# EVALUATION DER LEHRE
## - Erste Erfahrungen mit der Qualitätserfassung des chirurgischen Unterrichts -

F. Eitel, A. Bräth, L. Schweiberer*

Eine Möglichkeit, wie Evaluation methodisch gestaltet werden kann, zeigt der folgende Evaluationsansatz des Münchner Curricularen Innovationsprojektes (MCIP). Absicht der Autoren ist es, mit der vorliegenden Arbeit diesen Ansatz zur Diskussion zu stellen und zur weiteren Evaluationsentwicklung beizutragen. Mittels dieses Evaluationsansatzes wurde das Praktikum der Chirurgie im SS 1993 (2. klinischer Studienabschnitt, 9. Semester) evaluiert. Die als Längsschnitterhebung angelegte Untersuchung wurde mit einem Fragebogen durchgeführt, der jeweils zu Beginn und am Ende der Unterrichtsveranstaltung auszufüllen war. Ziel der Untersuchung war es u.a., empirische Entscheidungshilfen für die weitere Gestaltung der curricularen Organisation zu erhalten.

Die Ergebnisse der Untersuchung zeigen, daß das Innovationsprojekt die Lehre und den Lernerfolg verbessert hat. Ohne Reform und Evaluation hätte sich dieser Erfolg nicht erreichen lassen.

* An dieser Stelle sei allen beteiligten Dozenten Dank und Anerkennung ausgesprochen für ihre Kooperation bei der Evaluation. Eine namentliche Nennung ist aus Platzgründen nicht möglich. Es sei auf das Vorlesungsverzeichnis der LMU München verwiesen.

**Zusammenfassung**

Angesichts der zunehmenden Diskussion über Hochschullehre ist es mittlerweile nicht mehr fraglich, ob die Lehre systematisch bewertet werden sollte, sondern wie die Evaluation methodisch gestaltet werden kann. Die Evaluation des Münchener Curricularen Innovationsprojektes (MCIP) wird beschrieben. Sie ist als Längsschnitterhebung im quasi-experimentellen Panel-Design gestaltet. Als Untersuchungsinstrument dient ein Fragebogen mit 15 Items in didaktischen, organisatorischen und motivationspsychologischen Dimensionen. Außerdem wird die studentische Akzeptanz der Reform erfaßt. Die Analyse der regressionsanalytisch ermittelten Varianzen, Effektstärken und Zielerreichungsgrade ergibt eine vertretbare Variabilität bei Unterrichtsdurchführung und selbsteingeschätzten Lernerfolgen und dementsprechend zufriedenstellende Nutzwerte für den Unterricht. Damit ist der derzeitige Standort des MCIP in Sachen Lehrqualität empirisch geklärt. Insgesamt stellen die beschriebenen Evaluationsergebnisse wertvolle empirische Entscheidungshilfen für die weitere Gestaltung der curricularen Organisation dar. Die Definition von Evaluation wird auf empirischer Grundlage vorgenommen. Das weitere Vorgehen, insbesondere in Hinblick auf Kosten-Nutzen-Analyse der Lehre, wird im Ansatz aufgezeigt. Es erscheint notwendig, die Evaluationsmethodik im Sinne einer Evaluationsforschung weiter zu entwickeln, um so weitere Meßinstrumente für ein Qualitätsmanagement der Lehre zu bekommen, damit noch mehr der erforderlichen Daten (insbesondere Kosten) in den Organisationsprozeß zurückgekoppelt werden können. Der beschriebene Evaluationsansatz erscheint hierzu geeignet.
Das Statement, die Hochschullehre sei schlecht, ist politischer Natur; es entbehrt der empirischen Grundlage, erkenntlich daran, daß kaum jemand sagen kann, wie schlecht - quantitativ gesehen - sie tatsächlich ist. Nun hat aber angesichts der sich verschärfenden Verteilungskämpfe um Ressourcen das "große Zählen und Messen" (Hartmer 1993) begonnen. Somit ist nicht mehr die Frage zu stellen, ob die Lehrqualität empirisch überprüft, d.h. evaluiert werden soll, sondern auf welche Weise evaluiert werden kann. "Die Probleme der Lehrevaluation liegen vor allem in ihrer Methodik" (Hartmer 1993).

Vor diesem Hintergrund beschäftigt sich der vorliegende Artikel mit der Beschreibung eines prototypischen Evaluationsansatzes in der chirurgischen Ausbildung.

## 1 Datenquelle

Dargestellt wird ein Evaluationsansatz im Münchener Curricularen (Chirurgischen) Innovationsprojekt (MCIP): Die Evaluation des Praktikums der Chirurgie (2. klinischer Studienabschnitt, 9. Semester) im Sommersemester 1993 ist Gegenstand der vorliegenden Untersuchung.

Das MCIP ist gekennzeichnet durch einen pragmatischen Reformansatz (Eitel et al. 1992a). Charakteristika dieses Ansatzes sind

- die Reorganisation der chirurgischen Vorlesung, der Seminare und Praktika nach didaktischen Gesichtspunkten, wobei besonderer Wert auf die Lernmotivation der Studenten zugunsten eines sachbezogenen Lernverhaltens gelegt wird (siehe Eitel 1993a),

- die Evaluation, die prospektiv als Erhebung geplant wurde (Eitel 1992b).

Das Praktikum der Chirurgie ist seit dem Sommersemester 1989 nach diesen Gesichtspunkten reformiert. Es besteht aus 30 Arbeitsplätzen, an denen Studenten je nach Lehrgegenstand, Unterrichtsziel und -methode, in Gruppen à 3 bis Gruppen à 12 lernen. Die Arbeitsplätze sind nach sogenannten Lernebenen (Eitel et al. 1993a) organisiert: Ebene A dient dem Erwerb und der Rekapitulation kognitiver Inhalte durch Seminare, multimediale Selbstlernprogramme, Falldemonstrationen usw., Ebene B dient dem Erwerb chirurgischer Basisfertigkeiten durch praktische Übungen und Ebene C beinhaltet den Unterricht am Krankenbett mit dem Ziel der Anwendung und Übung der in Ebene A und B erworbenen Kompetenzen sowie der Sozialisation der Studenten im Berufsfeld. Die Studenten rotieren an 23 Praktikumstagen an den 30 Arbeitsplätzen.

## 2 Evaluationsmethodik im MCIP

### 2.1 Studiendesign

Die Evaluation wurde prospektiv nach dem CIPP-Modell (Stufflebeam et al. 1971) geplant (Abb. 1): Es handelt sich dabei um eine Programmevaluation; deren Gegenstand ist

- die Entwicklung des Unterrichtsprogrammes (Strukturevaluation),
- seine Implementierung in die Unterrichtspraxis (Prozeßevaluation)

- und das Unterrichtsergebnis in Form des Lernerfolges (Ergebnisevaluation).

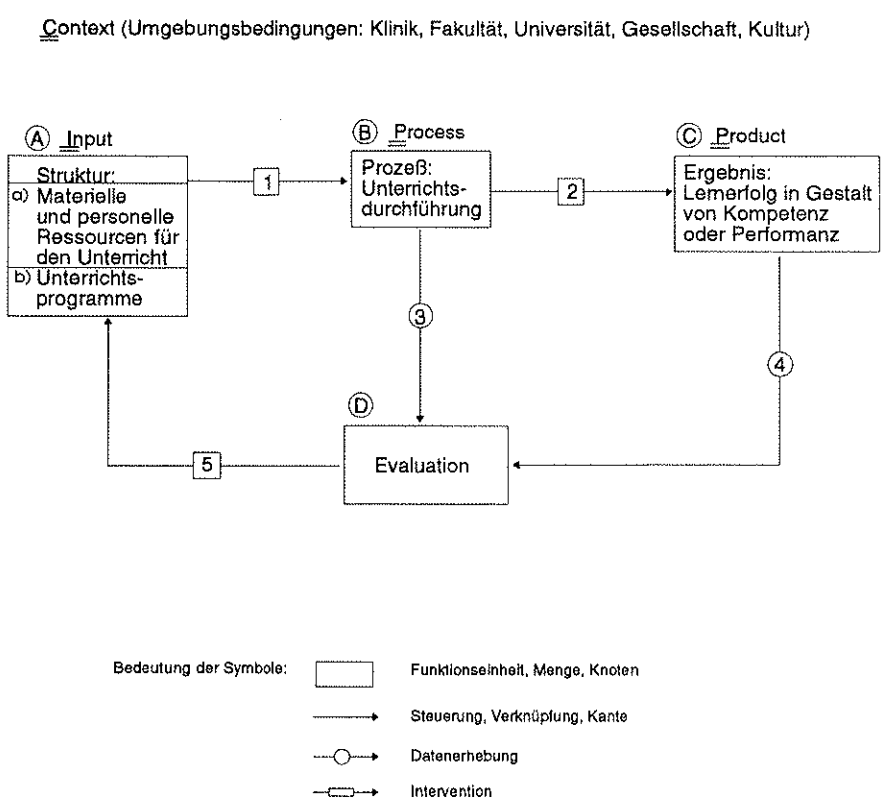
Dabei stand die Erfassung des Unterrichtsablaufes im Vordergrund. Prozeß- und Ergebnisevaluation dienen dazu, das ursprüngliche Unterrichtsprogramm optimierend zu ändern oder bei zufriedenstellenden Evaluationsergebnissen unverändert zu belassen. Das Charakteristikum dieses Evaluationsansatzes ist demnach die formative Rückkopplung der Prozeß- und Ergebnisbewertung zur Programmgestaltung. Diese evaluatorische Kopplung von Prozeß- und Ergebnisdaten an die curricularen Zielsetzungen ist geeignet für die qualitative Weiterentwicklung des Unterrichts. Dadurch kann im Laufe der evaluatorischen Programmentwicklung eine Programmoptimierung im Sinne eines Qualitätsmanagements der Lehre (Eitel et al. 1992a, Eitel 1993) erreicht werden. Die damit erreichte Entscheidungsorientierung der curricularen Reform (Cronbach 1963) stellt den gewählten Evaluationsansatz in den Rahmen der Qualitätssicherung bzw. des Qualitätsmanagements (Berwick et al. 1990).

Die vorliegende, zu berichtende Evaluation des Chirurgischen Praktikums wurde als prospektive Studie mit Totalerfassung geplant. Verglichen werden die über die Zeit erhobenen Evaluationsdaten (Längsschnittstudie). Eine Kontrollgruppe fehlt, so daß der Vergleich mit dem Status quo ante vorgenommen werden muß. Dieser Versuchsplan entspricht dem Panel-Design bzw. Pretest-Posttest-Design (Campbell et al. 1966). Die Studie wird über den Berichtszeitraum hinaus fortgeführt, so daß letztendlich eine Art Zeitreihenanalyse resultieren wird.

### 2.2 Datensammlung

Die Prozeß- und Ergebnisdaten für den vorliegenden Artikel wurden mittels Fragebogen im Längsschnitt über ein Semester mit Totalerfassung erhoben: Alle 150 Studenten des Praktikums erhielten für jede der 23 Sitzungen einen von den Praktikumsorganisatoren und Studenten gemeinsam entwickelten, seit 1988 erprobten und zwischenzeitlich mehrfach modifizierten Fragebogen (Tab. 1), auf dem zugleich vom Dozenten in der jeweiligen Veranstaltung die Anwesenheit des Studenten testiert werden mußte. Am Semesterende wurden die Evaluationsbögen in einer Abschlußbesprechung des Praktikums gegen Aushändigung des Scheins eingesammelt.

Abb.1: CiPP-Modell der Unterrichtsevaluation



Tab. 1: Original - Fragebogen zum Praktikum der Chirurgie

**Tab. 1: Original-Fragebogen zum Praktikum der Chirurgie**

Zu Beginn der Unterrichtsveranstaltung auszufüllen — Anleitung siehe letzte Seite

**Veranstaltungsort** *

(nähere Hinweise zu den einzelnen Veranstaltungsorten siehe letzte Seite)

| | | | | |
|---|---|---|---|---|
| ① | ⑦ | ⑬ | ⑲ | ㉕ |
| ② | ⑧ | ⑭ | ⑳ | ㉖ |
| ③ | ⑨ | ⑮ | ㉑ | ㉗ |
| ④ | ⑩ | ⑯ | ㉒ | ㉘ |
| ⑤ | ⑪ | ⑰ | ㉓ | ㉙ |
| ⑥ | ⑫ | ⑱ | ㉔ | ㉚ |

* Veranstaltungsorte aus Datenschutzgründen anonymisiert!

*Lernziel der heutigen Kursstunde*
(was sollen die Studierenden am Ende der Kursstunde können? - Dozenten fragen!) : .....

Wie weit bin ich zu Beginn vom obigen Lernziel entfernt?

1 — 2 — 3 — 4 — 5 — 6
Lernziel bereits erreicht — Sehr weit entfernt

**Gruppennr.:** .....

Am Ende der Unterrichtsveranstaltung auszufüllen:

**In der heutigen Unterrichtsveranstaltung . . .**

| | |
|---|---|
| . . . besaß ich bereits Vorwissen | 1 (ja) — 2 — 3 — 4 — 5 — 6 (nein) |
| . . . bestand für mich die Möglichkeit selbst tätig zu werden | 1 (sehr häufig) — 2 — 3 — 4 — 5 — 6 (nie) |
| . . . war der Kursaufbau | 1 (strukturiert) — 2 — 3 — 4 — 5 — 6 (chaotisch) |
| . . . ging der/die Dozent/in auf die Teilnehmer/innen ein | 1 (sehr häufig) — 2 — 3 — 4 — 5 — 6 (nie) |
| . . . hat sich der/die Dozent/in bemüht | 1 (ja) — 2 — 3 — 4 — 5 — 6 (nein) |
| . . . habe ich verstanden worum es ging | 1 (ja) — 2 — 3 — 4 — 5 — 6 (nein) |
| . . . bin ich gefordert worden | 1 (sehr häufig) — 2 — 3 — 4 — 5 — 6 (nie) |
| . . . waren Rückfragen möglich | 1 (sehr häufig) — 2 — 3 — 4 — 5 — 6 (nie) |
| . . . hat der/die Dozent/in mir gesagt was ich kann | 1 (sehr häufig) — 2 — 3 — 4 — 5 — 6 (nie) |
| . . . war ich | 1 (konzentriert) — 2 — 3 — 4 — 6 (abgelenkt) |

**Nach der heutigen Unterrichtsveranstaltung . . .**

| | |
|---|---|
| . . . möchte ich mich weiter mit dem Thema beschäftigen | 1 (ja) — 2 — 3 — 4 — 5 — 6 (nein) |
| . . . erteile ich der Veranstaltungsqualität insgesamt die Note | 1 (sehr gut) — 2 — 3 — 4 — 5 — 6 (ungenügend) |
| Wie weit bin ich am Ende der Veranstaltung vom obigen Lernziel entfernt? | 1 (Lernziel erreicht) — 2 — 3 — 4 — 5 — 6 (Sehr weit entfernt) |
| Die heutige Kursstunde hat mir: | 1 (sehr gut gefallen) — 2 — 3 — 4 — 5 — 6 (gar nicht gefallen) |

*Weitere Kommentare* (ev. Rückseite benützen)

**Unterschrift des Tutors zur Bestätigung der Anwesenheit des/der Studenten/in:** .....

### 2.3 Datenkodierung bzw. Dimensionalität des Fragebogens

Der Fragebogen enthält 15 Items (Tab. 1), davon 12 zur Erfassung des Unterrichtsprozesses und 3 zur Einschätzung des eigenen Lernerfolges durch die Studenten selbst. Die Einschätzungen der Studenten zu den einzelnen Items erfolgten mittels Schulnotenskala von 1 bis 6.

Die **Prozeßevaluation** bezog sich auf 7 vorab festgelegte didaktische Kriterien:

1. **Learning by doing (Item 3)**
2. **Intrinsische Lernmotivation (Item 6, 8, 11 und 12)**
3. **Veranstaltungsqualität (Item 4 und 13)**
4. **Interaktion Dozent/Student (Item 5, 7, 9 und 10)**
5. **Akzeptanz der Einzelveranstaltung seitens der Studenten (Item 15)**

Diese fünf Kriterien sind die Zielgrößen der MCIP-Evaluation. Die Kriterien 6 und 8 bestehen in einer anderen Gruppierung von Items der ersten 5 Kriterien entsprechend anderen didaktischen Gesichtspunkten, worauf aber im vorliegenden Rahmen nicht weiter eingegangen wird.

6. **Lehrqualität (Item 3, 4, 5, 9, 10 und 12)**
8. **Feedback (Item 3, 5, 9 und 10).**

Die **Ergebnisevaluation** erfolgte bei dieser Erhebung nur in Form der Einschätzung des Vorwissens bzw. des subjektiven Lernerfolges **(Item 1, 2 und 14; Kriterium 7).**

### 2.4 Datenreduktion

Die von jedem Studenten für jede der von ihm besuchten 23 Veranstaltungen abgegebenen Noten $i_n$ wurden auf einem 386er PC mit dem Tabellenkalkulations-Programm Excel® 4.0 datenbankmäßig in einer Rohwert-Tabelle für jedes der m = 15 Items erfaßt. Die einzelnen Noten der n Studenten wurden für jeden der p Veranstaltungsorte (Lernplätze, siehe Eitel et al. 1993a) mit dem genannten Programm durch Mittelwertbildung (mit Standardabweichung SD zur Beurteilung der Streuung bzw. Median zur Beurteilung der Verteilungsform) so reduziert, daß für jedes Item i bzw. Kriterium K eine Durchschnittsnote $J_m$ bzw. $K_m$ pro Lernplatz bzw. Veranstaltung o zur Verfügung stand (Tab. 2, Spalte 1 - 5).

Tab. 2: Werte der durchschnittlichen Kriterieneinschätzungen und Parameter an den einzelnen Veranstaltungsorten

| Tab 2 : Werte der durchschnittlichen Kriterieneinschätzungen und Parameter an den einzelnen Veranstaltungsorten | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | K 1-5 | | |
| Orte | K1 | K2 | K3 | K4 | K5 | K 1-5 | SD K1-5 | J 1,2 | J 14 | Wo | Zo | Vo |
| 1 | 2,2 | 2,2 | 2,2 | 2 | 2,2 | 2,20 | 0,20 | 2,84 | 2,20 | 0,63 | 0,118 | 0,091 |
| 2 | 2,5 | 2,4 | 2,2 | 1,9 | 2,1 | 2,17 | 0,15 | 3,30 | 2,29 | 0,89 | 0,132 | 0,068 |
| 3 | 2,4 | 2,3 | 2 | 1,7 | 2 | 2,10 | 0,17 | 3,08 | 2,19 | 0,86 | 0,158 | 0,080 |
| 4 | 3,5 | 2,4 | 2,3 | 1,9 | 2,2 | 2,30 | 0,22 | 3,52 | 2,61 | 0,86 | 0,079 | 0,095 |
| 5 | 2,4 | 2,5 | 2,6 | 2,4 | 2,7 | 2,50 | 0,15 | 2,86 | 2,23 | 0,64 | -0,001 | 0,060 |
| 6 | 2,4 | 2,3 | 2,2 | 1,9 | 2,2 | 2,20 | 0,17 | 2,90 | 2,11 | 1,97 | 0,121 | 0,080 |
| 7 | 2,2 | 2,3 | 2 | 1,7 | 2 | 2,05 | 0,22 | 3,16 | 2,15 | 0,90 | 0,180 | 0,108 |
| 8 | 2,7 | 2,4 | 2,4 | 1,9 | 2,5 | 2,33 | 0,20 | 3,12 | 2,30 | 0,72 | 0,066 | 0,086 |
| 9 | 2,4 | 2,1 | 1,8 | 1,6 | 1,8 | 1,95 | 0,28 | 2,62 | 1,98 | 0,75 | 0,221 | 0,145 |
| 10 | 3 | 2,5 | 2,6 | 2,1 | 2,6 | 2,51 | 0,20 | 2,90 | 2,37 | 0,58 | -0,005 | 0,080 |
| 11 | 3,7 | 2,7 | 2,6 | 2,2 | 2,6 | 2,60 | 0,19 | 3,72 | 2,72 | 0,90 | -0,041 | 0,074 |
| 12 | 2,7 | 2,5 | 2,4 | 2,1 | 2,6 | 2,42 | 0,18 | 2,60 | 2,13 | 0,47 | 0,031 | 0,076 |
| 13 | 1,1 | 2 | 1,8 | 1,8 | 2 | 1,79 | 0,21 | 3,31 | 1,99 | 0,81 | 0,283 | 0,120 |
| 14 | 1,3 | 2,2 | 2,2 | 1,9 | 2,3 | 1,97 | 0,21 | 3,93 | 1,97 | 1,26 | 0,213 | 0,107 |
| 15 | 4,1 | 2,8 | 2,8 | 2,4 | 3 | 2,84 | 0,16 | 3,10 | 2,44 | 0,63 | -0,136 | 0,056 |
| 16 | 2,5 | 2,4 | 2,2 | 1,7 | 2,2 | 2,18 | 0,21 | 3,31 | 2,28 | 1,01 | 0,127 | 0,097 |
| 17 | 3,7 | 2,3 | 2 | 1,7 | 2 | 2,24 | 0,34 | 2,98 | 2,11 | 0,94 | 0,104 | 0,153 |
| 18 | 3,7 | 2,4 | 2,2 | 1,9 | 2,1 | 2,32 | 0,27 | 3,19 | 2,35 | 1,04 | 0,073 | 0,116 |
| 19 | 2,3 | 2,3 | 2,1 | 1,7 | 2,1 | 2,13 | 0,16 | 2,77 | 2,05 | 0,87 | 0,147 | 0,077 |
| 20 | 2,3 | 2,3 | 2,3 | 1,9 | 2,3 | 2,19 | 0,26 | 2,64 | 2,11 | 0,50 | 0,123 | 0,117 |
| 21 | 2,3 | 2,5 | 2,2 | 2,1 | 2,3 | 2,29 | 0,14 | 3,08 | 2,28 | 0,95 | 0,083 | 0,063 |
| 22 | 2,4 | 2 | 1,7 | 1,5 | 1,9 | 1,81 | 0,23 | 3,16 | 2,29 | 0,70 | 0,276 | 0,130 |
| 23 | 2,6 | 2,5 | 2,5 | 2 | 2,4 | 2,37 | 0,18 | 3,12 | 2,22 | 1,04 | 0,051 | 0,074 |
| 24 | 3,1 | 2,4 | 2,1 | 2 | 2,1 | 2,43 | 1,03 | 2,89 | 2,26 | 0,67 | 0,026 | 0,424 |
| 25 | 2,4 | 2,2 | 1,9 | 1,5 | 1,5 | 1,90 | 0,34 | 3,23 | 2,30 | 0,88 | 0,238 | 0,177 |
| 26 | 4,8 | 3,1 | 3,7 | 2,9 | 3,3 | 3,31 | 0,19 | 3,70 | 3,17 | 0,54 | -0,323 | 0,056 |
| 27 | 3,1 | 2,5 | 2,5 | 1,9 | 2,3 | 2,35 | 0,24 | 3,53 | 2,47 | 0,99 | 0,059 | 0,103 |
| 28 | 4 | 2,6 | 2,5 | 2 | 2,5 | 2,54 | 0,28 | 3,30 | 2,42 | 0,77 | -0,015 | 0,109 |
| 29 | 2,3 | 2,1 | 1,9 | 2,7 | 2,2 | 2,19 | 0,43 | 2,72 | 2,09 | 0,68 | 0,122 | 0,196 |
| 30 | 3 | 3,2 | 2,7 | 4,7 | 3,4 | 3,57 | 0,29 | 3,15 | 2,25 | 0,88 | -0,428 | 0,082 |
| MW | 2,77 | 2,41 | 2,28 | 2,05 | 2,31 | 2,33 | 0,25 | 3,12 | 2,28 | 0,84 | 0,069 | 0,110 |

## 2.5 Datenanalyse

Die Datenanalyse erstreckt sich auf die parametrische Bearbeitung der Rohwerte $J_m$ bzw. $K_m$ (statistische Verarbeitung der Noteneinschätzungen der Studenten) für jeweils eine Veranstaltung, also eine spezifische Unterrichtsmaßnahme, sowie auf die inhaltliche Analyse unter Berücksichtigung der Antworten auf die offene Frage im Fragebogen.

### 2.5.1 Erfassung der Effektstärke der Unterrichtsdurchführung (Wirkungsanalyse)

Im vorliegenden Kontext wird als Wirkungsanalyse die Beschäftigung mit der Art (Richtung und Größe) des Zusammenhanges (Kante 1 und 2 in Abb. 1) zwischen Variablen (dargestellt durch Fragebogen-Items) bzw. Kriterien des Unterrichts (Knoten A, B und C in Abb. 1) definiert. Statistisch signifikante Unterschiede zwischen dem Verhalten vor der Instruktion und nach der Instruktion (Kriterium 7) verweisen auf einen Wirkungszusammenhang (siehe Kante 2 in Abb. 1) zwischen den Variablen Unterrichtsdurchführung und -ergebnis (Knoten B und C in Abb. 1; Test: Wilcoxon matched pairs zweiseitig auf 0,05-Niveau mit dem Statistikprogramm BIAS 3.0, siehe Akkermann 1992). Falls in diesem ersten Schritt der Wirkungsanalyse signifikante Unterschiede zu verzeichnen sind, wird der gefundene Wirkungszusammenhang durch seine Effektstärke w weiter spezifiziert, d.h. es wird die Wirksamkeit einer Unterrichtsmaßnahme untersucht.

Sie definiert sich aus dem Verhältnis der Differenz der Gruppenmittelwerte $J_m$ (beobachtetes Verhalten bzw. Einschätzung des eigenen Wissens vor ($J_{1.2}$ = Durchschnittsnote Item 1 und 2) bzw. nach Instruktion ($J_{14}$ = Durchschnittsnote Item 14)) zur Standardabweichung SD des Mittelwertes $J_{1.2}$ vor Instruktion. Wir bezeichnen den Parameter w als Wirkungsgrad einer Unterrichtsmaßnahme.

$$[1] \qquad w = \frac{J_{1.2} - J_{14}}{SD_{J_{1.2}}}$$

Die Wirkungsanalyse erfaßt in der MCIP-Evaluation die Ergebnisqualität des Unterrichtsprogrammes (Tab. 2, Spalte $w_o$). Je größer w, desto stärker steuert die Unterrichtsdurchführung den Lernerfolg.
Die Feststellung der Effektivität einer Instruktion durch quantitative Wirkungsanalyse, wie hier vorgeschlagen, ist notwendige aber nicht hinreichende

Bedingung der Evaluation. Denn es kann sein, daß eine Unterrichtsmaßnahme, obgleich an sich effektiv, nicht das von den Unterrichtsorganisatoren gesteckte Lernziel erreicht, z.B. weil die Wirkung fehleingeschätzt wurde bei der Implementierung der Maßnahme, oder das Unterrichtsziel unverhältnismäßig hoch angesetzt war, oder verzerrende Störgrößen die Wirksamkeit der Maßnahme minderten. Letzteres ist z.B. der Fall, wenn die Studenten in einen didaktisch wirksamen Unterricht gehen, ohne sich vorzubereiten, weil vielleicht eine bevorstehende Klausur in einem anderen Fach sie ablenkt, oder wenn die Wissensvoraussetzungen oder erforderliches Lernverhalten nicht vorhanden sind.

### 2.5.2 Zielerreichungsanalyse

Deshalb wird die Zielerreichungsanalyse als eigentliche Untersuchung der Qualität einer Maßnahme vorgeschlagen. Prinzip ist der Vergleich eines Ist-Wertes ($J_m$ Item 3 - 13 und 15) mit einem Soll-Wert s der von den Studenten im Durchschnitt eingeschätzten Skalenpunkte (Tab. 2, Spalte $z_o$).

$$[2] \qquad z = \frac{\text{Sollwert sw - Istwert } J_m}{\text{Sollwert sw}}$$

Wird das vorab festgelegte Unterrichtsziel von einer Instruktion (angenähert) erreicht, so ist deren Qualität als hoch einzuordnen. Negative z-Werte (Soll-Wert < Ist-Wert) sind bei Verwendung der Notenskala nicht zielführend, gute Qualität wird durch positive z-Werte repräsentiert (Soll-Wert > Ist-Wert). Der Soll-Wert wurde prospektiv für alle Items auf 2,5 Punkte der Notenskala festgelegt. Im Fall der Wertgleichheit von Soll- und Ist-Wert (Zähler = 0 in Formel [2]) wird ohne weitere Berechnung Zielerreichung mit einem Wert von z = 0,04 angenommen. Die Spanne möglicher z-Werte liegt mit einem Soll-Wert von sw = 2,5 Skalenpunkten zwischen z maximal +0,6 und z minimal -1,4. Als Grenzwert wurde z > 0,04 festgelegt.

### 2.5.3 Erfassung der Varianz

Verhaltensbeobachtungen und erst recht Einschätzungen unterliegen naturgemäß großer Variation. Als Parameter für die Variabilität des Eingeschätzten (Knoten B und C in Abb. 1) und der Variabilität beim Einschätzen (Kanten 3 und 4 in Abb. 1) wurde der Variationskoeffizient v

herangezogen, der durch das Verhältnis von Standardabweichung (SD) zum Mittelwert $J_m$ bzw. $K_m$ der durch die Studenten jeweils eingeschätzten Skalenpunktwerte (Noten 2) definiert ist (Tab. 2, Spalte $v_o$).

$$[3] \qquad v = \frac{SD_{Jm}}{Jm} \quad ; \quad v_o = \frac{SD_{Km}}{K_m}$$

Die Analyse der Variabilität wird bei der MCIP-Evaluation mit dem Ziel der Erfassung der Zuverlässigkeit der Unterrichtsdurchführung herangezogen. Je kleiner die Variabilität v ausfällt, desto homogener und damit zuverlässiger sind die erhobenen Daten für Organisationsentscheidungen. Als Grenzwert wurde prospektiv $v < 0{,}33$ festgesetzt.

### 2.6 Bewertung der Lehrqualität

#### 2.6.1. Qualitätskriterium: Varianz der Unterrichtsdurchführung

Die Zuverlässigkeit der Zielerreichung des Unterrichts läßt sich dadurch erfassen, daß man die Variationskoeffizienten (siehe [3])

$$[4] \qquad vo = \frac{SD_{K1\text{-}5}}{K1\text{-}5}$$

für die einzelnen Unterrichtsmaßnahmen, - mit anderen Worten: Veranstaltungsorte o (Knoten B in Abb. 1), - in Beziehung bringt zu den jeweiligen Zielerreichungsgraden $z_o$, die aus der Formel [2] für jeden Veranstaltungsort o zu errechnen sind. Hierzu werden die mit den Formeln [2] und [4] gewonnenen Parameterwerte aus allen von den Studenten eingeschätzten Prozeßqualitäts-Items pro Veranstaltungsort mit dem Statistikprogramm BIAS 3.0 (Ackermann 1992) regressionsanalytisch (Spearman-Rang-Regression) untersucht, wobei die Abszisse durch den jeweiligen Variationskoeffizienten $v_o$ (unabhängige Variable) und die Ordinate durch den jeweiligen Zielerreichungsgrad $z_o$ (abhängige Variable) belegt wird. Die für die einzelnen Veranstaltungsorte berechneten $v_o$-Werte und $z_o$-Werte ergeben in diesem Diagramm eine Punktwolke. Die Steilheit der Regressionsgerade

(Steigung = Z = Tangens-Alpha) läßt sich aus den Daten nach der folgenden Formel ermitteln:

$$[5] \qquad Z = \frac{Z_{30}}{V_{30}} - a$$

Der Achsenabschnitt a ist definiert als der Nullabstand des Schnittpunktes der Regressionsgeraden mit der Ordinate. Die Formel gilt nur, wenn Abszisse und Ordinate die gleichen Einslängen aufweisen. Die Steilheit der Regressionsgerade wird in der vorliegenden Untersuchung als ein Maß für die Homogenität bzw. Unzuverlässigkeit der Zielerreichung aufgefaßt: Je kleiner die Variationskoeffizienten $v_o$ und je höher die Zielerreichungsgrade $z_o$, desto gleichmäßiger die Qualität der Unterrichtsdurchführung. Die Steigung dieser Funktion repräsentiert einen Nutzenaspekt Z des Unterrichts. Er wurde bei der Evaluationsplanung nicht wertmäßig festgelegt, da keine empirischen Daten vorhanden waren; vielmehr wurde definiert, daß sein Wert positiv sein soll, damit der Unterrichtsdurchführung zufriedenstellende Qualität attestiert werden kann.

### 2.6.2. Qualitätskriterium: Relevanz der Unterrichtsdurchführung

Der zweite Nutzenaspekt, die Wertigkeit der Unterrichtsdurchführung, läßt sich als Funktion von Wirkungsgrad w [1] (Kriterium 7) und Zielerreichungsgrad z [2] der Unterrichtsdurchführung (Kriterium 1 - 5) für jede Veranstaltung fassen. Beide können wiederum in einem Diagramm für alle Veranstaltungsorte regressionsanalytisch (siehe oben) dargestellt werden. Abhängige Variable ist $z_o$ der Einzelveranstaltungen, unabhängige Variable deren Wirkungsgrad $w_o$. Der Nutzwert wird nach folgender Formel errechnet:

$$[6] \qquad N = \frac{Z_{30}}{W_{30}} - a$$

Auch hier wurde festgelegt, daß der Nutzenwert N positiv sein sollte. Eine Festlegung von Kontrollgrenzen konnte aufgrund fehlender empirischer Daten noch nicht vorgenommen werden.

Wir bezeichnen die regressionsanalytische Erstellung des v-z- bzw. w-z-Diagramms als Nutzenanalyse, wobei bei gegebener geringer Varianz im allgemeinen nur der Nutzen N berechnet werden muß.

### 2.6.3 Qualitätskriterium Effizienz

Die Erfassung der Kosten s der Lehre zusammen mit den Nutzwerten n ergibt quantitativ die "Effizienz" q von Unterrichtsmaßnahmen:

$$[7] \qquad q = \frac{n}{s} = \frac{z_o}{w_o \cdot s_o}$$

Wenn man die einzelnen Nutzenwerte (abhängige Variable $n_o = z_o : w_o$) mit den Kostenwerten (unabhängige Variable: $s_o$ = Input, Ressourcen, Strukturqualität) wiederum regressionsanalytisch in Zusammenhang bringt, erhält man eine Regressionsgerade, deren Steigung Q die "Effizienz" des Unterrichts repräsentiert.

$$[8] \qquad Q = \frac{n_o}{s_o} - a$$

Die Kosten-Nutzen-Analyse vergleicht den Knoten B oder C in Abb. 1 (Prozeß oder Ergebnis) mit dem Knoten A (Input).

Der Kosten-Nutzen-Analyse zugrunde liegt immer ein Vergleich von Alternativen. Im curricularen Bereich wird als Alternative bzw. Kontrollgruppe implizit und ohne weitere Berechnung häufig der Status quo angenommen, so daß eine echte Kontrolle nicht stattfindet. Es muß aber betont werden, daß die Evaluation E im Prinzip den expliziten Vergleich zweier Datensätze von Effizienzwerten Q darstellt, wobei Q1 die Kontroll- , Q2 die Experimentalgruppen-Effizienz und Z1 die Zielerreichungsvarianz der Kontrollgruppe darstellt.

$$[9] \qquad E = \frac{Q_1 - Q_2}{Z_1}$$

Falls keine Kontrollgruppe ($Q_1$) einbezogen werden kann wie im vorliegenden Fall, kann man sich dadurch behelfen, daß man einen Vergleich in sich vornimmt, also beispielsweise die Kosten-Nutzen-Werte vor und nach einer curricularen Reform (Intervention) vergleicht (quasi-experimentelles Design, siehe Campbell et al. 1966).

Wir definieren demnach "Evaluation" für das MCIP als Effizienzerhebung mit statistischem Effizienz-Vergleich in Form einer Längsschnittuntersuchung von Prozeß- und Ergebnisdaten, wobei je nach Datenlage unterschiedliche Versuchspläne zur Anwendung kommen (Pretest-Posttest-Design, Panel-Design, Zeitreihenanalyse: siehe Campbell et al. 1966, Wittmann 1985).

Evaluation in Form einer Kosten-Nutzen-Analyse ist die grundlegende Verfahrensweise, auf der die Sicherung bzw. das Management der Lehrqualität beruht.

## 3 Ergebnisse

Von den 150 Studenten des Praktikums im Sommersemester 1993 gaben 130 ausgefüllte Fragebogen ab, was einer Rücklaufquote von 86 % entspricht. Von den damit insgesamt 2.990 auszuwertenden Fragebögen waren schließlich 2.391 komplett auswertbar (79 % der zurückgelaufenen Bögen). Eine Totalerfassung hätte 3.450 Fragebögen ergeben. Somit wurden 69 %. aller möglichen Daten erfaßt. Die Datenbank enthält damit 35.865 Rohdaten, d.h. Einschätzungen bei den 15 Items auf der Notenskala für die einzelnen Veranstaltungsorte durch die Studenten.

### 3.1 Prozeßevaluation

Der erreichte Notendurchschnitt für alle Veranstaltungsorte beträgt 2,33 ± 0,25 und liegt damit zielführend unter dem Soll von 2,5 Skalenpunkten. Für alle Veranstaltungsorte beträgt der gemittelte Variationskoeffizient $v = 0,11$. Damit läßt sich im Gesamtpraktikum eine tolerierbare Varianz ($v < 0,33$) beobachten. In bezug auf den Zielerreichungsgrad beträgt der Durchschnitt $z = 0.07$ Skalenpunkte. Das vor Beginn der Evaluation für das Praktikum

gesteckte Ziel, einen positiven Zielerreichungsgrad über z = +0,04 Skalenpunkten zu erreichen, wurde damit erfüllt. Diese globalen Mittelwerte sind jedoch zu unspezifisch für Organisationsentscheidungen bzw. Interventionen zur Änderung des Unterrichtsprogrammes (siehe Abb. 1, Kante 5 und Knoten Ab). Deshalb wurde eine spezifischere Analyse der Unterrichtsdurchführung vorgenommen.

### 3.1.1 Inhaltliche, qualitative Analyse

Hierzu wurden anhand der z-Einzelwerte die auffälligen Veranstaltungsorte (Nr. 11, 15, 26, 28 und 30) identifiziert. Besonders eindrucksvoll sind die Veranstaltungsorte Nr. 26 und 30, da hier für den Variationskoeffizienten über alle Veranstaltungen die niedrigste Ausprägung und zugleich die geringste Zielerreichung festgestellt wird, so daß hier eine klare Indikation zu einer noch weitergehenden Analyse besteht, die dann inhaltlicher Art ist. Hierzu werden die im Fragebogen ausgefüllten Antworten auf die offene Frage herangezogen. Weitere Einzelheiten können hier aus Datenschutzgründen nicht dargestellt werden. Es sei nur soviel gezeigt: Am Arbeitsplatz 11 wurde das angekündigte Programm nicht komplett durchgeführt, am Platz 15 und 26 fiel der Unterricht zu oft aus und ein Dozent war insuffizient, am Platz 30 wurde ein Tutor gewünscht, der von den Organisatoren an diesem Platz mit Selbstlernprogrammen nicht vorgesehen war. Am besten beurteilt wurde im Praktikum die Interaktion Student/Dozent ($K_4$ = 2,05), die niedrigste Beurteilung erhielt das Kriterium 1 (Eigentätigkeit am Lernplatz).

### 3.1.2 Parametrische, quantitative Analyse

Exemplarisch für die Vorgehensweise sei die regressionsanalytische Untersuchung aller 5 Kriterien ($K_{1-5}$) für jeden Veranstaltungsort o = 1-30 dargestellt. Sie ergibt einen positiven Nutzwert von Z = 4,03, d.h. die Varianz der Zielerreichung ist im tolerablen Bereich. Der Nutzwert $N = z_o : w_o$ beträgt N = 7,20. Auch dieser Wert liegt im tolerablen Bereich, der Unterricht ist relevant. Damit liegen für das Praktikum der Chirurgie erstmals empirische Daten zur Varianz der Unterrichtsdurchführung vor.

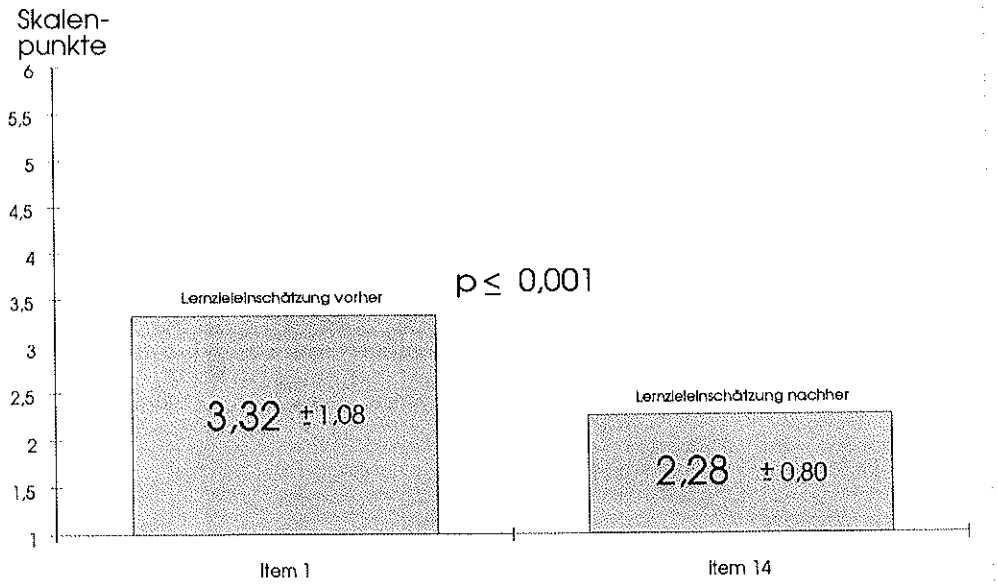
## 3.2 Ergebnisevaluation

Die Auswertung des Item 1 (Einschätzung der Kompetenz vor der Instruktion) und Item 14 (Erreichung des Lernzieles am Ende der Instruktion) zeigt einen signifikanten Unterschied ($p < 0,05$) des subjektiv wahrgenommenen

Lernerfolges im Vergleich zum Eingangswissen (Abb. 2). Der Wirkungsgrad $w_{1,14}$ (vergleiche [1]) beträgt 0,96; $z_{14}$ = 0,08 (siehe [2]) und $v_{14}$ = 0,35. Der Variationskoeffizient für Item 14 liegt damit kurz außerhalb des intendierten Grenzwertes von 0,33, die anderen Werte liegen im Toleranzbereich. Die Varianz Z14 der Lernzielerreichung liegt mit Z = 34,26 im Normbereich. Der Nutzwert N = 3,61 zeigt, daß das Kursziel in der subjektiven Einschätzung der Studenten (Kriterium 7) erreicht wurde.

Eine Kosten-Nutzen-Analyse wurde bislang aus Kostengründen bzw. wegen fehlender Manpower nicht systematisch im Praktikum der Chirurgie für die einzelnen Veranstaltungsorte durchgeführt.

Abb. 2: Studentische Einschätzung der Lernzielerreichung für alle Veranstaltungen im Praktikum Chirurgie SS 93



## 4 Diskussion

### 4.1 Diskussion der Methodik

#### Indikation der vorliegenden Studie

Die Lehrevaluation ist grundsätzlich erforderlich, da empirische Daten für qualitätsverbessernde Entscheidungen in der Lehre bislang fehlten. Die gefundenen Nutzwerte zeigen, daß die Reform der MCIP greift, daß bei weiterem untätigen Zuwarten ohne Reform und ohne Evaluation weiter Zeit verschwendet worden wäre. Damit wären weiter unnötige Kosten aufgelaufen. Insofern erscheint die vorliegende Fragestellung nach der Methodik der Evaluation gerechtfertigt. Auch angesichts des allgemeinen Interesses an der Lehre (Der Spiegel 1993, Friedman et al. 1990) erscheint sie als hochrelevant.

#### Methodik der vorliegenden Studie

Das vorliegende Untersuchungsdesign entspricht dem einer retrospektiven Fallbeschreibung des Evaluationsansatzes im MCIP. Ziel ist nicht, präskriptiv Evaluationsmethoden auf ihren Wert hin zu testen. Die vorliegende Arbeit ist also nicht als Meta-Evaluation zu verstehen. Die Meta-Evaluation des MCIP-Evaluationsansatzes wird an anderer Stelle erfolgen (Eitel et al. 1993b). Vielmehr soll die vorliegende Arbeit den Ansatz zur Diskussion stellen und im Sinne einer Hypothesendarstellung zur weiteren Evaluationsentwicklung beitragen. Damit ist das deskriptive Vorgehen gerechtfertigt. Weiteres Ziel dieser Studie ist, die Grundlage für die Generalisierbarkeit des gewählten Evaluationsansatzes durch die Diskussion zu erreichen. Insgesamt erscheint der Zugang in Form einer qualitativen Datenanalyse (Miles et al. 1985, Holman 1993) hiermit gerechtfertigt.

#### Diskussion der vorgestellten Evaluationsmethodik

Die Effektivität des aus Studentenratings stammenden Feedbacks für die Verbesserung von Curricula ist meta-analytisch nachgewiesen (Cohen 1980).

Da in der MCIP-Evaluation die Einführung einer Experimentalgruppe nicht praktikabel war, dient als impliziter Vergleich der Status quo der Lehre vor der Reform. Die Reliabilität des Fragebogens und die Konstruktvalidität der Items 3 bis 13 des Fragebogens werden an anderer Stelle überprüft (Eitel et al. 1993b).

Messung der Effektstärke, des Variationskoeffizienten und des Zielerreichungsgrades sind gängige Methoden (Hedges 1980, Thacker 1988, Drummond et al. 1992, siehe auch Eichhorn 1977), die allerdings in dieser Kombination bisher zur Untersuchung der Lehrqualität nicht angewandt wurden, wie auch die Verknüpfung von Strukturen des Qualitätsmanagements (siehe Abb. 1) mit Organisationsentscheidungen in der Lehre in der deutschsprachigen Medizin bislang nicht üblich ist. Die angewandte Evaluationsmethodik basiert auf Literaturmodellen (economic evaluation, Drummond et al. 1992). Die Zielgrößen (Qualitätskriterien) der MCIP-Evaluation entsprechen den Literaturkonzepten (problem based learning, Tresolini et al. 1990; Fallmethode, Renschler 1987; Selbstgesteuertes Lernen, Eitel 1992b, Eitel et al. 1993a). Da Effizienz der Endpunkt der MCIP-Evaluation sein wird, ist sowohl die Entwicklung der Wirkungsanalyse (cost-effectiveness-analysis, Drummond et al. 1992) als auch der Zielerreichungsanalyse (cost-utility--analysis, Drummond et al. 1992) angezeigt. Beide Verfahren sind notwendige Bedingung einer suffizienten Evaluation, da zweifellos die Zielerreichungsanalyse eine Wirkungsanalyse voraussetzt.

Der vorliegende Evaluationsansatz erfaßt bislang nur den Nutzenaspekt, nicht die Kosten der Reform. Die Evaluation ist demnach unvollständig. Es ist ein Manko der vorliegenden Studie, daß die Effizienz im pädagogischen Feld des MCIP derzeit aus Kostengründen nicht quantifiziert bzw. experimentell geprüft werden kann. Andererseits ist eine unvollständige, aber methodisch korrekte Evaluation von einem pragmatischen Standpunkt aus gesehen nützlicher als gar keine Evaluation. Der berichtete Evaluationsansatz befindet sich demnach noch in einem Entwicklungsstadium, ist aber so weit fortgeschritten, daß er diskutiert werden sollte. Die methodische Qualität der Evaluation kann nur empirisch mit quantitativen Verfahren letztlich gesichert werden.

Es sollten außerdem die Rahmenbedingungen im Sinne der Einführung einer Experimentalklausel in die ärztliche Approbationsordnung geschaffen werden, damit derartige Entwicklungen in den Fakultäten gefördert werden (Netzwerk 1993).

Dazu gehört, daß eine ins Einzelne gehende Kostenanalyse des Lehraufwandes erfolgt. Ein Vergleich der Höhe der Drittmitteleinwerbung im Vergleich zu den evaluierten, reformierten Lehrveranstaltungen in einem anderen Projekt ergibt, daß im MCIP doppelt so viele Lehrveranstaltungen bei einem Drittel an Drittmittelaufwendungen reformiert wurden, was zumindest ein qualitativer Hinweis auf die Effizienz der MCIP-Organisation ist.

Zu berücksichtigen sind neben den Kosten des Lehrprogrammes die Qualitätskosten, die durch die Evaluation als Mehrkosten entstehen. Sie werden pro Chirurgischer Klinik bei 8 Lehrpflichtveranstaltungen auf eine Sekretärin, einen technischen Assistenten und einen ärztlichen, in Epidemiologie ausgebildeten Evaluator geschätzt. Verbrauchsmaterial für Evaluation und einmalige Investitionskosten für die Datenverarbeitung als direkte Kosten dürften im Vergleich zu den Personalkosten gering sein. Die Grundausstattung muß als indirekter Kostenfaktor anteilig auf die Evaluationskosten umgelegt werden.

Diese genannten Qualitätskosten sind schätzungsweise gering im Vergleich zu dem ineffizienten Aufwand der bisherigen Lehre. Quantitative Studien der Qualitätskosten stehen allerdings noch aus.

### 4.2 Diskussion der Ergebnisse

Die Prozeßevaluation ergibt eine vertretbare Variabilität bei der Unterrichtsdurchführung. Das gesteckte Ziel, einen positiven Nutzwert zu erreichen, wurde erfüllt. Die Höhe des Nutzwertes wurde vorab nicht festgelegt, da eine empirische Grundlage für eine derartige Zielbestimmung fehlte. Weit wichtiger ist, daß der qualitative Standort des MCIP jetzt empirisch bestimmt ist und als Grundlage für einen Qualitätsfortschritt durch gezielte Organisationsmaßnahmen verwendet werden kann. Insbesondere können aufgrund der erhaltenen Nutzwerte die zukünftigen Zielwerte nun geplant werden.

Die Ergebnisevaluation zeigt einen positiven Nutzwert, womit das Ziel in bezug auf den von den Studenten selbst eingeschätzten Lernerfolgen ebenfalls erreicht ist. Die Frage ist, was dieser Befund wert ist, da er auf der subjektiven und damit unzuverlässigen Selbsteinschätzung beruht. Dieses Ergebnis sollte durch Messungen des objektiven Lernerfolges auf einer empirischen Basis geprüft werden. Hierzu käme eine Abschlußklausur in Form einer multiple-choice-Prüfung in Frage. Diese wurde deshalb nicht vorgenommen, weil Zweifel an der Validität derartiger Tests bestehen (Pauli 1985, Norcini et al. 1985). Zudem ergab ein Vergleich der Noten im II. Staatsexamen 1990 für die MCIP-Studenten tendenziell bessere Ergebnisse als in einem traditionellen Curriculum (Eitel et al. 1993b). Darüber hinaus ist das Lernziel des Kurses nicht nur der Erwerb faktischen, deklarativen Wissens, was mit den MC-Prüfungen abgeprüft werden könnte, sondern auch prozeduralen Wissens, was besser mit anderen Prüfungsmethoden erfaßt würde. Gültigere Meßmethoden wie die objective structured clinical examination (OSCE, Harden et al. 1979) oder andere Performanzmessungen

sind derzeit aus Kostengründen im MCIP nicht durchführbar. Ein entsprechender Antrag beim bisherigen Drittmittelgeber fiel durch.

Wenngleich unbestreitbar ist, daß die Endpunkte jeder medizinischen Lehrevaluation im objektivierten, deklarativen und prozeduralen Wissen bzw. in der klinischen Performanz besteht, kann doch nach unseren Beobachtungen der subjektiven Einschätzung des Lernerfolges durch die Studenten selbst eine gewisse Zuverlässigkeit bzw. Gültigkeit nicht völlig abgesprochen werden. Vor allem im fortgeschrittenen Ausbildungsstand entspricht die subjektive Selbsteinschätzung der erworbenen Kompetenz (self-assessment) objektiven Befunden (Woodward et al. 1983). Für die Beurteilung der eigenen Fertigkeiten allerdings ist die Selbsteinschätzung unzuverlässiger (Rezler 1989). Im vorliegenden Artikel sind die Ergebnisse in Form des subjektiven Lernerfolges quasi als Platzhalter mit Erklärungsfunktion aufgeführt worden, um einmal die Nachvollziehbarkeit der MCIP-Evaluation zu gewährleisten und zum anderen den Weg für die Zukunft der MCIP-Evaluation zu Protokoll zu geben.

Angesichts der Defizite in der Evaluation medizinischer, und insbesondere chirurgischer Curricula (Calhoun et al. 1986) sowie fehlender Rückkopplung von Evaluationsdaten in die curriculare Organisation (Steiner 1993) stehen auch die Fakultäten insgesamt angesichts der zunehmenden Diskussion der Hochschullehre vor der Herausforderung, hier methodische Ansätze zu entwickeln (Netzwerk 1993).

Zur Entwicklung der Evaluation gehört auch eine Änderung des Prüfungssystems, welches sich derzeit insofern nachteilig auf das Lernverhalten auswirkt, als die Medizinstudenten sich sehr um den Scheinerwerb, aber wenig um ihre Berufsqualifikation kümmern.

Die Entwicklung der Evaluation erfordert weiterhin, daß Lernziele operationalisiert werden. Hierbei wird der latente Wertkonflikt, unter dem vor allem die Medizinfakultäten stehen, zu lösen sein: Lehre mit dem Ziel "Berufsqualifikation" und Forschung mit dem Ziel "wissenschaftliche Bildung", also das Brot- und Bildungsstudium, müssen als Einheit gesehen und in den Lernzieldefinitionen der beruflichen Qualifikation und ärztlichen Bildung zusammengebracht werden. Denn eine Trennung in forschende Medizinuniversitäten und lehrende Medizin-Fachhochschulen (Akademien) ist sachfremd und zudem unbezahlbar.

### 4.5 Schlußfolgerungen

Die weitere Entwicklung der Lehrqualität auf der jetzt gewonnenen empirischen Basis wird Aufgabe eines zukünftigen Qualitätsmanagements (Berwick et al. 1990) der Lehre sein (Eitel et al. 1993b). Die Nützlichkeit der Evaluation für die weitere Entwicklung des Unterrichtsprogrammes im MCIP ist evident; diese Funktion der Evaluation stimmt mit Ergebnissen des Schrifttums überein (Marsh 1984, Cohen 1980).

Die Einrichtung von Studiendekanaten als Organe des Qualitätsmanagements für die Lehre ist das Gebot der Stunde (Eitel 1993, Netzwerk 1993). Sie könnten die notwendige Unterstützung der reformwilligen Hochschullehrer, von denen es mehr gibt als gemeinhin angenommen, vor allem in bezug auf die zur Programmentwicklung erforderliche Evaluation gewährleisten.

LITERATUR


1. Ackermann H (1992)
Bias 3.0
Biometrische Analyse von Stichproben
3. Aufl. Epsilon Verlag, Hochheim Darmstadt

2. Berwick DM, Godfrey AB, Roessner J (1990)
Curing Health Care - New Strategies for Quality Improvement
San Francisco: Jossey-Bass

3. Calhoun JG, Ten Haken JB, Bakosa D, Zelenock GB (1986)
Evaluating Performance in Surgical Education
In: Medical Education, A Surgical Perspective
RH Bartlett, GB Zelenock, WE Strodel, ML Harper, JG Turcotte (Eds.)
pp. 437-478, Chelsea (Michigan): Lewis Publishers 1986

4. Campbell DT, Stanley JG (1966)
Experimental and quasi-experimental designs for research
Rand McNally, Chigago

5. Cohen PA (1980)
Effectiveness of student-rating feedback for improving college instruction: a metaanalysis of findings
Research in Higher Education 13:321-41

6. Cronbach LJ (1963)
Course improvement through evaluation
Teach. Coll. Rec. 64:672-83

7. Drummond FM, Stoddart GL, Torrance GW (1992)
Methods for the Economic Evaluation of Health Care Programmes
Oxford Medical Publications, Oxford University Press, Oxford, New York, Toronto

8. Eichhorn S (1977)
Qualitäts- und Effizienzbeurteilung in der Krankenversorgung
Dtsch. Ärztebl. 42:2529-2533

9. Eitel F (1992)
Die Ausbildungsmisere
In: L Schweiberer, JR Izbicki (Hrsg.), Akademische Chirurgie - Aus-, Weiter- und Fortbildung - Analysen und Perspektiven
Springer: Berlin, Heidelberg, New York , S. 123-132

10. Eitel F (1992)
Wege zur problemorientierten studentischen Ausbildung und deren Evaluation
In: L Schweiberer, JR Izbicki (Hrsg.), Akademische Chirurgie
Springer: Berlin, S. 235-250

11. Eitel F (1993)
Die Studienreform ist tot, es lebe die Studienreform
Med. Ausbildung 10/2:114-122

12. Eitel F, Kanz KG, Seibold R, Sklarek J, Feuchtgruber G, Steiner B, Neumann A, Schweiberer L, Holzbach R, Prenzel M (1993)
Verbesserung des Studentenunterrichts - Sicherung der Strukturqualität medizinischer Versorgung
In: Habeck D, Schagen U, Wagner G (Hrsg.), Reform der Ärzteausbildung
Berlin: Blackwell

13. Eitel F, Prenzel M, Schweiberer L, Lyon HC (1993)
Quality assurance of education in surgery
Theoretical Surgery, in Druck

14. Friedman PC, de Blick R, Greer DS, Mennin SP, Norman GR, Sheps CG, Swanson DB, Woodward CA (1990)
Charting the Winds of Change: Evaluation Innovative Medical Curricula
Acad. Medicine 1:8-14

15. Harden RM, Gleeson FA (1979)
Assessment of clinical competence using an objective structured clinical examination (OSCE)
Med. Educ. 13:41-54

16. Hartmer M (1993)
Das große Zählen und Messen
Süddeutsche Zeitung (11.10.1993) 235, 47

17. Hedges LV (1980)
Unbiased estimation of effect size
Eval. Educ. 4:25-30

18. Holman RH (1993)
Qualitative Injury in Medical Research
J. Clin. Epidemiol. 46 (1):29-36

19. Marsh HW (1984)
Students' Evaluations of University Teaching:
Dimensionality, Reliability, Validity, Potential Biases and Utility
J. Educational Psychology 76 (5):707-754

20. Miles MB, Huberman AM (1984)
Qualitative Data Analysis
Sage, New York Park

21. Netzwerk (1993)
Positionspapier zur Reform der Medizinischen Ausbildung
Med. Ausbildung 10/2:123-138

22. Norcini JJ, Swanson DB, Grosso LJ, Webster GD (1985)
Reliability, validity and efficiency of multiple choice questions and patient management problem item formats in assessment of clinical competence
Med. Educ. 19:238-247

23. Pauli HG (1985)
Die Multiple-Choice-Methode in der Evaluation medizinischer Ausbildungsresultate
Schweiz. Ärztezeitung 66 (26):1193-1202

24. Prenzel M, Eitel F, Holzbach R, Schoenheinz RJ, Schweiberer L (1994)
Lernmotivation im studentischen Unterricht in der Chirurgie
Zeitschr. Päd. Psychologie 8, in Druck

25. Renschler HE (1987)
Definition der Fallmethode aus ihrer geschichtlichen Entwicklung in den Medizinschulen Europas
Schweiz. Rundschau Med. (Praxis) 76:981-996

26. Rezler AG (1989)
Self-assessment in Problem-based Groups
Medical Teacher 11/2:151-156

27. Der Spiegel (1993)
Welche Uni ist die beste?
Spiegel Spezial 3, Hamburg: Spiegel-Verlag

28. Steiner S (1993)
Evaluationsansätze in Medizinischen Curricula
Inauguraldissertation in Vorbereitung

29. Stufflebeam DL, Foley WJ, Gephart WJ, Guba EG, Hammond RL, Merriman HO, Provus MM (1971) Educational Evaluation and Decision Making, Peacock, Itasca, Illinois

30. Thacker SB (1988)
Meta-analysis
JAMA 259 (11):1685-89

31. Tresolini CP, Stritter FT, Hannum WH (1990)
Problem Based Learning and Instructional Theories
Medizinische Ausbildung 7:75-82

32. Wittmann WW (1985)
Evaluationsforschung
Springer, Berlin - Heidelberg - New York - Tokyo

33. Woodward CA, Ferrier BM (1983)
The contest of the medical curriculum at McMaster University: Graduates' evaluation of their preparation for postgraduate training
Med. Educ. 17:54-60

Anschrift des Verfassers:

Prof. Dr. Doria Eitel
Alfred Bräth
Prof. Dr. Leonhard Schweiberer
Ludwig-Maximilians-Universität München
Klinikum Innenstadt
Chirurgische Klinik und Poliklinik
Nußbaumstr. 20

80336 München