

Grundkonzept der formativen Evaluation des reformierten Dresdner Medizinstudiengangs am Beispiel des POL-Kurses „Nervensystem und Psyche“

I. Nitsche¹
B. Bergmann²

Fundamental Conception of Formative Evaluation of the Reformed Dresden Curriculum – The Example of the PBL-Course „Nervous System and Psyche“

Zusammenfassung

Die Medizinische Fakultät Carl Gustav Carus der Technischen Universität Dresden begann 1998 mit der Einführung problemorientierter Lehre. Für die begleitende formative Evaluation wurde ein Konzept entwickelt, welches auf einem Mehrebenenansatz beruht: Akzeptanz-, Lern- und Transfer Ebene. Auf jeder Ebene werden mehrere Zielkriterien durch standardisierte und freie Befragungen der Lernenden und der Lehrenden erhoben. Außerdem werden die Noten der Studierenden als Kriterium für Lernerfolg herangezogen. Auf diese Weise ist es möglich, sowohl die einzelnen Lehrveranstaltungen und Lehrfälle als auch den POL-Kurs in seiner Gesamtheit und im Verhältnis zu anderen Kursen des Curriculums mit einem vergleichbaren Messinstrumentarium zu bewerten. Den Programmgestaltern und -teilnehmern wird eine Rückmeldung zur Zielerreichung gegeben, wobei Stärken und Schwächen der neuen Lehrform konkret ausgewiesen sind. Auf der Basis der daraus abgeleiteten Optimierungsvorschläge werden die POL-Kurse und das gesamte Reformcurriculum schrittweise verbessert. Dieses Grundkonzept der formativen Evaluation wird exemplarisch am Beispiel des POL-Kurses „Nervensystem und Psyche“ demonstriert.

Schlüsselwörter

Formative Evaluation · Mehrebenenevaluation · problemorientiertes Lernen · Lehrfälle · POL-Kurse

Abstract

In 1998 the Medical Faculty of the Dresden University of Technology started implementing problem-based learning. At the same time a concept for an affiliated formative evaluation was developed. This is based on several levels: acceptance, success of learning, transfer. At each level defined criteria are evaluated by standardized and „open“ questions which are answered by students and teachers. The grades of the students in the examinations are an additional criterion, especially for the learning-level. With this approach it is possible to evaluate each of the paper-cases, the lectures, and the skills trainings as well as the whole PBL-course in relation to other courses of the curriculum using a comparable measurement. The planning teams and participants of the PBL-courses receive a feedback about the strengths and the weaknesses of the new learning model. This feedback includes recommendations for optimizing the further development of the individual courses and the whole reform program. The basic elements of the formative PBL-evaluation are demonstrated here by using our course „Nervous system and psyche“ as an example.

Key words

Formative evaluation · multi level evaluation · problem-based learning · paper-cases · PBL-courses

Institutsangaben

¹ Klinik und Poliklinik für Psychiatrie und Psychotherapie, Medizinische Fakultät Carl Gustav Carus der TU Dresden (komm. Direktor: Prof. Dr. med. W. Felber)

² Institut für Allgemeine Psychologie, Biopsychologie und Methoden der Psychologie der TU Dresden (Leiterin: Prof. Dr. rer. nat. habil. B. Bergmann)

Korrespondenzadresse

Prof. Dr. rer. nat. habil. Bärbel Bergmann · Institut für Allgemeine Psychologie · Biopsychologie und Methoden der Psychologie der TU Dresden · 01062 Dresden · E-mail: bergmann@psychologie.tu-dresden.de

Bibliografie

Med Ausbild 2003; 20: 212 – 217 © Georg Thieme Verlag Stuttgart · New York · ISSN 0176-4772

Grundstruktur des formativen Evaluationskonzepts

Mit der Einführung des problemorientierten Lernens in die Medizinausbildung der TU Dresden wurde ein sehr komplexes, innovatives Programm in Angriff genommen, das allen Akteuren einen beträchtlichen Lernprozess abverlangt. Eine schrittweise Optimierung mittels formativer Evaluation ist deshalb unumgänglich, wozu auf das 4-Ebenen-Modell von Kirkpatrick [1,2] zurückgegriffen wurde. Dessen vier hierarchisch angeordnete Ebenen – Reaktionen, Lernerfolg, Transfer und Resultate – erfassen wesentliche Programmkomponenten in Form einer Wirkungskette. Entsprechend der konstitutiven Elemente und Ziele problemorientierter Lehr-Lern-Konzepte wurden daraus folgende Bewertungskriterien zur formativen Evaluation des reformierten Dresdner Medizinstudiengangs (DIPOL) abgeleitet:

Reaktionen: Akzeptanz der neuen Lehrform durch die Teilnehmer; organisatorische und personelle Rahmenbedingungen.

Lernerfolg: Aneignung von Fachwissen und von Fähigkeiten/Fertigkeiten; Motivation zum Selbststudium.

Transfer: Transfererwartung bezüglich des in den Blockkursen angeeigneten Wissens bzw. der Fähigkeiten/Fertigkeiten; Leistungsentwicklung im Praktischen Jahr.

Die vierte Ebene fällt in das Ressort der summativen Evaluation und ist erst realisierbar, wenn das neue Lehr-Lern-Programm vollständig implementiert und optimiert wurde. Insofern liefert die formative Evaluation die Voraussetzung für eine spätere Outcomeprüfung.

Zur Erfassung der formulierten Bewertungskriterien wurde ein multimethodales Vorgehen gewählt. Selbsteinschätzungen durch die Lernenden stellen auf allen drei Evaluationsebenen einen wichtigen methodischen Zugang dar. Zusätzlich erfolgt eine Befragung der Lehrenden mit vergleichbaren Messinstrumenten. Das hierzu entwickelte Fragebogeninstrumentarium besteht jeweils aus einem standardisierten Teil (Items mit 6-stufiger Antwortskala: Wert 6 = vollkommene Zielerreichung; Wert 1 = Programmziel überhaupt nicht erreicht) und freien Antwortmöglichkeiten. Außerdem werden die studentischen Leistungen in den kursinternen Erfolgskontrollen und den Staatsexamina dokumentenanalytisch ermittelt und als Indikatoren für Lernerfolg einbezogen.

Mit einem solchen Evaluationskonzept ist es möglich, jeden POL-Kurs bereits nach seinem ersten Durchlauf umfassend zu bewerten, da zu diesem Zeitpunkt erfahrungsgemäß besonders starker Informationsbedarf seitens der Programmgestalter zum Grad der Kurszielerreichung besteht. Am Beispiel des ersten Blockkurses „Nervensystem und Psyche“ soll im Folgenden demonstriert werden, dass hierbei sowohl die einzelnen Lehrveranstaltungen des Kurses als auch der Kurs in seiner Gesamtheit und im Vergleich zu anderen Blockkursen evaluiert werden. Der 7,5-wöchige Kurs „Nervensystem und Psyche“ lief erstmals im Wintersemester 2001/2002 für insgesamt 142 Dresdner Medizinstudierende des siebten Fachsemesters, die in 19 Tutoriumsgruppen eingeteilt waren. Für die Mehrheit der Studierenden (81%) war dies bereits der fünfte POL-Kurs in Folge seit Eintritt in den klini-

schen Studienabschnitt, während die meisten der insgesamt 23 Tutoren (73%) erstmals in dieser Funktion tätig waren. Bewertet wurden die sechs Lehrfälle (paper-cases) des Kurses, die kursinternen Vorlesungen und Praktika sowie der gesamte Kurs unmittelbar zu seinem Abschluss.

Die Fallevaluation als Beispiel für die Lehrveranstaltungs-ebene

Zur Lehrveranstaltungsebene gehört die Evaluation der in den Tutorien erarbeiteten Lehrfälle sowie die in den POL-Kurs integrierten Vorlesungen und Praktika. Zielgruppe der Ergebnispräsentation sind in erster Linie die Fallautoren, die Tutoren sowie die Dozenten in den Vorlesungen, Praktika und Seminaren. Diese Gruppe der Programmgestalter ist in hohem Maße für die Feinplanung der Kursinhalte verantwortlich und benötigt daher eine möglichst detaillierte Rückmeldung. Deshalb erfolgt die quantitative und qualitative Datenauswertung sowie Ergebnispräsentation hier vorzugsweise auf Itemebene bzw. als wortwörtliche Wiedergabe der freien Anmerkungen. Dies soll im Folgenden an zwei kontrastierenden Fallbeispielen des POL-Kurses „Nervensystem und Psyche“ 01 demonstriert werden.

In Tab. 1 sind zunächst die einzelnen Bewertungsaspekte entsprechend ihrer Zuordnung zu den Evaluationsebenen aufgelistet, die mit dem Fallfragebogen erfasst werden. Die einfachste Auswertung ist die Darstellung und Interpretation der deskriptiven Statistik zu den einzelnen Items des Fallfragebogens. Hierzu werden die Mittelwerte, deren Konfidenzintervalle sowie die Standardabweichungen für die interessierenden Urteilergruppen berechnet. In Tab. 2 sind die Studentenwerte angegeben, aus denen hervorgeht, wie der erste Lehrfall des Kurses – der die Thematik Schlaganfall beinhaltete – beurteilt wurde. Ein Blick auf die Mittelwerte zeigt, dass die Studierenden im Allgemeinen mit der Praxisrelevanz des Themas sehr zufrieden waren. Auf der anderen Seite springt der niedrige Wert von 3,15 für die Bewertung des Verhältnisses von Stoffmenge zur verfügbaren Zeit sofort ins Auge. Fast alle anderen Indikatoren für die Qualität der

Tab. 1 Zuordnung der Fallfragebogenitems zu den Evaluationsebenen

Evaluationsebene	zu beurteilende Items	Abkürzung
Reaktionen	Verhältnis Stoffmenge zurzeit	G1 Stoffmenge – Zeit
	Nachvollziehbarkeit ärztlicher Verhaltensweisen	G2 Authentizität
	Möglichkeiten zur eigenständigen Gruppenarbeit	G3 Teamtauglichkeit
	Realitätsgehalt des Schreibstils	G4 Präsentationsstil
	Spannungsgehalt des Falles	G5 Spannungsgehalt
Lernerfolg	Klarheit der medizinischen Fachgrundprinzipien	F1 fachliche Grundprinzipien
	Schulung differenzialdiagnostischer Fähigkeiten	F2 diagnostischer Problemgehalt
	Repräsentativität des Krankheitsbildes	F3 Praxisrelevanz
	Ersichtlichkeit der Lernziele	F4 Lernzielpersichtlichkeit
	Wissenszuwachs durch die Fallbearbeitung	F5 Wissenszuwachs

Tab. 2 Deskriptive Statistik zum Fall „Schlaganfall“ (repräsentative Studentenchprobe n = 54)

	MW	KI _{uGr}	KI _{oGr}	SD
G1 Stoffmenge – Zeit	3,148	2,609	3,687	1,975
G2 Authentizität	3,778	3,375	4,180	1,475
G3 Teamtauglichkeit	4,389	4,024	4,754	1,338
G4 Präsentationsstil	4,204	3,867	4,541	1,234
G5 Spannungsgehalt	3,741	3,422	4,060	1,169
F1 fachliche Grundprinzipien	3,833	3,535	4,132	1,095
F2 diagnostischer Problemgehalt	3,426	3,101	3,751	1,191
F3 Praxisrelevanz	5,000	4,714	5,286	1,046
F4 Lernzielsichtigkeit	3,852	3,511	4,193	1,250
F5 Wissenszuwachs	3,759	3,424	4,094	1,228

Anmerkung: Es werden immer die 95%igen Konfidenzintervalle (KI) des Mittelwertes berechnet. KI_{uGr} ist der Wert für die untere und KI_{oGr} der für die obere Grenze der Intervallschätzung

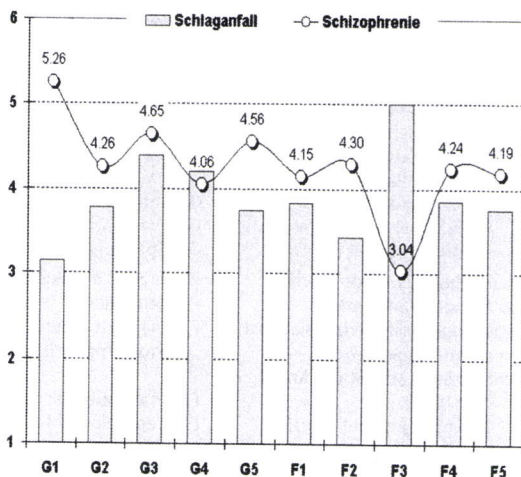
Lehrfälle liegen über 3,5 und damit im Bereich tendenzieller Zielerreichung. Dieses Ergebnis spricht zunächst dafür, dass der Fall insgesamt geeignet ist für problemorientiertes Lernen. Bei der Dateninterpretation wird jedoch immer auch die Urteilsvarianz berücksichtigt, die im vorliegenden Beispiel mit einer Standardabweichung von über 1,0 in sämtlichen zehn Bewertungskriterien angesichts einer 6-stufigen Urteilsskala relativ hoch ist. Mit Ausnahme der bereits erwähnten Praxisrelevanz ergibt sich somit für alle anderen neun Kriterien klarer Verbesserungsbedarf, da deren Bewertungs-„Spielraum“ in den Bereich der Nichtzielerreichung hineinreicht (< 3,5). Von allen sechs paper-cases des POL-Kurses „Nervensystem und Psyche“ wurde der Schlaganfallfall am kritischsten beurteilt. Die Inhaltsanalyse der freien Anmerkungen zeigte, dass der paper-case vor allem wegen seiner Anhäufung verschiedener Krankheitsbilder sowie einer wenig gelungenen inhaltlichen Verknüpfung der einzelnen Teile von Studierenden und Tutoren kritisiert worden war. Dadurch wirkte der Fall insgesamt undurchsichtig, nicht authentisch und überladen. Er konnte innerhalb der verfügbaren Zeit schlecht bewältigt werden, zumal zu Beginn des Kurses noch kein inhaltlicher Vorlauf durch die Vorlesungsreihe gegeben war. Beim inferenzstatistischen Vergleich (Willcoxon-Test) mit dem besten Fall des Kurses „Schizophrenie/Morbus Wilson“ schnitt „Schlaganfall“ in

fünf Urteilkriterien signifikant schlechter ab. In Abb. 1 sind diese Unterschiede veranschaulicht, wobei die nebenstehende Legende auch die Ergebnisse der Signifikanztests enthält. Neben dem bereits erwarteten Unterschied im Verhältnis Stoffmenge – Zeit (G1) zeigt sich unter anderem, dass Authentizität (G2), Spannungsgehalt (G5) und differenzialdiagnostische Potenzen (F2) beim Paper-case-„Schlaganfall“ noch stark ausbaufähig sind.

Aus der gesamten Datenanalyse wurde den Fallschreibern bzw. der Kursleitung folgendes Feedback gegeben: Prinzipiell ist der Schlaganfallfall als Einstieg in den POL-Kurs geeignet, da er ein typisches neurologisches Krankheitsbild schildert. Das Bestreben, damit möglichst viele weitere Diagnosen interdisziplinär zu verknüpfen, ist jedoch für den Anfangsfall eines POL-Kurses nicht zu empfehlen, da sich die Studierenden erst in das neue Themengebiet einarbeiten. Aus diesem Grund muss der Fall inhaltlich vereinfacht und stilistisch überarbeitet werden, was zugleich seine Authentizität erhöht und für ein ausgewogeneres Verhältnis von Stoffmenge und verfügbarer Zeit sorgt. Dazu erhielten die Fallschreiber die gesamte quantitative Statistik sowie die wortwörtliche Wiedergabe der freien Anmerkungen. Auf der Basis dieser Empfehlungen wurde der paper-case überarbeitet und lief im WS 2002/03 in optimierter Version. Derzeit erfolgt die Auswertung der Evaluationsergebnisse, die zeigen wird, inwieweit die vorgenommenen Verbesserungen zielführend waren. Die formative Evaluation der Vorlesungsreihe sowie der Praktika jedes POL-Kurses basiert auf den gleichen methodischen Prinzipien.

Die Kursabschlussevaluation als Beispiel für die Kursebene

Zur Kursebene gehören die Evaluation der einzelnen POL-Blockkurse mittels Kursabschlussfragebogen (KAF) und die Auswertung der studentischen Leistungen in den kursinternen Erfolgskontrollen (Triple Jump, Klausuren u. ä.). Zielgruppe der Ergebnispräsentation sind in erster Linie die Kursdirektoren und -koordinatoren, die hauptsächlich für die konzeptionelle Ausarbeitung des gesamten Blockkurses verantwortlich sind. Deshalb erfolgt die quantitative und qualitative Datenauswertung hier vorzugsweise auf Faktorebene bzw. als zusammenfassende Inhaltsanalyse der freien Anmerkungen.



Eignung für Gruppenarbeit

- G 1 Stoffmenge – Zeit
Z = -5,016 **
G 2 Authentizität
Z = -2,138 *
G 3 Teamtauglichkeit
G 4 Präsentationsstil
G 5 Spannungsgehalt
Z = -3,726 **

Fachlich-inhaltliche Eignung

- F 1 Fachl. Grundprinzipien
F 2 Diagnost. Problemgehalt
Z = -4,153 **
F 3 Praxisrelevanz
Z = -5,930 **
F 4 Lernzielsichtigkeit
F 5 Wissenszuwachs
Z = -2,499 *

Abb. 1 Vergleich der Fälle „Schlaganfall“ und „Schizophrenie/Morbus Wilson“ (Studentenmittelwerte; n = 54).

Tab. 3 Zuordnung der Kursfragebogenfaktoren zu den Evaluations-ebenen

Evaluations-ebene	KAF-Faktoren	Abkürzung
Reaktionen	Allgemeinbewertung des Kurses allgemeine Einstellung zu POL allgemeine organisatorische Rahmenbedingungen organisatorische Rahmenbe- dingungen im Tutorium Unterstützung der Tutortätigkeit Tutorperformance	F01 Globalurteil
		F08 Einstellung
		F10a Org.-Allgemein
		F10b Org.-Tutorium
		F09 Org.-Tutortätigkeit
		F06 Tutorperformance
Lernerfolg	Erwerb von Fachwissen Erwerb von kognitiven Fähig- keiten/Fertigkeiten Kommunikation/Kooperation in der Tutoriumsgruppe studentisches Engagement im Tutorium Motivation zum Selbststudium	F02a Wissen
		F02b Fähigkeiten
		F05 Kooperation
		F07 Engagement
		F03 Motivation
Transfer	Transfererwartung	F04 Transfererwartung

Anmerkung: Unterfaktor F10a wird nur von Studierenden und Faktor 09 nur von Tutoren beurteilt. Für alle anderen Bewertungsfacetten können Studenten- und Tutoreinschätzungen verglichen werden (F03 und F04 nur deskriptiv)

Tab. 3 gibt zunächst einen zusammenfassenden Überblick über die Faktorenstruktur des Kursabschlussfragebogens und deren Einordnung in das Evaluationsmodell, wie von Bergmann, Dieter u. Nitsche [3] ausführlich dargestellt. Im Gegensatz zum Fallfragebogen wird mit dem KAF immer eine Kompletterhebung durchgeführt. Die Rücklaufquote auswertbarer Fragebogen für „Nervensystem und Psyche“ 01 betrug 91% bei den Studierenden und 96% bei den Tutoren. Die Kursbewertung durch beide Urteilergruppen fiel mit Ausnahme der allgemeinen Einstellung zu POL (Tutoren mit höheren Werten; $F = 4,891^*$; $Eta^2 = 0,034$) gleich aus, weshalb im Folgenden nur auf die Evaluation durch die Studierenden Bezug genommen wird. Die grafische Veranschaulichung der Studentenbewertung (Abb. 2) informiert auf einen Blick über Stärken und Schwächen des ersten POL-Kurses

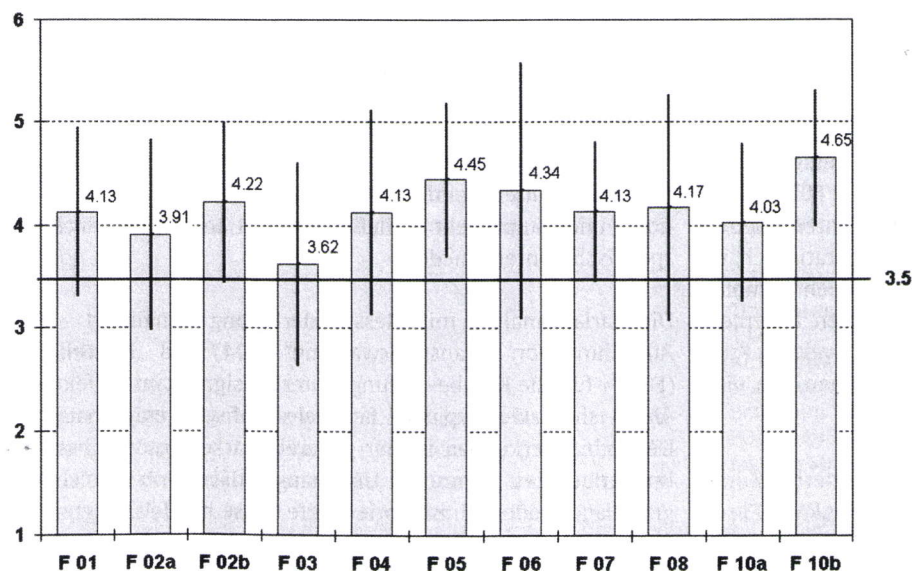


Abb. 2 Kursbewertung „Nervensystem und Psyche“ 01 durch die Studierenden ($n = 129$; Mittelwerte und Standardabweichungen).

„Nervensystem und Psyche“. Alle Faktormittelwerte liegen über 3,5. Bei Berücksichtigung der Standardabweichungen zeigt sich aber auch hier viel Verbesserungsspielraum, insbesondere in den motivationalen Kriterien (F03 und F08). Klar positiv wurden hingegen Kommunikation und Kooperation im Tutorium (F05) und die allgemeinen organisatorischen Rahmenbedingungen (F10b) beurteilt. Hierbei dürften auch Transfereffekte eine Rolle spielen, da sich deskriptiv eine stetige Verbesserung mit wachsender Anzahl absolvierter POL-Kurse nachweisen lässt. Auch die Varianzanalyse mit Messwiederholung für die 70 Studierenden, die neben „Nervensystem und Psyche“ 01 auch alle vier vorausgehenden POL-Kurse bewertet hatten, zeigt für beide Faktoren einen signifikanten Haupteffekt:

F05: $F = 6,374^{**}$, $Eta^2 = 0,105$

F10b: $F = 2,950^*$, $Eta^2 = 0,152$.

Die zusammenfassende Inhaltsanalyse der freien Anmerkungen ergab ein weitgehend komplementäres Bild zur quantitativen Datenanalyse. Gelobt wurden vor allem das Bestreben um interdisziplinäre Verknüpfung der drei im Kurs vertretenen Fachbereiche sowie das reichhaltige und zumeist gut gestaltete Angebot an Praktika und Seminaren. Verbesserungsbedarf wurde vor allem bei der inhaltlichen und organisatorischen Abstimmung der einzelnen Lehrmodule gesehen und dementsprechende Empfehlungen zur Optimierung des Kurses gegeben.

Der Semesterüberblick als Beispiel für die Curriculumsebene

Zur Curriculumsebene gehört die vergleichende Evaluation der POL-Blockkurse unter Berücksichtigung der verschiedenen Studentenjahrgänge. Dazu werden vor allem die Ergebnisse aus dem Kursabschlussfragebogen herangezogen. Zielgruppe der Ergebnispräsentation ist in erster Linie das für die gesamte Curriculumsplanung verantwortliche Studiendekanat. Die quantitative und qualitative Datenauswertung erfolgt fast ausschließlich auf Faktorebene bzw. als zusammenfassende Inhaltsanalyse zu den allgemeinen organisatorischen und personellen Rahmenbedingungen.

Tab. 4 Deskriptive Statistik zu den vier POL-Kursen des WS 2001/2002 (Studentenwerte)

		PM2 (n = 156) M	PT3 (n = 164) M	NP1 (n = 129) M	HK2 (n = 126) M	PM2 SD	PT3 SD	NP1 SD	HK2 SD
F01	Globalurteil	4,176	3,982	4,133	4,400	0,618	0,644	0,811	0,704
F02a	Wissen	3,988	4,009	3,900	4,130	0,783	0,817	0,927	0,805
F02b	Fähigkeiten	4,377	4,066	4,225	4,370	0,645	0,777	0,774	0,797
F03	Motivation	3,700	3,216	3,622	3,237	0,796	0,756	0,980	0,899
F04	Transfererwartung	4,172	4,184	4,128	4,300	0,673	0,793	0,987	0,901
F05	Kooperation	4,324	3,949	4,452	4,185	0,628	0,761	0,739	0,790
F06	Tutorperformance	4,534	4,428	4,348	4,645	0,961	0,932	1,237	0,862
F07	Engagement	3,981	3,753	4,133	3,839	0,558	0,633	0,671	0,711
F08	Einstellung	4,426	4,441	4,175	4,304	0,771	0,846	1,090	0,997
F10a	Org.-Allgemein	3,669	4,093	4,026	3,667	0,705	0,798	0,764	0,834
F10b	Org.-Tutorium	4,765	4,684	4,654	4,532	0,589	0,647	0,652	0,608

Anmerkung: PM = Pathomechanismen; PT = Pharmakotherapie; NP = Nervensystem – Psyche; HK = Herz – Kreislauf – Lunge

Als ein Beispiel für die formative Evaluation der Curriculumsebene dient die Darstellung zum Wintersemester 2001/2002 (Semesterüberblick), in dem mit „Pathomechanismen“ 02 und anschließend „Grundlagen der Pharmakotherapie“ 03 zwei eher grundlagenorientierte POL-Kurse für die Studierenden im ersten klinischen Studienjahr liefen. Zugleich standen für die Studierenden des zweiten klinischen Studienjahres folgende beide eher klinisch orientierte POL-Kurse nacheinander auf dem Programm: „Nervensystem und Psyche“ 01 und „Herz – Kreislauf – Lunge“ 02. Wie den Nummerierungen hinter den Kursbezeichnungen zu entnehmen ist, fanden fast alle Kurse bereits in zweiter bzw. dritter optimierter Version statt. Nur „Nervensystem und Psyche“ hatte seine „Premiere“ als POL-Blockkurs. Für den einfachen Semesterüberblick interessiert zunächst eine Gegenüberstellung der Bewertung für die einzelnen Kurse. Dies soll wiederum am Beispiel der studentischen Urteile illustriert werden (Tab. 4). Rein deskriptiv sind einige Gemeinsamkeiten, aber auch Unterschiede zwischen den Kursen feststellbar. So fällt beispielsweise auf, dass die allgemeine Bewertung (F01) des Herz-Kreislauf-Lunge-Kurses deutlich besser ist als für den Pharmakotherapie-Kurs. Auch beim Erwerb von Fachwissen (F02a) und Fähigkeiten/Fertigkeiten (F02b) sowie der Transfererwartung (F04) liegt „Herz – Kreislauf – Lunge“ vorn. Man könnte schlussfolgern, dass dieser bereits optimierte und klinisch orientierte Kurs besser für problemorientierte Lehre geeignet scheint als beispielsweise der grundlagenorientierte Kurs Pharmakotherapie. Es ist aber auch zu erkennen, dass die erwähnten Mittelwertsunterschiede beim Fachwissenserwerb und der Transfererwartung nur gering sind. Bei den – im Übrigen allgemein nur mäßig positiv eingeschätzten – motivationalen Bewertungskriterien (F03 und F07) gibt es auf den ersten Blick gar keine bemerkenswerten Unterschieden zwischen allen vier Kursen. Deshalb ist inferenzstatistisch zu prüfen, ob es allgemeine Unterschiede zwischen den grundlagenorientierten und den klinisch orientierten Kursen, allgemeine Unterschiede zwischen jeweils erstem und zweitem Kurs (Zeitabfolge) und spezifische Unterschiede zwischen den vier Kursen gibt.

Für den Vergleich grundlagen- vs. klinisch orientierter Kurse werden zunächst die Daten von PATHOM 02 und PHARMA 03 sowie NEUPSY 01 und HERZKL 02 über Mittelwertbildung zusammengefasst. Außerdem ist vorab zu prüfen, ob sich die beiden

Studentenjahrgänge im Ausgangsleistungsstand (Noten der Ärztlichen Vorprüfung) unterscheiden. Da dies nicht der Fall ist ($T = 0,437$ n.s.), kann vorausgesetzt werden, dass etwaige Bewertungsunterschiede nicht durch ein verschiedenes Leistungs-niveau der Studierenden verursacht sind. Der Globaltest (MANOVA) zeigt hochsignifikante Unterschiede zwischen den grundlagen- und klinisch orientierten Kursen des Wintersemesters 2001/02 an ($F = 6,050^{**}$; $Eta^2 = 0,211$). Sie konzentrieren sich jedoch nur auf folgende vier Faktoren:

- Globalurteil (F01): $F = 5,362^{*}$; $Eta^2 = 0,017$
- Kooperation (F05): $F = 8,974^{**}$; $Eta^2 = 0,028$
- Engagement (F07): $F = 4,725^{*}$; $Eta^2 = 0,015$
- Einstellung (F08): $F = 4,405^{*}$; $Eta^2 = 0,014$.

Dabei wurden die klinisch orientierten Kurse jeweils besser bewertet. Einzige Ausnahme bildet die allgemeine Einstellung zu POL, was ein Hinweis auf generelle Unterschiede zwischen den Studentengruppen ist. Hierzu kann eine Varianzanalyse mit Messwiederholung weiteren Aufschluss geben, da sie gleichzeitig überprüft, ob generelle Positionseffekte auftreten (erster vs. zweiter Kurs im Semester) und ob Wechselwirkungen zwischen der zeitlichen Kursabfolge und dem Kursinhalt feststellbar sind. Im Falle eines alleinigen signifikanten Haupteffekts kann davon ausgegangen werden, dass sich grundlagenorientierte und klinisch orientierte Kurse nicht unterscheiden. Lässt sich hingegen nur ein statistisch bedeutsamer Interaktionseffekt feststellen, ist – je nach Art der Wechselwirkung – von der Existenz solcher Unterschiede auszugehen. Dabei muss natürlich auch die verschiedene POL-Erfahrung der Studentengruppen (5. vs. 7. Fachsemester) diskutiert werden. Gleichzeitiges Auftreten von Interaktions- und Haupteffekt schließlich spricht am ehesten für kurs-spezifische Unterschiede.

Die Varianzanalyse mit Messwiederholung signalisiert – mit Ausnahme von „Transfererwartung“ (F04) und „Einstellung“ (F08) – für alle Kursbewertungskriterien signifikante Effekte. In Abb. 3 sind zwei typische Beispiele grafisch veranschaulicht. Deutlich zu erkennen ist der generell starke Motivationsabfall im Verlauf des Semesters. Unabhängig davon, ob es sich um grundlagen- oder klinisch orientierte Kurse handelt, zeigten die Studierenden bei den ersten Kursen noch wesentlich mehr Initiative zum Selbststudium als gegen Ende des Semesters. Da

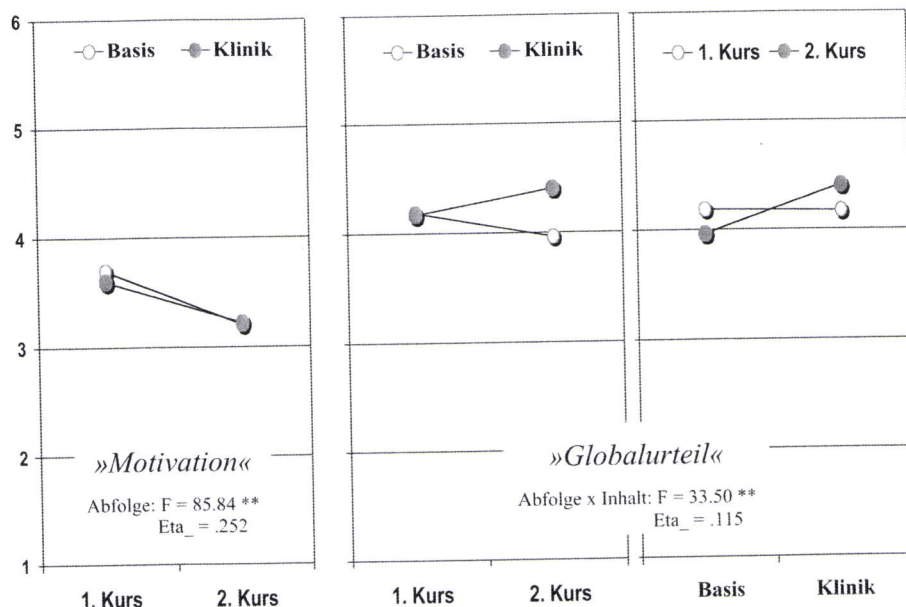


Abb. 3 Ausgewählte Ergebnisse der Varianzanalyse mit Messwiederholung (Abfolge × Inhalt).

sich ein solcher Effekt auch bei dem zweiten Motivationskriterium („Engagement“: $F = 41,93^{**}$; $Eta^2 = 0,143$) findet, deutet alles auf „POL-Müdigkeit“ im Verlauf des Semesters hin. Umso mehr sollten also seitens der Programmgestalter geeignete Maßnahmen zur Auflockerung der gewohnten Tutoriumsabfolge gegen Ende eines Semesters erwogen werden. Für das Globalurteil (F01) gilt, dass sich die jeweils ersten Kurse (PATHOM und NEUPSY) nicht unterscheiden, während der nachfolgende klinisch orientierte Kurs (HERZKL) signifikant besser und der grundlagenorientierte Kurs (PHARMA) schlechter bewertet wurde. Unter Berücksichtigung des ebenfalls signifikanten MANOVA-Ergebnisses für Faktor 01 spricht einiges dafür, dass klinisch orientierte POL-Kurse besser akzeptiert werden, zumal „Nervensystem und Psyche“ zum ersten Mal lief. Angemerkt sei, dass für den Erwerb von Fachwissen (F02a) und Fähigkeiten/Fertigkeiten (F02b) jeweils signifikante Haupt- und Interaktionseffekte feststellbar sind, so dass hier spezifische Unterschiede zwischen den vier Kursen angenommen werden können. Während bei „Fertigkeiten“ PHARMA schlechter als die anderen drei Kurse abschneidet, liegt der wesentliche Unterschied für „Wissen“ zwischen den beiden klinisch orientierten Kursen (HERZKL deutlich besser als NEUPSY). Die Gründe hierfür sind nur aus der Kenntnis jedes einzelnen Kurses und der allgemeinen Rahmenbedingungen zu beurteilen und werden auch in einer solch differenzierten Form mit den Programmgestaltern diskutiert. Mit dieser Vorgehensweise und unter Einbeziehung der Sommersemesterkurse ist es möglich, kursübergreifende Trends im Rahmen der formativen Evaluation aufzuzeigen und daraus Optimierungsvorschläge abzuleiten.

Resümee

Für innovative Programme wie die Reform des Dresdner Medizinstudiengangs steht das Ziel der Optimierung vor dem des Effektnachweises [4, 5]. Formative Evaluation ist hierfür die angemessene Verfahrensweise. Am Beispiel des POL-Kurses „Nervensystem und Psyche“ 01 wurde gezeigt, wie dabei die verschiede-

nen Evaluationsgegenstände auf den Evaluationsebenen des Kirkpatrick'schen Modells nach POL-spezifischen Kriterien bewertet werden. Für jeden Evaluationsgegenstand wurden Messinstrumente bzw. Erhebungsmethoden entwickelt, mit denen der Grad der Zielerreichung feststellbar ist. Die formative Evaluation beginnt bei der Beurteilung einzelner Lehrfälle bzw. -veranstaltungen, führt weiter über die Bewertung von POL-Blockkursen und mündet schließlich auf der Curriculumsebene in einen Kurs- und Jahrgangsvergleich. Schrittweise werden dabei alle bereits vorhandenen Ergebnisse integriert und zu einer Gesamteinschätzung verdichtet. Das heißt auch, dass die grundlegenden quantitativen und qualitativen Auswertungsmethoden je nach Evaluationsebene und -gegenstand dieselben sind, jedoch die Möglichkeiten statistischer Vergleiche mit zunehmender Fülle des Datenmaterials immer vielfältiger werden. Letztlich erfolgt eine zielgruppenspezifische Ergebnisrückmeldung an die Programmgestalter und -teilnehmer. Dabei werden positive und kritikwürdige Aspekte zu den Lehrveranstaltungen, Kursen und zum gesamten Curriculum mitgeteilt, konkrete Verbesserungsvorschläge unterbreitet und mit den Verantwortlichen diskutiert. Anzumerken ist, dass der gesamte Evaluationsprozess für die einzelnen Blockkurse wiederholt durchgeführt wird, um feststellen zu können, ob die vorgenommenen Optimierungen ihren Niederschlag in verbesserter Akzeptanz, höherem Lernerfolg und stärkerem Transfer finden.

Literatur

- Kirkpatrick DL. Evaluation of training. In: Craig RL, Bittel LR (eds): Training and development handbook. New York: AS TD/Mc Graw Hill, 1976
- Kirkpatrick DL. Evaluation training programs. The four levels (2nd ed). San Francisco: Berrett-Koehler Publishers, 1998
- Bergmann B, Dieter P, Nitsche I. Evaluation des Dresdner Modells DIPOL (Dresdner integratives problemorientiertes Lernen) des reformierten Medizinstudiums. Med Ausbild 2002; 19: 44–49
- Rossi PH, Freeman HE. Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung. Stuttgart: Enke, 1988
- Wottawa H, Thierau H. Lehrbuch Evaluation. Bern: Huber, 1998