

EVALUIERUNG DER QUALITÄT DER LEHRE AUS DER SICHT DER STUDIERENDEN. EIN PILOTPROJEKT DER MEDIZINISCHEN FAKULTÄT TÜBINGEN¹

W. STANGL, H. MOELLER

Zusammenfassung

In dieser Studie wurden die Teilnehmer an 10 Pflichtpraktika nach der Hälfte („Intervention“) und am Ende des Semesters zur Qualität der Lehre aus ihrer Sicht befragt. Die Ergebnisse standen den Lehrenden zur Verfügung.

Die Studie hat aus der Sicht der Studierenden zu einer Verbesserung der Qualität der Lehre geführt. Dies ist vor allem an der Steigerung eines Gesamtscores (Qualitätsindex) ersichtlich, der aus verschiedenen Qualitätsparametern über alle Kurse und Befragungen hinweg gebildet wurde.

Dabei gab es bezüglich der einzelnen Kurse und Praktika wesentliche Unterschiede: in drei Kursen stieg der Score kontinuierlich an. Hier hat sich sehr wahrscheinlich nicht nur die Beurteilung durch die Studierenden, sondern auch die Qualität der Kurse kontinuierlich gebessert. In zwei weiteren Kursen gab es die gleiche Tendenz, dagegen fiel der Score in drei anderen Kursen zeitweise oder sogar kontinuierlich ab.

Auch hinsichtlich der Bewertung der Lehre der einzelnen Dozenten gab es erwartungsgemäß teilweise drastische Unterschiede. Hier bietet sich an, daß die Kliniken und Institute die Bewertungen der Studierenden als Grundlage für konstruktive Diskussionen über hochschuldidaktische Probleme heranziehen.

Das Interventionsprinzip als konzeptioneller Ansatz der Evaluation ist offensichtlich geeignet, die Qualität der Lehre allmählich zu verbessern. Den Lehrenden kann empfohlen werden, in Abständen von einigen Semestern Befragungen der Studierenden nach dem hier beschriebenen Beispiel durchzuführen.

Problematisch ist außer einer mit der Zeit auftretenden Befragungsübersättigung der Umstand, daß es im Bereich der Lehre keine Sollwerte oder Standards gibt, die zum Vergleich mit den eruierten Daten herangezogen werden können.

Evaluation of the teaching quality according to student opinion.

A pilot project of the Medical Faculty of Tübingen

Summary

In this study students of ten different obligatory courses were questioned about the teaching quality after half of („intervention“) and at the end of the semester. The results were at the teacher's disposal.

The students had the impression the study led to an improvement of the teaching quality. This becomes obvious by regarding the improvement of a general score (index of quality) which consisted of different criteria over all courses.

With regard to the different courses there appeared to be fundamental differences: in three courses the score rose continuously. It is quite probable that in these cases not only the opinion of the students but also the quality of the courses improved continuously. In two further courses the same tendency could be observed, whereas the score went down temporarily or even continuously in three other courses.

As expected there appeared to be substantial differences in the judgement of the individual teachers. It would be favorable for the clinics and the institutes to take the judgement of the students as a basis for constructive discussions about didactical problems of university teaching.

The intervention principle as a conceptual approach of evaluating is obviously suitable to improve the teaching quality step by step. It can be recommended to the lecturers to carry out evaluations of the students in intervals of several semesters as practiced in this study.

One of the problems we encountered in the study was that the students became tired of filling out questionnaires. Also, we had no standards of teaching quality to compare the compiled data with.

1 Einführung

1991 stellte der Minister für Wissenschaft und Forschung des Landes Baden-Württemberg Mittel für Projekte zur Verfügung, die der Verbesserung der Lehre an den Hochschulen dienen sollten. Zu den geförderten Projekten gehörte auch die vorliegende Studie. Unter der Federführung des Studiendekans war in Zusammenarbeit mit dem Institut für Medizinische Informationsverarbeitung und der Studienkommission der Medizini-

schen Fakultät Tübingen die Idee zu einer Evaluierung mit neuartiger Konzeption entstanden. Dabei ging man von folgenden Grundvorstellungen aus: 1. Studierende sollten befragt werden, 2. die Befragung sollte mit Fragebögen durchgeführt werden, 3. als konstruktive und vertrauensbildende Maßnahme sollte einer Befragung mit interventionellem Charakter der Vorzug gegeben werden. Das nun im weiteren Verlauf vorgestellte Befragungsprojekt wurde von Wintersemester 1992/93 bis Sommersemester 1994 durchgeführt.

¹ Diese Arbeit enthält Teile der Dissertation von W. Stangl

2 Methoden

2.1 Auswahl der Lehrveranstaltungen:

Die evaluierten Lehrveranstaltungen waren scheinpflichtige Veranstaltungen gleicher Struktur in verschiedenen Stoffgebieten: sie erstreckten sich über ein ganzes Semester, und die Studierenden waren in gleichbleibende Gruppen eingeteilt, die während der gesamten Zeit von denselben Dozenten betreut wurden. Dies waren im vorklinischen Teil des Studiums die Kurse der Mikroskopischen und der Makroskopischen Anatomie, im klinischen Teil des Studiums der Internistische Untersuchungskurs und der Kurs der Biomathematik (1. klinisches Semester), der Augenspiegelkurs und der Neurologische Untersuchungskurs (2. klinisches Semester), das Praktikum der Neurologie, das Praktikum der Ophthalmologie, der Kurs der Psychosomatik (4. klinisches Semester) und der Kurs der Speziellen Pharmakologie (5. klinisches Semester).

Da die evaluierten Lehrveranstaltungen anwesenheitspflichtig waren, konnten alle Studierenden der jeweiligen Semester befragt werden und nicht - wie bei Evaluierungen von Vorlesungen üblich - nur eine Auswahl. In vier Semestern wurden so circa 9.000 Befragungen durchgeführt.

Die Aufteilung der Lehrveranstaltungen in Kursgruppen begünstigte die persönliche Rückkopplung zwischen Lehrenden und Lernenden, zumal dieselbe Lehrkraft das ganze Semester über ihre Kursgruppe beibehielt.

2.2 Die Intervention:

Am Ende der ersten Semesterhälfte füllten die Studierenden einen Fragebogen zur Qualität der Lehrveranstaltungen aus ihrer Sicht aus. Die Dozenten erhielten die anonym ausgefüllten Fragebögen bis zum nächsten Kurstermin zur Einsicht, um die Beurteilung zu erfahren. Das Ergebnis der Befragung wurde eine Woche später von den Lehrenden mit ihren Gruppen besprochen. Die Intervention sollte so ein konstruktives Instrument darstellen, das zu einem vertrauensbildenden Dialog zwischen Lehrenden und Studierenden führen und zu einer gemeinsamen Verbesserung der Lehrveranstaltungen motivieren sollte.

2.3 Die Evaluation:

Sie bestand aus drei Teilen: der Intervention mit schriftlicher Befragung, der Diskussion eine Woche später (s.o.) und einer weiteren schriftlichen Befragung am Ende der Lehrveranstaltung. Bei der zweiten Befragung wurden die gleichen Fragen wie bei der Intervention gestellt, ergänzt durch zwei Fragen zur Wirksamkeit der Intervention und zur Akzeptanz der Studie bei den Studierenden.

Die Fragebögen wurden anonym ausgefüllt, jedoch so codiert, daß die Fragebögen der einzelnen Befragungen zusammengeführt werden konnten, um eine Auswer-

tung sowohl im Quer- als auch im Längsschnitt zu ermöglichen.

Die Fragebögen enthielten Fragen mit vorgegebenen Antwortmöglichkeiten (geschlossene Fragen) und die Aufforderung zu weiteren Anmerkungen (offene Fragen). Die Fragebögen wurden vom Studiendekan, Mitarbeitern des Institutes für Medizinische Informationsverarbeitung und Studierenden formuliert und von den einzelnen Abteilungen, die für die Praktika verantwortlich waren, genehmigt bzw. modifiziert. Abbildung 1 zeigt einen solchen Fragebogen.

Die ausgefüllten Fragebögen wurden datentechnisch von einer für diesen Zweck angeworbenen wissenschaftlichen Hilfskraft erfaßt und ausgewertet. Die Ergebnisse wurden in graphischer Form - Umsetzung der Auswertungsergebnisse in ein „gut-mittel-schlecht“-Schema und Darstellung mittels Balkendiagrammen - sowohl den einzelnen Kliniken und Instituten als auch dem Studiendekan vorgelegt.

Um eine effektive und vergleichende Auswertung der Frage nach der Qualität der Lehre gewährleisten zu können, wurden die Fragebogen-Vorderseiten für alle Kurse gleich gestaltet. So war es über alle evaluierten Veranstaltungen hinweg möglich, vergleichende Analysen durchzuführen. Die Fragebogen-Rückseiten hingegen wurden kursspezifisch ausgerichtet, um so auch eine individuelle Kursproblematik erkennen zu können. Die Auswertung der Wirksamkeit des Interventionsprinzips wurde durch die zweimalige Befragung pro Semester mit denselben Fragen ermöglicht.

Abgesehen von kursspezifischen Ergebnisdarstellungen, vergleichenden Darstellungen aller Kurse und Praktika und Darstellungen des gemeinsamen Ergebnisses aller Kurse und Praktika wurden ab dem zweiten Befragungssemester (SS 1993) zusätzlich anonymisierte, dozentenspezifische Ergebnispräsentationen erstellt, die den einzelnen Dozenten zugeschickt wurden: jeder Dozent erhielt eine anonyme Übersicht über alle Dozenten des betreffenden Kurses, die so codiert war, daß er nur seine eigenen Ergebnisse persönlich zuordnen konnte. Durch diese vergleichende, aber Namen nicht öffentlich nennende Dozentenkritik versprach man sich eine zusätzliche Motivation der beurteilten Lehrenden sowie einen Anstoß zur kritischen Selbstbetrachtung. Auch die Leiter der Kliniken und der Institute erhielten dozentenvergleichende Übersichten, allerdings ebenfalls ohne Aufschlüsselung der Namen. Abbildung 2 zeigt das Studiendesign noch einmal in Übersicht.

**Befragung der Studierenden nach der Qualität
der Lehrveranstaltungen aus ihrer Sicht**

Personen-Code*

Beispielfragebogen eines Untersuchungskurses SS 94 2. Befragung

1. Wieviel Studierende wurden von einem (einer) Dozenten(in) betreut?
2. War die Zahl der Studierenden pro Dozent(in) angemessen zu groß viel zu groß ?
3. Waren die Unterrichtsräume gut geeignet eingeschränkt geeignet schlecht geeignet ?
ggfs. : Warum? -----
4. Waren die Lehr- und Lernmittel (z. B. Augenspiegel, Skripte, Videos ...) ausreichend oder nicht ?
quantitativ : ausreichend nicht ausreichend --> warum nicht? -----
qualitativ : ausreichend nicht ausreichend --> warum nicht? -----
5. Ließ der Dozent (die Dozentin) sich vertreten ?
ja, mehr als 5 Mal 3-5 Mal 1-2 Mal nein, war immer selbst anwesend
6. War der Dozent (die Dozentin) pünktlich ?
immer meistens selten nie
7. War der Dozent (die Dozentin) auf die einzelnen Kurstage gut vorbereitet oder nicht ?
immer meistens selten nie
8. Wie war das Lernklima ?
motivierend teils / teils nicht motivierend
ggfs. : Warum? -----
9. Waren Sie auf die einzelnen Kurstage gut vorbereitet oder nicht ?
immer meistens selten nie
10. Haben Sie in diesem Kurs bisher genügend gelernt ?
ja naja nein --> warum nicht? -----
11. War das vorausgesetzte Fachwissen dem Wissensstand Ihres Semesters angemessen oder nicht ?
angemessen nicht angemessen --> warum nicht? -----

* Die Fragebögen der 1. und der 2. Befragung (Ende des Semesters) sollen für die einzelnen Studierenden so zusammengeführt werden können, daß die Anonymität gewahrt bleibt. Bitte geben Sie deshalb als Code den Anfangsbuchstaben des Vornamens und das Geburtsdatum Ihrer Mutter an. **bitte wenden**

12. Wieviel Studierende befragen bzw. untersuchen in der Regel einen (eine) Patienten (Patientin) zur gleichen Zeit ?
13. Wurden Sie bei Ihren praktischen Übungen konkret angeleitet ?
ja, völlig ausreichend ja, einigermaßen nicht ausreichend
14. Wurden Ihre Befunde überprüft und besprochen ?
ja, völlig ausreichend ja, einigermaßen nicht ausreichend
15. Hatten Sie ausreichend Gelegenheit, die Patienten (Patientinnen) mit unterschiedlichen Erkrankungen zu untersuchen, oder gab es dazu nicht ausreichend Gelegenheit ?
ausreichend nicht ausreichend <-- warum nicht ? _ _ _ _ _
_ _ _ _ _
16. War der Kurs für die Patienten (Patientinnen) zumutbar oder eher nicht ?
ja, ohne Bedenken ja, mit Bedenken eher nein
17. Wollen Sie die Erhebung von Normalbefunden mehr oder weniger als bisher durch die gegenseitige Untersuchungen von Studierenden lernen ?
mehr weniger so lassen
Warum ? _ _ _ _ _
_ _ _ _ _
18. Hat sich nach der ersten Befragung in der Lehrveranstaltung etwas geändert oder nicht?
ja nein
Falls ja, was? _ _ _ _ _
_ _ _ _ _
19. Halten Sie eine Befragung der Studierenden zur Beurteilung der Qualität von Lehrveranstaltungen in der vorliegenden Form für sinnvoll oder nicht?
ja nein
Falls nein, warum nicht? _ _ _ _ _
_ _ _ _ _
20. Wurden die Ergebnisse der ersten Befragung vom Dozenten im Kurs diskutiert oder nicht?
ja nein

Abbildung 1: Beispiel für einen Fragebogen (hier: Untersuchungskurs)

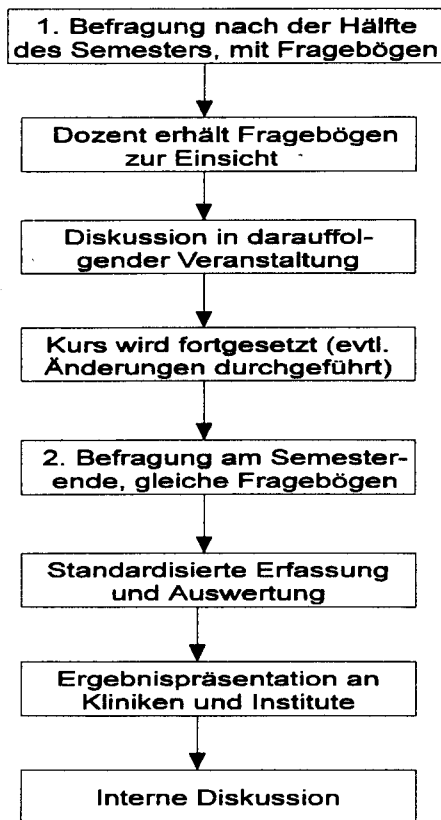


Abb. 2: Studiendesign

3 Fragestellung

Die Tübinger Studie wurde durchgeführt, um

- die generelle Machbarkeit eines solchen Evaluationsprojektes im Hinblick auf den technisch organisatorischen Aufwand, auf die Kooperation seitens der Lehrenden und der Lernenden, aber auch im Hinblick auf eine eventuelle Weiterempfehlung zu untersuchen,
- die Qualität einzelner Lehrveranstaltungen im Medizinstudium aus der Sicht der Studierenden beurteilen zu lassen,
- die Auswirkung einer Intervention nach Ablauf der ersten Hälfte der Lehrveranstaltungen auf den Kursablauf, gemessen an den Beurteilungen der befragten Studierenden, beurteilen zu können („direkte Intervention“) und
- die Auswirkung einer viersemestrigen Dauerevaluations („indirekte Intervention“) auf die Qualität der Lehre, wiederum gemessen an den Beurteilungen der Studierenden, zu eruieren.

4 Ergebnisse

Die Fragebögen enthielten Fragen nach der Qualität der strukturellen Rahmenbedingungen der Kurse (Angemessenheit der Kursgruppengröße, Räume, Lehrmittelquantität und -qualität, Vorbereitetheit der

Dozenten), hier Strukturqualität genannt, Fragen nach der Qualität der Durchführung der Kurse (Pünktlichkeit der Dozenten, Lernklima), hier Prozeßqualität genannt, und Fragen nach der Qualität des durch den Kurs Gelernten (Lernerfolg im Kurs), hier Ergebnisqualität genannt.

Aus allen drei Qualitätsdimensionen wurde ein Gesamtscore gebildet, der stellvertretend für die allgemeine Qualität der Lehre im entsprechenden Kurs stand (Qualitätsindex). Die Ergebnisse der kursspezifischen Fragen (Fragebogen-Rückseiten) wurden nicht in den Gesamtscore des jeweiligen Kurses aufgenommen: Erstens ließen sie sich untereinander nicht vergleichen (klinisch-praktische Kurse mit theoretischen wie Pharmakologie oder Biomathematik) und hätten somit die einheitliche Bewertung der Vorderseiten verzerrt. Zweitens wurden auf den Rückseiten bei einigen Kursen fast ausschließlich Fragen nach der Ergebnisqualität, in anderen hingegen nur Fragen nach der Qualität des kursprozessualen Geschehens gestellt - auch dies hätte die Wertigkeit des aus allen drei Qualitätsdimensionen gleichermaßen zusammengesetzten Gesamtscores verschoben. Drittens gab es bei vielen Fragen der Fragebogen-Rückseiten zu hohe missing-Raten - das ist die Quote der Befragten, die keine Angaben zur Frage machten -, um das Ergebnis werten zu können (als Obergrenze wurde 15% gesetzt). Die Ergebnisse der Fragen der Rückseiten dienten aber der Verdeutlichung der kursspezifischen Problematik und ihres Ausprägungsgrades.

In den zweiten Befragungen der Semester wurde zusätzlich gefragt, ob sich seit der ersten Befragung in der Lehrveranstaltung etwas geändert hätte (Auswirkung der Intervention) und ob die befragten Studierenden die Evaluation für sinnvoll erachteten (Akzeptanz).

4.1 Allgemeine Ergebnisse:

Rücklaufquoten: Da die ausgeteilten Fragebögen und die ausgefüllt zurückerhaltenen getrennt gezählt wurden, ließen sich für jede einzelne Befragung präzise Rücklaufquoten ermitteln. Durch das Procedere der Befragung - Austeilen der Fragebögen zu Beginn der Veranstaltung und sofortiges Wiedereinsammeln nach Ausfüllen - waren die Rücklaufquoten in der Studie zu keinem Zeitpunkt ein Problem und betragen in allen vier Semestern deutlich über 90%.

Änderungen im Kursverlauf: Ob zwischen erster und zweiter Befragung eine Änderung im Kursverlauf aufgetreten sei, bejahten im ersten Semester der Befragung im Durchschnitt 26,8% der Befragten, im zweiten Semester durchschnittlich 17,3%, im dritten Semester insgesamt 15,3% und im vierten Semester im Schnitt 11,7%.

Befragung sinnvoll: Die Befragung für sinnvoll hielten zu Beginn der Studie 93,9% der Befragten, im SS 1993 76,7%, im WS 1993/94 77,9% und im SS 1994 73,3%. Insgesamt war die Kooperation von seiten der befragten Studierenden über alle vier Semester außerordentlich gut.

Befragungsergebnisse im Kurs besprochen: Im letzten Semester der Studie wurde allen Fragebögen noch eine dritte zusätzliche Frage hinzuge stellt: ob die Er-

gebnisse der ersten Befragung von der Lehrkraft im Kurs diskutiert worden waren oder nicht; die befragten Studenten bejahten eine kurs(gruppen)interne Besprechung im Durchschnitt lediglich zu 31,0%.

4.2 Beurteilung der Strukturqualität:

Die folgenden Ziffern spiegeln den prozentualen Anteil der Befragten wider, die die Fragen mit einer guten Bewertung beantworteten (z.B. Zahl der Studierenden pro Dozent „angemessen“, Unterrichtsräume „gut geeignet“, Lehrmittelqualität/-quantität „ausreichend“, Dozent „immer“ pünktlich oder vorbereitet, Lernklima „motivierend“ etc.).

Mit durchschnittlich 60% bis 70% über alle Kurse hinweg (Schnitt aller acht Befragungen: 66,7%) wurde die strukturelle Qualität am besten von allen drei Qualitätsebenen (s.o.) bewertet. Während sie zu Beginn der Studie noch 62,2% betrug, wurde sie mit den Semestern bis zu einem Wert von 69,2% nach und nach besser beurteilt. Dabei wurde die Lehrmittelmenge mit 78,0% (Durchschnitt aller Befragungen) am besten beurteilt, es folgten die Vorbereitetheit der Dozenten (Durchschnitt insgesamt 72,6%) und die Lehrmittelqualität (68,2%). Deutlich weniger positive Stimmen hingegen erhielten die Kursgruppengrößen (57,9%) und die Unterrichtsräume (56,7%).

4.3 Beurteilung der Prozeßqualität:

Die aufgeführten Ziffern beziehen sich wie oben auf die „gut“-Bewertungen der Befragten. Über alle Kurse hinweg betrug der mit den Befragungen steigende Score der Prozeßqualität zwischen 53,1% im WS 92/93 und 62,8% im WS 93/94 (im SS 94 wieder ein geringfügiger Abfall). Der Durchschnitt aller Befragungen lag hier bei 58,5% und damit unter dem der Strukturqualität.

Die Qualität der Durchführung der Kurse wurde hinsichtlich der Dozentenpünktlichkeit mit durchschnittlich 72,2% recht gut beurteilt; das bewertete Lernklima trat demgegenüber mit durchschnittlich 44,7% deutlich zurück.

4.4 Beurteilung der Ergebnisqualität:

Auf den Fragebogen-Vorderseiten wurde nur eine Frage nach dem Lernerfolg gestellt („Haben Sie in diesem Kurs bisher genügend gelernt?“). Der Score der Ergebnisqualität (gleichgesetzt mit „genug gelernt?“ = „ja“) betrug über alle Kurse und Befragungen hinweg durchschnittlich 40,1% (32,6% zu Beginn der Studie, Anstieg bis auf 45,6% im WS 93/94 und 43,4% im SS 94). Die Spannweite in den einzelnen Kursen und Praktika war mit Werten zwischen 1,8% und 75,2% enorm. Damit wurde der Lernerfolg in den Kursen nach den strukturellen und kursprozessualen Gegebenheiten am schlechtesten beurteilt.

4.5 Verlauf der Gesamtscores (Qualitätsindex):

Der Gesamtscore über alle Kurse und Befragungen hinweg betrug 55,1%. Er hatte bei den beiden Befragungen des WS 92/93 die Werte 49,3% und 51,8% (Spannweite in den Kursen zwischen 32,9% und

65,9%), im SS 93 52,9% und 55,1% (Spannweite 34,0% bis 72,4%), im WS 93/94 57,8% und 59,3% (Spannweite 39,1% bis 76,1%) und im SS 94 56,8% und 57,7% (Spannweite 37,2% bis 80,1%), und konnte sich damit insgesamt im Verlauf der Studie nennenswert steigern.

4.6 Interventionsfragebogen:

Da nicht nur von Interesse war, ob sich die Qualität der Lehrveranstaltungen durch die Intervention in den einzelnen Semestern und im Verlauf der vier Evaluationssemester verbesserte, sondern auch die Frage nach dem Wie der Intervention, wurde im SS 94 an die Dozenten der befragten Kurse zusätzlich ein Fragebogen zur Intervention ausgegeben.

Die Auswertung der von den Dozenten anonym ausgefüllten und zurückgeschickten Bögen - von 68 verschickten Bögen gelangten 42 zur Auswertung, was einer Rücklaufquote von 61,8% entspricht - zeigte, daß nach eigenen Angaben 85,4% der Antwortenden die ausgefüllten Fragebögen „immer“ gelesen hätten, daß 63,4% die studentische Kritik verwertbar fanden und daß 85,4% mit den Studenten über die Ergebnisse im Kurs diskutiert hätten (mit Kollegen 70,7%, mit der Instituts-/Klinikleitung 14,6%).

Nur 43,9% waren der Meinung, daß die Studie die Interaktion zwischen Lehrenden und Studierenden verbessert hätte. Nur 39,0% der Antwortenden bestätigten der Studie eine Auswirkung auf die Qualität der Lehre; hingegen hielten 70,7% sie für sinnvoll.

Auch die Kooperation von seiten der Dozenten war über die vier Evaluationssemester hinweg sehr gut. Dies wurde u.a. dadurch erreicht, daß sie selbst kaum zusätzlichen Aufwand mit den Befragungen hatten: die Durchführung lag ausschließlich in studentischen Händen.

Die Art des Umganges der Lehrkräfte, Kliniken und Institute mit den Ergebnissen der Befragungen war bei dieser Studie nicht vorgegeben: dies war weder möglich noch intendiert. Daher gibt es insgesamt zuwenig Erkenntnisse über den Vorgang der Intervention selbst und über den Grad des Zusammenhanges von Intervention und Auswirkung auf die Lehrqualität.

5 Diskussion

Die in der Zusammenfassung angesprochene Verbesserung der (Beurteilung der) Qualität der Lehre kann man mit dreierlei Maß messen, wobei jedes Maß, für sich gesehen, seine Richtigkeit hat:

Erstens muß man untersuchen, ob die Qualitätsbeurteilungen der Kurse im Verlauf der vier Semester stiegen, gleichblieben oder fielen (der Durchschnitt der Scores, über alle acht Befragungen hinweg, sagt schließlich nichts über eventuelle Steigerungen der positiven Bewertungen innerhalb dieses Zeitraumes aus).

Hinsichtlich der drei Einzelscores Struktur-, Prozeß- und Ergebnisqualität gab es im Verlauf der vier Evaluationssemester zu 59,7% Verbesserungen der Beurteilungen, zu 37,5% Verschlechterungen und zu 2,8% gleichbleibende Bewertungen. Man kann somit einen Index erstellen, der den Wirkungsgrad der indirekten Intervention (s.o.) angibt: 43 von 72 möglichen, evtl.

durch die Intervention bedingten Verbesserungen entsprechen einem Prozentsatz von 59,7%.

Der Gesamtscore stieg in drei Kursen kontinuierlich an, von denen man so sagen kann, daß zwei Jahre Evaluation die Beurteilung durch die Studierenden verbessert haben. In zwei weiteren Kursen kann dies mit Vorsicht ebenfalls konstatiert werden. Bei drei Kursen kann man entweder kontinuierlich oder aber zwischendurch eher von einem Abfallen des Gesamtscores sprechen.

Zweitens ist von Bedeutung, wieviele Beurteilungsverbesserungen es innerhalb der zwei Befragungen jedes Semesters in den einzelnen Kursen gab. Hier war zu betrachten, wieviele der Einzelkriterien der Struktur-, Prozeß- und Ergebnisqualität sich von den ersten zu den zweiten Befragungen der einzelnen Semestern änderten. Es zeigte sich, daß es in 53,5% der möglichen Fälle zu einer Beurteilungsverbesserung innerhalb des Semesters kam.

Bei der Bildung eines Index aus diesen Daten ergaben 137 von 256 möglichen Verbesserungen einen vermutlichen Wirkungsgrad der direkten Intervention (s.o.) von 53,5%.

Drittens könnte man betrachten, welches durchschnittliche Qualitätsniveau hinsichtlich des Gesamtscores jeder Kurs über acht Befragungen hinweg erreichte. Dies wurde in Tübingen zwar durchgeführt, nicht aber im Sinne eines Rankings der evaluierten Lehrveranstaltungen. Eine Rangreihenfolge unterschiedlichster, miteinander letztlich schwer zu vergleichender, Kurse zu erstellen, wird von Tübinger Seite nicht für sinnvoll erachtet und daher auch nicht publiziert.

Zusammenfassend kann man sagen, daß die Intervention ein tendenziell geeignetes Mittel ist, die Qualität der Lehre allmählich zu verbessern. Zum einen gibt es aber Probleme mit einem gewissen „Übersättigungseffekt“, den Lehrende und Studierende dauernden Befragungen gegenüber entwickeln: so ging die Akzeptanz seitens der Studierenden - ablesbar an den Ergebnissen der semesterweise gestellten Frage, ob die Befragung für sinnvoll erachtet wurde - von anfänglich 93,9% auf zuletzt 73,3% zurück (s.o.), was nach vier Semestern zwar einen noch immer beachtlichen Wert darstellt, an der prinzipiellen Problematik aber nichts ändert. Hier wäre es hilfreich, den Studierenden die Zusammenhänge und Wege zwischen Befragung und Konsequenzen transparenter zu machen, was nicht ausreichend geschah.

Sonst aber, bei konstruktiver Zusammenarbeit zwischen Lehrenden und Studierenden, sind Befragungen ein sensibles und sinnvolles Meßinstrument. Von Tübinger Seite aus wird daher empfohlen, im Falle einer auf den Erfahrungen der durchgeführten Studie aufbauenden Weiterentwicklung entweder häufiger, zumindest aber dauerhaft kontinuierlich, zu evaluieren.

Zum anderen gibt es Probleme durch die Beurteilung struktureller Mängel, für die die Institute und Kliniken bzw. Dozenten nicht verantwortlich sind, wie z.B. Mangel an Unterrichtsräumen, zuwenig und/oder veraltete Untersuchungsgeräte und Medien, zuwenig Lehrkräfte und damit zu große Kursgruppen usw. Hier sind strukturelle Maßnahmen notwendig, die von höherer Ebene aus gesteuert werden müssen.

Problematisch erscheint außerdem der Umstand, daß es im Bereich der Lehre keine Sollwerte oder Standards gibt, mit denen die durch eine Evaluierung erhaltenen Daten verglichen werden könnten. Hier gäbe es z.B. die Möglichkeit, die Studierenden eingehend zu befragen, welche Kriterien ihnen in welchem Ausmaß an der Lehre wichtig sind, um so auf ein objektivierbares Niveau und möglicherweise sogar zu einer Standardisierung zu gelangen.

Ein Evaluationsprojekt wie dieses ist nicht nur mit relativ geringen Mitteln, die dafür zu einem umso größeren Erfolg führen, machbar. Es ist hinsichtlich seiner konzeptionellen Vorstellung einer Befragung interventionellen Charakters auch anderen Fakultäten und Hochschulen zu empfehlen. Die durch die Studie aufgedeckten, immanenten Schwächen können mit weiter ausgereiften Materialien und Methoden unter dem Versuch eines durch zunehmende Standardisierung der Lehrqualität erhöhten Objektivierungsgrades zukünftig abgebaut, eventuell sogar beseitigt werden.

Literatur

Basler, H.D. et al. (1993)
ein Fragebogen zur Akzeptanz der Lehre. Unveröffentlichtes Manuskript, Phillips Universität Marburg.

Endruweit, G. (1993)
Programmevaluation als Laienspiel. Bemerkungen über Meinungsforschung. Sozialforschung und Pfüsch bei Studentenbefragungen. Mitteilungen des Hochschulverbandes, 2, 129-132

Kromrey, H. (1980)
Empirische Sozialforschung. Methoden und Modelle der Datenerhebung und Datenauswertung. 3. Aufl., UTB, Opladen

Kromrey, H. (1983)
Lehrevaluation darf nicht auf Umfragen reduziert werden. Mitteilungen des Hochschulverbandes, 4, 253-256

Kromrey, H. (1993)
Studentische Vorlesungskritik. Empirische Daten und Konsequenzen für die Lehre. Soziologie, 1, 39-56

Lienert, G.A. (1989)
Testaufbau und Testanalyse. 4. Aufl., Psychologie-Verl.-Union, München

Mayntz, R. et al. (1978)
Einführung in die Methoden der empirischen Soziologie. 5. Aufl., Westdeutscher Verlag, Opladen

Mussgnug, R. (1992)
Gefährdet die Lehrevaluation die Freiheit der Wissenschaft? Mitteilungen des Hochschulverbandes, 4, 253-256

Protokoll
des Ordentlichen Medizinischen Fakultätentages der Bundesrepublik Deutschland am 12. Juni 1993 in Halle/Saale, 43-53

Süllwold, F. (1992)
Welche Realität wird in der Beurteilung von Hochschullehrern durch Studierende erfaßt? Mitteilungen des Hochschulverbandes, 1, 34f.

Webler, W.-D. (1991)
Kriterien für gute akademische Lehre. Das Hochschulwesen,
6, 243-249

Dank

Unser Dank gilt Herrn Prof. Dr. R. Bloch, Institut für Aus-, Weiter- und Fortbildung (IAWF), Medizinische Fakultät Bern, Herrn Prof. Dr. H.-K. Selbmann und Herrn C. Meisner, Institut für Medizinische Informationsverarbeitung, Medizinische Fakultät Tübingen, für ihre konstruktive Beratung. Unser Dank geht außerdem

an das Ministerium für Wissenschaft und Forschung des Landes Baden-Württemberg für die Finanzierung und damit für die Ermöglichung der durchgeführten Studie.

Anschrift des Verfassers:

cand. med. Wolfgang Stangl
Annabergstr. 7
D-70374 Stuttgart