
Annette Kluge, Kerstin Schüler (Hrsg.)

**Qualitätssicherung und -entwicklung
an Hochschulen:
Methoden und Ergebnisse**



PABST SCIENCE PUBLISHERS
Lengerich, Berlin, Bremen, Miami,
Riga, Viernheim, Wien, Zagreb

Beratung von Lehrenden auf der Grundlage studentischer Veranstaltungsbeurteilungen

Markus Dresel, Heiner Rindermann und Karen Tinsner

Zusammenfassung: Mit der Evaluation von Lehrveranstaltungen wird häufig das Ziel verfolgt, Lehrqualität zu entwickeln. Indes zeigt sich, dass deren ausschließliche Messung keine und die einfache Ergebnisrückmeldung nur geringe Qualitätsverbesserungen nach sich ziehen. Viel versprechender erscheint die Beratung von Lehrenden auf der Grundlage der Evaluationsergebnisse. Bisherige Studien sprechen für die Wirksamkeit des Ansatzes – allerdings war bislang sowohl die externe Validität als auch die Übertragbarkeit der Befunde auf die Rahmenbedingungen deutschsprachiger Hochschulen fragwürdig. Vorgestellt werden zwei Studien, in denen die Beratungswirkungen unter Maßgabe einer hohen externen Validität im deutschsprachigen Raum überprüft wurden. Dazu wurden insbesondere langfristige Wirkungen und Transferwirkungen untersucht. Die Ergebnisse sprechen mit moderaten bis großen Effekten dafür, dass mit der Beratung von Lehrenden nachhaltige Verbesserungen der Lehrqualität erzielt werden können. Abschließend werden Praxisempfehlungen zu deren Umsetzung genannt.

Schlagwörter: Beratung von Lehrenden, studentische Beurteilungen der Lehrqualität, Lehrveranstaltungsevaluation

Abstract: As a rule, the main objective behind course evaluations is the development of teaching quality. Nevertheless, it has been shown, that exclusive measurement has no effect, and that simple written feedback on evaluation results has little effect on the quality of teaching. In the existing literature, counseling of instructors based on feedback derived from student ratings has shown to be a fruitful approach in enhancing teaching quality. However, these studies are questionable in terms of external validity and generalizability to conditions outside North America. This chapter describes two studies in greater detail which examine the effects of counseling under the requirement of a high external validity in the German language area. Particular attention was paid to examining long-term and transfer effects. The results point out, with moderate to large effects, that well founded counseling of instructors can give rise to lasting improvements in teaching quality. Finally, some practical suggestions are made for effective counseling.

Key Words: Consultation of Instructors, Student Ratings Feedback of Teaching Quality

Einleitung

Die Evaluation von Lehrveranstaltungen unter der Verwendung studentischer Beurteilungen ist ein in der deutschsprachigen Hochschullandschaft mittlerweile sehr verbreitetes Verfahren zur Qualitätsmessung in der Lehre. Teilweise eingebettet in umfassendere Konzepte zur Sicherung der Qualität von Lehre und Forschung an Hochschulen, soll die Lehrqualität nicht nur verlässlich eingeschätzt, sondern auch verbessert werden. Häufig wird der Qualitätsverbesserung formal größte Priorität zugeschrieben.

Im Hinblick auf das erste Ziel, die Bestimmung von Lehrqualität, liegt mittlerweile eine Fülle an Befunden vor, die starke Vorbehalte gegenüber der Zuverlässigkeit und Inhaltsgültigkeit studentischer Veranstaltungsbeurteilungen weitgehend entkräften (Überblicke bei Marsh & Roche, 1997; Rindermann, 2001). Auch wenn die Debatte noch anhält und insbesondere die Rolle verzerrender Einflüsse von Drittvariablen nicht abschließend geklärt ist, kann davon ausgegangen werden, dass mit studentischen Beurteilungen eine hinreichend valide Einschätzung der Lehrqualität gewonnen werden kann.

Dass die Evaluation von Lehrveranstaltungen positive Effekte auf die Lehrqualität hat, das zweite mit ihr verfolgte Ziel, ist indes deutlich fragwürdiger. In der Praxis der Lehrveranstaltungsevaluation wird häufig davon ausgegangen, dass die Rückmeldung der Evaluationsergebnisse quasi automatisch zu den erwünschten Verbesserungen in der Lehrqualität führt, also ohne dass dazu weitere Maßnahmen ergriffen werden müssten. Ergebnisse von Wirksamkeitsstudien, in denen das „Feedbackmodell“ untersucht wurde, erbrachten entweder einen nicht statistisch nachweisbaren oder höchstens einen kleinen Effekt des Feedbacks auf die Lehrqualität (zusammenfassend Menges & Brinko, 1986; Rindermann, 2001). Die Metaanalyse von Menges und Brinko (1986), in die 23 Untersuchungen zur Wirkung der Rückmeldung studentischer Veranstaltungsbeurteilungen eingingen, erbrachte als durchschnittlichen Effekt lediglich $d=0.22$. Durch eine zeitferne Ergebnismrückmeldung erst nach Abschluss der evaluierten Lehrveranstaltungen (wodurch Transferprobleme entstehen können) und durch Rückmeldungen, die für statistische Laien nur schwer interpretierbar sind, dürften die Effekte des Feedbacks noch weiter reduziert werden (vgl. Dresel, Tinsner & Schröck-Pauli, 2006). Noch fragwürdiger ist die Annahme, dass allein schon die Erfassung der Lehrqualität (d. h. ohne deren Ergebnisse zurückzumelden) zu einer Sensibilisierung und damit zu einer Verbesserung der Lehrqualität führt – hier resultierten Effektgrößen um $d=0$ (Rindermann, 2001). Webler (1992) schlug zur Verbesserung der Wirksamkeit das „hochschuldidaktische Diskursmodell“ vor, nach dem Dozent(inn)en die Evaluationsergebnisse in der betreffenden Lehrveranstaltung vorstellen und mit den Teilnehmer(inne)n diskutieren. Aber auch dieses Modell erwies sich als kaum geeignet, substanzielle Verbesserungen in der Lehrqualität zu erzielen, was wohl auch daran liegt, dass es häufig nur von den positiv bewerteten Dozent(inn)en akzeptiert und umgesetzt wird (Rindermann, 2001). Betrachtet man die referierten Befunde zusammenfassend, steht zu befürchten, dass Hochschulen das mit der Lehrveranstaltungsevaluation angestrebte Ziel der Qualitätsverbesserung wohl nur in geringem Maße erreichen, sofern sie keine Maßnahmen in ihr Evaluationskonzept integrieren, die über die einfache Rückmeldung von Evaluationsergebnissen und deren veranstaltungsinterne Besprechung hinausgehen.

Beratungsansatz

Eine naheliegende und häufig befürwortete Maßnahme ist die Beratung von Lehrenden auf der Grundlage der studentischen Veranstaltungsbeurteilungen (Brinko & Menges, 1997; Dresel & Rindermann, 2006; Knapper & Piccinin, 1999; Penny & Coe, 2004; Rindermann & Kohler, 2003; Rindermann, Kohler & Meisenberg, 2006). Gegenüber dem herkömmlichen Evaluationsfeedback zielt der Beratungsansatz nicht nur auf die Rückmeldung von Evaluationsergebnissen, sondern explizit auch auf drei weitere Funktionen ab (vgl. Dresel & Rindermann, 2006):

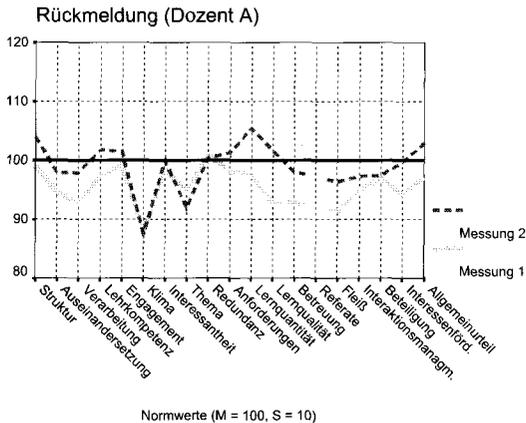
- *Verbesserung der Wahrnehmung von Diskrepanzen zwischen tatsächlichem und wünschenswertem Lehrhandeln.* Diese kann durch oberflächliche Registrierung, fehlerhafte Interpretation oder Abwertung der Ergebnisse reduziert sein.
- *Motivierung.* Diese kann aufgrund ungünstiger Ursachenerklärungen für die Evaluationsergebnisse, geringer Selbstwirksamkeitserwartungen oder eines geringen subjektiven Anreizes von Qualitätsverbesserungen erforderlich sein.
- *Vermittlung von spezifischem hochschuldidaktischem Wissen zur Verbesserung problematischer Aspekte des Lehrhandelns.* Dies kann allerdings keine hochschuldidaktischen Weiterbildung ersetzen, die möglicherweise indiziert ist.

Typische Elemente des konkreten Beratungsgesprächs sind die Informierung des bzw. der Lehrenden über die Evaluationsergebnisse, das Ausräumen möglicher Verständnisschwierigkeiten, die Aufklärung bei etwaigen Bedenken im Hinblick auf die Validität der Studierendenbeurteilungen, das Hinweisen auf bestehende Ist-Soll-Diskrepanzen, die Diskussion der Selbstsicht der betreffenden Lehrperson, die Diskussion von Ursachen für positive oder negative Evaluationsergebnisse, die Betonung von Anreizen verbesserten Lehrhandelns, die Erörterung von Möglichkeiten zur Verbesserung der Lehre, die Auswahl veränderbarer Aspekte des Lehrhandelns und die darauf aufbauende Erarbeitung konkreter Verbesserungsvorschläge, die Vereinbarung von Verbesserungszielen sowie die etwaige Empfehlung hochschuldidaktischer Weiterbildungsmaßnahmen.

In der Praxis werden Beratungsgespräche meist von einer fachlich ausgewiesenen Person („Expertenberatung“), manchmal aber auch von akademischen Kolleg(inn)en („kollegiale Beratung“) geführt. Beratungsgespräche können ausschließlich mit einer Lehrperson stattfinden, aber auch ergänzende Diskussionen in einer Gruppe mehrerer Dozent(inn)en beinhalten. Typischerweise dauern sie ein bis zwei Stunden, in manchen Konzepten auch länger. Ein Fallbeispiel ist in Kasten 1 dargestellt.

Kasten 1. Fallbeispiel einer schwierigen, aber erfolgreichen Beratung (nach Kohler & Rindermann, 2000)

Dozent A (Ergebnisse s. Abbildung) hatte sich für die Hilfestellung bei der Interpretation der Ergebnisse kaum interessiert. Er meinte, die Gründe für das offensichtlich eher negative Feedback bei den Studierenden gefunden zu haben: Sie hätten kein Interesse und wollten nicht mitarbeiten. In den offen erfassten Beurteilungen der Lehrveranstaltung wurde als Kritik durchweg sein absoluter Anspruch auf eine eigene Lehrmethodik genannt. Darauf entgegnete er, dass er nicht anders könne und von den Teilnehmer(inne)n erwarte, dass sie dies annähmen. Er rechtfertigte seine Methodik mit den Erfolgen, die er bei anderen Studierenden sonst hätte.



Der Berater versuchte, Erklärungen anzubieten, die einer einseitigen Schuldzuweisung entgegenwirken. Ein wesentlicher Grund schien darin zu liegen, dass die Studierenden einer anderen Institution freiwillig zu ihm kommen, während im vorliegenden Fall Teilnahme-pflicht bestand. Diese unterschiedlichen Ausgangsbedingungen verlangten womöglich unterschiedliche didaktische Herangehensweisen. Darauf fand Dozent A weitere Erklärungen, die diese angebotenen Thesen widerlegten. Der Berater hatte am Ende des Gesprächs den Eindruck, dass die negative Haltung gegenüber der Gruppe bestehen blieb.

Trotz des schwierigen Verlaufs der Beratung versuchte der Berater, Dozent A zu einer Wiederholungsmessung zu bewegen. Dabei erläuterte er nochmals eingehend das Ziel der Evaluation und verdeutlichte die Chance zur Optimierung des Unterrichts. Trotz Bedenken seitens des Lehrenden willigte dieser ein. In der Wiederholungsmessung zeigten sich verbesserte Evaluationsergebnisse, eine wichtige Dimension (soziales Klima) wurde aber ähnlich kritisch bewertet wie beim ersten Mal. Dozent A glaubte, dass die Verbesserungen nichts mit seinem Unterrichtsstil zu tun haben, sondern mit der Zusammensetzung der Gruppe. Die Studierenden, die ihn bei der Wiederholungsmessung bewertet hatten, hätten seinen Unterricht besser verstanden. Der Berater konnte dies entkräften und erarbeitete gemeinsam mit dem Dozenten Veränderungen, die dieser seit dem ersten Beratungsgespräch vorgenommen hatte und auf welche die verbesserten Bewertungen (insbesondere Lernquantität und -qualität) zurückgeführt werden konnten. Zum sehr negativ beurteilten Klima äußerte Dozent A, dass ihm der Umfang des Gelernten wichtiger sei als Beliebtheit. Er spräche das Unangenehme an und seine Didaktik sei „bohren“, „aufbrechen“ und die „Konfrontation“. Der Berater gab dem Lehrenden zu verstehen, dass er seine Methodik in ihrer Authentizität akzeptiere.

Befundlage zur Wirksamkeit des Beratungsansatzes

Experimentelle und quasi-experimentelle Studien im angloamerikanischen Raum

In einer Reihe von Studien im angloamerikanischen Raum erwies sich die Dozent(inn)enberatung der ausschließlichen Rückmeldung von Evaluationsergebnissen überlegen (Überblicke bei Penny & Coe, 2004; Rindermann, 2001). Beispielsweise untersuchte Wilson (1986) die Effekte der Beratung quasi-experimentell. Auf der Grundlage studentischer Beurteilungen in der Mitte des Semesters wurde eine Gruppe an Dozent(inn)en beraten und zwar im Hinblick auf Lehrtechniken, die bei besonders gut beurteilten Kollegen identifiziert wurden. Dozent(inn)en einer zweiten Gruppe erhielten eine schriftliche Ergebnisrückmeldung. Am Ende des Semesters waren verbesserte Bewertungen durch die Studierenden nur für die erste Gruppe nachweisbar. Im Rahmen einer ähnlichen Studie fokussierten Marsh und Roche (1993) in Beratungsgesprächen auf jene Lehrdimensionen, die bei den jeweiligen Dozent(inn)en den größten Entwicklungsbedarf indizierten. In der Studie, in der im Gegensatz zur Arbeit von Wilson (1986) eine randomisierte Versuchsgruppenzuweisung realisiert werden konnte, waren ebenfalls Verbesserungen gegenüber einer Kontrollgruppe nachweisbar, insbesondere auf den (individuellen) Zieldimensionen des Beratungsgesprächs. In einer jüngeren Studie konnten Hampton und Reiser (2004) unter Anwendung eines vergleichbaren experimentellen Designs Beratungseffekte auch für studentische Tutoren nachweisen.

Penny und Coe (2004) unterzogen 12 experimentelle Studien zu den Effekten der Beratung von Lehrenden einer Metaanalyse und fanden eine mittlere Effektgröße von $d=0.69$. Danach übersteigt die Wirksamkeit des Beratungsansatzes die – ebenfalls metaanalytisch bestimmte – durchschnittliche Wirkung der ausschließlichen Ergebnisrückmeldung deutlich (vgl. Menges & Brinko, 1986).

Einschränkungen der experimentellen und quasi-experimentellen Studien

Der Beratungsansatz erscheint vor dem Hintergrund der geschilderten Befundlage als viel versprechend. Allerdings muss bei deren Bewertung in Rechnung gestellt werden, dass die (quasi-)experimentellen Studien zwar eine hohe interne Validität, aber nur eine eingeschränkte externe Validität aufweisen. Die Einschränkungen ergeben sich vor allem aus der Außerachtlassung von langfristigen Wirkungen und Transferwirkungen auf andere als der Beratung zu Grunde liegenden Lehrveranstaltungen. Dies ergibt sich aus dem verbreiteten Untersuchungsparadigma, bei dem die Beratung auf der Grundlage einer Studierendenbefragung in der Mitte des Semesters durchgeführt wird und deren Wirkungen in einer zweiten Befragung am Ende des Semesters erfasst werden. Nach den von Hager und Hasselhorn (2000) vorgeschlagenen allgemeinen Qualitätsstandards an die Evaluation psychologischer Interventionsmaßnahmen sollten Maßnahmen erst dann als „effektiv“ bezeichnet werden, wenn langfristige Wirksamkeit und Transferwirkung nachgewiesen sind. Dieser Nachweis wurde in den bisherigen Untersuchungen nicht überzeugend erbracht. Insbesondere vom Transfer auf andere Veranstaltungen ist anzunehmen, dass er beispielsweise durch unterschiedliche Veranstaltungsthemen, -formen und -größen erschwert wird.

Speziell für den deutschsprachigen Raum stellt sich zusätzlich das Problem der fragwürdigen Generalisierbarkeit der im angloamerikanischen Raum erbrachten Evidenz. Dies folgt aus den unterschiedlichen Rahmenbedingungen der Lehre an

Hochschulen im deutschsprachigen Raum. Diese ist – im Gegensatz zu angloamerikanischen Institutionen – vielfach charakterisiert durch eine öffentliche Trägerschaft, einen hohen Anteil unkündbaren Lehrpersonals, höhere Lehrbelastung, gering entwickelte Sanktionierungs- bzw. Gratifikationssysteme, heterogene Zuständigkeiten, große Teilnehmerzahlen sowie einen höheren Stellenwert der Veranstaltungsform „Vorlesung“.

Im Folgenden stellen wir zwei eigene Studien eingehender vor, in denen wir die Wirkungen der Dozent(inn)enberatung unter Maßgabe einer möglichst hohen externen Validität an einem privaten Ausbildungsinstitut und an einer staatlichen Hochschule überprüften. Dabei betrachteten wir insbesondere langfristige Wirkungen und Wirkungen auf die Lehrqualität in anderen Veranstaltungen als jene, in Bezug auf die beraten wurde. Unserem Wissen nach sind diese beiden Studien die bisher einzigen, in denen die Wirkungen des Beratungsansatzes im deutschsprachigen Raum untersucht wurden.

Eine Untersuchung von längerfristigen Beratungswirkungen an einer privaten Bildungsinstitution im deutschsprachigen Raum

In der Studie von Rindermann und Kohler (2003) wurde ein Evaluations-Beratungs-Modell für ein privates Ausbildungsinstitut für Logopädie entwickelt und auf seine Effektivität hin überprüft. Die Auszubildenden an diesem Institut, die eine ähnliche Alters- und Qualifikationsstruktur aufwiesen wie Studierende an Hochschulen, zahlten einen Ausbildungsbeitrag (ca. 700 Euro pro Monat) und wurden im Quasiklassenverband in Kursen mit etwa 20 Teilnehmer(inne)n von befristet und nebenberuflich angestellten Berufspraktiker(inne)n unterrichtet.

In die Analyse gingen die Evaluationsergebnisse von 16 Dozent(inn)en ein, die in zwei aufeinander folgenden Semestern an der Evaluation ihrer Veranstaltungen teilnahmen (insgesamt 592 von den Auszubildenden ausgefüllte Fragebögen; thematisch ähnliche Veranstaltungen in beiden Semestern). Im Anschluss an den ersten Evaluationszeitpunkt fand ein etwa einstündiges Einzelgespräch zwischen dem Einrichtungsleiter, einem Diplom-Psychologen, und den jeweiligen Dozent(inn)en statt. Darin wurden das schriftliche Feedback erläutert (auch im Vergleich zu einer Selbst einschätzung der Lehrenden), Unterrichtsformen und einzelne didaktische Strategien besprochen und Verbesserungsmöglichkeiten erörtert.

Zur Evaluation der Lehrqualität der einzelnen Veranstaltungen kam das Heidelberger Inventar zur Lehrveranstaltungsevaluation II (*HILVE II*; Rindermann, 2001) in einer geringfügig adaptierten Fassung zum Einsatz. Mit diesem Fragebogen werden unter anderem studentischen Beurteilungen in Bezug auf mehrere *Dozenten(inn)enmerkmale* (Strukturierung des Stoffs, anschauliche Auseinandersetzung mit dem Thema, Anregung einer tiefenorientierten Verarbeitung des Stoffs, Lehrkompetenz, Engagement, soziales Klima, Betreuung außerhalb der Veranstaltung, Interaktionsmanagement) und mehrere Aspekte des *Lehrerfolgs* als Produkt der Lehrbemühungen (quantitativer Umfang des Gelernten, Sinnhaftigkeit des Gelernten, Förderung der Interessen am Fachgebiet und am Studium, Allgemeinbeurteilung der Veranstaltung) erfasst. Verbesserungen auf diesen Dimensionen stellten das Ziel der Beratung dar.

Die Dozent(inn)en erhielten im auf die Beratung folgenden Semester in den meisten der genannten Dimensionen verbesserte Beurteilungen, wobei die

Effektgrößen moderate bis große Effekte indizierten (Abbildung 1). Statistisch auf dem 5%-Niveau abgesichert werden konnte, dass die Lehrenden in der Sicht der Studierenden über umfangreichere Lehrkompetenzen verfügten, sich mehr für die Lehre engagierten, Veranstaltungen leiteten, in denen mehr gelernt wird und die insgesamt attraktiver sind. Ebenfalls auf dem 5%-Niveau signifikant war die Verbesserung beim zusammengefassten Wert für die Dozent(inn)enmerkmale. Aufgrund des geringen Stichprobenumfangs auf der Ebene der Lehrenden konnten einige weitere Effekte erst unter Anwendung des 10%-Niveaus statistisch abgesichert werden (vgl. Abbildung 1). Diese spiegeln nichtsdestotrotz wenigstens moderate Verbesserungen wider.

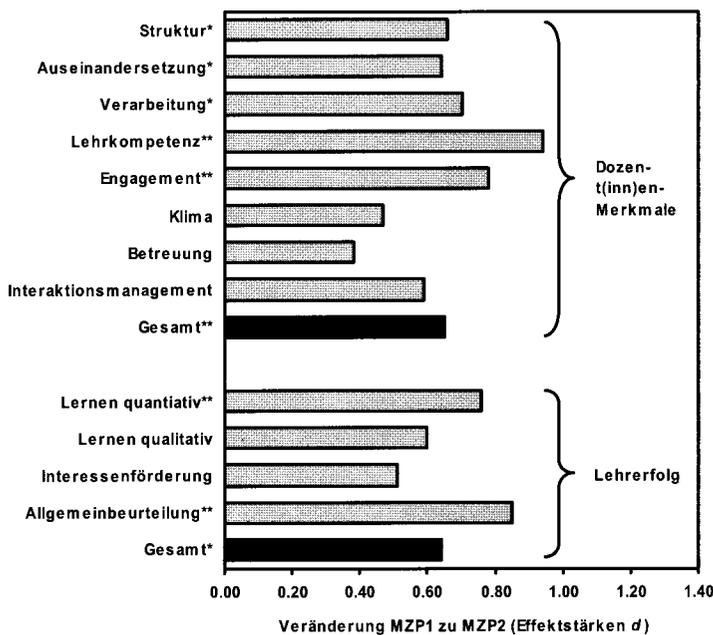


Abbildung 1. Ausgewählte Veränderungen bei beratenen Dozent(inn)en in der Studie von Rindermann und Kohler (2003). Die Abbildung basiert auf den in der Originalarbeit aufgeführten Zahlenwerten (* $p < .10$. ** $p < .05$)

Neben den genannten Effekten für die Gesamtstichprobe berichten Rindermann und Kohler (2003) Werte für jene Gruppe an Lehrenden, die nicht bereits zum ersten Evaluationszeitpunkt hervorragend bewertet wurden (Ausnahme von drei Lehrenden mit Deckeneffekten). Diese Gruppe repräsentiert die wichtigste Zielgruppe der Beratung, denn nur bei ihr können messbare Verbesserungen eintreten. Die Verbesserungen in dieser Gruppe waren im Vergleich zur Gesamtstichprobe erwartungsgemäß deutlich größer (Dozent(inn)enmerkmale gesamt: $d=0.96$. Lehrerfolg gesamt: $d=1.03$) und bis auf wenige Ausnahmen (Klima, Betreuung) auf dem 5%-Niveau statistisch absicherbar.

Zusammenfassend sprechen die Befunde mit der Betrachtung von Lehrveranstaltungen verschiedener Auszubildendengruppen in verschiedenen Semestern für einen wenigstens moderaten Effekt des Beratungsansatzes. Damit liefern sie ein Indiz für die Nachhaltigkeit der Wirkung der Dozent(inn)enberatung. Einschränkungen in Bezug auf die Transferwirkungen bestehen dahingehend, dass in die Studie größtenteils thematisch und strukturell gleiche Lehrveranstaltungen einbezogen wurden. Weiterhin gleicht das private Ausbildungsinstitut in mancherlei Hinsicht mehr den Bildungsinstitutionen, die in den angloamerikanischen Studien einbezogen wurden, als den überwiegend staatlichen Hochschulen im deutschsprachigen Raum: Die Lehrkräfte waren keine Professor(inn)en mit Lebenszeitstellen, die Auszubildenden hatten auch die Stellung von Kunden, die Veranstaltungen fanden in relativ kleinen Gruppen statt. Folglich kann auch aus den Befunden von Rindermann und Kohler (2003) nicht zweifelsfrei geschlossen werden, dass die Beratung unter den Rahmenbedingungen hiesiger staatlicher Hochschullehre effektiv ist.

Eine Untersuchung von langfristigen Wirkungen und Transferwirkungen der Beratung an einer staatlichen Hochschule im deutschsprachigen Raum

Um die Effektivität des Beratungsansatzes an staatlichen Hochschulen zu überprüfen, wurde in der Studie von Dresel und Rindermann (2006) an einer deutschen Fachhochschule das komplette Lehrangebot des Fachbereichs Architektur einer wiederholten Lehrveranstaltungsevaluation unterzogen (insgesamt 2.546 beantwortete Studierendenfragebögen). Zwölf Professor(inn)en, auf die 2.034 der Fragebögen entfielen, wurden anhand der Ergebnisse des ersten Evaluationszeitpunkts beraten. Mit einem Messwiederholungszeitraum von zwei Jahren und einer recht großen Variabilität im Hinblick auf die evaluierten Lehrveranstaltungen (im Durchschnitt leiteten die Dozent(inn)en nur 28% ihrer Lehrveranstaltungen wiederholt) wurde in dieser Studie der Fokus noch stärker auf langfristige Beratungseffekte und Transfer-effekte gelegt. Zur Erfassung der Lehrqualität kam wiederum das *HILVE II* zum Einsatz. Als abhängige Variable wurde der zusammengefasste Lehrerfolg betrachtet.

Die Beratung der Lehrenden fand im Rahmen eines Klausurtags statt. Zu Beginn nahmen die Professor(inn)en eine Selbsteinschätzung ihrer Lehrqualität vor. In einer sich anschließenden zweistündigen Arbeitseinheit wurden die Gesamtergebnisse des Fachbereichs vorgestellt und diskutiert. Der Schwerpunkt lag hier auf der Identifikation von generellen Stärken und Schwächen und möglichen Strategien zur Verbesserung der Lehrqualität. Es folgten 30 bis 45 Minuten lange Einzelberatungssitzungen, die von den beiden Autoren geleitet wurden. Hier wurde den Lehrenden ihre schriftliche Rückmeldung der individuellen Evaluationsergebnisse ausgehändigt und erläutert. Unter Fokussierung verhaltensnaher Prozessmerkmale der Lehre ließen sich individuelle Stärken und Schwächen identifizieren und eine Ursachenanalyse anregen. In Bezug auf vorab ausgewählte Aspekte wurden gemeinsam Strategien zur Verbesserung der Lehrqualität erarbeitet. Zum Abschluss der Einzelsitzungen wurden die Anreize verbesserter Lehrqualität fokussiert. Nach den Einzelberatungen fand im Plenum eine 60-minütige Abschlussdiskussion statt, in der Gelegenheit zum Austausch über die erarbeiteten Verbesserungsstrategien bestand.

Eine methodische Besonderheit dieser Studie gegenüber früheren Untersuchungen, die aufgrund des (im regulären Lehrbetrieb typischen) hohen Anteils variierender Lehrveranstaltungen erforderlich war, bestand in der Berücksichtigung potenzieller Biasvariablen, die einen verzerrenden Einfluss auf das Studierendenurteil

ausüben können. Dazu zählen Merkmale der Studierenden, die in unterschiedlichen Veranstaltungen unterschiedlich verteilt sein können (z. B. individuelles Vorwissen, Anstrengung), Veranstaltungsmerkmale (z. B. Veranstaltungsgröße, Vorwissensheterogenität) und Dozent(inn)enmerkmale (z. B. Alter, Geschlecht). Zur Berücksichtigung der auf unterschiedlichen Ebenen angesiedelten Biasvariablen fand ein Drei-Ebenen-Modell Verwendung, das mittels hierarchisch-linearer Modellierung analysiert wurde.

Die Analyse erbrachte einen statistisch auf dem 5%-Niveau absicherbaren Effekt des Evaluationszeitpunkts, der auf eine Verbesserung des Lehrerfolgs der Dozent(inn)en verweist. Je nach Berechnung der Effektgröße (die für die Mehrebenenanalyse von Veränderungen im vorliegenden Design noch nicht geklärt ist) resultierte eine moderate bis große Verbesserung des Lehrerfolgs im Bereich $d=0.68$ bis $d=0.91$.

Unter methodischen Gesichtspunkten zeigte sich zudem, dass die Außerachtlassung der verzerrenden Einflüsse von Biasvariablen zu Fehlern bei der Beurteilung der Veränderungen führen kann. So betrug die Korrelation der mit dem Dreiebenenmodell unter Berücksichtigung der Biasvariablen geschätzten individuellen Verbesserungen und der unkorrigierten, auf Lehrendenebene aggregierten Differenzwerte $r=.79$, was einer Fehlervarianz von 38% entspricht. In der Tat waren sowohl Dozent(inn)en identifizierbar, deren Verbesserungen durch die Verwendung unkorrigierter Differenzwerte stark überschätzt wird, als auch solche für die das Gegenteil galt.

Zusammenfassend konnten mit der Arbeit von Dresel und Rindermann (2006) die von Rindermann und Kohler (2003) gewonnenen Befunde repliziert werden. Dabei unterstreichen die Ergebnisse noch stärker, dass die Beratung von Dozent(inn)en sowohl zeitlich überdauernde Verbesserungen als auch Verbesserungen bei anderen Veranstaltungen als jenen, in Bezug auf die beraten wurde, nach sich zieht. Darüber hinaus wurden Indizien dafür erbracht, dass der Beratungsansatz auch im Kontext staatlich getragener Hochschullehre effektiv ist.

Zusammenfassende Beurteilung der Wirksamkeit des Beratungsansatzes

Werden die referierten Befunde zusammenfassend betrachtet, erscheinen die von uns unter der Maßgabe einer hohen externen Validität durchgeführten Untersuchungen als sinnvolle Ergänzung der eingangs vorgestellten (quasi-)experimentellen Studien. Einerseits weisen unsere Arbeiten einen deutlichen Zugewinn an externer Validität auf, da in den (quasi-)experimentellen Studien langfristige Wirkungen und Transferwirkungen weitgehend ausgeblendet wurden. Andererseits konnten wir in unseren Untersuchungen aufgrund der typischen Anforderungen in der Praxis des Lehrbetriebs keine Kontrollgruppe realisieren. Es war weder ethisch zu vertreten noch in den zuständigen Gremien durchzusetzen, dass eine zufällig ausgewählte Gruppe von Lehrenden über einen längeren Zeitraum zur Beratung verpflichtet wird, während sie einer anderen zufällig ausgewählten Gruppe vorenthalten wird. Damit weisen unsere Untersuchungen eine geringere interne Validität auf als die (quasi-)experimentellen Studien. Da Evaluationsforschung primär angewandte Forschung ist und damit die Bedingungen und Anforderungen des Anwendungskontexts nicht ausblenden kann, sondern vielmehr auf sie angewiesen ist, wenn sie praxisbezogene Aussagen generieren will, dürfte es vermutlich auch kaum möglich sein, alle wünschenswerten versuchsplanerischen Maßnahmen und gleichzeitig eine

hohe externe Validität zu realisieren. Das bekannte Dilemma zwischen interner und externer Validität erscheint hier verschärft. Auf Basis der vorliegenden Evidenz für die geringe Wirksamkeit anderer Ansätze (ausschließliche Messung, einfache Ergebnismeldung, veranstaltungsinterne Besprechung) kann dennoch angenommen werden, dass ein wesentlicher Anteil der von uns registrierten Verbesserungen Nettowirkungen der Beratung sind.

Bewertet man die nun vorliegende Befundlage insgesamt, erscheint die Beratung von Lehrenden auf der Grundlage studentischer Veranstaltungsbeurteilungen als effektive Maßnahme zur Verbesserung der Lehrkompetenzen und des Lehrhandelns. Hochschulen und anderen Bildungseinrichtungen steht damit ein Verfahren zur Verfügung, mit dem sie das Ziel der Qualitätsverbesserung in der Lehre effektiv erreichen können.

Praxisempfehlungen zur Umsetzung der Beratung von Lehrenden

Abschließend führen wir einige Empfehlungen zur praktischen Umsetzung der Beratung von Lehrenden auf der Grundlage studentischer Lehrveranstaltungsbeurteilungen an. Diese im Folgenden genannten Aspekte leiten sich sowohl aus unseren Erfahrungen in der Beratungspraxis und aus unseren Untersuchungen als auch aus der Metaanalyse von Penny und Coe (2004) ab.

- *Verwendung qualitativ hochwertiger Messinstrumente:* Die eingesetzten Studierendenfragebögen sollten theoretisch haltbar und psychometrisch überprüft sein. Sie sollten die angestrebten Produkte der Lehre (Lehrerfolg), deren Prozessmerkmale (didaktisches Handeln der Lehrenden) sowie relevante Studierendenmerkmale und Rahmenbedingungen abbilden. Nur wenn durch ein hochwertiges Messinstrument die Validität der Qualitätsschätzung gewährleistet ist, können in der Beratung sinnvolle Handlungsschritte erarbeitet werden. Empfehlenswert sind Instrumente, die handlungsnahe und damit der Veränderung gut zugängliche Aspekte abdecken.
- *Verwendung von Selbsteinschätzungen:* Deren Funktion liegt in der Aufdeckung von Diskrepanzen zwischen Fremd- und Selbstwahrnehmung, die zur Erarbeitung von Verbesserungsmöglichkeiten besonderen Anlass liefern. Am sinnvollsten wird dazu eine Dozent(inn)enversion des eingesetzten Studierendenfragebogens eingesetzt, mit der ein itemweiser Vergleich möglich ist.
- *Ergänzung um weitere Informationsquellen:* Die Beratung kann auf einer facettenreicheren Grundlage erfolgen, wenn die studentischen Beurteilungen um weitere Informationsquellen ergänzt werden. Neben den Selbsteinschätzungen kommen dafür vor allem Verfahren in Frage, mit denen der Prozess des Lehrens abgebildet wird (z. B. Videobeobachtung).
- *Modifikation ungünstiger Selbsteinschätzungen:* Nicht selten verfügen Lehrende über ungünstige Einschätzungen ihrer Lehrkompetenz. Dies kann im Rahmen der Ursachenanalyse erkannt werden, wenn sie problematische Ergebnisse auf personale Faktoren zurückführen, die sie als unveränderlich wahrnehmen (z. B. Temperament, rhetorische Fähigkeiten). Hier bietet es sich an, ungünstige Ursachenerklärungen durch motivationsförderliche Erklärungen zu ersetzen (z. B. Sprechtechnik) und aufzuzeigen, mit welchen Maßnahmen diese Faktoren

entwickelt werden können (z. B. Sprechtraining). Dabei kommen Methoden des Reattributionstrainings zum Einsatz, die sich bereits in anderen Anwendungskontexten bewährt haben (Dresel & Ziegler, 2006; Försterling & Stiensmeier-Pelster, 1994).

- *Formulierung von Verbesserungszielen*: Als effektiv hat es sich erwiesen, wenn einige wenige Aspekte der Lehre ausgewählt werden und sich die Dozent(inn)en dafür Verbesserungsziele setzen. Diese Ziele sollten so formuliert sein, dass sie sich auf spezifische Aspekte des Lehrhandelns richten, anspruchsvoll aber trotzdem erreichbar sind und einen zeitlichen Rahmen beinhalten.
- *Ermöglichung der Diskussion und des Erfahrungsaustauschs zwischen Dozent(inn)en*: Über die Expertenberatung in Eins-zu-Eins-Gesprächen hinaus erscheint die Ergänzung um Diskussionsmöglichkeiten in einer Gruppe von Lehrenden sinnvoll. Dies fördert nicht nur die Identifikation und Aneignung von effektiven didaktischen Handlungsstrategien, sondern auch ein Klima, in dem die Lehrqualität und die individuelle Entwicklung der Lehrkompetenzen aller wertgeschätzt werden. Eine Möglichkeit, Expertenberatung und Gruppendiskussionen zu verbinden, besteht darin, beides im Rahmen eines Klausurtags zur Lehrqualität zu implementieren wie in der Studie von Dresel und Rindermann (2006).

Literatur

- Brinko, K. T. & Menges, R. J. (Eds.) (1997). *Practically speaking: A sourcebook for instructional consultants in higher education*. Stillwater, OK: New Forum.
- Dresel, M. & Rindermann, H. (2006). *Beratung von Lehrenden auf der Grundlage von studentischen Beurteilungen der Lehrqualität: Mehrebenenanalytische Wirksamkeitsprüfung unter Berücksichtigung von Biasvariablen. Manuskript unter Begutachtung.*
- Dresel, M., Tinsner, K. & Schröck-Pauli, C. (2006). *Universitätsweite Lehrveranstaltungsevaluation an der Universität Ulm. Abschlussbericht der Arbeitsgruppe Lehrveranstaltungsevaluation*. Ulm: Universität Ulm.
- Dresel, M. & Ziegler, A. (2006). Langfristige Förderung von Fähigkeitsselbstkonzept und impliziter Fähigkeitstheorie durch computerbasiertes attributionales Feedback. *Zeitschrift für Pädagogische Psychologie*, 20, 49-63.
- Försterling, F. & Stiensmeier-Pelster, J. (Hrsg.) (1994). *Attributionstheorie. Grundlagen und Anwendungen*. Göttingen: Hogrefe.
- Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 41-85). Bern: Huber.
- Hampton, S. E. & Reiser, R. A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education*, 45, 497-527.
- Kohler, J. & Rindermann, H. (2000). Evaluation der Logopädieausbildung: Beratung der Dozenten und institutionelle Verankerung des Evaluierungsprozesses als Bedingungen für effektive Evaluationen. *L.O.G.O.S. interdisziplinär*, 8, 244-252.

- Knapper, C. & Piccinin, S. (Eds.) (1999). *Using consultants to improve teaching*. San Francisco: Jossey-Bass.
- Marsh, H. W. & Roche, L. A. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30, 217-251.
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Menges, R. J. & Brinko, K. T. (1986). *Effects of student evaluation feedback: A meta-analysis of Higher Education research*. San Francisco: Paper presented at the meeting of the American Educational Research Association.
- Penny, A. R. & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research*, 74, 215-253.
- Rindermann, H. (2001). *Leherevaluation – Einführung und Überblick zur Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen*. Landau: VEP.
- Rindermann, H. & Kohler, J. (2003). Lässt sich Lehrqualität durch Evaluation und Beratung verbessern? Überprüfung eines Evaluations-Beratungs-Modells. *Psychologie in Erziehung und Unterricht*, 50, 71-85.
- Rindermann, H., Kohler, J. & Meisenberg, G. (2006). Quality of instruction improved by evaluation and consultation of instructors. *International Journal for Academic Development*. Manuscript in press.
- Webler, W.-D. (1992). Evaluation der Lehre: Praxiserfahrungen und Methodenhinweise. In D. Grünh & H. Gattwinkel (Hrsg.), *Evaluation von Lehrveranstaltungen. Überfrachtung eines sinnvollen Instruments* (S. 143-161)? Berlin: FU-Dokumentationsreihe.
- Wilson, R. C. (1986). Improving faculty teaching. Effective use of student evaluations and consultants. *Journal of Higher Education*, 57, 196-211.