hinaus an die Eignungskommission gesandt werden.

Durch die hochschulweite Verpflichtung von Neuberufenen zur Teilnahme an Kursen des hdw-nrw und zur Evaluation ihrer Lehrveranstaltungen wurde eine neue Schicht von Professoren und Professorinnen gewonnen bzw. aufgebaut, die schon früh Erfahrungen mit hochschuldidaktischen Workshops und der Weiterentwicklung ihrer eigenen Veranstaltung gemacht haben. Entsprechend der eigenen Alltagserfahrungen sowie auch der Ergebnisse von Fendler (2004) ist davon auszugehen, dass die persönliche Erfahrung mit hochschuldidaktischer Weiterbildung und Beratung eine positivere Haltung ihr gegenüber zur Folge hat. Die Hemmschwelle, sich an Prozessen der Weiterentwicklung von Studium und Lehre zu beteiligen und selbstständig innovative Lehr- & Lernansätze auszuprobieren und zu evaluieren, ist insgesamt bereits gesunken. Wiederum gestiegen ist die Bereitschaft, mit Studierenden, Fachkolleg/innen und Vertretern der Hochschuldidaktik über kritische Ergebnisse der eigenen Lehrveranstaltungsevaluation zu sprechen. Nicht zuletzt sind für die Rolle als Evaluationsbeauftragte solche Professoren und Professorinnen besonders geeignet und füllen diese in besonders engagierter Weise aus, die eigene Erfahrungen mit hochschuldidaktischen Kursen und Empfehlungen zur Verbesserung der Lehrveranstaltung weitergeben und als Multiplikator fungieren können und wollen. Was in Folge formaler Regelungen als individuelle Maßnahme begonnen hat, kann so die Qualitätskultur eines ganzen Fachbereichs maßgeblich stärken.

Auf Grundlage der beschriebenen Zahlen, Fakten und Erfahrungen geben die Autoren abschließend ganz im Sinne von Boris Schmidt (2010) eine Antwort auf seine Abschlussformel ‚Qualitätsentwicklung = f (Evaluation, Hochschuldidaktik und deren Verknüpfung)‘. Unsere Formel heißt: ‚Qualitätsentwicklung = f (Hochschuldidaktik, Evaluation und Qualitätskultur in den Fachbereichen)‘. Erst wenn die beiden, bewusst in dieser Reihenfolge, erstgenannten Aspekte ihre Verknüpfung im dritten finden, kann dies zu den gewünschten Erfolgen und Folgen in der Qualitätsentwicklung von Studium und Lehre führen.

Literaturverzeichnis

- Bargel, T./El Hage, N. (2000): Evaluation der Hochschullehre: Modelle, Probleme und Perspektiven. In: Helmke, A./Hornstein, W./Terhart, E. (Hg.): Qualität und Qualitätssicherung im Bildungsbereich; Schule, Sozialpädagogik, Hochschule. Weinheim. S. 207-224.
- Böttcher, W./Grewe, C.M. (2010): Eine Untersuchung zur Wirksamkeit der studentischen Lehrveranstaltungskritik am Beispiel der Westfälischen Wilhelms-Universität Münster. In: Pohlenz, P./Oppermann, A. (Hg.): Lehre und Studium professionell evaluieren. Bielefeld. S. 73-82.
- Braun, E. (2007): Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen. (BEvaKomp), Göttingen.
- Craanen, M. (2011): Fakultätsübergreifende Qualitätsentwicklung von Lehrveranstaltungen am Karlsruher Institut für Technologie (KIT). In: Das Hochschulwesen – HSW, 59. Jg./H. 5, S. 153-161.
- DeGEval – Gesellschaft für Evaluation e.V. (2008): Standards für Evaluation. 4. unveränderte Auflage. Mainz.
- FH Aachen (2004): Evaluationsordnung der Fachhochschule Aachen. Teil A: Studium, Lehre und Weiterbildung vom 22.01.2004.

- Deutscher Bildungsrat (1971): Lernziele der Gesamtschule. Gutachten und Studien der Bildungskommission. Bd. 12, Stuttgart.
- Dresel, M./Tinsner, K. (2008): Onlineevaluation von Lehrveranstaltungen. Zeitschrift für Evaluation. 7 Jg./H. 2, S. 183-211.
- Flender, J. (2004): Optimierung ja – Weiterbildung nein? Zur Motivation von Lehrenden, ihre Lehre zu verbessern. In: Das Hochschulwesen – HSW, H. 1, S. 19-24.
- Gollwitzer, M./Schlotz, W. (2003): Das "Trierer Inventar zur Lehrveranstaltungsevaluation" (TRIL): Entwicklung und erste testtheoretische Erprobungen. In: Krampen, G./Zayer, H. (Hg.): Psychologiedidaktik und Evaluation IV. Neue Medien, Konzepte, Untersuchungsbefunde und Erfahrungen zur psychologischen Aus-, Fort- und Weiterbildung. Bonn: Deutscher Psychologen Verlag, S. 114-128.
- Hacker, W. (1978): Allgemeine Arbeits- und Ingenieurpsychologie. Psychische Struktur und Regulation von Arbeitstätigkeiten. 2. überarb. Auflage, Bern/Stuttgart/Wien.
- Heger, M./Mundiyapurath, S. (2004c): Studieren lernen und lehren. Schlüsselkompetenzen für Ingenieure/-innen. Veröffentlicht am 19.05.2004, unter: www.lehrdee.de/docs/erfolgsmodelle/db51.html (letzter Zugriff am 7.7.2014).
- Heger, M./Callhoff, A. (2005): Das Netzwerk Hochschuldidaktische Weiterbildung der nordrhein-westfälischen Fachhochschulen: hdw-nrw-fh. In: Brendel, S./Kaiser, K./Macke, G. (Hg.): Hochschuldidaktische Qualifizierung: Strategien und Konzepte im internationalen Vergleich (Blickpunkt Hochschuldidaktik Bd. 115). Bielefeld, S. 237-247.
- Heger, M./Sell, R. (1986): Längsschnittstudie zur Analyse der Studierfähigkeit im Ingenieurstudium an der RWTH Aachen und der FH Aachen. Abschlußbericht zu einem Forschungsprojekt des Ministers für Wissenschaft und Forschung des Landes Nordrhein-Westfalen, erschienen unter dem Titel: Studieren lernen im Ingenieur-Studium. Eine Begleituntersuchung zu Seminaren zur Verbesserung des Lern- und Arbeitsverhaltens an der RWTH Aachen. In: Brandt, D./Sell, R. (Hg.): Arbeitsbericht Nr. 22 aus dem HDZ der RWTH Aachen.
- Heger, M. (2004a): Konzept hochschuldidaktischer Aktionsforschung am Beispiel INGMEDIA. In: Dorothee M. Meister, Sigmar-Olaf Tergang, Peter Zentel (Hg.): Evaluation von E-Learning. Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven. Münster, S. 74-83.
- Heger, M. (2004b): Studieren lehren. In: Welbers U./Gaus G. (Hg.): The shift from Teaching to Learning. Konstruktionsbedingungen eines Ideals (Blickpunkt Hochschuldidaktik Bd. 116). Bielefeld, S. 158-164.
- Heger, M. (1986): Lehrbegleitende Beratung zu Massenveranstaltungen in der Hochschullehre. Darstellung eines Beratungskonzeptes und Fallbeispiel. In: Zeitschrift für Hochschuldidaktik, 10. Jg./H. 4, S. 484-512.
- Heger, M. (2009): Lehrveranstaltungsevaluation und lehrbegleitende Beratung an der Fachhochschule Aachen. In: Richthofen, Anja von und Lent, Michael (Hg.): Qualitätsentwicklung in Studium und Lehre (Blickpunkt Hochschuldidaktik Bd. 119). Bielefeld, S. 170-183.
- Heger, M. (1991): Lehrveranstaltungskritik als angewandte Hochschuldidaktik in: Sturm, M. / Csanyi G. (Hg.): Feedback versus Evaluation. Teil I. Zeitschrift für Hochschuldidaktik, Nr. 1/2. Wien, S. 43-69.
- Heger, M. (1997): Quantität und Qualität – Hochschuldidaktik mit Professoren und Studierenden Lernen aus lehrbegleitenden Beratungen an der Universität Kaiserslautern. In: Das Hochschulwesen, H. 4, S. 211-222.
- Heger, M. (1985): Verbesserung der Lehr- und Lernsituation in ingenieurwissenschaftlichen Studiengängen – eine Evaluation Hochschuldidaktischer Aus- und Fortbildung. Dissertation TU Berlin.
- Heger, M. (1984): Voraussetzungen, Prozesse und Ergebnisse von HDZ-Veranstaltungen für Lehrende: Skizzen handlungsorientierter Evaluation. In: Brandt, D./Sell R. (Hg.): Angewandte Hochschuldidaktik – Konzeption, Praxis, Bewertung. Blickpunkt Hochschuldidaktik, Nr. 76, Weinheim/Basel.
- Hof, C. (2001): Wie lässt sich soziale Kompetenz konkreter bestimmen? In: GdWZ, Grundlagen der Weiterbildung; Praxis, Forschung, Trends, H. 2, S. 151-154.
- Koch, E. (2004): Gute Hochschullehre. Theoriebezogene Herleitung und empirische Erfassung relevanter Lehraspekte. Hamburg.
- Kromrey, H. (2003): Qualität und Evaluation im System Hochschule. In: Stockmann, Reinhard (Hg.): Evaluationsforschung, Opladen: Leske+Budrich (2. Aufl.), S. 233-258.
- Leontjew, A.N. (1977): Tätigkeit, Bewußtsein, Persönlichkeit. Stuttgart.
- Rindermann, H. (2003): Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. In: Zeitschrift für Evaluation, 3. Jg./H. 2., S. 233-256.
- Rindermann, H. (2009): Lehrevaluation – Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Mit einem Beitrag zur Evaluation computerbasierten Unterrichts. Landau: Empirische Pädagogik.
- Schmidt, B. (2008): Warum oft wirksam? Und warum manchmal wirkungslos? – Subjektive Erklärungen zur Wirkung von Lehrveranstaltungsevaluation aus Sicht von Nutzern und Anbietern. In: Zeitschrift für Evaluation, 7. Jg./H. 1, S. 9-32.
- Schmidt, B. (2010): Qualitätsentwicklung = Evaluation + Hochschuldidaktik? Von einer Rechnung, die nicht aufgeht. In: Pohlenz, P./Oppermann, A. (Hg.): Lehre und Studium professionell evaluieren. Bielefeld, S. 179-192.
- Schmidt, B./Loßnitzer, T. (2010): Lehrveranstaltungsevaluation: State of the Art, ein Definitionsvorschlag und Entwicklungslinien. In: Zeitschrift für Evaluation, H. 1, S. 49-72.
- Souvignier, E./Gold, A. (2002): Fragebögen zur Lehrevaluation: Was können sie leisten? Zeitschrift für Evaluation, H. 2, S. 265-280.
- Stake, R.E. (1972): Verschiedene Aspekte pädagogischer Evaluation. In: Wulf, Ch. (Hg.): Evaluation, Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. München, S. 92-112.
- Sturm, M./G. Csanyi (1991): Feedback versus Evaluation. Teil I. Zeitschrift für Hochschuldidaktik, 2. Jg./H. 1, Wien.
- Tillmann, A./Reiß, S./Moosbrugger, H./Krömker, D./Schweizer, K./Gold, A. (2011): Qualitätssicherung der Lehre an großen Universitäten: Psychometrische Studien zum Frankfurter Studierendenfragebogen zur Evaluation von Lehrveranstaltungen (STUD-FEL). In: Qualität in der Wissenschaft, 5. Jg./H. 3, S. 79-89.
- Tinsner K./Thumser-Dauth K. (2010): Zusammenwirken von Evaluation und Hochschuldidaktik. Chancen für die Verbesserung der Lehrqualität. In: Pohlenz, P./Oppermann, A. (Hg.): Lehre und Studium professionell evaluieren. Bielefeld, S. 193-207.
- Trautwein, C./Merkt, M. (2013): Struktur und Entwicklung von Lehrkompetenz im Spannungsfeld von Überzeugungen, Konzepten und Praxis von Lehren und Lernen. ProfiLe-Teilprojekt Hamburg. In: Heiner, M./Wildt, J. (Hg.): Professionalisierung der Lehre. Perspektiven formeller und informeller Entwicklung von Lehrkompetenz im Kontext der Hochschulbildung (Blickpunkt Hochschuldidaktik Bd. 123). Bielefeld, S. 179-210.
- Volpert W. (1980): Beiträge zur psychologischen Handlungstheorie. Schriften zur Arbeitspsychologie Nr. 28, Bern/Stuttgart/Wien.
- Webler, W.D. (1991): Kriterien für gute akademische Lehre. In: Das Hochschulwesen – HSW, 39 Jg./H. 6, S. 243-249.
- Wolbring, T. (2013): Fallstricke der Lehrevaluation. Möglichkeiten und Grenzen der Messbarkeit von Lehrqualität (Hochschule und Gesellschaft Bd.1). Frankfurt am Main/New York: Campus.
- Zumbach, J./Spinath, B./Schahn, J./Friedrich, M./Kögel, M. (2007): Entwicklung einer Kurzskaala zur Lehrevaluation. In: Krämer, M./Preiser, S./Brusdeylins, K. (Hg.): Psychodidaktik und Evaluation Bd. 6, Göttingen. S. 317-325.

■ Jörg Jörissen, M.A., wissenschaftlicher Mitarbeiter, Zentrum für Hochschuldidaktik und Qualitätsentwicklung in Studium und Lehre (ZHQ), FH Aachen, Lehrbeauftragter der Katho NRW, E-Mail: joerissen@fh-aachen.de

■ Prof. Dr. Michael Heger, Geschäftsführer des Zentrums für Hochschuldidaktik und Qualitätsentwicklung in Studium und Lehre (ZHQ), FH Aachen, E-Mail: heger@fh-aachen.de

Benjamin Ditzel



Benjamin Ditzel

Evaluationsverfahren als Ausgangspunkt für Diskussions- und Reflexionsprozesse. Erfahrungen mit dem Aufbau eines hochschulweiten Qualitätsmanagements an der Universität Hildesheim

In der Debatte um die Qualität an Hochschulen ist ein Trend zu institutionellen Ansätzen des Qualitätsmanagements (QM) zu beobachten (vgl. Nickel 2008; Loukola/Zhang 2010). Ausgehend von der Erkenntnis, dass Evaluationsverfahren in der Praxis häufig bei dem Schritt der Erhebung und Analyse stehen bleiben und zu selten in konkrete Verbesserungsmaßnahmen münden (vgl. Mittag et al. 2003, S. 130ff; Heublein et al. 2012, S. 20) werden bislang eher unverbundene Einzelmaßnahmen und -instrumente (vgl. Schmidt 2010) aufeinander bezogen und in ein konsistentes Managementsystem integriert. Mit Prozessen der Evaluation ist dabei in zunehmendem Maße eine Steuerungsabsicht verbunden (vgl. Lühje 2004, S. 5; HRK 2006), so dass die Ergebnisse nicht mehr nur Eingang in Informations- und Kommunikationsprozesse finden, sondern systematisch für Entscheidungsprozesse herangezogen werden.

Mit Blick auf diese Entwicklung soll im vorliegenden Artikel zwei grundlegenden Fragestellungen nachgegangen werden: Zum einen will der Artikel einen Beitrag dazu leisten, den bislang noch diffusen Begriff des QM an Hochschulen durch die Benennung relevanter Elemente bzw. Gestaltungsdimensionen konkreter zu fassen (siehe Abschnitt 2). Und zum anderen soll der Frage nachgegangen werden, was es heißen kann, die Schwierigkeiten einer Steuerung von Hochschulen ernst zu nehmen und Erkenntnisse aus der Management- und Organisationsforschung für das QM an Hochschulen fruchtbar zu machen (siehe Abschnitte 1, 3 und 4).

1. Herausforderungen organisationaler Qualitätsansätze

Gerade in jüngerer Zeit wurden erfolgreiche Praxisbeispiele für die Integration von Einzelmaßnahmen der Qualitätssicherung und -entwicklung, zur Implementierung kontinuierlicher Regelkreise und zur organisationsweiten Verankerung des Qualitätsmanagements beschrieben (vgl. Nickel 2007; Pasternack/Hölscher 2008; Winde 2010). In der Praxis stellen sich diese Beispiele als äußerst heterogen dar und beziehen sich zumeist auf einzelne Leuchtturmprojekte; in der Breite ist eine ähnliche Tendenz noch nicht erkennbar.

Auch sind viele Fragen der konkreten Ausgestaltung eines QM nach wie vor offen (vgl. Scheytt 2005, S. 136; Nickel 2008, S. 17). Die Verfahren der externen Qualitätssicherung liefern nur wenige Anhaltspunkte; die

Adaption und Anwendung von Ansätzen aus der Wirtschaft (ISO 9001, EFQM-Modell) stellt sich als schwierig dar und hat sich bislang nicht durchgesetzt. Insgesamt fehlt es nicht nur an einem allgemein verbindlichen Grundverständnis für ein hochschuladäquates QM (Nickel 2008, S. 26), sondern überhaupt an relevantem Orientierungswissen zu dessen Ausgestaltung.

Vor diesem Hintergrund stellt sich die Frage, was sich hinter QM konkret verbirgt, welche Elemente ein derartiges Managementsystem umfassen sollte und wie es organisational verankert werden kann. Dies soll im vorliegenden Beitrag am Beispiel des Qualitätsansatzes der Universität Hildesheim anhand von Gestaltungsdimensionen diskutiert werden. Ein besonderer Schwerpunkt wird dabei auf die Etablierung von Diskussions- und Reflexionsprozessen als Ausgangspunkt für Veränderungen in Studium und Lehre gelegt.

Aus einer theoretisch-reflexiven Perspektive stellt sich darüber hinaus die Frage, inwiefern die Hochschule überhaupt in der Lage ist, „sich selbst als Organisation zu begreifen“ (Pellert 2000, S. 39), wie ihre Managementfähigkeit gefördert werden kann bzw. ob sich eine Organisation in der intendierten Weise überhaupt steuern lässt.

Hochschulbezogene Analysen betonen regelmäßig die Besonderheiten der Organisationsform (Musselin 2007; Wilkesmann/Schmid 2012): Das Modell der professionellen Bürokratie (vgl. Mintzberg 1983) lenkt den Blick auf die starke Stellung der Profession; das Modell der organisierten Anarchie (vgl. Cohen et al. 1972) stellt rationales und planbares Entscheidungsverhalten in Bildungsorganisationen in Frage und das Konzept lose gekoppelter Systeme (vgl. Weick 1976) betont die Unabhängigkeit der Teilsysteme einer Organisation. Insgesamt wird die Hochschule als äußerst komplex und in ihren Zielsystemen paradox beschrieben (vgl. Barnett 2000; Meister-Scheytt/Scheytt 2005).

Die organisationalen Spezifika stellen nicht nur besondere Anforderungen an managementorientierte und auf die Organisation abzielende Interventionen (Scheytt 2005, S. 5), sie stellen auch die Wirksamkeit von Management grundsätzlich infrage (Krücken 2008; Hechler/Pasternack 2012, S. 38). Die paradoxe Herausforderung besteht darin, dass die Hochschule steuernd auf die akademischen Leistungsprozesse des Forschens, Lehrens und Lernens einwirken soll, obwohl ihr diese im Kern eigentlich nicht zugänglich sind (vgl. Huber 2012, S. 249). Neuere organisationssoziologische Arbeiten weisen

zwar darauf hin, dass sich Hochschulen zunehmend in „einheitlich handlungs-, entscheidungs- und strategie-fähige Akteure“ wandeln (Krücken 2011, S. 109). Doch trotz der Stärkung von Kopplungsmechanismen beispielsweise durch die Etablierung von Strukturen und Prozessen der Qualitätssicherung bleiben die grundsätzlichen Eigenarten der Organisation Hochschule erhalten (vgl. Kloke/Krücken 2012, S. 320).

Insgesamt lässt sich darüber hinaus in der Management- und Organisationsforschung eine generelle Steuerungsskepsis beobachten (vgl. Krücken 2008; Habersam 2000; Berthold 2011). Denn es könne weder von rational handelnden Individuen (vgl. March 1978; Simon 1991) noch von rational handelnden Organisationen (vgl. Luhmann 1973) ausgegangen werden. Vielmehr handeln Organisationen aufgrund von operationaler Geschlossenheit und Selbstreferenz nach ihrer eigenen Funktionslogik (vgl. Luhmann 2006), so dass eine gezielte, direkte oder lineare Steuerung von außen nur bedingt möglich ist (Willke 2001, S. 31).

Neo-institutionalistische Analysen werfen darüber hinaus die Frage auf, inwieweit veränderte Steuerungsmechanismen und Leitungsstrukturen tatsächliche Veränderungen in einer Organisation insbesondere auf der realen Handlungsebene bewirken. Empirische Erkenntnisse legen die Vermutung nahe, dass es dabei eher zu einer Entkopplung von Formal- und Aktivitätsstruktur (vgl. Meyer/Rowan 1977) bzw. Organisationsstruktur und Schauseite einerseits und Handlungsroutrinen andererseits (vgl. Brunsson 1991; Kühl 2011, S. 95ff) kommt. Gerade für Reflexionen zum QM sind solche Positionen von besonderer Bedeutung. Denn die Möglichkeiten für die gezielte Intervention in die akademischen Kernprozesse stellen sich dadurch als wesentlich beschränkter dar, als es mitunter in der Debatte um das Management von Hochschulen anerkannt wird (vgl. Krücken 2008; Habersam 2000).

Angesichts der Komplexität der Umwelt gilt Komplexitätsreduktion nicht mehr als adäquate Managementstrategie (vgl. Baecker 2003; Malik 2008). Die Organisation ist auf die Intelligenz dezentraler Einheiten angewiesen (Willke 1989, S. 19) und muss „viel stärker als bislang mit vorläufigen, provisorischen, ungewissen und hypothetischen Prämissen“ arbeiten (Willke 2007, S. 112). Da Planungsprozesse nicht eindeutig determinierbar und immer nur vorübergehend gültig sind, verliert die Planung ihr Primat für den Steuerungsprozess (vgl. Krücken 2008, S. 356; Steinmann et al. 2013, 125ff.). Auch auf Dokumentation, Standards und Konformität abzielende Managementpraktiken gelten als wenig geeignet, die Lebens- und Lernfähigkeit von Organisationen zu gewährleisten (vgl. Beck/Walgenbach 2003; Ditzel 2013a). Damit aber ist ein Paradigmenwechsel vom klassischen, auf Planung, Konformität und Kontrolle setzenden Ansatz hin zu einem Managementverständnis verbunden, das Steuerung als Kombination sensibler, zentraler Mechanismen der Kontextsteuerung mit gleichzeitiger Anregung und Förderung dezentraler Selbststeuerungsprozesse versteht (vgl. Willke 1989; Ditzel 2013a).

Doch obwohl viel zum QM und noch mehr zur Qualitätssicherung an Hochschulen publiziert wird, bleibt

weitgehend unklar, welche Auswirkungen derartige Überlegungen auf die Konzeption von QM-Systemen haben. Die zahlreichen Positionspapiere und Erfahrungsberichte weisen diesbezüglich in der Regel ein geringes Reflexionspotenzial auf. Analysen von Qualitätsansätzen (vgl. Webler 2002; Pasternack 2004; Nickel 2007) beziehen sich weitgehend auf eine deskriptive und vergleichende Darstellung von Instrumenten. Es findet zwar eine intensive und auch kritische Auseinandersetzung mit der Methodik und dem Beitrag einzelner Instrumente zum QM statt. Aber es wird selten hinterfragt, inwiefern die Ansätze für die Hochschule als Expertenorganisation (vgl. Pellert 1999) überhaupt geeignet sind und welche Funktion die Instrumente vor diesem Hintergrund spielen können.

Eine Konsequenz aus den skizzierten Überlegungen zur Steuerbarkeit der Hochschule als Organisation kann für das QM darin bestehen, das Steuerungsparadigma der Selbstorganisation in den Mittelpunkt zu stellen und die dezentrale Verantwortung für Qualität anzuerkennen, ohne diese allerdings im Sinne traditioneller akademischer Selbststeuerungsmechanismen sich selbst zu überlassen:

- Im Sinne von Selbststeuerung geht es darum, die Verantwortung für Steuerungsprozesse möglichst dezentral zu verorten, gleichzeitig aber die Selbststeuerungs- und Selbstreflexionsfähigkeit zu fördern.
- Im Sinne von Kontextsteuerung scheint es für eine Anbindung und Rückbindung der Teilsysteme an die Gesamtorganisation andererseits erforderlich, die dezentrale Selbststeuerung durch wenige, allgemeine Rahmenvereinbarungen auf die Organisation zu beziehen.
- Anstelle von Konformität und daraus abgeleiteter Kontrollanliegen geht es eher um die Anregung von Lern- und Veränderungsprozessen.
- Auf diese Weise gewinnen Rückkopplungsprozesse und die Beteiligung der unterschiedlichen Akteure an Bedeutung.
- Es wird wichtiger für Akzeptanz von Veränderungen zu werben, als sie von oben durchzusetzen.

Wie sich ein solcher konzeptioneller Bezugsrahmen auf die konkrete Implementierung an einer Hochschule auswirken kann, wird am Beispiel der Universität Hildesheim nachgezeichnet. Aspekte der Selbstreflexion und Selbststeuerung werden beispielhaft vorgestellt. Als Irritationen auslösende Momente kommen insbesondere Evaluationsverfahren zum Einsatz, die in organisational verankerte Diskussions- und Reflexionsprozesse münden.

2. Elemente des Qualitätsmanagements der Universität Hildesheim

Im Jahr 2007 wurde an der Universität Hildesheim mit der Einrichtung der Stabstelle QM damit begonnen, die bis dahin weitgehend unabhängig voneinander durchgeführten Einzelmaßnahmen der Qualitätssicherung und -entwicklung in einem hochschulweiten QM zu systematisieren, miteinander zu verknüpfen und auf unterschiedliche für die Qualität von Studium und Lehre relevanten Bereiche und Aspekte auszuweiten. Ziel war es, die zu diesem Zeitpunkt anstehende Welle an Akkredi-

tierungs- bzw. Reakkreditierungsverfahren systematisch zu begleiten und mit Blick auf eine mögliche Prozess- bzw. Systemakkreditierung den Aufbau eines hochschulweiten QM-Systems voranzutreiben.

Das dabei aufgebaute QM-System lässt sich anhand der nachfolgenden Elemente beschreiben. Die Darstellung dieser Elemente erfolgt am Beispiel der Universität Hildesheim, spannt gleichzeitig aber mögliche Gestaltungsfelder des QM an Hochschulen auf. Dabei beziehen sich die strategische und die strukturelle Dimension auf die Rahmenbedingungen zur Einbettung in die Organisation, während die Prozessdimension, die Instrumente, der Qualitätsregelkreis sowie das Dokumentations- und Berichtssystem als instrumentelle Gestaltungsdimensionen Aufschluss über die methodische Ausgestaltung geben:

Die **strategische Dimension** beschreibt die strategische Verankerung und Ausrichtung des QM innerhalb der Hochschule. Im Sinne von Kontextsteuerung erscheint es sinnvoll, einzelne Aktivitäten, Maßnahmen und Prozesse der Qualitätssicherung und -entwicklung hinsichtlich übergeordneter Ziele zu fokussieren und aufeinander zu beziehen. Dabei ist es wichtig, geeignete Orte und Mechanismen der Kopplung zu identifizieren. Anknüpfungspunkte für das QM ergeben sich beispielsweise über ein Leitbild, einen Entwicklungsplan oder Ziel- und Leistungsvereinbarungen. Eine Operationalisierung kann über Ziele, Standards und Leistungsindikatoren erfolgen. Dabei ist jedoch zum einen die Komplexität des Qualitätsbegriffs zu beachten, der sich einer klaren Definition und quantifizierbaren Messung weitgehend entzieht (vgl. Pellert 2002, S. 24; Kromrey 2004, S. 240; Harris-Hümmert 2011, S. 35ff.). Zum anderen sollte es dabei um die Definition gemeinsamer, hinreichend abstrakt und allgemein gehaltener Rahmenvorgaben für die Gesamtorganisation gehen (vgl. Orton/Weick 1990, S. 212; Pellert 2002, S. 28). Eine starre, kleinteilige Detailsteuerung sollte vermieden werden, damit genug Spielraum für Interpretationen und Anpassungen an lokale Erfordernisse bleibt.

Das 2008 vom Senat der Universität Hildesheim beschlossene Leitbild benennt eine grundsätzliche Selbstverpflichtung der Hochschulmitglieder, die Qualität in Lehre und Forschung zu sichern und weiterzuentwickeln. Das QM steht unter dem Motto einer *diskursiven und partizipativen Qualitätskultur*. Damit soll zum Ausdruck gebracht werden, dass es in erster Linie darum geht, durch das Anregen von Diskussions- und Reflexionsprozessen organisationales Lernen zu fördern und zu einer Veränderung der Werte und Haltungen beizutragen. Im Sinne von Partizipation sollen alle relevanten Akteure eingebunden werden. Eine darüber hinausgehende Konkretisierung von Qualitätszielen hat bis auf die Identifikation wesentlicher Handlungsfelder zur Weiterentwicklung von Studium und Lehre bislang nicht stattgefunden, da der Fokus auf der Etablierung von Evaluationsverfahren und Reflexionsprozessen lag. Eine Diskussion des zugrundeliegenden Qualitätsverständnisses erfolgte im Zusammenhang mit der Einigung auf Qualitätsmerkmale und Frageitems.

Die **strukturelle Dimension** beschreibt die strukturelle Verankerung des QM in der Organisation durch Kommunikations- und Entscheidungsstrukturen. Die Defini-

tion von Zuständigkeiten und Verantwortlichkeiten bezieht sich einerseits auf Prozesse in Studium und Lehre allgemein und andererseits auf solche der Qualitätssicherung und -entwicklung. Die Verantwortung für Qualität sollte möglichst dort verortet werden, wo die Leistungserstellung stattfindet. Dass über Qualität in den dezentralen Einheiten diskutiert wird, kann zwar nicht garantiert werden. Durch die Etablierung von Gremien, die sich mit Qualitätsthemen auseinandersetzen, durch unterschiedliche Veranstaltungen, durch das Etablieren von Reflexionsprozessen, durch Evaluationsverfahren und die Pflicht zur Berichterstattung kann die Aufmerksamkeit jedoch immer wieder auf qualitätsrelevante Themen gelenkt werden. Die Entwicklung einer entsprechenden Qualitätskultur wird dadurch nicht zwangsläufig sichergestellt, aber sie wird – im Laufe der Zeit – wahrscheinlicher.

Das QM der Universität Hildesheim zeichnet sich durch einen solchen dezentralen Ansatz aus. Die Einbindung der Lehrenden, der Studierenden sowie der Mitarbeiterinnen und Mitarbeiter der Verwaltung und unterstützender Einrichtungen in die Weiterentwicklung der Qualität von Studium und Lehre geschieht durch eine enge Verzahnung des zentralen und dezentralen QM. Während das zentrale insbesondere dafür zuständig ist die notwendigen Rahmenbedingungen für Qualität zu schaffen, konzentriert sich das dezentrale QM auf die inhaltliche Ausgestaltung.

Für das *zentrale QM* wurde eine Senatskommission eingerichtet. Sie begleitet konzeptionell den Prozess der Entwicklung und Weiterentwicklung des Qualitätsmanagements der Universität Hildesheim und bereitet wichtige Entscheidungen in Senat und Hochschulleitung vor. Die Stabstelle QM hat eine unterstützende, beratende und durchführende Funktion. Für das dezentrale QM wurde auf die im Hochschulgesetz verankerten, paritätisch mit Studierenden besetzten Studienkommissionen zurückgegriffen. Sie sind für die Qualitätssicherung und Weiterentwicklung der Studiengänge ihres Fachbereichs zuständig und bereiten die entsprechenden Entscheidungen der Fachbereichsräte vor.

Den Fachbereichen kommt eine weitgehende Autonomie bei der Entwicklung und Umsetzung von Maßnahmen zur Qualitätssicherung und -entwicklung zu. Zugleich wird in den Evaluationsverfahren darauf geachtet, dass sich über die Bedürfnisse und Ansprüche einzelner Disziplinen und Studiengänge hinaus eine Vergleichbarkeit der Ergebnisse herstellen lässt, über die dann immer auch universitätsweit diskutiert werden kann.

Die **Prozessdimension** sensibilisiert für den Einfluss von Prozessabläufen auf die Ergebnisqualität. Im Sinne des Konformitätsansatzes werden besonders qualitätsrelevante Prozesse mit ihren Verantwortlichen, Abläufen und Schnittstellen betrachtet und – wo dies im Sinne der Qualitätssicherung des Ergebnisses notwendig erscheint – dokumentiert und hinsichtlich verbindlicher Prozesskomponenten fixiert. Dabei interessieren in erster Linie jene die Qualität in besonderem Maße beeinflussenden Prozessschritte sowie Schnittstellen zwischen Organisationseinheiten. Im Sinne des Lern- und Verbesserungsansatzes geht es aber auch darum, die Leistungsfähigkeit der Prozesse im Hinblick auf die zu erzielenden Ergeb-

nisse immer wieder (selbst-)kritisch zu hinterfragen und abteilungsübergreifend insbesondere bezogen auf die Schnittstellen nach Möglichkeiten zur Optimierung zu suchen (vgl. Ditzel 2013a).

Ausgehend von einer Prozesslandkarte werden an der Universität Hildesheim die besonders qualitätsrelevanten Prozesse definiert. Sie geben Hinweise, an welchen Stellen die Qualität besonders beeinflusst wird. Wesentliche Kernprozesse sind in einen Regelkreis von Rückkopplungs- und Veränderungsprozessen eingebettet. Um die Prozessqualität zu steigern, wurden in ausgewählten, v.a. service-orientierten Bereichen wie Immatrikulation, Prüfungsorganisation, Studienberatung und Bibliothek Optimierungsprojekte durchgeführt (vgl. Ditzel 2013b).

Instrumente: Indem Einzelinstrumente systematisiert und untereinander verknüpft werden, tragen sie wesentlich zur Etablierung eines QM-Systems bei. Vor dem Hintergrund der Überlegungen zur Steuerbarkeit von Hochschulen spielen Fragen der Funktion von Evaluationsverfahren (Legitimation, Kontrolle, Steuerung, Feedback, Erkenntnis, vgl. Kromrey 2004) sowie ihrer organisationalen Einbettung in Kommunikations- und Entscheidungsprozesse eine wichtige Rolle. In den Abschnitten 3 und 4 werden diese beiden Aspekte am Beispiel der Universität Hildesheim ausführlicher diskutiert. Der sog. **Qualitätsregelkreis** nach Deming liefert eine methodische Orientierung, um das eher abstrakte Qualitätsdenken für konkrete Prozesse und Probleme übersetzbar zu machen. Er gibt Hinweise darauf, welche Aufgaben der Qualitätssicherung und -entwicklung in den unterschiedlichen Anwendungsfeldern und auf den unterschiedlichen Abstraktionsebenen relevant sind und welche Schritte jeweils durchlaufen werden sollten – dass Qualität definiert wird, dass Ziele entwickelt und überprüft werden, dass die zur Schaffung von Qualität notwendigen Rahmenbedingungen bereitgestellt werden, dass Feedbackschleifen zur Beurteilung der Qualität eingerichtet werden, dass die Qualität diskutiert und reflektiert wird und schließlich, dass aus diesen Erkenntnissen gelernt und bei Bedarf Maßnahmen zur Weiterentwicklung ergriffen werden (Follow-up). Dieses Schema kann auf den Bereich Studium und Lehre allgemein, zugleich aber auch auf einzelne Teilbereiche oder Prozesse angewendet werden (vgl. Ditzel 2013b).

Der Qualitätsregelkreis wurde an der Universität Hildesheim als grundsätzliches Orientierungsschema und als Reflexionsschema zur kontinuierlichen Verbesserung etabliert; eine kleinteilige und deterministische Definition und Überprüfung konkreter Qualitätsziele findet vor dem Hintergrund der eingangs diskutierten Steuerungsskepsis nur eingeschränkt statt (vgl. Abschnitt 1). Eine Dynamisierung und Verbindlichkeit von Follow-up-Prozessen wird durch regelmäßige Anlässe zur Diskussion und Reflexion, die Definition klarer Zuständigkeiten sowie die Verknüpfung von zentralem und dezentralem QM (vgl. Abschnitt 4) sichergestellt.

Dokumentations- und Berichtssystem: Dokumentation spielt bei der Etablierung prozessorientierter Managementsysteme eine wichtige Rolle, nicht nur um hochschulintern Transparenz über Verantwortlichkeiten, Prozessabläufe und Instrumente zu schaffen und externe

Anforderungen zur Legitimation und Rechenschaftslegung zu befriedigen, sondern auch, um gezielt Daten und Informationen zur Leistungsfähigkeit von Prozessen bereitzustellen und auf dieser Grundlage Reflexions- und Lernprozesse anzuregen bzw. zu unterstützen (vgl. Ditzel 2013b).

3. Evaluationsverfahren als Feedbackinstrumente

Für die Einbindung von Evaluationsverfahren in das QM-System stellt sich zunächst die Frage, welche Funktion sie einnehmen sollen. Durch Hochschulgesetze und Akkreditierungsverfahren sind Hochschulen zunehmend aufgefordert, Verfahren zur Beurteilung der Qualität ihrer Leistungsprozesse durchzuführen. Solche externen Anforderungen im Sinne einer Legitimationsfunktion zu bedienen, ohne die Ergebnisse der Qualitätssicherung zur Weiterentwicklung von Studium und Lehre zu verwenden, trägt allerdings wenig zu einem funktionsfähigen QM bei. Hochschulintern sollen Evaluationsverfahren darüber hinaus insbesondere aus Leitungssicht eine Datengrundlage zur Kontrolle und Steuerung liefern, um beispielsweise Entscheidungen der Personalentwicklung oder des Ressourceneinsatzes daran auszurichten. Angesichts der Komplexität des Qualitätskonzeptes (Pellert 2002, S. 24) und der eingeschränkten Möglichkeiten für eine objektive Messung (Kromrey 2004, S. 240) sind Evaluationen dafür allerdings nur bedingt geeignet. Aus-sichtsreich erscheint vor diesem Hintergrund und angesichts der in der Organisations- und Managementforschung verbreiteten Steuerungsskepsis (vgl. Abschnitt 1) insbesondere die Feedbackfunktion.

Entsprechend liegt an der Universität Hildesheim der Fokus von Evaluation darauf, eine Gesprächs- und Reflexionskultur zu fördern. Evaluationsverfahren werden als Feedbackinstrumente genutzt, um Irritationen zu erzeugen und damit Lern- und Veränderungsprozesse anzuregen (vgl. Willke 1989). Es geht darum, Anlässe zu schaffen, um über die Qualität von Studium und Lehre zu diskutieren. Ziel ist es nicht, die Qualität exakt zu messen oder die Lehrenden zu kontrollieren.

Einen solchen Ansatz hochschulintern glaubhaft zu vermitteln, stellt allerdings eine große Herausforderung dar. Denn insbesondere seitens der Lehrenden besteht eine große Skepsis gegenüber der Implementierung von qualitätssichernden Maßnahmen, mit denen häufig „Kontrolle“ und „Bürokratisierung“ assoziiert werden oder Angst vor Transparenz und Machtverlust verbunden ist. Ein von diesen Vorurteilen abweichendes QM zu etablieren setzt eine vertrauensstiftende und sensible Herangehensweise voraus und bedeutet viel Zeit, Ausdauer und Verständnis für die Ängste der Betroffenen aufzubringen. Um die Akzeptanz der Evaluationsverfahren zu verbessern, wurde daher großer Wert auf die Ausdifferenzierung unterschiedlicher Evaluationsinstrumente, eine partizipative Gestaltung sowie eine hohe Transparenz der Verfahren gelegt.

Differenziertes Evaluationssystem

Aufgabe der Evaluationsverfahren ist es, relevante Informationen zur Qualität von Studium und Lehre bereitzu-

stellen und damit Impulse für Reflexionsprozesse zu geben. Die Verfahren können dabei immer nur einen Teilausschnitt der Qualitätsrealität repräsentieren, sie können kein objektives Qualitätsbild erzeugen. Insofern sollten die Verfahren in ihrer Gesamtheit die Vielschichtigkeit des Qualitätsbegriffs berücksichtigen und alle relevanten Dimensionen, Ebenen, Interaktionsphasen und Interessengruppen adressieren (vgl. Ditzel 2011). Nicht für jede Perspektive oder für jedes Kriterium sind die zur Verfügung stehenden Instrumente in gleicher Weise geeignet. Eine wesentliche Herausforderung besteht darin, den richtigen Mix zu finden, um ein möglichst vollständiges und vielschichtiges Bild von der Qualität in Studium und Lehre zu zeichnen (Pellert 2002, S. 28). An der Universität Hildesheim wurde dazu ein mehrstufiges Evaluationssystem aufgebaut:

- Die durch eine zentral organisierte *Evaluation von Lehrveranstaltungen* gewonnenen Daten dienen den Lehrenden als subjektive Rückmeldung der Studierenden zu wichtigen Qualitätsmerkmalen der Lehre. Die Befragung wird in der Regel in der sechsten Semesterwoche durchgeführt, so dass die Lehrenden noch im laufenden Semester mit Studierenden über die Ergebnisse ins Gespräch kommen und die Studierenden bereits von etwaigen Veränderungen profitieren können, zumindest aber eine direkte Rückmeldung erhalten.
- Eine *Studiengangsevaluation* untersucht den übergeordneten Zusammenhang der Studiengänge hinsichtlich ihres strukturellen und inhaltlichen Aufbaus. Neben dem Lehrangebot und den Prüfungsmodalitäten werden Faktoren erhoben, die aus Sicht der Studierenden die Studierbarkeit beeinträchtigen.
- Eine *Evaluation der Studienbedingungen* nimmt übergeordnete Rahmenbedingungen des Studierens in den Blick. Dabei geht es um Dienstleistungs- und Organi-

sationsprozesse, also um Qualitätsaspekte der Betreuung und Beratung von Studierenden (durch Lehrende, durch die Studienberatung, durch das Prüfungsamt) sowie um die Qualität der bereitgestellten Infrastruktur (Räume, Ausstattung, Universitätsbibliothek oder elektronische Serviceangebote).

Um die Evaluationsbelastung zu reduzieren, die Evaluationsverfahren besser aufeinander abzustimmen und genug Zeit für das Einleiten von Maßnahmen einzuräumen, werden diese drei Verfahren alternierend im Semesterrhythmus durchgeführt (siehe Abbildung 1). Ergänzt werden diese Verfahren durch regelmäßige Absolventen-, Erstsemester- sowie anlassbezogene Befragungen zu ausgewählten Themen- und Funktionsbereichen. Die Etablierung hochschulweiter Evaluationsverfahren wurde mit der Lehrveranstaltungsevaluation begonnen. Rückblickend erscheint es sinnvoller, zunächst mit übergeordneten Verfahren zu starten, bei denen die Lehrenden nicht persönlich im Fokus stehen. Evaluationsverfahren zu den Studienbedingungen oder zu den Studiengängen relativieren die Bedeutung der Lehrveranstaltungsevaluation und zeigen, dass sich das Qualitätsbild nicht allein aus der Bewertung der Lehrenden als Person speist. Widerstände gegen Evaluationsverfahren können auf diese Weise abgebaut werden.

Partizipative Gestaltung der Evaluationsinstrumente

Die Entwicklung der Evaluationsinstrumente erfolgte in einem mehrstufigen Prozess, an dem unterschiedliche Gremien beteiligt waren. Initiiert wurde der Prozess durch den zuständigen Vizepräsidenten und die Stabstelle QM. Ausgehend von der Senatskommission wurden alle dezentralen Studienkommissionen in die Planung der Verfahren und die Entwicklung der Fragebögen

Abbildung 1: Übersicht über regelmäßige interne Evaluationsverfahren

Evaluationsinstrumente		Erhebungsrhythmus					Fragebogen
		SoSe	WiSe	SoSe	WiSe	SoSe	
Lehrveranstaltungs-evaluation	▪ Unterschiedliche Fragebögen je nach Veranstaltungsart ▪ Für individuelle Veranstaltungsarten können alternative Instrumente angewendet werden						▪ zentrale Fragebögen je Veranstaltungsart ▪ dezentrale Ergänzung je Fachbereich/Institut
Studiengangsevaluation	▪ Befragung von Studierenden ab dem zweiten Hochschulsemester						▪ zentraler Fragebogen ▪ dezentrale Ergänzung je Studiengang
Evaluation der Studienbedingungen	▪ Befragung von Studierenden ab dem zweiten Hochschulsemester						▪ zentraler Fragebogen
Erstsemesterbefragung	▪ Befragung von Studierenden im ersten Hochschulsemester						▪ zentraler Fragebogen
Absolventenbefragung	▪ Befragung von Absolventinnen und Absolventen frühestens ein Jahr nach Beendigung des Studiums	i.d.R. für Jahrgänge gebündelt					▪ dezentrale Fragebögen

involviert. Dabei kam der Beteiligung von Studierenden eine besondere Rolle zu.

Der Diskussionsprozess zu einer hochschulweiten Evaluation der Lehrveranstaltungen wurde zunächst mit einem Fragebogenentwurf gestartet, der vornehmlich unter methodischen Gesichtspunkten entwickelt worden war. Die einzelnen Items stießen jedoch bei Lehrenden anderer Fächerkulturen und bei Studierenden auf Kritik. Es zeigte sich schnell, dass es zwar wichtig ist, Aspekte der Forschungsmethodik bei der Entwicklung zu berücksichtigen, dass aber ein partizipativer Aushandlungsprozess ebenso notwendig ist, um die Akzeptanz bei Lehrenden und Studierenden nicht zu gefährden.

Transparenz über Verfahren, Ergebnisse und Konsequenzen

Für die Akzeptanz von Evaluationsverfahren sowohl bei Studierenden als auch bei Lehrenden spielt Transparenz eine wichtige Rolle; Transparenz bezüglich der zugrunde liegenden Ziele, bezüglich des Verfahrens und bezüglich der Ergebnisse. Lehrende beziehen ihre Forderung nach Transparenz in erster Linie auf den Verfahrensablauf und die Verwendung der Ergebnisse. Für sie ist besonders wichtig, dass festgelegt ist, wem welche Daten zur Verfügung gestellt werden. Dabei reicht es nicht, die Evaluationsverfahren nachvollziehbar z.B. auf einer Internetseite darzustellen und die Lehrenden zu informieren. Die Skepsis sitzt tief, so dass über die Zeit erst Vertrauen aufgebaut werden muss. Dabei wirkt die Nutzung personenbezogener Evaluationsergebnisse bei Personalentscheidungen oder die Diskussion in Leitungsgremien einer vertrauensstiftenden Herangehensweise entgegen. Für Studierende steht im Vordergrund, was mit den Ergebnissen passiert, welche Maßnahmen zur Weiterentwicklung von Studium und Lehre tatsächlich ergriffen werden und welcher konkrete Nutzen sich für sie ergibt. In der Praxis zeigt selbst ein breites Informations- und Partizipationsangebot wenig Wirkung, wenn die Studierenden darauf nicht zugreifen (vgl. Ditzel/Bergt 2013). Insofern müssen aufmerksamkeitsbindende Maßnahmen schon viel früher starten und setzen eine konstruktive Feedbackkultur voraus. Denn bei den Studierenden hat sich vielfach – berechtigt oder nicht – die Wahrnehmung gefestigt, dass Evaluationen ohne Konsequenzen bleiben. Aus studentischer Sicht besteht die wesentliche Herausforderung darin, Transparenz über die Verwendung der Evaluationsergebnisse und die Ableitung von Maßnahmen zu schaffen. Und das gelingt besser, wenn Studierende direkt in die Weiterentwicklung von Studium und Lehre und in die Entwicklung und Auswertung von Evaluationsinstrumenten eingebunden werden.

4. Etablierung von Diskussions- und Reflexionsprozessen

Evaluationsinstrumente sind zwar eine notwendige, aber nicht hinreichende Bedingung für ein Qualitätsmanagement. Zur Stärkung der Selbstreflexionsfähigkeit der Hochschule gilt es darüber hinaus eine Gesprächs- und Reflexionskultur zu fördern. Damit soll eine Auseinandersetzung mit Qualitätsthemen und Organisationsbelangen angeregt werden. Durch das Anstoßen von

Lern- und Veränderungsprozessen soll darüber hinaus aber auch konkret an der Weiterentwicklung von Studium und Lehre gearbeitet werden, denn es geht nicht um Reflexion zum Selbstzweck.

Dazu braucht es zunächst institutionalisierte oder informelle Orte, an denen solch eine Diskussion und Reflexion stattfinden kann. Bei der Ausgestaltung des QM an der Universität Hildesheim wurde deshalb darauf geachtet, auf allen für Studium und Lehre relevanten Ebenen entsprechende Orte zu etablieren und gleichzeitig die vorhandenen Strukturen weitgehend zu nutzen. Im Zentrum der Qualitätsdiskussionen stehen die Gremien des zentralen (Senatskommission QM) und v.a. dezentralen QM (Studienkommissionen der Fachbereiche). Darüber hinaus finden Versammlungen der Institute bzw. Studiengänge mit ihren Studierenden statt. Auch dort werden Themen der Qualität von Studium und Lehre diskutiert und es wird nach Möglichkeiten zur Weiterentwicklung gesucht. Und schließlich dient der direkte Kontakt zwischen Lehrenden und Studierenden der Bearbeitung von Qualitätsthemen.

Damit diese Orte tatsächlich zur Diskussion und Reflexion genutzt werden, bedarf es darüber hinaus unterschiedlicher Anlässe der Reflexion. Die Diskussionen können sich aus der Bearbeitung von Evaluationsergebnissen ergeben, sie können aber auch spontan erfolgen oder im Rahmen institutionalisierter Gremien und Veranstaltungsformen stattfinden und beispielsweise durch die Gestaltung der Tagesordnung durch einen festen Tagesordnungspunkt gelenkt werden. Häufig werden auch Akkreditierungs- und Reakkreditierungsverfahren zum Anlass genommen, Studienreformprozesse in Gang zu bringen.

Diskussion und Reflexion von Evaluationsergebnissen

Damit es gelingt, Evaluationsergebnisse sowohl als Grundlage für Diskussions- und Reflexionsprozesse zu nutzen als auch daraus Maßnahmen zur Verbesserung abzuleiten und umzusetzen, wird an der Universität Hildesheim großer Wert darauf gelegt sie denjenigen Akteuren bereitzustellen, die als Lehrende, als Entscheidungsträger in den Gremien oder als Serviceeinrichtungen einen Beitrag zur Verbesserung leisten können. Durch die Vernetzung der Gremien des zentralen und dezentralen QM wird sichergestellt, dass Reflexionsprozesse angestoßen werden.

Für ein derartiges Follow-up wurden an der Universität Hildesheim unterschiedliche Prozesse definiert:

- Die Ergebnisse der *Lehrveranstaltungsevaluation* gehen den Lehrenden spätestens zwei Wochen nach der Befragung zu. Sie sind dann aufgefordert, mit den Studierenden darüber ins Gespräch zu kommen. Da Lehrende sehr unterschiedlich mit den Ergebnissen umgehen und insbesondere diejenigen einer Auseinandersetzung aus dem Weg gehen, die besonders in der Kritik stehen, kommt einer aktiven, Diskussionen einfordernenden Haltung der Studierenden hohe Bedeutung zu.
- Die *Studiengangsevaluation* und die *Evaluation der Studienbedingungen* liefern Anhaltspunkte, was auf übergeordneter Ebene der Studieninhalte, Studienorganisation und Studienbedingungen verbessert wer-

den kann. Dadurch, dass keine personenbezogenen Daten enthalten sind, wird es möglich, die Ergebnisse in den Gremien zu thematisieren. Gesamtauswertungen werden der Senatskommission, dem Senat sowie der Hochschulleitung bereitgestellt und dort bezogen auf hochschulweite Aspekte diskutiert. Den Studienkommissionen stehen die Auswertungen für die sie betreffenden Studiengänge zur Verfügung. Das ist der Ort, an dem konkrete Verbesserungspotenziale diskutiert und Maßnahmen zur Weiterentwicklung der Studiengänge angestoßen werden können. Um eine Auseinandersetzung mit den Evaluationsergebnissen zu initiieren, werden die Fachbereiche aus der Senatskommission heraus aufgefordert, sich mit den Ergebnissen auseinanderzusetzen und in einer Stellungnahme die eingeleiteten Maßnahmen darzulegen. Für die Evaluation der Studienbedingungen werden die Serviceeinrichtungen der Universität aus der Senatskommission heraus mit der Ableitung von Handlungspotenzial und mit der Umsetzung von Verbesserungsmaßnahmen jeweils für die sie betreffenden Leistungsbereiche beauftragt. Im Sinne der Kontextsteuerung erfolgt somit auf übergeordneter Ebene ein Impuls, sich mit den Evaluationsergebnissen auseinanderzusetzen. Die inhaltliche Auseinandersetzung erfolgt dann vor Ort bezogen auf einzelne Fächer und Studiengänge.

Der eigentliche Mehrwert eines Evaluationsverfahrens besteht in dem, was nach der Befragung und Auswertung an Diskussionen, Reflexionen und Veränderungen angestoßen wird. Die Entwicklung von Evaluationsverfahren fokussiert im Alltag aber meist auf das Erhebungsinstrument und verwendet viel Zeit auf methodische Aspekte und die Aushandlung konkreter Items. Auch an der Universität Hildesheim wurde der Diskussion der Fragebögen viel Raum gewährt, wodurch der anschließende Prozess des Follow-up weniger Beachtung fand. Es erscheint sinnvoll, bereits frühzeitig auch darüber nachzudenken, in welcher Form die Daten ausgewertet werden sollen, wer welche Auswertungen zu welchem Zeitpunkt zur Verfügung gestellt bekommen soll und in welcher Form die ergriffenen Maßnahmen dokumentiert und kommuniziert werden. Denn daran misst sich am Ende nicht nur der Nutzen des QM, davon hängt auch ab, inwiefern sich Studierende überhaupt in zukünftige Qualitätsdiskussionen einbringen werden.

Schaffung offener Diskussionsforen

Einen wichtigen Meilenstein für das QM der Universität Hildesheim setzte ein Bologna-Tag im Januar 2010, bei dem einen Tag lang Studierende und Lehrende gemeinsam ohne die üblichen Seminarverpflichtungen über Möglichkeiten der Weiterentwicklung von Studium und Lehre in zahlreichen Workshops und einer abschließenden Vollversammlung diskutierten. Die im Rahmen des Bologna-Tages vereinbarten Handlungsfelder bildeten den Ausgangspunkt für umfangreiche Überarbeitungen der Studiengänge und eine kontinuierliche Berichterstattung in den QM-Gremien.

Der Bologna-Tag hat nicht nur geholfen, neben die standardisierten Verfahren der Evaluation ein v.a. für qualita-

tive Fragen offenes Diskussionsforum zu stellen, in dem ein konstruktiver Austausch zwischen Studierenden und Lehrenden entstehen kann. Er hat auch dazu beigetragen, Qualitätsaspekte weniger aus einer formalen, sondern vielmehr aus einer inhaltlichen Perspektive zu diskutieren.

5. Fazit

Ein pragmatischer Blick auf das Qualitätsmanagement an Hochschulen hat zur Identifikation wesentlicher Elemente bzw. Gestaltungsdimensionen geführt: Neben einer strategischen und strukturellen Verankerung geht es im QM um die Gewährleistung und Verbesserung von Qualität in unterschiedlichen Prozessen. Dazu werden einzelne Instrumente in einem Qualitätsregelkreis zusammengeführt. Ein Dokumentations- und Berichtssystem kann nicht nur unter Konformitätsgesichtspunkten dazu dienen, Prozesse und Verantwortlichkeiten zu definieren, sondern auch die Basis für Reflexions- und Lernprozesse zu schaffen.

Aus einer theoretischen Perspektive heraus erscheint es notwendig, die Implementierung von QM an Hochschulen einerseits an den Besonderheiten der Hochschule als Expertenorganisation und andererseits an der Komplexität des Qualitätsbegriffs zu orientieren. Für Prozesse der Qualitätssicherung und -entwicklung bedeutet das, ein differenziertes und aufeinander abgestimmtes Instrumentarium bereitzustellen, das den vielen Facetten von Qualität, den unterschiedlichen Interessengruppen und den teils widersprüchlichen Zielen begegnet. Für eine übergeordnete Systematik eines Qualitätsmanagements bedeutet es, das Steuerungsparadigma der Selbstorganisation in den Mittelpunkt zu stellen und die dezentrale Verantwortung für Qualität anzuerkennen, ohne diese im Sinne traditioneller akademischer Selbststeuerungsmechanismen allein sich selbst zu überlassen. Auf diese Weise werden Diskussions- und Reflexionsprozesse zu einem Kernelement des QM. Wie sich ein solcher konzeptioneller Bezugsrahmen auf die konkrete Implementierung an einer Hochschule auswirken kann, wurde am Beispiel der Universität Hildesheim nachgezeichnet.

Ein Fokus lag zunächst darauf, Evaluationsverfahren als Feedbackinstrumente zu etablieren, um die Aufmerksamkeit auf qualitätsrelevante Themen zu lenken sowie Lern- Veränderungsprozesse und anzuregen. Da Evaluationsverfahren kein objektives Qualitätsbild erzeugen, sondern immer nur einen Teilausschnitt der Qualitätsrealität repräsentieren können, wurde darauf geachtet, ein System sich ergänzender Verfahren bereitzustellen. Der Beteiligung der Betroffenen an der Entwicklung der Verfahren kam durch Aushandlungsprozesse auf unterschiedlichen Ebenen eine besondere Bedeutung zu.

Die Entwicklung von Evaluationsverfahren und die Etablierung von Gremien und Verantwortlichkeiten für Qualitätsbelange dienen dazu, Diskussions- und Reflexionsprozesse zu ermöglichen und zu fördern. Ein solcher Ansatz stieß insbesondere in Fachbereichen und Instituten mit einer ausgeprägt diskursiven Fächerkultur auf fruchtbaren Boden. Die Herausforderung in diesen Bereichen bestand weniger darin, Diskussionsprozesse

in Gang zu bringen als eher diejenigen Orte zu identifizieren und als solche zu erkennen, an denen eine Auseinandersetzung mit „Qualitätsthemen“ bereits stattfindet. Dabei musste teilweise erst gelernt werden, dass existierende Diskussionsprozesse der Weiterentwicklung von Studiengängen dienen ohne dass sie bisher als QM-Maßnahmen verstanden wurden. Das weit verbreitete Bild von QM als Dokumentation, Formalisierung und Kontrolle verstellt erst einmal den Blick für solch eine Betrachtungsweise.

Der in diesem Artikel beschriebene Qualitätsansatz vertraut zunächst auf das In-Gang-Kommen dezentraler Diskussions- und Reflexionsprozesse und verzichtet auf eine übermäßige zentrale Steuerung und Kontrolle. Qualitätsmanagement steht immer in einem solchen Spannungsverhältnis zwischen zentraler Steuerungsin-tention einerseits und dezentraler Selbststeuerung andererseits. Letzteres stärker zu betonen bedeutet zwar in Kauf zu nehmen, dass mehr Zeit für Aushandlungs- und Veränderungsprozesse verwendet wird und auch einmal Gefahr zu laufen, dass Evaluationsergebnisse nicht genutzt und Veränderungsprozesse nicht angestoßen werden. Dieser Ansatz hat vor dem Hintergrund der Hochschule als Expertenorganisation jedoch mehr Aussicht auf Akzeptanz und damit langfristig auf Erfolg. Indem Evaluationsergebnisse bereitgestellt werden, indem sich Studierende aktiv für ihre Belange einsetzen, indem die Hochschulleitung gezielt Projekte fördert und auf eine Verständigung auf allgemeine Vereinbarungen hinwirkt, kann das Projekt QM selbst an einem Ort wie der Expertenorganisation Hochschule gelingen.

Literaturverzeichnis

- Baecker, D. (2003): Organisation und Management, Frankfurt am Main.
- Barnett, R. (2000): Realizing the University in An Age of Supercomplexity, Philadelphia, Open University Press.
- Beck, N./Walgenbach, P. (2003): ISO 9000 and Formalization - How Organizational Contingencies Affect Organizational Responses to Institutional Forces, in: Schmalenbach Business Review, 55. Jg./Nr. 4, S. 293-320.
- Berthold, C. (2011): Als ob es einen Sinn machen würde... Strategisches Management an Hochschulen, Gütersloh.
- Brunsson, N. (1991): The Organization of Hypocrisy. Talk, Decisions and Actions in Organizations, New York, John Wiley & Sons.
- Cohen, M.D./March, J.G./Olsen, J.P. (1972): A Garbage Can Model of Organizational Choice, in: Administrative Science Quarterly, 17. Jg./Nr. 1, S. 1-25.
- Ditzel, B. (2011): Evaluationsverfahren und ihre Rolle im Qualitätsmanagement: am Beispiel der Stiftung Universität Hildesheim, in: Kopinke, L./Materlik, D. (Hg.): Blubbsoft-Anwendertagung am 8. April 2011, Berlin.
- Ditzel, B. (2013a): Der prozessorientierte Ansatz an Hochschulen: eine organisationstheoretische Betrachtung, in: Zeitschrift für Hochschulentwicklung, 8. Jg./Nr. 2, S. 110-124.
- Ditzel, B. (2013b): Prozessqualität an der Universität Hildesheim: Erfahrungen mit dem prozessorientierten Ansatz im Bereich Studium und Lehre, in: Qualität in der Wissenschaft, 7. Jg./Nr. 3, S. 59-67.
- Ditzel, B./Bergt, T. (2013): Studentische Partizipation als organisationale Herausforderung - Ergebnisse einer explorativen Studie, in: Weber, S.M. et al. (Hg.): Organisation und Partizipation, Wiesbaden. S. 177-186.
- Habersam, M. (2000): Die Idee des Universitätscontrolling. Qualitätsentwicklung durch Evaluation, in: Laske, S. et al. (Hg.): Qualitätsentwicklung in Universitäten. München. S. 151-174.
- Harris-Hümmert, S. (2011): Evaluating Evaluators. An Evaluation of Education in Germany, Wiesbaden.
- Hechler, D./Pasternack, P. (2012): Hochschulorganisationsanalyse zwischen Forschung und Beratung, die hochschule, Sonderband.
- Heublein, U. et al. (2012): Maschinenhaus - Campus für Ingenieure. Zusammenfassender Bericht zur Qualitätssicherung im Maschinenbau- und Elektrotechnikstudium, Frankfurt/Main.
- HRK (2006): Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung, Beiträge zur Hochschulpolitik, Bd. 1/2006, Teil I, Bonn.
- Huber, M. (2012): Die Organisation Universität, in: Apelt, M./Tacke, V. (Hg.): Handbuch Organisationstypen, Wiesbaden. S. 239-252.
- Kloke, K./Krücken, G. (2012): „Der Ball muss dezentral gefangen werden.“ - Organisationssoziologische Überlegungen zu den Möglichkeiten und Grenzen hochschulinterner Steuerungsprozesse am Beispiel der Qualitätssicherung in der Lehre, in: Wilkesmann, U./Schmid, C.J. (Hg.): Hochschule als Organisation, Wiesbaden. S. 311-324.
- Kromrey, H. (2004): Qualität und Evaluation im System Hochschule, in: Stockmann, R. (Hg.): Evaluationsforschung, Opladen. S. 233-258.
- Krücken, G. (2008): Lässt sich Wissenschaft managen? in: Wissenschaftsrecht, 41. Jg./Nr. 4, S. 345-358.
- Krücken, G. (2011): Soziologische Zugänge zur Hochschulforschung, in: die hochschule, 20. Jg./Nr. 2, S. 102-116.
- Kühl, S. (2011): Organisationen. Eine sehr kurze Einführung, Wiesbaden.
- Loukkola, T./Zhang, T. (2010): Examining Quality Culture: Part 1 - Quality Assurance Processes in Higher Education Institutions, Brussels, EUA.
- Luhmann, N. (1973): Zweckbegriff und Systemrationalität: Über die Funktion von Zwecken in sozialen Systemen, Frankfurt/Main.
- Luhmann, N. (2006): Organisation und Entscheidung, 2. Aufl., Wiesbaden.
- Lüthje, J. (2004): Qualitätsmanagement an Hochschulen, in: HRK (Hg.): Evaluation - ein Bestandteil des Qualitätsmanagements an Hochschulen, Beiträge zur Hochschulpolitik, Bd. 9, Bonn. S. 5-6.
- Malik, F. (2008): Strategie des Managements komplexer Systeme. Ein Beitrag zur Management-Kybernetik evolutionärer Systeme, Bern, Haupt.
- March, J. G. (1978): Bounded Rationality, Ambiguity, and the Engineering of Choice, in: The Bell Journal of Economics, 9. Jg./Nr. 2, S. 587-608.
- Meister-Scheytt, C./Scheytt, T. (2005): The Complexity of Change in Universities, in: Higher Education Quarterly, 59. Jg./Nr. 1, S. 76-99.
- Messner, M.T./Scheytt, T./Becker, A. (2007): Messen und Managen: Controlling und die (Un-)Berechenbarkeit des Managements, in: Mennicken, A./Vollmer, H. (Hg.): Zahlenwerk. Kalkulation, Organisation und Gesellschaft. S. 87-104.
- Meyer, J.W./Rowan, B. (1977): Institutionalized Organizations: Formal Structure as Myth and Ceremony, in: American Journal of Sociology, 83. Jg./Nr. 2, S. 340-363.
- Mintzberg, H. (1983): Structure in Fives: Designing Effective Organizations, Englewood Cliffs/New Jersey: Prentice Hall International.
- Mittag, S./Bornmann, L./Daniel, H.-D. (2003): Evaluation von Studium und Lehre an Hochschulen. Handbuch zur Durchführung mehrstufiger Evaluationsverfahren, Münster.
- Musselin, C. (2007): Are Universities Specific Organisations? in: Krücken, G./Kosmützky, A./Torka, M. (Hg.): Towards a multiversity? Universities between global trends and national traditions, Bielefeld. S. 63-84.
- Nickel, S. (2007): Institutionelle QM-Systeme in Universitäten und Fachhochschulen. Konzepte - Instrumente - Umsetzung. Eine empirische Studie, Arbeitspapier Nr. 94, Gütersloh.
- Nickel, S. (2008): Qualitätsmanagementsysteme an Universitäten und Fachhochschulen: Ein kritischer Überblick, in: Beiträge zur Hochschulforschung, 30. Jg./Nr. 1, S. 16-39.
- Orton, J.D./Weick, K.E. (1990): Loosely Coupled Systems: A Reconceptualization, in: The Academy of Management Review, 15. Jg./Nr. 2, S. 203-223.
- Pasternack, P. (2004): Qualitätsorientierung an Hochschulen. Verfahren und Instrumente, HoF-Arbeitsberichte 5/04, Wittenberg.
- Pasternack, P./Hölscher, M. (2008): Mehr Qualität wagen, in: Qualität in der Wissenschaft, 2. Jg./Nr. 1, S. 19-25.
- Pellert, A. (1999): Die Universität als Organisation, Wien, Böhlau.
- Pellert, A. (2000): Expertenorganisationen reformieren, in: Hanft, A. (Hg.): Hochschulen managen? Zur Reformierbarkeit der Hochschulen nach Managementprinzipien, Neuwied. S. 39-55.
- Pellert, A. (2002): Hochschule und Qualität, in: Reil, T./Winter, M. (Hg.): Qualitätssicherung an Hochschulen: Theorie und Praxis, Bielefeld. S. 21-29.
- Scheytt, T. (2005): Potenziale der DIN EN ISO 9000 ff-Norm für die Qualitätsentwicklung in Hochschulen, in: HRK (Hg.): Qualität messen - Qualität managen. Leistungsparameter in der Hochschulentwicklung, Beiträge zur Hochschulpolitik, Bd. 6, Bonn. S. 136-142.
- Schmidt, B. (2010): Qualitätsentwicklung = Evaluation + Hochschuldidaktik? Von einer Rechnung, die nicht aufgeht, in: Pohlenz, P./Oppermann, A. (Hg.): Lehre und Studium professionell evaluieren, Bielefeld. S. 179-192.
- Simon, H. A. (1991): Bounded Rationality and Organizational Learning, in: Organization Science, 2. Jg./Nr. 1, S. 125-134.
- Steinmann, H./Schreyögg, G./Koch, J. (2013): Management. Grundlagen der Unternehmensführung, Wiesbaden.
- Weblar, W.-D. (2002): Vergleich von fünf wichtigen Evaluationsmodellen in Deutschland - Grundlagen einer Meta-Evaluation, in: Neumann, K./Osterloh J. (Hg.): Gute Lehre in der Vielfalt der Disziplinen, Weinheim.

- Weick, K.E. (1976): Educational Organizations as Loosely Coupled Systems, in: Administrative Science Quarterly, 21. Jg./Nr. 1, S. 1-19.
- Wilkesmann, U./Schmid, C.J. (2012): Organisationssoziologie der Hochschule versus Hochschul-Organisationssoziologie, in: Wilkesmann, U./Schmid, C.J. (Hg.): Hochschule als Organisation, Wiesbaden. S. 7-13.
- Willke, H. (1989): Controlling als Kontextsteuerung – zum Problem dezentralen Entscheidens in vernetzten Organisationen, in: Eschenbach, R. (Hg.): Supercontrolling: vernetzt denken, zielgerichtet entscheiden, Wien. S. 62-92.
- Willke, H. (2001): Systemtheorie III: Steuerungstheorie, Stuttgart.
- Willke, H. (2007): Einführung in das systemische Wissensmanagement, Heidelberg.

Winde, M. (2010): Von der Qualitätsmessung zum Qualitätsmanagement. Praxisbeispiele an Hochschulen, Essen.

■ **Benjamin Ditzel**, Dipl.-Ing., Qualitätsmanager der Fakultät Technik und Informatik, Hochschule für Angewandte Wissenschaften Hamburg, bis Ende 2011 Qualitätsmanager der Universität Hildesheim,
E-Mail: benjamin.ditzel@gmx.de

Annette Nauwerth, Ursula Walkenhorst, Renate von der Heyden, Simone Rechenbach (Hg.) Hochschuldidaktik in Übergängen - Eine forschende Perspektive

Übergänge in ein unbekanntes System sind mit Unsicherheiten und Lernbedarfen verbunden. Am Beispiel von Studiengängen für Gesundheitsfachberufe an der FH Bielefeld wird das Erleben der Studierenden und Absolventen in den Übergängen in das Studium bzw. in den Beruf dargestellt.

Die Erfahrungen wurden genutzt, spezifische Unterstützungsangebote zu entwickeln. Die entsprechenden Konzeptionen und Evaluationsergebnisse werden für die Einführungswochen mit den integrierten Tutorien und der Arbeit mit Portfolios beschrieben. Im Hinblick auf den Übergang in den Beruf werden ein Mentoring- sowie ein Berufseinsteiger -Programm diskutiert. Das vorliegende Buch beruht auf Ergebnissen eines Forschungsprojektes im Rahmen der BMBF Förderlinie „empirische Bildungsforschung“.

3-937026-79-7, Bielefeld 2013, 360 Seiten, 39.80 Euro

Erhältlich im Fachbuchhandel und direkt beim Verlag – selten im Versandbuchhandel (z.B. nicht bei Amazon).
Bestellung - E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22



Liebe Leserinnen und Leser,

nicht nur in dieser lesenden Eigenschaft (und natürlich für künftige Abonnements) sind Sie uns willkommen.

Wir begrüßen Sie im Spektrum von Forschungs- bis Erfahrungsberichten auch gerne als Autor/in. Wenn das Konzept der „Qualität in der Wissenschaft“ Sie anspricht - wovon wir natürlich überzeugt sind - dann freuen wir uns über Beiträge von Ihnen in den ständigen Sparten

- Qualitätsforschung,
- Qualitätsentwicklung/-politik,
- Anregungen für die Praxis/Erfahrungsberichte, aber ebenso
- Rezensionen,
- Tagungsberichte,
- Interviews.

Die Hinweise für Autorinnen und Autoren finden Sie unter: www.universitaetsverlagwebler.de.

Hauptbeiträge der aktuellen Hefte Fo, HSW, HM, P-OE und ZBS

Auf unserer Homepage www.universitaetsverlagwebler.de erhalten Sie Einblick in das Editorial und Inhaltsverzeichnis aller bisher erschienenen Ausgaben.

Fo

Forschung

Politik - Strategie - Management

Fo 1+2/2014
Europäische Wissenschaftspolitik
(Vorschau, u.a.):

Reinhard F. Hüttl
Innovationspolitische Leitthemen in der neuen Legislaturperiode – Die Schnittstellen von Politik- und Wissenschaftssystem

Marc-Denis Weitze
Frühzeitiger Dialog statt nachträglicher Akzeptanzbeschaffung: Perspektiven der Technikkommunikation

Thomas Lange/Mirco Kaesberg:
„No Innovation without Education“ – MINT-Bildung ist der kritische Erfolgsfaktor des Innovationsstandorts Deutschland

Christoph Egle/Christian Kobsda/Patrick Pfister
Politikberatung im Multi-Stakeholder-Format – Der Innovationsdialog zwischen Bundesregierung, Wirtschaft und Wissenschaft

Sicco Lehmann-Brauns
Eine neue Ordnung des Wissens und der Aufstieg der Wissenswirtschaft – Herausforderungen und Chancen für den FuE-Standort Deutschland

Oliver Pfirrmann
Schumpeters Erben – wie aus technologischen Entwicklungen eine neue Gründungsdynamik in Deutschland entstehen kann

Henrike von Lyncker/Ralf Behn
Der Beitrag der Wissenschaft im Kontext der Energiewende: ESYS und Forschungsforum Energiewende

Martin Carrier
Wahrheitsfindung unter Zeitdruck. Auswirkungen der Beschleunigung in der Wissenschaft

Peter Biegelbauer & Thomas Palfinger
Verschiedene Verfahren der Auswahl von Forschungsprojekten: Ein Vergleich von neun angewandten Forschungsförderungsorganisationen

Thomas Stehnken
Innovationspolitik im Spannungsgefüge des europäischen Mehrebenensystems

HSW

Das Hochschulwesen

Forum für Hochschulforschung, -praxis und -politik

HSW 3/2014
Innovative Konzepte und Ideen in der Mathematik-Lehre

Anke Hanft & Stefanie Kretschmer
Studiengestaltung und -organisation für heterogene Studierende

Leonhard Riedl, Daniel Rost & Erwin Schörner
Die Übergangsproblematik von der Schule zur Hochschule im Fach Mathematik am Beispiel der Lehramtsausbildung einer Universität

Ingrid Ahrenholtz & Andrea Ruf
Akzeptanz und Erfolg von zusätzlichen Maßnahmen in der Studieneingangsphase in Studiengängen der Mathematik und Naturwissenschaften

Christian Hochmuth
Eine Analyse des Konfliktumfeldes Hochschule

Jörn Schnieder
Mathematiklernen lernen mit philosophischen Denkmethoden – ein Werkstattbericht

HM

Hochschulmanagement

Zeitschrift für die Leitung, Entwicklung und Selbstverwaltung von Hochschulen und Wissenschaftseinrichtungen

HM 1/2014
Die Hochschullandschaft wird (noch) vielfältiger
(Vorschau)

Susanne in der Smitten & Michael Jäger
Ziel- und Leistungsvereinbarungen in der Hochschulfinanzierung

Roland Königsgruber
Organisatorische Innovationen im europäischen Hochschulsektor. Niederländische Liberal Arts Colleges und französische multi-Campus Hochschulen

Klaus Palandt
Die Entwicklung der privaten und kirchlichen Hochschulen; wann erklärt der WR Hochschulen zu Fachhochschulen, wann zu Universitäten?

Wolff-Dietrich Webler
Anmerkungen zur Seniorprofessur – Konzepte von Hochschulen zum Umgang mit Emeriti bzw. pensionierten ProfessorInnen

P-OE

Personal- und Organisationsentwicklung

in Einrichtungen der Lehre und Forschung

Ein Forum für Führungskräfte, Moderatoren, Trainer, Programm-Organisatoren

POE 1+2/2014
(Vorschau, u.a.)

Martin Mehrrens

Förderung mit Perspektive und Organisationsbezug Personal- und Organisationsentwicklung zwischen aktiver Positionierung und systemischer Bescheidenheit

Jael Fuck & Ute Symanski

Kongress Personalentwicklung an Hochschulen – ein externer Blickwinkel

Hildegard Guderian

Funktionen als Mittel des Organisationsmanagements

Cornelia Ruppert

Netzwerk „PE-NRW“ gegründet

Susanne Schulz

Keine OE ohne PE – ganzheitliche Personalentwicklung als Schlüsselgröße für die systemische Organisationsentwicklung an Universitäten

Franziska Schmidt & Tobias Seidl

Gründe Lehrender zur Teilnahme an hochschuldidaktischen Weiterbildungsveranstaltungen

Judith Hoffmann, Frank Meier & Martin Schultze

Potenziale und Handlungsempfehlungen für die hochschuldidaktische Weiterbildung

Stefan Schohl & Kristin Unnold

Der Beitrag des Gesundheitsmanagements an der Universität Bielefeld zum Inplacement

ZBS

Zeitschrift für Beratung und Studium

Handlungsfelder, Praxisbeispiele und Lösungskonzepte

ZBS 2/2014

Der Weg an die Hochschule

Beratungsentwicklung/-politik

Ulf Bade

Ideen zu einem besseren Zulassungswesen an deutschen Hochschulen

Kurt Bunke

Das Integrierte Portal für den Hochschulzugang (IPH) – künftige Hilfe in einem komplexen Entscheidungsprozess?

Jessica Heibült & Moritz Müller

Der dritte Bildungsweg an die Universität – Übergangserfahrungen von beruflich qualifizierten Studierenden

Maik Gattermann-Kasper

Menschen mit Beeinträchtigungen beim Übergang an die Hochschule – Anforderungen an Informations- und Beratungsangebote

Matthias Lehmann

Wissen, was kommt?! – Empirische Befunde zu Beratungs- und Betreuungsangeboten an Musikhochschulen

Empirie und Praxis

Annett Kupfer, Romy Simon

& Sandra Wesenberg
„Wohnen mit Kommilitonen“ – Ein Wohnkonzept zur Unterstützung von Studierenden

Martin Kattmann &

Sebastian Wieschowski
Hürden abbauen, Hochschulen öffnen: Vorarbeiten zu einer Bestandsaufnahme von Studienorientierungs- und -einführungsmaßnahmen an deutschen Universitäten



Für weitere Informationen

- zu unserem Zeitschriftenangebot,
- zum Abonnement einer Zeitschrift,
- zum Erwerb eines Einzelheftes,
- zum Erwerb eines anderen Verlagsproduktes,
- zur Einreichung eines Artikels,
- zu den Autorenhinweisen

oder sonstigen Fragen, besuchen Sie unsere Verlags-Homepage:

www.universitaetsverlagwebler.de

oder wenden Sie sich direkt an uns:

E-Mail:

info@universitaetsverlagwebler.de

Telefon:

0521/ 923 610-12

Fax:

0521/ 923 610-22

Postanschrift:

UniversitätsVerlagWebler
Bünder Straße 1-3
33613 Bielefeld

**Gutenberg Lehrkolleg der Johannes Gutenberg-Universität Mainz (Hg.):
GLK-Tagungsband
Teaching is Touching the Future – Emphasis on Skills**

Am 29. und 30. November 2012 veranstaltete das Gutenberg Lehrkolleg der Johannes Gutenberg-Universität Mainz die internationale Tagung "Teaching is Touching the Future – Emphasis on Skills".

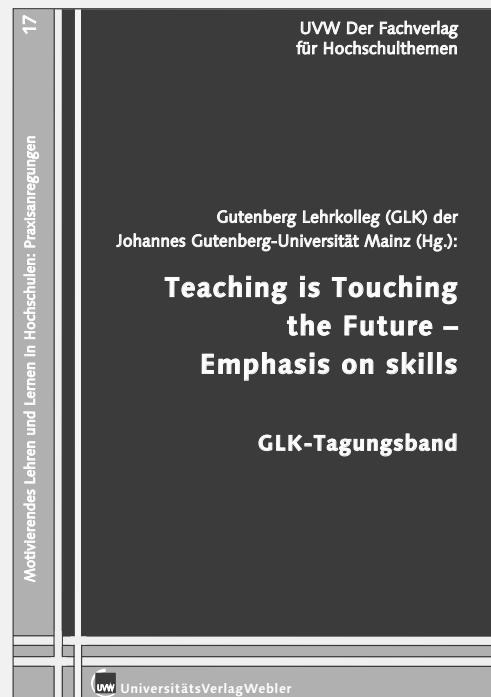
Im Rahmen dieser Tagung wurde die Neuorientierung der akademischen Lehr- und Lernformen an deutschen Hochschulen diskutiert, bei der die Lernerzentrierung in den Fokus rückt.

Mit Vorträgen und Postern wurden Forschungsergebnisse und Umsetzungsbeispiele zum "shift from teaching to learning" vorgestellt und fachspezifisch wie fachübergreifend erörtert.

Der vorliegende Sammelband beinhaltet die Tagungsbeiträge in schriftlicher Form. Zu Themen wie Kompetenzmessung/-modellierung, Kompetenzen der Lehrenden, Kompetenzorientiertes Prüfen oder Vermittlung von Schlüsselqualifikationen/überfachliche Kompetenzentwicklung werden verschiedene Ansätze einer Kompetenzorientierung im Kontext von Studien- und Lehrveranstaltungsplanung präsentiert.

Auch werden neue Herausforderungen deutlich, die sich durch die notwendige Abstimmung von Lernzielen, Lehr- und Lernmethoden sowie Prüfungsformen ergeben.

Bielefeld 2014, ISBN 13: 978-3-937026-85-5, 435 Seiten, 49.50 Euro



**Gutenberg Lehrkolleg der Johannes Gutenberg-Universität Mainz (Hg.):
Gute Lehre – von der Idee zur Realität
Innovative Lehrprojekte an der JGU**



Exzellenz in der Lehre ist ein Schlüsselfaktor, wenn es um die Attraktivität einer Hochschule geht. Steigende Studierendenzahlen und die Anforderungen der Wissensgesellschaft – gekennzeichnet durch Informationsflut, Globalisierung und Wettbewerb – bewirken einen Wandel an den Hochschulen und verlangen eine Neuorientierung in den Lehr- und Lernformen sowie eine Optimierung von Lernprozessen.

In diesem Sammelband werden innovative methodisch-didaktische Konzepte, die vom Gutenberg Lehrkolleg der Johannes Gutenberg-Universität Mainz gefördert wurden, vorgestellt, ihr Modellcharakter und ihre Wirkung für die Lehrpraxis evaluiert: von der Trainingsapotheke am Institut für Pharmazie und Biochemie über die Konzeption neuartiger E-Übungsaufgaben für mathematische Service-Lehrveranstaltungen bis hin zur Entwicklung eines Klang-Licht-Bootes für die Luminale 2012 in Frankfurt. So entsteht ein Überblick über die Vielfältigkeit kreativer Lehrideen sowie deren Nachhaltigkeit, Übertragbarkeit und Potential für hochschulweite Strukturveränderungen.

Bielefeld 2013, ISBN 13: 978-3-937026-86-2, 205 Seiten,
38.60 Euro zzgl. Versand

Erhältlich im Fachbuchhandel und direkt beim Verlag – selten im Versandbuchhandel (z.B. nicht bei Amazon).
Bestellung - E-Mail: info@universitaetsverlagwebler.de, Fax: 0521/ 923 610-22