
**Inauguraldissertation
zur Erlangung des akademischen Doktorgrades (Dr. phil.)
im Fach Psychologie
an der Fakultät für Verhaltens- und
Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg**

Titel der Dissertation
*Evaluation einer hochschuldidaktischen Weiterbildung
an der Medizinischen Fakultät Heidelberg*

vorgelegt von
Dipl.- Psych. Gerald Wibbecke

Jahr der Einreichung
2015

Dekanin: Prof. Dr. Birgit Spinath
Beraterinnen: Prof. Dr. Martin Kadmon
Prof. Dr. Birgit Spinath

Danksagung

Professorin Martina Kadmon danke ich für die Möglichkeit diese Arbeit umzusetzen. Durch die angenehme Arbeitsatmosphäre unter ihrer Leitung fühlte ich mich stets motiviert und unterstützt. Weiterhin danke ich ihr für ihre hilfreichen Rückmeldungen und ihre wertvolle Beratung bei der Entstehung dieser Dissertation.

Professorin Birgit Spinath danke ich für die Einladung in die Forschungsgemeinschaft der Pädagogischen Psychologie und für ihre freundlichen und konstruktiven Rückmeldungen bei der Erstellung dieser Arbeit.

Allen Trainerinnen und Trainern der Dozentenschulung danke ich dafür, dass ich ihre Arbeit in dieser Dissertation untersuchen durfte. Besonderer Dank gilt in diesem Zusammenhang Professor Thorsten Steiner und Dr. Eginhard Koch, welche in dem Untersuchungszeitraum die Dozentenschulung leiteten und meine Arbeit unterstützten.

Ich bedanke mich bei allen, die Teile dieser Arbeit gelesen und mir viele wertvolle Rückmeldungen gegeben haben, insbesondere Janine Kahmann für hilfreiche Diskussionen und die Unterstützung bei den Lehrhospitationen in den Beobachtungsstudien.

Bei Martina Damaschke und Dagmar Schweinfurth bedanke ich mich für ihre Unterstützung bei den Lehrhospitationen, welche weit über ihre eigentlichen Aufgaben, als wissenschaftliche Hilfskraft, hinausging.

Meiner Ehefrau Anna-Lena Wibbecke möchte ich abschließend besonders für konstruktive Diskussionen, ihre Unterstützung bei zahlreichen Lehrhospitationen und der Erstellung der Lehrvideos danken, die ich bei der Testung des BEO-LV eingesetzt habe.

Anstatt stets das generische Maskulinum oder beide Genera zu nutzen, werden in dieser Arbeit geschlechtsneutrale Bezeichnungen bevorzugt. Wenn dies aus Gründen guter Lesbarkeit nicht möglich war, wird hier das generische Maskulinum verwendet.

Inhaltsverzeichnis

1	Einleitung.....	10
2	Lehre an Hochschulen	12
2.1	Die Ausbildung der Lehrenden an Hochschulen	12
2.2	Lehrende in der Medizin	12
3	Die Ziele hochschuldidaktischer Weiterbildungen.....	15
3.1	Hochschuldidaktische Angebote in Deutschland.....	15
3.2	Wirkung von Hochschullehre	16
3.3	Die allgemeine Lehrkompetenz an Hochschulen.....	18
3.4	Studierenden- und lehrendenzentrierte Lehre	19
3.5	Die Wirkungen studierendenzentrierter Lehre.....	24
4	Die Evaluation hochschuldidaktischer Weiterbildungen.....	28
4.1	Effektivität hochschuldidaktischer Weiterbildungen.....	28
4.2	Untersuchungspläne	31
4.3	Messinstrumente	34
4.4	Implikationen für die hochschuldidaktische Forschung	40
5	Dozentenschulung der Medizinischen Fakultät Heidelberg.....	42
5.1	Inhalte, Konzepte und Ziele	42
5.2	Hypothesen zur Evaluation	46
6	Empirische Studie 1: Die Selbsteinschätzung der Lehrenden	49
6.1	Hypothesen.....	49
6.2	Methodik	51
6.2.1	Durchführung.....	51
6.2.2	Fragebogen.....	54
6.2.3	Auswertung.....	59
6.3	Ergebnisse	61
6.3.1	Deskriptive Ergebnisse	61
6.3.2	Testung auf Gruppenunterschiede in den Kontrollvariablen	65
6.3.3	Hypothesentestung.....	70
6.4	Diskussion	78
7	Empirische Studie 2: Das Lehrverhalten	85
7.1	Basisdaten Beobachtungsstudie	85
7.1.1	Fragestellung.....	85
7.1.2	Methodik.....	85

7.1.3	Ergebnisse	88
7.1.4	Diskussion	88
7.2	Entwurf eines Beobachtungsbogens	89
7.2.1	Konstruktion	89
7.2.2	Vorstudie 1: Beurteilerübereinstimmung und Interraterreliabilität	96
7.2.3	Vorstudie 2: Beurteilerübereinstimmung und interne Validität	101
7.3	Beobachtung von Lehrverhalten durch Hochschuldidaktiker	105
7.3.1	Hypothesen	106
7.3.2	Methodik	106
7.3.3	Ergebnisse	110
7.3.4	Diskussion	117
7.4	Beobachtung von Lehrverhalten durch Lehrende	119
7.4.1	Hypothesen	119
7.4.2	Methoden	120
7.4.3	Ergebnisse	124
7.4.4	Diskussion	125
8	Empirische Studie 3: Die Lehre aus Sicht der Studierenden	128
8.1	Vorstudie	128
8.1.1	Methodik	128
8.1.2	Ergebnisse	131
8.1.3	Diskussion und Skalenüberarbeitung	133
8.2	Hauptstudie	135
8.2.1	Hypothesen	135
8.2.2	Methodik	136
8.2.3	Deskriptive Ergebnisse	139
8.2.4	Hypothesentestung	142
8.2.5	Diskussion	144
9	Zusätzliche Erhebung: Ziele der Dozentenschulung aus Sicht der Trainer	148
9.1	Fragestellung	148
9.2	Methodik	148
9.3	Ergebnisse	149
9.4	Diskussion	150
10	Gesamtdiskussion	151

Literaturverzeichnis	159
Tabellenverzeichnis	178
Abbildungsverzeichnis.....	180
Abkürzungsverzeichnis.....	181
Erklärung gemäß § 8 (1) c) und d) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften.....	183
Anhang.....	185

1 Einleitung

Der Bologna-Prozess hat den europäischen Hochschulraum verändert und in Deutschland zu tiefgreifenden Reformen an den Hochschulen geführt (Jaudzims & Schnitzer, 2005). Auf europäischer Ebene herrscht Konsens, dass die Qualität der Lehre gesteigert, (Pleschová et al., 2012) und die Kompetenzorientierung an den Hochschulen gestärkt werden sollen (van der Wende, 2000). In Deutschland fordert der Wissenschaftsrat die Qualitätsverbesserung von Lehre und Studium (Wissenschaftsrat, 2005, 2008). Lehrende sollen hohes Engagement für die Lehre zeigen, hochwertige Lehrveranstaltungen durchführen und Verantwortung für den Lernprozess der Studierenden übernehmen (Wissenschaftsrat, 2008). Der *Qualitätspakt Lehre* fördert seit 2011 Maßnahmen zur Betreuung der Studierenden, der Qualifizierung des Hochschulpersonals und zur Professionalisierung der Lehre. An 44 % der deutschen Hochschulen (186 von 423) werden Projekte durch die Mittel des Bundes (insgesamt zwei Milliarden Euro) gefördert. Weiterbildungen für Lehrende haben in diesem Zusammenhang an Bedeutung gewonnen.

Hochschuldidaktische Weiterbildungen gelten als wirksam darin, die Qualität der Lehre zu steigern (Steinert et al. 2006; Stes, Min-Leliveld, Gijbels & Van Petegem, 2010). Die bestehenden Angebote und deren Wirkungen müssen genauer analysiert werden. So ist nicht gesichert, dass Lehrende ihr Lehrverhalten nach dem Besuch didaktischer Weiterbildungen verändern (Steinert et al. 2006; Stes et al., 2010). Auch liegt wenig Wissen dazu vor, ob didaktische Weiterbildungen eine länger anhaltende Wirkung haben (Stes, Clement & van Petegem, 2007; Stewart, 2014). Diese Fragen sollen in der vorliegenden Arbeit am Beispiel einer hochschuldidaktischen Weiterbildung für Lehrende an der Medizinischen Fakultät Heidelberg untersucht werden.

Die Lehre in der Medizin hat sich in den vergangenen Jahren verändert. Insbesondere die Rahmenvorgaben der Approbationsordnung für Ärzte (ÄApprO, Bundesministerium der Justiz und für Verbraucherschutz, 2002) haben das Medizinstudium reformiert und zu einer intensiveren praktischen Ausbildung geführt. In Heidelberg entstand zwischen 1999 und 2002 der reformierte medizinische Studiengang, das *Heidberger Curriculum Medicinale* (HeiCuMed). Neben der Umstellung auf ein modulares Kursrotationsprogramm, wurden neue praktische Lehr- und Prüfungsformate eingeführt (Steiner et al., 2003). Begleitend findet seitdem regelmäßig eine *Dozentenschulung* (DOS) der Medizinischen Fakultät Heidelberg statt. Die DOS soll die Lehrkompetenz der Lehrenden steigern und sie dazu motivieren, vielfältige didaktische Methoden in

ihrer Lehre umzusetzen. Damit das gelingen kann, muss die DOS selbst kritisch evaluiert werden.

Evaluationsforschung verwendet Kriterien und Standards der Wissenschaft, um die Qualität von Programmen zu bestimmen. Im Vordergrund steht der Nachweis „kausaler Aussagen über Ursachen und Wirkungsmechanismen und über die Generalisierbarkeit der Ergebnisse“ (Wittmann, 2009, S. 60). Die wissenschaftliche Evaluation didaktischer Weiterbildungen soll feststellen, ob eine Weiterbildung zu einer bestimmten Wirkung führt und analysieren, wie die gewonnenen Erkenntnisse das Wissen über die Gestaltung und die Evaluation von Weiterbildungen bereichern. In dieser Arbeit ist eine solche Programmevaluation beschrieben.

Die vorliegende Arbeit gliedert sich inhaltlich in zehn Kapitel. Der Einleitung folgen theoretische Grundlagen der Lehre an Hochschulen (Kapitel 2), hochschuldidaktischer Weiterbildungen (Kapitel 3) und der Evaluation didaktischer Weiterbildungen (Kapitel 4). Das fünfte Kapitel beschreibt die Dozentenschulung der Medizinischen Fakultät Heidelberg und die in dieser Arbeit untersuchten Hypothesen. Danach werden die durchgeführten Studien vorgestellt. In Kapitel 6 wird die Wirkung der Dozentenschulung aus Sicht der Lehrenden analysiert. In Kapitel 7 wird der Entwurf eines Beobachtungsinstrumentes dargestellt, mit dem das didaktische Verhalten Lehrender gemessen wurde. Schließlich wird in Kapitel 8 beschrieben, welche Lehre die Studierenden in der Humanmedizin bevorzugen. Daraus sollen praktische Schlussfolgerungen für die optimale Umsetzung der Lehre und der Gestaltung der DOS abgeleitet werden. Dem folgt die Befragung der Trainer zu den wichtigsten Zielen der DOS (Kapitel 9), mit der die Inhaltsvalidität der in dieser Arbeit untersuchten abhängigen Variablen belegt werden soll. In der abschließenden Gesamtdiskussion (Kapitel 10) werden die Ergebnisse zusammengefasst und in ihrer Bedeutung für Forschung und Praxis diskutiert.

2 Lehre an Hochschulen

2.1 Die Ausbildung der Lehrenden an Hochschulen

Mit der Diskussion über die Qualität der Studienbedingungen, ist die Ausbildung der Hochschullehrer in den Mittelpunkt gerückt. Lehrende werden nicht systematisch auf ihre Lehrtätigkeit vorbereitet (Helmke & Schrader, 2010; Kane, Sandretto & Heath, 2002; Michael, 2007; Webler, 2000; Winteler, 2008). Weder in Deutschland (Fendler & Gläser-Zikuda, 2013) noch in den meisten anderen Ländern (Parsons, Hill, Holland & Willis, 2012) besteht eine Verpflichtung zur didaktischen Weiterbildung. Hochschullehrer werden primär über ihre wissenschaftliche Leistung ausgewählt, unter der Annahme, dass gute Forschungsleistungen auch zu einer Weitergabe des Wissens qualifizieren (Dany, 2007; Roche & Marsh, 2000). Jedoch steht die Lehrkompetenz in keinem korrelativen Zusammenhang zu der Forschungsleistung (Hattie & Marsh, 1996). So können Lehrende ihre Lehrkompetenz nicht direkt von ihrer Forschungskompetenz ableiten (Roche & Marsh, 2000). Ihre Lehrkompetenz eignen sich Lehrende selbstständig in ihrem persönlichen Kontext an. Für diesen autodidaktischen Prozess prägte Lortie den Begriff „apprenticeship of observation“ (1975, S. 61). Angehende Lehrende in Schulen lehren so, wie sie es selbst in der Schulzeit erlebt haben. Was sie aus Beobachtung und Erfahrung gelernt haben, ist vorrangig intuitiv und wenig an didaktischen Überlegungen orientiert (Lortie, 1975). Auch Lehrende an Hochschulen führen tendenziell die Lehr- und Lernkultur fort, die sie selbst im Studium erlebt haben (Kane et al, 2002; Michael, 2007). Die individuelle Ausgestaltung der Lehre erfolgt vorrangig durch die Orientierung an Rollenmodellen und durch Selbstreflexion, stellenweise auch durch die Unterstützung von Kollegen (Futter, 2011; Winteler, 2008). Dies trifft auch auf Lehrende in der Medizin zu.

2.2 Lehrende in der Medizin

Ärzte an Universitätskliniken sind mit drei wesentlichen Dienstaufgaben beauftragt: Krankenversorgung, Forschung und Lehre (McLean, Cilliers & van Wyk, 2008). Lehrende in der Medizin bereiten sich weitgehend autodidaktisch auf ihre Lehraufgabe vor (Irby, 1994). Sowohl deutsche Ärzte an Universitätskliniken (Huwendiek, Hahn, Tönschhoff & Nikendei, 2013; Lammerding-Köppel, Fabry, Hofer, Ochsendorfer & Schirlo, 2006) als auch Ärzte in anderen Ländern (Huwendiek et al., 2010) berichten dabei von einem vergleichsweise geringeren Stellenwert der Lehraufgabe gegenüber der Krankenversorgung und der Forschung.

Die berufsqualifizierende und praktische Ausrichtung des Medizinstudiums hat zu einer höheren Vielfalt von Lehrformaten geführt, die wiederum erweiterte Anforderungen an die Lehrenden stellen (Harden & Crosby, 2000). Neben der Vorlesung und dem Seminar, welche die vorrangigen Lehrformate an deutschen Hochschulen darstellen (Seidel & Hoppert, 2011), stehen in medizinischen Fakultäten *Skills Labs* (Lernstudios), *Bedside-Teaching* (Unterweisung am Patienten), *Problemorientiertes Lernen*, *Praktika* und *Kommunikations- und Interaktionstrainings* auf dem Stundenplan (Steiner et al., 2003). Simulationsverfahren wie das Skills Lab gelten als effektiv, um klinisch-praktische Fähigkeiten zu erlernen (McGaghie, Issenberg, Petrusa & Scalese, 2010). Studierende erlernen und vertiefen hier Fähigkeiten an Modellen oder Simulationen. Dies kann frei oder unter Anleitung von Lehrenden geschehen. Beim Problem-orientierten Lernen (POL) arbeiten Studierende in Kleingruppen an praxisorientierten klinischen Problemfällen. POL wurde für die medizinische Ausbildung entwickelt und hat sich weltweit etabliert (Neville, 2009). Es gilt als effektiv darin, das Verständnis grundlegender Prinzipien zu fördern (Gijbels, Dochy, Van den Bossche & Segers, 2005). In Kommunikations- und Interaktionstrainings für Studierende der Medizin (Medi-KIT) soll Gesprächsführung im medizinisch-klinischem Kontext geübt werden (Kurtz, Silverman, Draper, van Dalen & Platt, 2005). Geschulte Schauspielpatienten konfrontieren die angehenden Ärzte mit schwierigen Situationen, in denen sie ihre Fachkenntnisse anwenden sollen. Anschließend erhalten die Studierenden von den Patienten, Kommilitonen und dem Lehrenden, der die Gruppe (ca. sechs Studierende je Gruppe) moderiert, ein Feedback zu ihrer Leistung.

Anwendungsorientierte Lehrformate wie POL und Medi-KIT betonen die Aktivierung der Studierenden. Dagegen beinhaltet eine Vorlesung in der Medizin die „zusammenhängende Darstellung und Vermittlung von wissenschaftlichen und methodischen Kenntnissen durch den Vortrag von Lehrkräften“ (ÄApprO, 2002, S. 3). Hier steht also der Impuls durch die Lehrenden im Vordergrund. Seminare sollen dagegen ausdrücklich studentische Beiträge ermöglichen: „Die Studierenden haben durch eigene Beiträge vor allem fächerübergreifende Probleme und Beziehungen zwischen medizinischen Grundlagen und klinischen Anwendungen zu verdeutlichen“ (ÄApprO, 2002, S. 3).

Kontinuierliche Weiterbildungen sollen Lehrende auf die Herausforderungen in der Lehre vorbereiten (Wissenschaftsrat, 2008). Das Interesse am medizinischer Aus-, Weiter- und Fortbildung hat im deutschsprachigen Raum deutlich zugenommen (Hu-

wendiek et al., 2013). Angesichts einer Vielzahl von neuen Lehrformaten sowie als „Folge von öffentlichem Druck, Empfehlungen des Wissenschaftsrats, Änderungen der Rahmenbedingungen (z. B. Habilitationsordnungen) und Vorleistungen in bestimmten Bereichen (Bsp. Baden-Württemberg) beginnen didaktische Qualifizierungsmaßnahmen an Wert zu gewinnen. Sie sind auf dem Weg, zur Voraussetzung für eine erfolgreiche Hochschullaufbahn zu werden“ (Lammerding-Köppel, Fabry, Hofer, Ochsdendorf & Schirlo, 2006). Didaktische Weiterbildungsangebote werden im folgenden Kapitel beschrieben.

3 Die Ziele hochschuldidaktischer Weiterbildungen

3.1 Hochschuldidaktische Angebote in Deutschland

Die Hochschuldidaktik ist eine anwendungsorientierte Disziplin, mit dem Ziel „Unterrichtspraxis zu verbessern und ein dafür geeignetes Wissens- und Handlungsrepertoire bereitzustellen“ (Helmke & Schrader, 2010, S. 273). Wissenschaftliche Hochschuldidaktik ist in verschiedene Disziplinen eingebettet und wurde bislang eher der Pädagogik zugeordnet (Leitner, 1984). Sie kann als Teil der Pädagogischen Psychologie verstanden werden (Helmke & Schrader, 2010), auch wenn sich diese im deutschsprachigen Raum vorrangig mit den Lehr- und Lernprozessen an Schulen beschäftigt hat (Lübeck, 2009; Winteler, 2006). Die empirische Fundierung der Hochschuldidaktik ist dagegen immer noch gering (Helmke & Schrader, 2010). Der Tätigkeitsschwerpunkt der Hochschuldidaktiker liegt darin, Service-Angebote im Bereich der Lehr- und Lernprofessionalisierung zu machen, nicht in der wissenschaftlichen Arbeit (Battaglia, 2010). Die Deutsche Gesellschaft für Hochschuldidaktik strebt aktuell eine stärkere wissenschaftliche Institutionalisierung mit eigenständiger Nachwuchsförderung an (Battaglia, 2010).

Forschungsbedarf besteht sowohl bei der Konzeption hochschuldidaktischer Angeboten als auch bei deren Wirkung. Die Deutsche Gesellschaft für Hochschuldidaktik (2014) berichtet auf ihrer Internetpräsenz von 87 institutionellen Mitgliedern in Deutschland, an denen Weiterbildungen entworfen und durchgeführt werden. Hochschuldidaktische Weiterbildung beinhaltet alle Vorgehensweisen, welche die Lehrkompetenz einer Person steigern können, also beispielsweise kollegiale Hospitationen, Mentorenverhältnisse durch erfahrene Kollegen, studentische Lehrevaluationen oder die Arbeit an Lehrportfolios (Webler, 2000). In dieser Arbeit wird die Bezeichnung auf Angebote beschränkt, die in Form von Trainings- oder Schulungsmaßnahmen stattfinden. Der Umfang der Angebote variiert, viele sind als Workshops von ein bis zwei Tagen konzipiert. Sie umfassen häufig eng umgrenzte Themen wie Rhetorik, Prüfungsformate und Konfliktbewältigungsstrategien, daneben existieren Angebote, die eine umfangreiche Veränderung der lehrbezogenen Einstellungen und Verhaltensweisen anstreben (vgl. Kapitel 3.4)

Der Besuch von hochschuldidaktischen Weiterbildungen ist in der Regel freiwillig (Helmke & Schrader, 2010). Die Nutzung und die Akzeptanz hochschuldidaktischer Angebote war in der Vergangenheit eher gering (Webler, 2000). Das Interesse an Weiterbildungen ist bei Lehrenden mit geringer Lehrerfahrung höher (Ulrich, Hannover,

Vervecken & Braun, 2008; Schwaiger & Salzgeber, 2004), wobei die Teilnahmeabsicht nicht unbedingt zu einer tatsächlichen Teilnahme führt (Spiel & Fischer, 1998).

Dagegen haben hochschuldidaktische Aus- und Weiterbildungsangebote in der Medizin seit den 1990er Jahren an Bedeutung gewonnen (Cohen, Murnaghan, Collins & Pratt, 2005; McLean, et al., 2008; Steinert 2014; Steinert et al., 2006). Weiterbildungen sollen sich an Lehrende verschiedenster Karrierestufen richten (Benor, 2000), selbst erfahrene Lehrende sehen für sich Qualifizierungsbedarf (Huwendiek et al., 2008). An mehreren medizinischen Fakultäten in Deutschland (z. B. in Tübingen, Mannheim und Heidelberg) sehen die Habilitationsordnungen die Teilnahme an einem zertifizierten Programm zur medizindidaktischen Qualifikation vor. Lehrende sollten eine inhaltlich breite Weiterbildung von insgesamt mindestens 200-240 Unterrichts- oder Arbeitseinheiten (von je 45 Minuten) belegen (Lammerding-Köppel, Fabry, Hofer, Ochsendorf & Schirlo, 2006). Entsprechende Angebote bestehen bereits an den meisten medizinischen Fakultäten, jedoch noch nicht in diesem Umfang. Die zugrundeliegenden Annahmen der Weiterbildungen werden in den folgenden Kapiteln beschrieben.

3.2 Wirkung von Hochschullehre

Effektive hochschuldidaktische Weiterbildungen sollten auf empirischen Erkenntnissen aus der Lehr- und Lernforschung an Hochschulen basieren. Wirkungsmodelle zur Hochschullehre (Dunkin & Barnes, 1986; Helmke & Schrader, 2010) unterscheiden vier Gruppen von Variablen: *Prozess-, Produkt und Kontextvariablen*, sowie die *Merkmale der Lehrenden und der Lernenden*. Prozessvariablen beinhalten die Verhaltensweisen der Lehrenden, Produktvariablen die Wirkungen der Lehre, die Untersuchung von Zusammenhängen derselben wird als *Prozess-Produkt-Forschung* bezeichnet (Brophy & Good, 1986). Die Kontextvariablen beinhalten die Rahmenbedingungen, in denen Lehre stattfindet, die Merkmale der Lehrenden und der Lernenden bezeichnen die individuellen Eigenschaften der handelnden Personen. Wie Abbildung 1 zeigt, ist das Feld der Lehr- und Lernforschung komplex (Helmke & Schrader, 2010).

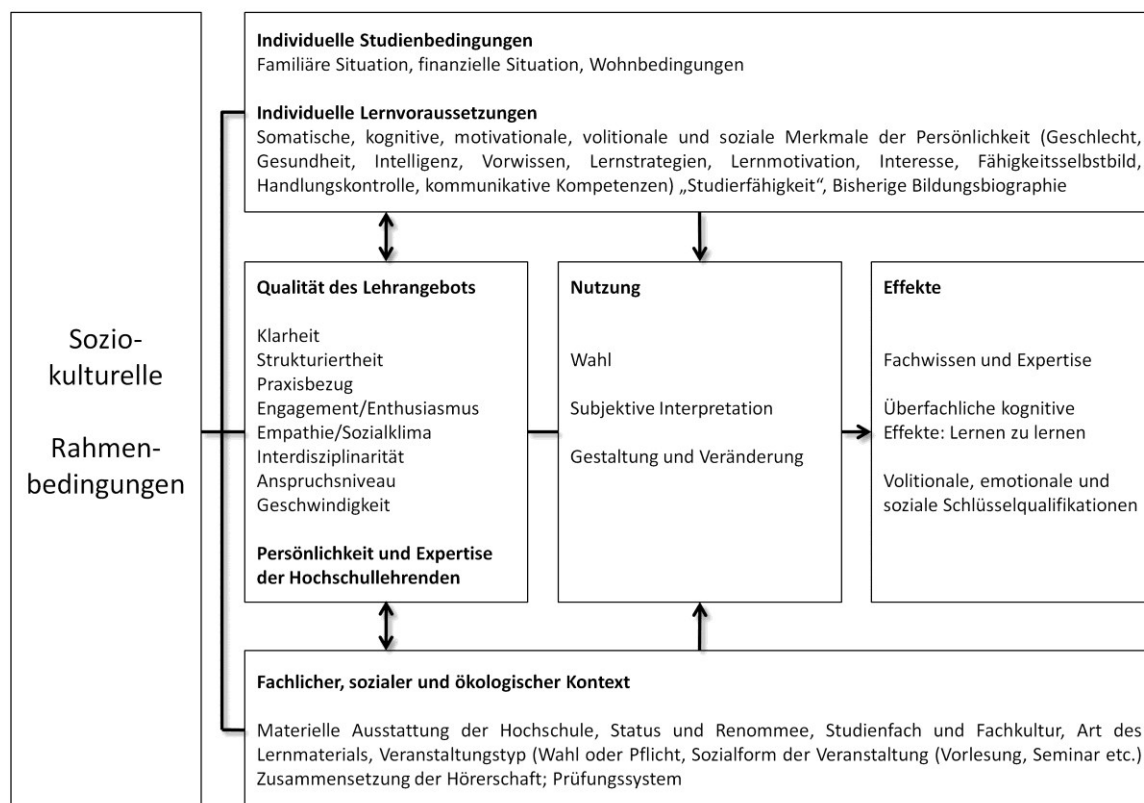


Abbildung 1. Rahmenmodell zur Wirkungsanalyse des Hochschulunterrichts nach Helmke und Schrader (2010, S. 275).

Die Qualität des Lehrangebots wird hier vorrangig durch die Lehrenden bestimmt und ist somit der Ansatzpunkt für Interventionen in der Ausbildung der Lehrenden. Lehrende bewirken Effekte, wenn ihre Angebote seitens der Studierenden angenommen werden. Ob und wie Angebote genutzt werden, wird durch die Eigenschaften der Studierenden sowie durch die Rahmenbedingungen der Lehre beeinflusst. Die primär schulische Unterrichtsforschung konnte Merkmale erfolgreicher Lehre identifizieren (Dubs, 2009; Hattie, 2009; Helmke, 2009). Dabei wurde die Bedeutung der Lehrperson für die Qualität der Lehre betont. Die Ergebnisse aus der Schulforschung sind jedoch nicht unreflektiert auf die Hochschullehre übertragbar (Helmke & Schrader, 2010). Lehrende an Hochschulen unterrichten erwachsene Studierende und haben neben der Lehre weitere Aufgaben zu erfüllen. Das folgende Kapitel beschreibt spezielle Annahmen zur allgemeinen Lehrkompetenz an Hochschulen.

3.3 Die allgemeine Lehrkompetenz an Hochschulen

Zu der allgemeinen Lehrkompetenz an Hochschulen wurden viele Kompetenzmodelle entwickelt (siehe Al-Kabbani, Trautwein & Schaper, 2013; Paetz, Ceylan, Fiehn, Schworm & Harteis, 2011). Lehrkompetenz an Hochschulen umfasst nach Paetz et al. (2011) die Lehre (v.a. Didaktische Methodenkenntnis und Fachwissen), Prüfungen (v.a. Prüfungskompetenz und die Fähigkeit Feedback zu geben) und die Akademische Selbstverwaltung (v.a. Kooperationsfähigkeit). Kompetenzmodelle sollen Grundlage für die Planung hochschuldidaktischer Weiterbildungen sein und Lehrende zur Reflexion anregen (Al-Kabbani et al., 2013). Bestehende Modelle sind eher theoretisch (z. B. Chur, 2005; Wildt, 2004) bzw. empirisch abgeleitete (z. B. Benz, 2005; Reichmann, 2008) Aufzählungen von Teilkompetenzen, die für Lehrende an Hochschulen im weitesten Sinne als wichtig erachtet werden. Diese sind beschreibend, wenig trennscharf und nehmen keine Gewichtung der Teilkompetenzen vor (Paetz et al., 2011). Neueren Konzeptionen (Trautwein und Merkt, 2012, S. 91; Paetz et al., 2011, S. 108) gelingt zwar eine genauere Differenzierung als vorherigen Modellen, detailliertere Aussagen zu Operationalisierungen von Teilkompetenzen oder über Ursache- und Wirkungsbeziehungen treffen sie jedoch nicht. Speziell für die Medizin liegt ebenfalls kein umfassendes Modell vor (Milner, Gusic & Thorndyke, 2011). Lehrkompetenz ist ein nicht einheitlich operationalisiertes Konstrukt. Offen bleibt, wie sich Lehrkompetenz entwickelt, insbesondere ob es zu verschiedenen Zeitpunkten in der Karriere von Hochschullehrern typische Herausforderungen und Bedürfnisse gibt. Dazu müssen präzise Modelle für Hochschullehrer entwickelt werden, mit denen Verläufe der Lehrkompetenz abgebildet werden können (Al-Kabbani et al., 2013).

Selbsteinschätzungen der Lehrkompetenz korrelieren positiv mit studentischen Evaluationen (Feldman, 1989; Stehle, 2011). Roche und Marsh (2000) haben das didaktische Selbstkonzept von Lehrenden an Hochschulen mit einer adaptierten Form des *Student Evaluation of Educational Quality Questionnaire* (Marsh, 1982) gemessen und dabei neun Faktoren und die Gesamtbewertung mit insgesamt 32 Items unterschieden. Gesamtbewertungen der eigenen Lehrkompetenz stehen dabei in niedrigerem Zusammenhang mit studentischen Evaluationen als spezifische Komponenten der Lehrleistung (Roche & Marsh, 2000). Wenn Lehrende bereits mehrere studentische Evaluationen ihrer Lehre erhalten haben, sind die Zusammenhänge höher (Roche & Marsh, 2000). Nach Roche und Marsh sind die Selbstkonzepte der Lehrenden bedeutend für ihre Leistung und Motivation in der Lehre und müssen daher noch differenzierter untersucht

werden. Besser als die allgemeine Lehrkompetenz sind Modellvorstellungen zu der Entwicklung von Lehr- und Lernkonzeptionen (Kember, 1997; Kugel, 1993) erforscht.

3.4 Studierenden- und lehrendenzentrierte Lehre

Lehrverhalten wird von den subjektiven Überzeugungen zu der eigenen Lehre beeinflusst (siehe Hativa & Goodyear, 2002). Diese werden in der Literatur unterschiedlich bezeichnet (siehe Lübeck, 2009; Trautwein, 2013). Etabliert haben sich die Bezeichnungen *approaches to teaching* (Lehransätze; Trigwell & Prosser, 1996) und *conceptions of teaching* (Lehrkonzeptionen; Kember, 1997). Basierend auf Lübecks Darstellung (2009) wird in dieser Arbeit folgender Zusammenhang angenommen: „Lehransätze setzen sich aus Motiven bzw. Intentionen (internale Faktoren) und Strategien (beobachtbare Anteile) zusammen und speisen sich aus Lehrkonzeptionen (den ihnen zugrunde liegenden Überzeugungen). Lehrkonzeptionen wiederum sind feinere Abstufungen von allgemeineren Lehrorientierungen“ (S. 40). Lehrorientierungen sind allgemeinere Vorstellungen der Lehrenden über Wissenserwerb und Wissensvermittlung (Braun & Hannover, 2008; Kember, 1997; Lübeck, 2009), während Lehrkonzeptionen spezifischere Auffassungen darstellen, wie Lehre in bestimmten Situationen funktioniert (Pratt, 1992; Lübeck, 2009). Die Begriffe werden nicht von allen Autoren trennscharf verwendet, Abbildung 2 zeigt den Zusammenhang von Lehrkonzeption, Lehransatz und Lehrverhalten nach Lübeck (2009, S. 40).

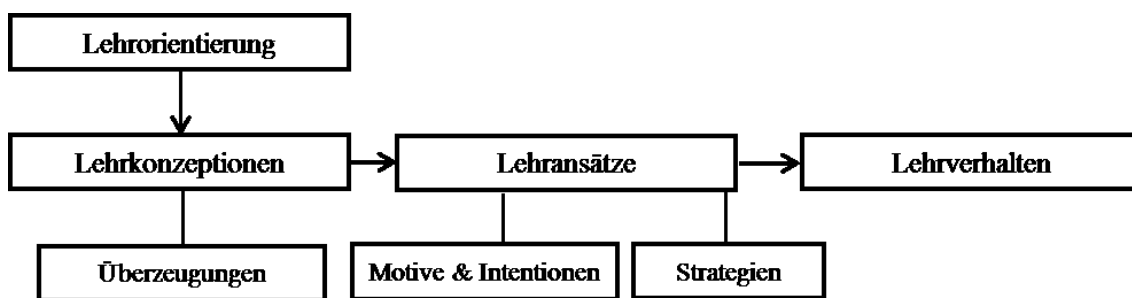


Abbildung 2. Zusammenhang von Lehrkonzeption, Lehransatz und Lehrverhalten (nach Lübeck, 2009, S. 40).

Lehransätze von Lehrenden an Hochschulen wurden relativ unabhängig von den Erkenntnissen aus der Schulforschung erforscht (Kane et al., 2002). Primär auf Interviewstudien basierend (Kember & Kwan, 2000; Samuelowicz & Bain, 1992; Trigwell, Prosser & Taylor, 1994), werden zwei Lehransätze unterschieden. Kember und Kwan

(2000, S.476) stellen die beiden Pole *Inhaltsorientierung* (*content-centered*, auch als *teacher-centered* oder *contentoriented* bezeichnet, Kember, 1997) und *Lernorientierung* (*learning-centered*, auch als *student-centered* oder *learning-oriented* bezeichnet, Kember, 1997) gegenüber: “The content-centred approach is characterised by a focus upon the material or content which is to be taught. The learning-centred approach concentrates upon the student and ensuring that appropriate learning takes place” (S. 475). Lehransätze sind nach Kember und Kwan (2000) komplexe Konstrukte, sie beinhalten eine motivationale Facette und mehrere strategische Facetten (Abbildung 3). Lehrende unterscheiden sich darin, ob sie die Motivierung der Studierenden als ihre Aufgabe betrachten und inwiefern sie Raum für studentische Individualität, Erfahrungen und Bedürfnisse in ihrer Lehre schaffen. Welche Ausprägung Lehrende zwischen den beiden Polen einnehmen, ist nach Kember und Kwan (2000) primär von der Lehrkonzeption abhängig.

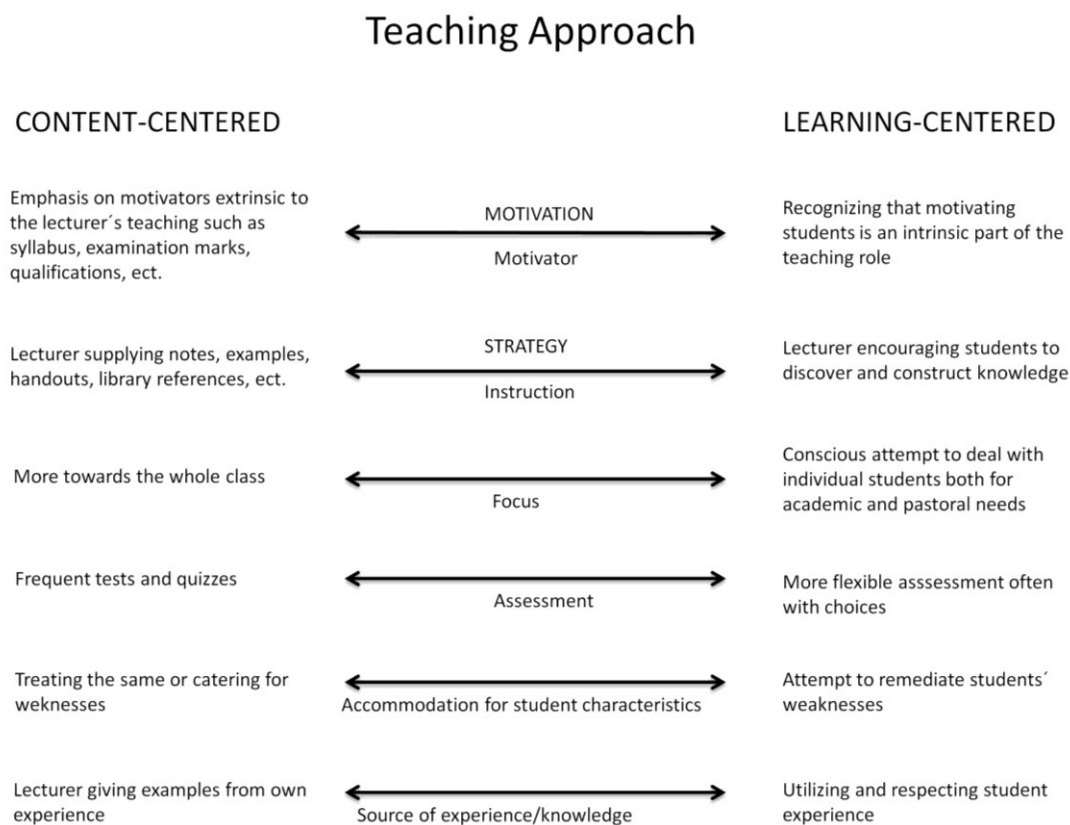


Abbildung 3. Komponenten und Dimensionen von Lehransätzen (Kember & Kwan, 2000, S. 476)

Lehrverhalten soll sich durch Lehrkonzeptionen und Lehransätze erklären, vorhersagen und modifizieren lassen (Abb. 2). Wie im Modell zu der Wirkung des Hochschulunter-

richts (Abb. 1) dargestellt, müssen Überzeugungen nicht zwingend mit Lehransätzen und Lehrhandlungen übereinstimmen. Die Rahmenbedingungen, wie das Studienfach, die Fachkultur oder das Lehrformat, können zu Inkonsistenzen zwischen Überzeugungen, Intentionen und tatsächlichem Verhalten führen (Murray & MacDonald, 1997). Die Kontextabhängigkeit von Lehransätzen entspricht den theoretischen Annahmen (Trigwell & Prosser, 2004) und konnte empirisch bestätigt werden (Lübeck, 2009). Murray und MacDonald (1997) sowie Kane et al. (2002) diskutieren als weitere Erklärung die Handlungstheorie von Argyris und Schön (1996). Danach gibt es häufig Unterschiede zwischen dem, was Mitarbeiter in Organisationen, in diesem Fall die Lehrenden, in Befragungen vorgeben zu tun und dem, was sie tatsächlich tun. Lehransätze sind insgesamt noch nicht ausreichend erforscht (Lübeck, 2009). In Kapitel 4.3.3 werden die damit verbundenen Schwierigkeiten bei der Messung von Lehransätzen dargestellt.

Didaktische Weiterbildungen sollen häufig lernorientierte bzw. studierendenzentrierte Konzeptionen und Lehransätze fördern (Postareff, Lindblom-Ylänne & Nevgi, 2007; Steinert et al., 2006; Stes et al., 2010; Winteler, 2006), damit Lehrende häufiger studierendenzentrierte Lehrmethoden einsetzen (Entwistle & Walker, 2000; Lübeck, 2009; Norton, Richardson, Hartley, Newstead & Mayes, 2005; Prosser, Trigwell & Taylor, 1994). Gleichzeitig soll eine Abkehr von einer ausschließlich inhaltszentrierten bzw. lehrendenzentrierten Wissensvermittlung erfolgen (Biggs & Tang, 2007; Macke, Hanke & Viehmann, 2008; Wahl, 2006; Winteler, 2008). Studierendenzentrierte Lehre soll selbstgesteuertes Lernen und höhere Aktivität der Lernenden in Lehrveranstaltungen bewirken (Johannes & Seidel, 2012; Kember, 1997; Lea, Stephenson & Troy, 2003; Prince, 2004). Manche Autoren erweitern den Begriff auf Bestrebungen, den Studierenden mehr grundsätzliche Wahlfreiheit und Mitsprache im universitären Lehrkontext einzuräumen (vgl. O'Neill & McMahon, 2005). In dieser Arbeit wird Studierendenzentrierung enger gefasst und auf Lehrhandlungen beschränkt, die vorrangig durch die Lehrenden initiiert und strukturiert werden (siehe Tabelle 1): Interaktion zwischen Lehrenden und Studierenden sowie Einsatz von Übungen und Arbeitsphasen in den Lehrveranstaltungen. Studierendenzentrierte Lehrhandlungen sind so in jedem Veranstaltungsformat an Hochschulen möglich.

Tabelle 1. *Beispiele für die Umsetzung studierendenzentrierter Lehre*

Beschreibung	Nennung bei
Aktives Lernen, Betonung von tiefem Lernen und Aufbau von Verständnis, Eigenverantwortung der Lernenden, Lehrende und Lernende reflektieren gemeinsam den Lernprozess.	Lea et al. (2003)
Murmelgruppen (zu zweit oder zu mehreren), Gruppenarbeit, Quiz, kurze geschriebene Selbstreflexionen, Präsentationen, Rollenspiele, Entwurf von Mindmaps, Gruppendiskussionen.	O'Neill und McMahon (2005)
„Ein hoher Sprechanteil von Studierenden, ein Schwerpunkt auf der didaktischen Phase der Elaboration sowie studierendenzentrierte Arbeitsformen (Gruppenarbeit, Stationsarbeit oder mehrere Arbeitsformen gleichzeitig)“	Johannes und Seidel (2012, S. 240)

Lehrendengesteuerte Lehre umfasst didaktische Handlungen, bei denen die Studierenden rezeptiv sind. Lehrende, die diese bevorzugen, sehen es als ihre Aufgabe, Wissen so zu präsentieren, dass es möglichst gut von den Studierenden gelernt werden kann. In der zusätzlichen individuellen Aktivierung der Studierenden sehen sie wenig Notwendigkeit (Reinmann & Mandl, 2006). Seidel und Johannes (2012) operationalisieren Lehrendenzentrierung folgendermaßen: „ein hoher Sprechanteil der Lehrperson, ein Schwerpunkt auf der didaktischen Phase der Wissensvermittlung und lehrendenzentrierte Arbeitsformen (Plenumsarbeit, Referate)“ (S.240).

Lehrhandlungen sind zu jedem Zeitpunkt entweder studierendenzentriert oder lehrendenzentriert. Die Abgrenzung von studierendenzentrierter und lehrendenzentrierter Lehre ist in der Literatur nicht immer eindeutig. Schwierig ist sie besonders dann, wenn studierendenzentrierte und lehrendenzentrierte Elemente vermengt werden. So wird beispielsweise der *fragend-entwickelnde Unterricht* als eher lehrendenzentriert betrachtet, obwohl hier relativ viel Interaktion stattfindet (vgl. Wahl, 2006). Auch die *direkte Instruktion*, als Oberbegriff für verschiedene didaktische Vorgehensweisen (Hattie, 2009), wird häufig nicht der studierendenzentrierten Lehre zugeordnet (z. B. in Moniz, Fine & Bliss, 2008), obwohl sie Elemente wie Interaktion und Übungsphasen enthalten kann. In Evaluationsstudien zu Lehransätzen müssen daher exakte Definitionen vorgenommen werden.

Die Einführung methodisch abwechslungsreicher und intensiver studierendenzentrierter Lehre wird als Paradigmenwechsel in der Lehre bezeichnet, als „Shift from teaching to learning“ (Barr & Tagg, 1995). Der Wandel „von der Wissensvermittlung zur studierendenzentrierten Lehre“ ist ein erklärtes hochschulpolitisches Ziel (Attard, Di Iorio, Geven & Santa, 2010; Hochschulenrektorenkonferenz, 2008). Dieses basiert auf der Annahme, dass studierendenzentrierte Lehre, in Einklang mit konstruktivistischen Auffassungen zum Lernprozess, den Aufbau von Wissen erleichtert und tieferes Lernen seitens der Studierenden unterstützt (Biggs, 1999; Trigwell & Prosser, 1996; Trigwell, Prosser & Waterhouse, 1999). Konstruktivistische Annahmen des Lernens basieren auf Theorien, die den aktiven Anteil eines Individuums am Lernprozess besonders hervorheben (Dewey, 1902, übers. 2002; Piaget, 1978). Auch wenn jede Form des Lernens konstruktivistisch erklärt werden kann (Renkl, 2005), gelten Lehrhandlungen dann als besonders lernförderlich, wenn sie die individuelle und selbstgesteuerte Aktivität der Lernenden unterstützen oder bewirken (Reinmann & Mandl, 2006).

Als weiteres Argument für studierendenzentrierte Lehre wird die zunehmende Bedeutung von allgemeinen Schlüsselkompetenzen angeführt. Im Qualifikationsrahmen für Deutsche Hochschulabschlüsse (Kultusminister-Konferenz, 2005) werden Qualifikationen beschrieben, die Absolventen erworben haben sollen. Studierende sollen neben dem Wissenserwerb auch kommunikative, instrumentale und systemische Kompetenzen erwerben. Das beinhaltet die Fähigkeit, Wissen anzuwenden, zu bewerten und weiterzugeben. Mit dieser „Umorientierung von Input- zu Outputorientierung“ (S. 3) rücken die zu erwerbenden Qualifikationen in den Mittelpunkt. So kann der Erwerb kommunikativer Kompetenzen voraussetzen, dass Studierende mehr Gelegenheit erhalten, eigene Wortbeiträge in Lehrveranstaltungen zu formulieren. Es besteht die Annahme, dass überfachliche Kompetenzen durch die Hochschullehre gefördert werden können (Braun & Hannover, 2011). Studierendenzentrierte Lehre kann beispielsweise soziale Kompetenzen fördern, indem sie die Studierenden zu mehr Zusammenarbeit anregt.

Wissen soll angewendet werden können. Wenn studierendenzentrierte Lehre dazu führt, dass sich auch die praktischen Kompetenzen steigern lassen, sind Absolventen besser auf das Berufsleben vorbereitet. In der Medizin soll bessere Lehre eine positive Wirkung auf die Patienten der angehenden Ärzte haben (Sullivan & Irby, 2014). Integriert in das Rahmenmodell zur Wirkungsanalyse des Hochschulunterrichts (Kapitel 3.2), liegt didaktischen Weiterbildungsangeboten die Annahme zugrunde, dass Veränderungen im Lehrverhalten, trotz der komplexen Wechselbeziehungen im Feld Hoch-

schule, zu einer signifikanten Verbesserung der studentischen Ergebniskriterien führen kann. Inwiefern diese Annahme wissenschaftlich belegt ist, wird im folgenden Kapitel dargestellt.

3.5 Die Wirkungen studierendenzentrierter Lehre

Kognitionspsychologische Befunde zeigen, dass eine aktive und tiefe Verarbeitung (Elaboration) von neuen Informationen zu besserem Behalten führt (Craik & Tulving, 1975; Slamecka & Graf, 1978). Auch das wiederholte Üben von Inhalten steigert die Gedächtnisleistung (vgl. Anderson, 2007; Goldstein, 2011). Die Kombination von studierendenzentrierter und lehrendenzentrierter Lehre sollte daher zu besseren Lernergebnissen führen, als lehrendenzentrierte Lehre alleine. Ruiz-Primo, Briggs, Iverson, Talbot und Shepard (2011) berechneten insgesamt mittlere Effektstärken auf die Lernleistung für verschiedene Formen von studierendenzentrierten Lernaktivitäten (Glass's $\Delta = 0.47$). Sie schränken ein, dass viele Studien aufgrund fehlender statistischer Kennwerte nicht berücksichtigt werden konnten. Die einbezogenen Studien weisen eher schwache Versuchspläne auf. Vortestungen fehlen häufig, Randomisierung findet meistens nicht statt und die Validität der selbstentwickelten Messinstrumente bleibt weitgehend offen. Sie fanden bei quasi-experimentellen Versuchsplänen geringere Effektstärken als bei randomisierten. Zudem traten Unterschiede bei den vier einbezogenen Fachbereichen auf, die Effekte bei Biologie und Physik waren höher als bei Chemie und Ingenieurwissenschaften. Freeman et al. (2014) berechneten in der aktuell umfassendsten Metaanalyse ebenfalls eine mittlere Effektstärke (Hedge's $g = 0.47$) auf die Prüfungsleistungen und geringere mittlere Durchfallquoten (21.8 % gegenüber 33.8 %). Sie fanden keine fachspezifischen Effekte, aber höhere Effekte in Lehrveranstaltungen mit weniger als 50 Studierenden. Zudem sahen sie höhere Leistungen in *Force Concept Inventory*-Tests (Tests zum Verständnis physikalischer Grundbegriffe) gegenüber herkömmlichen Prüfungsfragen als Beleg dafür, dass aktivierende Lehre insbesondere tieferes Verständnis begünstigt. Die methodische Qualität der Studien stand insgesamt in keinem Zusammenhang zu den Ergebnissen. Einen Publikationsbias schlossen Freeman et al. (2014) aus, da 114 Studien ohne Effekt nötig wären, um insgesamt nur kleine Effektstärken auf Prüfungsleistungen vorzufinden.

Welche Intensität und welcher Zeitumfang der Aktivierungen besonders wirksam sind, konnten Freeman et al. (2014) nicht aus den Studien ableiten, da hierzu häufig genaue Angaben fehlen. In der Regel nehmen die eingesetzten Aktivierungen nur etwa

10 – 20 % der gesamten Unterrichtszeit in Anspruch (z. B. Booth & James, 2001; Cavanagh, 2011; Deslauriers, Schelew & Wieman, 2011; Gerbig-Calcagni, 2010), gelegentlich aber auch mehr als zwei Drittel der Zeit (Saville, Zinn, & Elliot, 2005). Ebenso ist oft unklar, wie die individuellen Lern- und Arbeitsphasen in die Gesamtstruktur der Lehrveranstaltung eingebettet sind. Nach Wahl (2013) sind individuelle Lern- und Arbeitsphasen besonders wirksam, wenn sie sich in einer Lehrveranstaltung mit Phasen der kollektiven Vermittlung abwechseln und zeitlich überwiegen.

Auch die konkrete Umsetzung kann deutlich variieren. Turpen und Finkelstein (2009) stellten fest, dass eine Methode (hier: Partnergespräche in Verbindung mit Abstimmungsverfahren) von verschiedenen Lehrenden unterschiedlich umgesetzt wird. Auch Fragen durch Lehrende können ganz verschieden eingesetzt werden. Studien an Schulen zeigen, dass Fragen in der Regel eine moderat positive Wirkung auf den Lernzuwachs haben, ohne dass sich dabei ein bestimmtes Anspruchsniveau als grundsätzlich vorteilhaft erwiesen hat (siehe Good & Brophy, 1986; Hattie, 2009; Dubs, 2009). Rowe (1986) stellte fest, dass der Einsatz von Fragen durch Lehrende effektiver ist, wenn Lehrende nach einer Frage und nach einer Antwort von Lernenden mindestens drei Sekunden, statt typischerweise nur einer, vergehen lassen. Die Wirkung von Fragen im Hochschulunterricht wurde bislang jedoch wenig untersucht (Levin, 2005).

Angesichts der Vielfalt aktivierender Methoden (vgl. Kapitel 3.4) und der komplexen Interaktion verschiedener Faktoren (vgl. Kapitel 3.2) stellt sich die Frage, unter welchen Umständen studierendenzentrierte Lehre besonders effektiv ist. Andrews, Leonard, Colgrove und Kalinowski (2011) sind der Ansicht, nachdem sie selbst keine positive Wirkung feststellen konnten, dass viele Lehrenden aktivierende Methoden nicht effektiv umsetzen. Publierte Studien seien positiv verzerrt, da die Lehrenden, welche die Interventionen durchführen, über eine überdurchschnittliche Expertise in der Didaktik verfügten. Lehrende, die weniger erfahren sind, erzielen nur geringe Effekte mit aktivierenden Methoden. Die Autoren schlussfolgerten, dass aktivierende Lehre nur dann erfolgreich ist, wenn sie gut umgesetzt wird.

Studierendenzentrierte Lehre sollte nur eine höhere Wirksamkeit gegenüber traditionellen Unterricht haben, wenn sie die Aufmerksamkeit und die aktive Auseinandersetzung mit den Lerninhalten erhöht. Es gibt wenige Studien, die das Eintreten dieser Voraussetzung überprüft haben. Die Variable *Aufmerksamkeit in Lehrveranstaltungen* ist nicht leicht zu erfassen (Wilson & Korn, 2007). Gerbig-Calcagni (2010) untersuchte die Aufmerksamkeit von Studierenden in Abhängigkeit davon, ob sie mit oder ohne

aktive Verarbeitungsphasen unterrichtet wurden. Weder die beobachtete noch die selbstberichtete Aufmerksamkeit war in der Gruppe mit Aktivierungen signifikant höher. Die Studierenden, die mit aktivierenden Lehrmethoden unterrichtet wurden, erzielten auch keine besseren Behaltensleistungen, allerdings schätzten sie ihre subjektive Behaltensleistung vergleichsweise etwas höher ein. Die Annahme, dass studierendenzentrierte Lehre zu tieferer kognitiver Verarbeitung führt, wurde in Fragebogenstudien zu dem Einsatz von Lernstrategien untersucht und ließ sich nicht eindeutig nachweisen. Booth und James (2001) konnten beispielsweise nach dem Einsatz aktivierender Lehrmethoden keine tiefere Verarbeitung bei den Studierenden messen. Mehrere Studien zeigen sogar gegenläufige Befunde auf, wie, dass studierendenzentrierte Lehre mit mehr oberflächlichem Lernen (Baeten, Dochy & Struyven, 2008; Gijbels & Dochy, 2006) oder weniger tiefen Lernansätzen (Groves, 2005; Reid, Duvall & Evans, 2005) einherging.

Baeten, Kyndt, Struyven und Dochy (2010) sehen eine Vielzahl von Faktoren, die einen Einfluss auf den Zusammenhang von studierendenzentrierten Lehrangeboten und den studentischen Lernansätzen haben. Sie unterscheiden in einem Review Kontextfaktoren (z. B. Lehrmethoden, Fachdisziplin, Prüfungsformate), die Wahrnehmung der Lehre durch die Studierenden (z. B. die subjektive Schwierigkeit und Klarheit der Anforderungen) und Faktoren auf Seiten der Studierenden (z. B. Alter, Lernmotivation und Persönlichkeitseigenschaften). Beispielsweise sind die wahrgenommenen Prüfungsanforderungen relevant: Wenn die Studierenden mit oberflächlichen Lernstrategien gute Erfolge erzielen können, sei der Einsatz tiefer Lernstrategien weniger wichtig.

Neben der effektiven Umsetzung und passenden Rahmenbedingungen ist die Akzeptanz der Studierenden eine Voraussetzung für die Wirksamkeit studierendenzentrierter Lehre. Lehrende sehen in der Ablehnung studierendenzentrierter Lehre durch die Studierenden ein Hindernis bei deren Umsetzung (Michael, 2007). Studierende der Medizin, der Wirtschaftswissenschaften und der Psychologie lehnen sowohl stark lehrendenzentrierte Lehrformate (wie Vorlesungen ohne studentische Beteiligungsmöglichkeit) als auch stark studierendenzentrierte Formate (Rollenspiele und studentische Präsentationen) eher ab (Sander, Stevenson, King & Coates, 2000). Am meisten Zustimmung erhielten interaktive Vorlesungen, in denen Lehrende präsentieren, gelegentlich Fragen stellen und Aufgaben zur Bearbeitung geben. Nach Elen, Clarebout, Léonarda und Lowyck (2007) befürworten Studierende sowohl studierendenzentrierte, als auch lehrendenzentrierte Merkmale der Lehre. Bei Pauler und Jucks (2013) beendeten Stu-

dierende den Satzanfang „Für gute Lehre ist es aus meiner Sicht wichtig, dass ...“ häufiger mit einer lehrerzentrierten Formulierung weiter als mit studierendenzentrierten. In einer Studie von Smith und Cardaciotto (2011) bewerteten Studierende Lernaktivitäten mit höherer Aktivierung weniger positiv als Lernaktivitäten mit geringerer Aktivierung, trotz besserer Behaltensleistungen. Konzepte wie *inter teaching* (Textbasierte Kleingruppenarbeit in Seminaren, vgl. Boyce & Hine, 2002) wurden dagegen von Studierenden gegenüber traditionellen Lehrmethoden in Untersuchungen bevorzugt (Saville, Zinn & Elliott, 2005; Saville, Zinn, Neef, Norman & Ferreri, 2006).

Verschiedene Formen studierendenzentrierter Lehre werden unterschiedlich bewertet. Böttger, Gien und Müller (2011) fanden, dass Studierende in Vorlesungen den Vortrag der Lehrperson, ebenso wie reflexive Phasen und Diskussionen als eher sinnvoll ansahen. Gruppen-, Partner- und Einzelarbeit sowie Referate sahen sie dagegen als eher nicht sinnvoll an. In Seminaren waren dagegen Diskussionen und reflexive Phasen höher angesehen, während Vorträge, Einzelarbeiten und Referate als weniger sinnvoll betrachtet wurden. Angesichts der breiten Auslegung des Begriffs Studierendenzentrierung, ist es also wichtig zu unterscheiden, welche Formen studierendenzentrierter Lehre positiv oder negativ innerhalb bestimmter Lehrformate bewertet werden.

Der aktuelle Forschungsstand lässt offen, wie die studentischen Wahrnehmungen aktivierender Lehrmethoden erklärt werden können, wie veränderbar sie sind und ob es zwischen Studierenden unterschiedlicher Studiengänge bedeutsame Unterschiede gibt. Die Befunde deuten darauf hin, dass studierendenzentrierte Lehre für den jeweiligen Kontext passend eingesetzt werden muss. Ist das nicht der Fall, kann die Ablehnung studierendenzentrierter Lehre deren intendierte Wirkung verhindern. Studien zur Evaluation hochschuldidaktischer Weiterbildungen sollten daher die Wahrnehmung der Studierenden berücksichtigen.

4 Die Evaluation hochschuldidaktischer Weiterbildungen

4.1 Effektivität hochschuldidaktischer Weiterbildungen

Mit der steigenden Zahl von hochschuldidaktischen Weiterbildungen hat deren Begleitforschung an Bedeutung gewonnen. Evaluationen von Weiterbildungen untersuchen, ob die angestrebten Ziele erreicht werden. Auch die möglicherweise unbeabsichtigten Wirkungen sollen erfasst werden. Als *formative* Evaluation (Scriven, 1972) können sie Hinweise zur Optimierung der Weiterbildung geben. Auch sollten Evaluationsstudien Ziele und Inhalte von Weiterbildungen einer kritischen Prüfung unterziehen (Scriven, 1972).

Häufig fehlen methodisch angemessene Evaluationen von Personalentwicklungsprogrammen: „Wenn Evaluationsstudien durchgeführt werden, dann enthalten vorherrschende Designs meist einfache Prä-Post-Trainings ohne Kontrollgruppen oder basieren auf subjektiven Einschätzungen der Teilnehmer nach einem Training“ (Sonntag, 2002, S. 70). Dieser Befund lässt sich auf hochschuldidaktische Weiterbildungen generalisieren. Reviews (Levinson-Rose & Menges, 1981; McAlpine, 2003; Parsons, Hill, Holland & Willis, 2012; Prebble et al., 2004; Steinert et al. 2006; Stes et al., 2010; Weimer & Lenze, 1997) zeigen seit 1981, dass die Durchführung von Evaluationsstudien mit starken Forschungsdesigns eine Herausforderung für die Hochschuldidaktik darstellt. Die folgenden Kapitel stellen den aktuellen Forschungsstand dar, auf dessen Basis Implikationen für die Evaluation didaktischer Weiterbildungen abgeleitet werden.

Ob eine Weiterbildung erfolgreich ist, kann anhand unterschiedlicher Kriterien bewertet werden. Verschiedene Ebenen von Ergebniskriterien werden zur Evaluation von Weiterbildungsmaßnahmen herangezogen (Steinert et al.; 2006; Stes et al., 2010). Steinert et al. (2006) haben das Konzept von Kirkpatrick (1967, 1994) erweitert und auf didaktische Weiterbildungen übertragen. Die erste Ebene, die *Reaktion*, beschreibt die Wahrnehmungen der Programmteilnehmer zu den Rahmenbedingungen, Inhalten, Methoden und den Anbietern der Weiterbildung. Hier geht es insgesamt um die Zufriedenheit mit einer Weiterbildung. Die Ebene *Lernen* betrachtet den Zuwachs an Wissen und Fähigkeiten seitens der Teilnehmer und mögliche Einstellungsänderungen. Die dritte Ebene, das *Verhalten*, bzw. der Transfererfolg, überprüft die Umsetzung des Gelernten in die Praxis. Steinert et al. (2006) erweitern diese Kategorie um die Bereitschaft der Teilnehmer, neues Wissen und neue Fähigkeiten praktisch umzusetzen. Die Ebene der *Resultate* beinhaltet die Auswirkung der Weiterbildung auf die Hochschule als Organisation und auf die Studierenden. Kirkpatrick nimmt an, dass ein Zusammenhang zwi-

schen den Ebenen besteht. So sei bereits die Zufriedenheit eine wichtige Voraussetzung für den Erfolg der Weiterbildung (Kirkpatrick & Kirkpatrick, 2006). Wenn sich das Wissen, die Fähigkeiten oder die Einstellungen positiv verändert haben, sind Verhaltensänderungen und bessere Resultate wahrscheinlicher. Diese Wirkungszusammenhänge konnten metaanalytisch jedoch nicht nachgewiesen werden (Alliger, Tannenbaum, Bennett & Traver, 1997). Steinert et al. (2006) verwenden die Ebenen in ihrem Review daher lediglich für eine umfassende Darstellung der Programmwirkungen.

Stes et al. (2010) lehnen ihre Darstellung ebenfalls an das Konzept von Kirkpatrick an. Sie verzichten aber auf die Analyse von Studien, welche die Ebene Reaktion erhoben haben, da diese für sie kaum Relevanz auf die Auswirkungen der Weiterbildungen haben. Die empirischen Studien in ihrem Review ordnen sie den Kriterien *Change within teachers*, *Change within students* und *Institutional impact* zu. Sie berichten von 36 Studien, in denen 108 Analysen auf den unterschiedlichen Unterebenen durchgeführt wurden (Tabelle 2).

Tabelle 2. Zielebenen, Forschungsmethodik und Ergebnisse in Studien zur Wirksamkeit didaktischer Weiterbildungen (Angaben nach Stes et al., 2010)

Zielebenen	Analysen insgesamt	Forschungsmethodik		
		Quantitativ	Qualitativ	MM
<i>Change within teachers:</i>				
Change in attitudes	16 (14)	6 (4)	6 (6)	4 (4)
Change in conceptions	8 (7)	4 (4)	4 (3)	0
Change in knowledge	12 (12)	5 (5)	4 (4)	3 (3)
Change in skills	13 (13)	5 (5)	5 (5)	3 (3)
Behavior	31 (29)	12 (11)	13 (12)	6 (6)
<i>Institutional impact</i>				
	9 (9)	1 (1)	7 (7)	1 (1)
<i>Change within students:</i>				
Change in perceptions	7 (7)	5 (5)	2 (2)	0
Change in study approaches	4 (3)	3 (2)	1 (1)	0
Change in learning outcomes	8 (8)	5 (5)	2 (2)	1 (1)

Anmerkungen. MM = Mixed-Methods. Die Zahl in Klammer steht für die Anzahl der Studien, die laut Stes et al. (2010) zumindest teilweise Evidenz aufgezeigt haben.

Es finden verschiedene Forschungsansätze Anwendung. Mehrere Studien wählen qualitative Methoden oder einen Mixed-Methods Ansatz. Der verbreitete Einsatz qualitativer Forschung lässt sich mit den Besonderheiten der Studien begründen. Häufig liegen hier geringe Fallzahlen vor, so dass die Voraussetzungen für quantitative Analyse nicht gegeben sind, bzw. nur eine sehr geringe statistische Teststärke vorliegt. Auch existieren kaum standardisierte, quantitative Messinstrumente. Es besteht die Annahme, dass qualitative Forschung gut geeignet ist, individuelle Veränderungsprozesse abzubilden. Dementsprechend wird die Bedeutung qualitativer Methodik in der Evaluation didaktischer Programme wiederkehrend betont (Levinson-Rose & Menges, 1981; Steinert et al., 2006; Stes, et al., 2010).

Stes et al. (2010) fanden in den Studien mehrheitlich Belege für die Wirksamkeit der Weiterbildungen. Tendenziell fanden sie weiterhin eine höhere Wirksamkeit von Weiterbildungen, die sich über mehrere Tage erstrecken. Es machte keinen Unterschied ob die Weiterbildungen in Kursform oder in alternativen Formaten stattfanden, oder ob sie fachspezifisch bzw. fachübergreifend ausgerichtet waren. Spencer (2014) fasst die Ergebnisse von vier Reviews zusammen (Amundsen & Wilson, 2012; Steinert et al., 2006, Steinert, Naismith & Mann, 2012; Stes et al., 2010), in denen die Merkmale effektiver Weiterbildungen abgeleitet wurden (Tabelle 3). Diese Aufstellung dient als Anhaltspunkt, wie die Merkmale der Programme in Studien beschrieben werden können, häufig fehlt eine exakte Beschreibung der Interventionen (Stes et al., 2010).

Tabelle 3. *Schlüsselmerkmale effektiver Weiterbildungen*

Merkmal
Zeitlich ausgedehnte Weiterbildungen
Institutionelle Unterstützung des Programms
Bedarfsorientierte Interventionen
Vielfältiger Methodeneinsatz innerhalb einer Weiterbildung
Projektarbeit ermöglichen (Individuell oder in Gruppen)
Möglichkeit schaffen, neue Methoden praktisch anzuwenden
Kollegiale Zusammenarbeit ermöglichen
Kollegiale Hospitationen und kollegiales Feedback ermöglichen
Feedback zu der Lehrkompetenz geben

Anmerkung. Aufzählung nach Spencer (2014, S. 358).

Die Aufzählung nach Spencer (2014) gibt einen allgemeinen Hinweis auf effektive Elemente von Weiterbildungen, jedoch bleibt offen, wie sich die Elemente ergänzen sollten und welcher zeitliche Umfang dabei notwendig ist. Die Darstellung legt nahe, dass die in dieser Arbeit untersuchte Weiterbildung (vgl. Kapitel 5) eine größere Wirksamkeit erzielen sollte, wenn sie möglichst viele erfolgreiche Elemente nach Spencer (2014) enthält. Um die zu erwartenden Effekte zu bestimmen, soll davor die empirische Grundlage der bestehenden Wirksamkeitsforschung dargestellt werden. Evaluationsstudien unterscheiden sich stark in ihrer theoretischen Fundierung, der Forschungsmethodik, den Untersuchungsplänen, den abhängigen Variablen und den verwendeten Messinstrumente. Deshalb geht das folgende Kapitel genauer auf die Unterschiede der Untersuchungspläne ein.

4.2 Untersuchungspläne

Neben der Beschreibung der Zielebenen und der Wahl der Methodik unterscheiden sich Studien darin, welche Untersuchungspläne Verwendung finden. Rost (2007, S.116) unterscheidet vier Ansätze (Tabelle 4) zur Beschreibung von Untersuchungsplänen.

Tabelle 4. *Untersuchungspläne nach Rost (2007)*

Mindestens zwei unterschiedliche Gruppen	Mit Randomisierung:	Experimentelles Design mit EG und KG
	Ohne Randomisierung:	„Quasi“- experimentelles Design mit VSG und VGG
Nur eine behandelte/mehrfach gemessene Gruppe:		„Vor“-experimentelles Design
Eine einmalige Erhebung bei einer Gruppe:		Deskriptives Design

Anmerkungen. EG = Experimentalgruppe. KG = Kontrollgruppe. VSG = Versuchsgruppe. VGG = Vergleichsgruppe.

Ein experimentelles Design ist vorhanden, wenn es gelingt, alle potentiellen Teilnehmer nach dem Zufallsprinzip entweder der Weiterbildung oder einer Kontrollbedingung zuzuweisen. Die Gruppen sollten für eine gute interne Validität etwa 50 Personen umfassen (Rost, 2007). In der Evaluation didaktischer Weiterbildungen sind diese Voraussetzungen jedoch kaum zu erfüllen. Der Zugang zu einer Weiterbildungsmaßnahme lässt sich nicht verweigern, erst recht nicht, wenn diese verpflichtend ist. Zudem ist die freie

Nachfrage nach Weiterbildungen selten ausreichend groß für eine effektive Randomisierung.

Problematisch an einem quasiexperimentellen Versuchsplan ist, dass sich keine kausalen Schlüsse ziehen lassen (Rost, 2007). Wenn Teilnehmer in der Versuchsgruppe nach der Weiterbildung beispielsweise eine höhere Studierendenzentrierung berichten, muss das keine Wirkung darstellen. Möglicherweise haben die Personen diese aufgrund ihres Interesses an studierendenzentrierter Lehre besucht und weisen von Anfang an eine höhere Ausprägung auf. Eine Verbesserung des Vorgehens besteht darin, beide Gruppen einer Vortestung zu unterziehen. Besonders bei kleinen, nicht randomisierten Stichproben, können Unterschiede bezüglich der interessierenden Variablen auftreten. Diese können so bei der Interpretation der Ergebnisse berücksichtigt werden. Sie dienen bei Verlusten von Studienteilnehmern zusätzlich als Grundlage für Vergleiche zwischen denen, die ausscheiden und solchen, die bleiben.

Doch eine Vortestung reicht nicht aus. Bei fehlender Randomisierung können die Merkmale der nicht randomisierten Gruppenzugehörigkeit mit der Entwicklung zwischen den Messzeitpunkten interagieren (Rost, 2007). So kann die Studierendenzentrierung bei zwei Gruppen vor einer Weiterbildung identisch sein. Wenn in einer Studie zu Trainingseffekten ein Effekt bei der Versuchsgruppe beobachtet wird, dann kann diese Entwicklung unabhängig von der Weiterbildung stattgefunden haben. Engagierte Lehrende können sich selbst Wissen angeeignet haben oder zu der Entscheidung gekommen sein, einen anderen Lehransatz zu verfolgen, während sich die Vergleichsgruppe in diesem Zeitraum nicht intensiv mit der Weiterentwicklung der eigenen Lehre auseinandergesetzt hat. Dies würde zu einer Überschätzung der Wirksamkeit der Weiterbildung führen.

Zusätzlich können sich Unterschiede zugunsten der Versuchsgruppe durch sozial erwünschte Antworten ergeben. Da Studien zur Weiterbildungsevaluation häufig durch die Programmgestalter selbst durchgeführt werden, gefährdet dies die interne Validität der Ergebnisse. Wenn Studierendenzentrierung positiv dargestellt wird, ist die Intention entsprechender Fragebögen leicht zu durchschauen (Desimone, 2009). Diese Limitation muss durch möglichst wertneutrale Messinstrumente minimiert werden.

Nichtrandomisierte Zwei-Gruppenpläne mit Vortest sind trotz dieser Limitationen in den anwendungsbasierten Feldern der Psychologie weit verbreitet (Rost, 2007). Sie sind dort die beste Lösung, wo keine Randomisierung möglich ist.

Deskriptive und vor- experimentelle Untersuchungspläne, also Untersuchungen ohne Vergleichsgruppe, sind nach Rost „wissenschaftlich wertlos und können, wenn überhaupt, nur rein illustrativen Zwecken dienen“ (2007, S. 117–118). Wenn in einer Studie nur die Teilnehmer der Weiterbildung getestet werden, ist es schwierig zu beurteilen, inwieweit die gemessenen Ergebnisse auf die Weiterbildung zurückzuführen sind. Wie oben ausgeführt, können von der Weiterbildung unabhängige Reifungsprozesse stattgefunden haben. Bei einer Untersuchung, in der zusätzlich noch die Vortestung fehlt, sind die Ergebnisse ebenfalls schwierig zu interpretieren.

Nach den hier dargestellten Empfehlungen bei Rost (2007), basieren die meisten Studien zu didaktischen Weiterbildungen auf schwachen Versuchsplänen. Das wird auch in den Reviews kritisch angemerkt (Parsons et al., 2012; Spencer, 2014; Steinert et al., 2006; Stes et al., 2010). Dies betrifft vor allem die qualitativen Analysen (Stes et al. 2010), nur die wenigsten beinhalten Kontrollgruppen oder Vor- und Nachtestungen. Auch die Mehrheit der quantitativen Analysen basiert nicht auf robusten Untersuchungsplänen (Tabelle 5). Die zu erwartenden Effekte didaktischer Weiterbildungen konnten daher bislang kaum präzisiert werden.

Tabelle 5. Anteil der Studien bei Stes et al. (2010) mit Vortestung und Kontrollgruppe

	Forschungsmethodik			
	Insgesamt	Quantitativ	Qualitativ	MM
	<i>N</i> = 108	<i>N</i> = 46	<i>N</i> = 44	<i>N</i> = 18
Analysen mit Vortestung	34	24	4	6
Analysen mit Kontrollgruppe	15	13	0	2

Anmerkung. MM = Mixed-Methods.

Bereits Levinson-Rose und Menges (1981) forderten die Verwendung stärkerer Untersuchungspläne. Stes et al. (2010) sehen bei dem Vergleich von Studien, die vor, oder nach dem Jahr 2000 veröffentlicht wurden, allerdings keinen Trend zu besseren Versuchsplänen. Zukünftige Studien sollten zumindest quasi-experimentelle Versuchspläne mit Vor- und Nachtestungen beinhalten (Parsons et al., 2012; Stes et al., 2010).

4.3 Messinstrumente

Messinstrumente sind von hoher Bedeutung für die Validität der Ergebnisse. Sie müssen auf die Wirkungen der Weiterbildung abgestimmt und gleichzeitig auch für die Vergleichsgruppe angemessen sein. Sie müssen veränderungssensitiv sein, möglichst keine Boden- oder Deckeneffekte hervorrufen, die Vergleichbarkeit von Individuen ermöglichen und die Gefahr sozialer Erwünschtheit minimieren. Um eine hohe Teilnahme bei Befragungen zu gewährleisten, sollen sie zudem zeitökonomisch sein (Rost, 2007). Im Anschluss an die Erhebung müssen die wissenschaftlichen Gütekriterien erhoben und berichtet werden. In der Mehrheit der vorliegenden Studien wurden diese Anforderungen nicht erfüllt, es liegen wenige wissenschaftlich geprüfte Instrumente vor (Steinert et al., 2006; Stes et al., 2010). Die vorliegenden Instrumente für die verschiedenen Zielebenen (nach Steinert et al., 2006) werden nun dargestellt.

Messung der Zielebene Lernen. Die Zielebene Lernen erfasst den Zuwachs an Wissen und Fähigkeiten sowie mögliche Einstellungsänderungen. Wissen und Fähigkeiten müssen entsprechend der Inhalte bestimmt werden. Einstellungen können psychologische Konstrukte wie Lehransätze, Selbstwirksamkeitserwartungen, Engagement für die Lehre und Freude bei der Lehre umfassen.

In Studien zu Lehransätzen wird in der Regel auf das *Approaches to Teaching Inventory (ATI; Prosser & Trigwell, 1999)* oder dessen revidierte Form zurückgegriffen (*ATI-R; Trigwell, Prosser & Ginns, 2005*). Trotz häufiger Verwendung des *ATI/ATI-R* (z. B. Gibbs & Coffey, 2004; Hanbury, Prosser & Rickinson, 2008; Postareff, Lindblom-Ylänne & Nevgi, 2007 sowie als deutsche Übersetzung bei Lübeck, 2009; Kröber, 2010; Ulrich, 2013) ist dessen Einsatz umstritten. Meyer und Eley (2006) halten das Instrument für ungeeignet, um Lehransätze valide zu erfassen. Sie kritisieren die theoretische und messtheoretische Fundierung sowie die Operationalisierung der Items. Lübeck (2009) schlägt eine konzeptionelle Überarbeitung vor.

Die Schwierigkeiten bei der Messung von Lehransätzen betreffen vor allem die Frage zu der Dimensionalität. Ursprünglich waren studierendenzentrierte und lehrendenzentrierte Lehre als Endpunkte auf einer Dimension gegenübergestellt (Kember, 1997; Kember & Kwan, 2000; Kugel, 1993; Winteler, 2008). Andere Autoren verstehen Lehransätze dagegen als zweidimensional (Postareff und Lindblom-Ylänne, 2008; Prosser & Trigwell, 1999; Lübeck, 2009) oder mehrdimensional (Pratt, 2002). Das *ATI/ATI-R* beinhaltet zwei Skalenwerte: *Studierendenzentrierung* und *Lehrendenzentrierung*. Bei einer zweidimensionalen Messung mit dem *ATI/ATI-R* können Lehrende sowohl eine

starke Lehrendenzentrierung und starke Studierendenzentrierung als auch niedrige Lehrendenzentrierung und niedrige Studierendenzentrierung aufweisen. Solche Profile sind möglich, da die mit dem *ATI/ATI-R* gemessenen Einstellungen und Verhaltensweisen nicht alle Facetten der Konstrukte Lehrenden- und Studierendenzentrierung abbilden. Zudem ist das *ATI/ATI-R* so formuliert, dass die Zustimmung zu einem Item keine Aussage zu der Quantität einer Lehrhandlungen macht (Beispielitem studierendenzentrierte Lehre: „Ich ermutige die Studierenden, ihr bestehendes Wissen so zu re-strukturieren, dass sie sich die Inhalte auch unter Berücksichtigung neuer Denkweisen erschließen können“. Übersetzung nach Lübeck, 2009). Es ist daher unklar, wie die Skalenwerte im *ATI/ATI-R* in Zusammenhang zu Lehrverhalten stehen, aus hohen Werten bei der Studierendenzentrierung lässt sich nicht der Anteil studierendenzentrierter Lehre in Lehrveranstaltungen ableiten. In bisherigen Studien, in denen das *ATI/ATI-R* eingesetzt wurde, traten schwierig zu interpretierende Profile von Lehrenden auf. So stieg bei Gibbs und Coffey (2004) zwar die Studierendenzentrierung der Teilnehmer nach einer Weiterbildung, deren Ausprägung war jedoch geringer als bei der Kontrollgruppe zum ersten Messzeitpunkt. Postareff, Lindblom-Ylänne und Nevgi (2007) berichteten von relativ konstanten Werten vor und nach einer didaktischen Weiterbildung, während sich bei Johannes und Seidel (2011) Lehranfänger nach einem Training zur Studierendenzentrierung sogar lehrendenzentrierter einschätzten. Dabei ließ sich kein Zusammenhang von selbstberichteter Lehrendenzentrierung und dem beobachteten Lehrverhalten finden. Insgesamt konnten mit dem *ATI/ATI-R* bislang kaum bedeutsame Veränderungen der Studierendenzentrierung nach einer didaktischen Weiterbildung nachgewiesen werden. Die Veränderungssensitivität des Instruments ist für diesen Zweck möglicherweise zu gering. Prosser und Trigwell (2006) betonen, dass der *ATI/ATI-R* nicht für Veränderungsmessung konzipiert wurde, sondern vorrangig für Zusammenhangsanalysen von Lehransätzen und Merkmalen der Lehr- und Lernumgebung.

Weiterhin weisen Prosser und Trigwell (2006) auf die Kontextabhängigkeit von Lehransätzen und deren Messung mit dem *ATI/ATI-R* hin (vgl. Lindblom-Ylänne, Trigwell, Nevgi & Ashwin, 2006; Lübeck, 2009). Vorlesungen sind häufiger mit einem lehrendenzentrierten, und Seminare häufiger mit einem studierendenzentrierten Ansatz assoziiert (Lübeck, 2009). Zu weiteren Lehrveranstaltungsformen liegen keine Daten vor. Normwerte existieren zu keinem Lehrformat. Aufgrund der Itemformulierungen ist das Instrument zudem nicht optimal auf Kleingruppenformate wie POL oder Bedside-Teaching anwendbar. Für die Messung der Lehransätze in dieser Arbeit musste daher

ein neues Messinstrument entworfen werden. Dieses sollte das Konstrukt möglichst vollständig abbilden und veränderungssensitiv sein. Die genaue Vorgehensweise wird in Kapitel 6 beschrieben.

Postareff et al. (2007) nahmen an, dass Lehrende mit höherer Selbstwirksamkeitserwartung (Bandura, 1982) mehr engagiert sind, ihr didaktisches Methodenspektrum erfolgreich einzusetzen. Die Hypothese, dass die Selbstwirksamkeitserwartungen der Lehrenden mit längeren didaktischen Weiterbildungen steigen, mussten sie jedoch verwerfen. Die Lehrenden in ihrer Befragung wiesen schon vor der Weiterbildung relativ hohe Selbstwirksamkeitserwartungen auf. Überraschenderweise stiegen diese zunächst, sanken aber nach weiteren Weiterbildungen wieder etwas ab. Möglicherweise besteht die Gefahr, dass Lehrende zunächst unrealistisch hohe Erwartungen bezüglich ihrer Selbstwirksamkeit haben, die dann im Rahmen didaktischer Weiterbildungen korrigiert werden. Aufgrund zunehmend geringerer Fallzahlen in den Versuchsgruppen bei Postareff et al. (2007), ist diese Annahme zunächst nur spekulativ. Da die DOS ebenfalls eine zeitlich umfassende Weiterbildung ist (vgl. Kapitel 5), ließ sich daraus die Frage ableiten, ob die Zuversicht von Lehrenden zu- oder abnimmt, selbst eine positive Wirkung auf den Lernerfolg der Studierenden ausüben zu können.

Messung von Lehrverhalten. Bislang stützt sich die Forschung zu didaktischen Weiterbildungen vorrangig auf selbstberichtetes Verhalten der Lehrenden, da Beobachtungsstudien ressourcenintensiv sind (Desimone, 2009; Hora & Ferrare, 2013). Es mangelt an wissenschaftlich geprüften Beobachtungsinstrumenten (Levinson-Rose & Menges, 1981; Stes et al., 2010). Beobachtungsinstrumente, die für Forschungszwecke konzipiert wurden, basieren häufig auf Instrumenten aus dem schulischen Kontext (siehe z. B. Sawada et al., 2002; Walkington et al., 2011), es existieren aber auch Entwürfe für die Hochschule (Braun & Hannover, 2011; Murray, 1983; Wainwright, Flick & Morell, 2003). Wainwright, et al. (2003) sehen in Beobachtungsinstrumenten, die aus dem schulischen Kontext stammen, die Limitation, dass sich die darin enthaltenen didaktischen Kontexte nicht angemessen auf die Hochschule übertragen lassen, insbesondere die Interaktion zwischen Lehrenden und Lernenden.

In den Instrumenten werden Beobachtungskategorien bezüglich ihres Auftretens oder ihrer Qualität bei der Umsetzung zusammenfassend beurteilt. Beobachtungsbögen, die rein deskriptive Daten generieren, bieten den Vorteil, dass sie weniger von theoretischen Annahmen abhängig sind, höhere Übereinstimmungen aufweisen sollten als Bewertungen und außerdem günstiger für den Feedbackprozess sind (Centra, 1979; Hora

& Ferrare, 2013; Wehr, 2008). Die Beobachtungsinstrumente, die in der hochschuldidaktischen Praxis eingesetzt werden, erfüllen vorrangig den Zweck, den Lehrenden eine strukturierte Rückmeldung zu geben. Sie sind eher als Beobachtungsleitfäden angelegt (siehe z. B. Diez, Fischer, Rühmann & Weiß, 2005; Reiber, 2005; Wehr, 2008), und nicht für quantitative Auswertungen konzipiert.

In den wenigen Evaluationsstudien didaktischer Weiterbildungen, in denen Lehrverhalten gemessen wurde, kamen inhaltlich reduzierte Beobachtungsbögen in Kombination mit Videoverfahren zum Einsatz. Die umfangreichste Studie stammt von Mahler und Benor (1984), welche die Wirkung einer viertägigen Weiterbildung auf 60 Lehrende der Medizin untersuchten. Sie analysierten jeweils drei Lehrveranstaltungen vor und bis zu drei Lehrveranstaltungen nach der Weiterbildung. Als abhängige Variablen untersuchten sie die Aktivierung (Kategorisierung von „Teacher lectures“ bis zu „Student initiates“, S. 267) und kognitiven Leistungen (Kategorisierung von „all talking is on a knowledge level“ bis zu „all talking is on a evaluation level“, S. 267) der Studierenden. Als Beobachtungsinstrument wurde das *Technion Diagnostic System* (Perlberg et al., 1974) eingesetzt. Jeweils ein geschulter Beobachter kategorisierte in einer Lehrveranstaltung drei fünfminütige Videosequenzen (in Drei-Sekunden-Intervallen) bezüglich einer der beiden abhängigen Variablen. Sowohl bei der Aktivierung als auch bei den Leistungen fanden Mahler und Benor (1984) signifikant höhere Effekte nach dem Training. Auch bei Lehrenden, deren Weiterbildung mehr als ein Jahr zurücklag (bis zu 500 Tage), wurden signifikante Unterschiede für die Aktivierung der Studierenden, aber nicht für die Leistungen der Studierenden gefunden. Das Lehrformat, die Anzahl der Studierenden und die Auswahl der fünfminütigen Sequenzen in den Lehrveranstaltungen wurden nicht beschrieben.

Gallos, van den Berg und Treagust (2005) analysierten das Lehrverhalten von 13 Lehrenden der Chemie nach einer Schulung. Ziel der Trainer war es, eine drei-Phasen Struktur mit Lehrendenvortrag, gefolgt von studentischer Aktivierung mit Einzel- und Gruppenarbeit und einer abschließenden Zusammenfassung in den Lehrveranstaltungen zu etablieren. Die Veranstaltungen der Lehrenden wurden zweimal vor und zweimal nach der Weiterbildung gefilmt. Anschließend analysierten drei Beobachter die Lehrveranstaltungen und maßen den zeitlichen Anteil jeder Unterrichtsphase. Die Beobachtungskategorien und deren Definitionen wurden nicht genauer beschrieben. Die Autoren schlussfolgerten auf der Basis ihrer Beobachtungen, dass neun von dreizehn Lehrenden ihr Lehrverhalten deutlich zugunsten einer drei-Phasen Struktur verändert haben.

Ein neues Instrument ist das videogestützte Verfahren von Seidel und Hoppert (2011). Mit einem Kodierschema lassen sich lehrendenzentrierte Merkmale („Aktive Beteiligung: Dozent“, „Didaktische Phase: Wissensvermittlung“, „lehrendenzentrierte Arbeitsformen“) und studierendenzentrierte Merkmale („Aktive Beteiligung: Studierende“, „Didaktische Phase: Elaboration“, „Studierendenzentrierte Arbeitsformen“) in Seminaren mit zufriedenstellender Beobachterübereinstimmung messen. Das Verfahren wurde eingesetzt, um die Beziehung zwischen Lehrverhalten und Lehransätzen zu untersuchen (Johannes & Seidel, 2012). Höhere Studierendenzentrierung ging mit mehr Elaborationsanteilen und aktiverer studentischer Beteiligung in den Lehrveranstaltungen einher (bei insgesamt neun Lehrenden). Es zeigte sich kein Zusammenhang von selbstberichteter Lehrendenzentrierung und dem beobachteten Lehrverhalten.

Die Beobachtungskriterien von Mahler und Benor (1984), Gallos et al. (2005) und Seidel und Hoppert (2011) sind inhaltlich eng umgrenzt, indem sie allgemein zwischen dozenten-zentrierter Wissensvermittlung und studentischer Aktivierung unterscheiden. Sie sind daher nicht optimal für Studien in anderen Kontexten anwendbar. Weitere Beobachtungsbögen, die in publizierten Studien zum Einsatz kamen, dienten eher der Beschreibung konkreter Lehrveranstaltungen und wurden nicht für die Evaluation didaktischer Weiterbildungen entworfen. Die *Classroom Observation Rubric* (Turpen & Finkelstein, 2009) ist auf eine konkrete Lehrsituation zugeschnitten (Lehrenden-Studierenden- Interaktion bei dem Einsatz interaktiver Abstimmungsverfahren). Relativ neu ist das *Teaching Dimensions Observation Protocol* (Hora & Ferrare, 2014), mit dem Lehrverhalten auf verschiedenen Dimensionen klassifiziert werden kann und für das gute Beobachterübereinstimmungen erzielt wurden (Hora & Ferrare, 2014). Steinert et al. (2006) und Stes et al (2010) empfehlen die Konstruktion von Instrumenten, die in verschiedenen Kontexten eingesetzt werden können. Für die umfassende Evaluation der DOS lag folglich kein validiertes und geeignetes Beobachtungsinstrument von Lehrverhalten vor. Daher wurde ein Beobachtungsbogen für die Messung von Lehrverhalten entworfen. Die genaue Vorgehensweise wird in Kapitel 7.2 beschrieben.

Messung von Veränderungen auf institutioneller Ebene. Zu dieser Zielebene gibt es wenig Empirie. Die Mehrheit der berichteten Studien bei Stes et al. (2010) beruht auf qualitativen Daten, keine berichtet von einer Vergleichsgruppe. Indikatoren für institutionelle Veränderungen sind beispielsweise stärkere kollegiale Zusammenarbeit (Sydow, 1998), stärkere Wertschätzung der Lehre auf dem Campus (Medsker, 1992) und Engagement in der Curriculumentwicklung (Snyder, 2001). Es existiert kein umfassen-

des Konzept, wie Veränderungen auf institutioneller Ebene zu messen sind. Stes et al. (2010) plädieren daher für eine Weiterentwicklung der Forschungsmethodik.

Messung von Veränderungen auf Seiten der Studierenden. Auf dieser Zielebene werden studentische Wahrnehmungen der Lehre, Lernansätze und Lernergebnisse untersucht. Die Wahrnehmung der Lehre kann durch studentische Bewertungen erhoben werden (Gibbs & Coffey, 2004; Nurrenbern, Mickiewics & Francisco, 1999). Objektive Lernergebnisse wurden in den vorliegenden Studien bislang nicht untersucht. Dazu müssten Studierende von Lehrenden mit oder ohne Weiterbildung zu dem gleichen Thema unterrichtet, und anschließend vergleichende Tests eingesetzt werden. Es ist sehr schwierig, diese Voraussetzungen im Rahmen der Evaluation einer didaktischen Weiterbildung zu schaffen. Stes et al. (2010) führen Studien auf, in denen der Lernzuwachs indirekt erhoben wurde, beispielsweise indem die Studierenden ihren Lernzuwachs selbst positiv beurteilten (Brauchle & Jerich, 1998) oder Lehrende mit Weiterbildung der Annahme waren, dass ihre Studierenden nun mehr gelernt hatten (McShannon & Hynes, 2005). Die Validität dieser Wirksamkeitsbelege ist eher kritisch zu bewerten.

Alternativ zu Lernerfolg kann das Lernverhalten erfasst werden. Marton und Säljö (1976) fanden heraus, dass Studierende qualitativ unterschiedlich lernen. Sie gaben Studierenden Texte zu lesen und kündigten an, ihnen dazu später Fragen zu stellen. Manche Studierende konzentrierten sich dann beim Lesen primär darauf, die notwendigen Antworten bei den späteren Fragen zu erkennen, während andere sich mehr auf die Bedeutung und die Zusammenhänge der Inhalte konzentrierten. Daraus leiteten Marton und Säljö (1976) zwei unterschiedliche Lernansätze ab: tiefes und oberflächliches Lernen. Die Messung studentischer Lernansätze als mögliche Wirkung didaktischer Weiterbildungen, basiert auf der Annahme, dass studierendenzentrierte Lehre zu tieferer Verarbeitung führen soll (vgl. Kapitel 3.5). Lehrende mit hoher Studierendenzentrierung regen Studierende möglicherweise stärker dazu an, Elaborationsstrategien beim Lernen einzusetzen. Starke Lehrendenzentrierung kann dagegen dazu führen, dass Studierende sich in Lehrveranstaltungen mehr auf das Auswendiglernen von Fakten konzentrieren und dadurch nicht die Zusammenhänge der Lehrinhalte verstehen (Biggs & Tang, 2007). Der Nachweis, dass Lehrende im Anschluss an eine Weiterbildung tiefere Verarbeitungsstrategien bei ihren Studierenden bewirken, konnte bisher allerdings noch nicht erbracht werden. Trigwell et al. (1999) fanden einen Zusammenhang zwischen Lehr- und Lernansätzen, lehrendenzentrierte Lehre ging mit oberflächlichem Lernen einher. Gibbs und Coffey (2004) fanden bei Studierenden, deren Lehrpersonen studie-

rendenzentrierter gelehrt hatten, weniger Verwendung von oberflächlichen Lernstrategien, jedoch keine Veränderung bei tiefen Lernstrategien.

4.4 Implikationen für die hochschuldidaktische Forschung

Die vorliegenden Ergebnisse zur Wirksamkeit didaktischer Weiterbildungen sind aus wissenschaftlicher Sicht wenig überzeugend. Die Untersuchungspläne der Studien sind eher schwach und basieren auf Messinstrumenten, deren Objektivität, Reliabilität und Validität gering ist oder unklar bleibt. Die Studien untersuchen die Wirkungen vorrangig aufgrund von Selbstberichten der Lehrenden, die an den Programmen teilgenommen haben. Den meisten Studien fehlt eine unmittelbare Beobachtung des Lehrendenverhaltens und der daraus resultierenden Veränderungen (Stes et al, 2009). Der Kontext, in dem Lehre stattfindet, wird kaum berücksichtigt (Wilson, 2012). Die Langzeitwirkungen der Weiterbildungen bleiben unklar (Stewart, 2014). Außerdem sind die Studien wenig vergleichbar, zu unterschiedlich sind die unabhängigen (die Programme) und abhängigen Variablen (die untersuchten Wirkungen). Auch ein Blick auf Studien, die nicht bei Stes et al. (2010) berücksichtigt wurden, sei es weil sie erst nach ihrer Analyse publiziert wurden (Barczyk, Buckenmeyer, Feldman & Hixon, 2011; Hanbury, Prosser & Rickinson, 2008; Rodgers, Christie & Wideman, 2014) oder weil sie aus dem deutschsprachigen Raum stammen (Borchard, 2002; Fendler, Seidel & Johannes, 2013; Heger, 1985; Hofer, Jansen & Soboll, 2005; Johannes & Seidel, 2012; Kröber, 2010; Thumser-Dauth, 2007; Ulrich, 2013; Wegner & Nückles, 2011), ändert nichts an dieser Schlussfolgerung.

Diese Schwächen werden jedoch von den Autoren der Studien und der Reviews diskutiert, dementsprechend liegt eine Vielzahl von Verbesserungsvorschlägen für die Forschungspraxis vor. Tabelle 6 fasst die Vorschläge aus der Literatur zusammen. Diese Empfehlungen sollen bei der Evaluation der didaktischen Weiterbildung in dieser Arbeit berücksichtigt werden. Die hier untersuchte Dozentenschulung der medizinischen Fakultät Heidelberg wird in den folgenden Kapiteln beschrieben und evaluiert.

Tabelle 6. *Implikationen für die hochschuldidaktische Forschung*

Verbesserungsvorschläge

- Verwendung stärkerer Versuchspläne
 - Stärkere theoretische Fundierung der Programmevaluation
 - Genauere Beschreibung der Weiterbildungen (Ziele, Inhalte, Methoden)
 - Zusätzliche Untersuchung von Langzeiteffekten
 - Verwendung multipler Methoden und Datenquellen
 - Einsatz wissenschaftlich geprüfter Messinstrumente
 - Prozess-orientierte Studien als Ergänzung zu Wirksamkeitsmessungen
 - Stärkerer Fokus auf Verhalten und Resultate, weniger auf Selbstberichte
 - Genaue Erfassung des Kontextes (Rahmenbedingungen der Lehre)
-

Anmerkung. Aufzählung nach Parsons et al. (2012), Spencer (2014), Steinert et al (2006) und Stes et al. (2010).

5 Dozentenschulung der Medizinischen Fakultät Heidelberg

5.1 Inhalte, Konzepte und Ziele

Die in dieser Programmevaluation untersuchte Form der DOS der Medizinischen Fakultät Heidelberg besteht seit 2006; davor bestand sie aus einem einzelnen fünftägigen Workshop. Die DOS gliedert sich in zwei Abschnitte (DOS 1 und 2). Die Teilnahme an der DOS 1 und 2 (bzw. alternativ an einer äquivalenten Weiterbildung) ist für Lehrende der medizinischen Fakultät mit Habilitationsabsicht verpflichtend. Andere Lehrende der Fakultät können freiwillig teilnehmen. Die DOS 1 und 2 sind im Rahmen des Baden-Württemberg-Zertifikats des Hochschuldidaktischen Zentrums (HDZ) Baden-Württemberg mit insgesamt 120 Unterrichtseinheiten anerkannt; von der Landesärztekammer Baden-Württemberg werden für die Teilnahme CME - (Continuing Medical Education) Punkte vergeben.

Jede DOS hat eine Gesamtteilnehmerzahl von bis zu 42 Personen. Diese rekrutieren sich aus verschiedenen Disziplinen der Humanmedizin und der Zahnmedizin. Die Mehrheit der Teilnehmer sind Mediziner, aber auch Naturwissenschaftler, Mathematiker und Psychologen nehmen teil. Die DOS fand bis zum Jahr 2013 etwa 50 Kilometer außerhalb von Heidelberg, in einem Tagungshotel statt. Tabelle 7 zeigt die zeitliche Ausdehnung der DOS. Die Schulung umfasst mehrere Präsenzphasen in Kursen sowie zwei Phasen der Selbstreflexion in Kleingruppen. Jeweils im Anschluss an die DOS 1 und 2 sollen die Lehrenden die gelernten Inhalte in der Lehrpraxis anwenden. Eine Lehrveranstaltung wird in diesem Zeitraum wechselseitig mit einem Kollegen aus der DOS hospitiert und die Beobachtungen schriftlich analysiert. Abschließend werden die Beobachtungen in einer Kleingruppe reflektiert, welche von einem der Trainer von DOS geleitet wird. Mit einer Gesamtdauer von neun bis zwölf Monaten zwischen dem ersten Termin von DOS 1 und der abschließenden Reflexion nach der DOS 2 ist die DOS zeitlich ausgedehnter als die meisten anderen hochschuldidaktischen Weiterbildungen (vgl. Wilson, 2012, S. 66).

Tabelle 7. Zeitlicher Rahmen der Dozentenschulung (DOS) 1 und 2

	Präsenztermine	UE
DOS 1		
Workshop	2	22
Praxisphase: Hospitation, Kollegiale		
Hospitation & Abschlussreflexion	3	20
DOS 2		
Workshop	5	58
Praxisphase: Hospitation, Kollegiale		
Hospitation & Abschlussreflexion	3	20
Gesamt	13	120

Anmerkung. UE = Unterrichts- bzw. Arbeitseinheiten (je 45 Minuten).

Insgesamt teilen sich etwa jeweils zehn bis zwölf Trainer der Fachbereiche Medizin, Psychologie, Pädagogik und Naturwissenschaften die Themenfelder untereinander auf. Es gibt geringe Variationen in den Abläufen, Inhalten, Methoden und den eingesetzten Trainern. Die Themen (Tabelle 8) decken ein weites Spektrum ab und sind auf die Anforderungen der medizinischen Lehre ausgerichtet. Verschiedene Facetten der Lehrkompetenz sollen gestärkt werden. Entsprechend den Empfehlungen des Wissenschaftsrats (2008) liegt der Fokus der DOS auf „Methoden des Lehrens und Lernens“ (S. 67). Die Lehrenden erhalten schriftliche Materialien und eine Veröffentlichung zu didaktischen Methoden in der Lehre (Strittmatter-Haubold & Ehlail, 2012). Es werden unterschiedliche Methoden eingesetzt: Vorträge durch die Trainer, Diskussionsrunden, Gruppenarbeits- und Einzelarbeitsphasen. Dazu gibt es Übungseinheiten zu verschiedenen Lehrformaten in der Medizin (vgl. Kapitel 2.2), Informationen zu medizinischen Prüfungsformaten und spezielle Übungen für die individuelle Präsentationstechnik wie das *Microteaching*. Bei dieser Methode werden Lehrsequenzen gefilmt und die Videoaufnahmen in einer Kleingruppe analysiert. Diese kritische Analyse des Lehrverhaltens in Verbindung mit konstruktivem Feedback, gilt als effektive Methode bei der Ausbildung von Lehrkräften (Hattie, 2009).

Tabelle 8. *Themen der Dozentenschulung (DOS)*

Thema	DOS
Grundlagen Lerntheorie	1 , 2
Kompetenzorientiertes Lehren/Lernziele	1 , 2
Lehrveranstaltungen strukturieren	1 , 2
Einstieg in Lehrveranstaltungen	1 , 2
Methodeneinsatz in Lehrveranstaltungen	1 , 2
Überarbeitung eigener Lehrveranstaltungen in Kleingruppen	1 , 2
Lernstile	2
Kommunikation & Feedback	2
Übungseinheit Micro-Teaching	2
Übungseinheit Medi-KIT	2
Übungseinheit Simulation	2
Übungseinheit Problemorientiertes Lernen	2
Prüfungen in der Medizin	2

Zwei theoretische Ansätze liegen der DOS zugrunde. Zum einen das *Sandwichprinzip* nach Wahl (2006), und zum anderen das *constructive alignment* (Biggs, 1996; Biggs & Tang, 2007). Das Sandwichprinzip sieht zu Beginn einer Lehrveranstaltung eine Einleitung vor, in der die zentralen Konzepte der Lehrveranstaltung dargestellt werden. Danach wechseln sich Phasen der kollektiven Vermittlung, also der Instruktion durch die Lehrenden, mit Phasen der subjektiven Auseinandersetzung ab. Diese beinhalten Gruppen-, Partner- oder Einzelarbeiten, in denen eine tiefere Verarbeitung der Inhalte stattfinden soll. Die Lehrenden werden bei der DOS motiviert, diese individuellen Phasen so zu planen, dass sich alle Studierenden aktiv mit den Inhalten der Lehrveranstaltung auseinandersetzen. Diese individuellen Lern- und Arbeitsphasen sind laut Wahl (2006) wichtig für den individuellen Lernprozess der Studierenden und sollen oberflächlichem Lernen entgegenwirken. In diesen Phasen ändert sich die Rolle der Lehrperson, sie sollte sich „aus dem Lernprozess heraushalten, wo sie nicht gebraucht wird bzw. wo sie stören würde. Sie sollte aber die Chance ergreifen, zu beraten, zu unterstützen, und zu coachen...“ (S. 109). Der schwierigste Zeitpunkt für Lehrende besteht darin, den „Sozialformwechsel so zu vollziehen, dass wieder gemeinsam weitergearbeitet werden kann“ (S.111). Lehrende müssen dann abwägen, wie die Ergebnisse der subjektiven Ausei-

nandersetzung im Plenum synchronisiert werden. Danach präsentieren sie wieder Inhalte (kollektive Lernphase). Das Sandwichprinzip beinhaltet also lehrendenzentrierte Phasen, die sich nach etwa 15 – 20 Minuten mit studierendenzentrierten Phasen abwechseln sollten. Die Länge der studierendenzentrierten Phasen ist von der verwendeten Methode abhängig (vgl. Strittmatter-Haubold & Ehlail, 2012), sollte insgesamt aber möglichst umfassend sein (Wahl, 2006). In der DOS werden die Lehrenden ermutigt, das Sandwichprinzip sowohl in Großgruppen-, als auch in Kleingruppenlehrveranstaltungen umzusetzen. Abschließend soll die Lehrveranstaltung mit einem inhaltlichen Ausstieg durch die Lehrenden beendet werden.

Die zweite zentrale Theorie ist das constructive alignment (Biggs, 1996; Biggs & Tang, 2007). Der Begriff steht für die wechselseitige Abstimmung von Lernzielen, Lehre und Prüfungen. Lernziele geben an, welche Kompetenzen die Studierenden erwerben sollen. Die Formulierung und Verfolgung klarer Lernziele gilt als wichtiger Prädiktor für den Lernerfolg bei Studierenden (Feldman, 2007) und wird vom Wissenschaftsrat (2008) gefordert. Im Idealfall orientieren sich sowohl die Lehre als auch die Prüfungen an den definierten Lernzielen. Klarheit und Transparenz von Lern- und Prüfzielen soll auch den Lehrenden die Unterrichtsplanung erleichtern. In der DOS wird den Lehrenden empfohlen, die Lernziele in Hinblick auf Prüfungen oder praktische Anforderungen hin zu formulieren, diese transparent zu machen (beispielsweise im Rahmen des Einstiegs in eine Lehrveranstaltung) und die Lehre daran auszurichten. Insbesondere die Phasen der subjektiven Auseinandersetzung sollen den Studierenden Zeit geben, die Lernziele innerhalb der Lehrveranstaltungen zu erzielen.

Zusammenfassend lassen sich aus den Inhalten der DOS folgende wichtige Ziele ableiten: die Lehrenden sollen studierendenzentrierte Lehre umsetzen und die Lernziele der Lehrveranstaltungen transparent machen. Die DOS ist so gestaltet, dass die Schlüsselmerkmale erfolgreicher Weiterbildungen (vgl. Kapitel 4.1) weitgehend verwirklicht werden konnten. Die DOS erstreckt sich über einen längeren Zeitraum und die Lehrenden erhalten die Gelegenheit bedarfsorientiert und individuell an ihrer Lehre zu arbeiten sowie dazu Feedback zu erhalten. Im folgenden Kapitel werden die abgeleiteten Hypothesen für die Evaluation der DOS dargestellt.

5.2 Hypothesen zur Evaluation

Eine erste Programmevaluation der DOS fand durch Roos et al. (2014) statt. Diese untersuchte 2006 die damals fünftägige DOS, deren Schwerpunkt auf den Grundlagen der Lerntheorie, der Strukturierung von Lehrveranstaltungen, den medizinischen Prüfungen, dem Simulationstraining, dem Medi-KIT, dem Micro-Teaching, dem POL sowie der Reflexion der Lehrendenrolle lag. Zwei Kohorten der DOS mit insgesamt 56 Personen wurden auf den Zielebenen nach Kirkpatrick (1994) untersucht. Es erfolgte eine Zufriedenheitsmessung (Ebene Reaktion), ein Multiple-Choice-Wissenstest vor und nach der DOS (Zielebene Lernen) sowie zehn Lehrhospitation und dreizehn Interviews zum Lehrverhalten (Zielebene Verhalten). Es zeigte sich eine hohe Zufriedenheit der Teilnehmer mit der Arbeitsatmosphäre und den Inhalten, insbesondere mit der Übungseinheit Micro-Teaching. Zudem stieg das didaktische Wissen der Teilnehmer nach der Weiterbildung signifikant an und konnte auch sechs Monate nach der DOS nachgewiesen werden. Die qualitative Analyse von Verhaltensbeobachtungen und dem selbstberichteten Handeln wies auf einen erfolgreichen Transfer des Gelernten in die Praxis hin. Die Limitationen der Studie waren eine kleine Stichprobe und das Fehlen einer Vergleichsgruppe. In dieser Arbeit sollen die Ergebnisse von Roos et al. (2014) für ein neues DOS-Konzept überprüft und die gezogenen Schlussfolgerungen zu der Evaluation didaktischer Weiterbildungen (vgl. Kapitel 4.4) berücksichtigt werden.

Angesichts der inhaltlichen Breite der DOS sollten in dieser Arbeit nicht alle intendierten und möglichen Wirkungen der DOS untersucht werden. Im Fokus standen die Zielebenen Lernen und Verhalten (nach Steinert et al., 2006). Insbesondere sollte hier die Annahme geprüft werden, dass Lehrende, welche die DOS besucht haben, eine höhere Studierendenzentrierung aufweisen und ihre Lernziele transparent machen. Die Entscheidung, den Fokus auf diese Variablen zu legen, geschah auch mit dem Ziel, eine ausreichende Teststärke zu erzielen. Bei einer höheren Zahl von abhängigen Variablen wäre eine Alpha-Korrektur unumgänglich. Weitere Variablen wurden lediglich mit einbezogen, um eine umfassendere Bewertung der DOS vornehmen zu können: (a) Die Zielebene Reaktion, unter der Annahme, dass eine hohe Zufriedenheit mit einer Weiterbildung zwar keine hinreichende, aber eine notwendige Bedingung für eine erfolgreiche Weiterbildung darstellt (vgl. Kapitel 4.1). (b) Die Selbstwirksamkeitserwartungen bezüglich der Lehre, um deren möglichen Zusammenhang mit Studierendenzentrierung besser zu verstehen (vgl. Kapitel 4.3) und zu prüfen, ob die DOS hier einen positiven Einfluss ausübt. (c) Die Selbsteinschätzungen der Lehrkompetenz, da sie als bedeutend

für Leistung und Motivation in der Lehre gelten (vgl. Kapitel 3.3). Dazu noch zwei Aspekte, die sich der Zielebene Resultate zuordnen lassen: (d) die Weiterbildungsbereitschaft der Lehrenden. Der Wissenschaftsrat (2008) fordert die Bereitschaft zur kontinuierlichen Weiterbildung für die Lehre. Daher sollte hier geprüft werden, ob die Lehrenden nach der verpflichtenden Teilnahme an der DOS, noch freiwillig bereit sind, didaktische Weiterbildungen zu besuchen. (e) Die Wahrnehmung der Studierenden gegenüber studierendenzentrierter Lehre. Wie in Kapitel 3.4 dargestellt, ist die Wirkung studierendenzentrierter Lehre von deren Umsetzung und Akzeptanz abhängig. Tabelle 9 stellt die Hypothesen zur Evaluation der DOS dar, diese sind hier wie bei Steinert et al. (2006) verschiedenen Zielebenen zugeordnet.

Andere Inhalte der DOS haben für diese Evaluationsstudie geringe Relevanz: Die Lehrformate POL, Medi- KIT und Simulation sollen beispielsweise alle Lehrenden bei der DOS nur kennenlernen, da diese nicht in allen Fächern eingesetzt werden. Ebenso trifft dies auf den Themenblock Prüfungen in der Medizin zu, da Lehrende in unterschiedlichem Ausmaß an der Erstellung und Durchführung von Prüfungen beteiligt sind. Auch aus dem Themenblock Lernstile (nach Kolb, 1985) wurden für diese Evaluationsstudie keine abhängigen Variablen abgeleitet, da deren theoretischen Implikationen in Frage gestellt wurden (vgl. Garner, 2000) und das Konzept von Lerntypen umstritten ist (Looß, 2001).

Tabelle 9. *Hypothesen zur Evaluation der Dozentenschulung (DOS)*

Zielebene	Hypothese
Lernen: Einstellungen	Die DOS steigert die Studierendenzentrierung Die DOS steigert die subjektive Lehrkompetenz Die DOS steigert die Selbstwirksamkeitserwartungen bezüglich der Lehre
Verhalten	Die DOS steigert die Studierendenzentrierung Die DOS bewirkt eine höhere Transparenz der Lernziele
Reaktion	Die Lehrenden sind mit der DOS zufrieden.
Resultate:	
Organisatorische Veränderungen	Die Lehrenden sind nach der DOS weiterhin motiviert, an didaktischen Weiterbildungen teilzunehmen.
Resultate: Studierende	Studierende sehen in der Umsetzung studierendenzentrierter Lehre einen Nutzen für ihren Lernerfolg

Die exakten Hypothesen werden in der Lehrendenbefragung (Kapitel 6), der Beobachtungsstudie (Kapitel 7.3) und der Studierendenbefragung (Kapitel 8) jeweils genauer ausformuliert. Um die Forschungsfragen inhaltlich zu validieren, wurden die Trainer der DOS befragt (Kapitel 9). Für die Durchführung der Studien lag eine Zustimmung der Ethikkommission Heidelberg vor (Antragsnummer S 163/2011).

6 Empirische Studie 1: Die Selbsteinschätzung der Lehrenden

Die Wirkung der DOS sollte sich durch veränderte Selbsteinschätzungen der Lehrenden zeigen. Dazu wurden die Lehrenden schriftlich befragt. Zuerst werden in diesem Kapitel die zentralen Hypothesen, danach in Kapitel 6.2 die Methodik dargestellt. In Kapitel 6.3 werden die zentralen Ergebnisse der Studie beschrieben und in Kapitel 6.4 diskutiert.

6.1 Hypothesen

Um die Wirkung von der DOS zu überprüfen, wurden bei der Lehrendenbefragung, entsprechend der Klassifikation bei Steinert et al. (2006), vier Zielebenen unterschieden (vgl. Kapitel 4.1). Schwerpunkt dieser Fragebogenstudie waren die Zielebenen Lernen (Einstellungen) sowie das selbstberichtete Verhalten. Wie in Kapitel 5.2 dargestellt, lag der Fokus dieser Arbeit darauf, zu untersuchen, ob Lehrende nach Besuch der DOS ihre Lehre stärker an den Lernzielen orientieren und eine höhere Studierendenzentrierung zeigen. Die Zielebene Lernen (Einstellungen) umfasste die Einstellung zur Studierendenzentrierung, die selbsteingeschätzte didaktische Kompetenz und die Selbstwirksamkeitserwartungen bezüglich der Lehre. Die nächste Zielebene Verhalten umfasst die Analyse der selbst eingeschätzten Lehrhandlungen. Dabei wurde geprüft ob die Lehrenden ihre Lehre nach der DOS stärker an den Lernzielen ausrichten und inwiefern sie studierendezentrierte Lehrhandlungen umsetzen. Die beiden Zielebenen Reaktion und Resultate wurden zusätzlich in die Auswertung mit einbezogen, um eine umfassendere Bewertung der DOS vornehmen zu können. Die allgemeinen Untersuchungshypothesen sind in Tabelle 10 dargestellt. Die Hypothesen für die einzelnen Gruppenvergleiche werden in Kapitel 6.3.3 beschrieben.

Tabelle 10. *Empirische Studie 1: Hypothesen*

Zielebene	Hypothesen
Lernen: Einstellungen	H1: Die Lehrenden berichten nach der DOS von einer höheren Studierendenorientierung als Lehrende ohne/vor Besuch der DOS. H2: Die Lehrenden schätzen ihre didaktische Kompetenz nach der DOS höher ein als Lehrende ohne/vor Besuch der DOS. H3: Die Lehrenden berichten nach der DOS von höheren Selbstwirksamkeitserwartungen bezüglich ihrer Lehre als Lehrende ohne/vor Besuch der DOS.
Verhalten	H4: Die Lehrenden berichten nach der DOS von einer höheren Umsetzung studienorientierter Lehre als Lehrende ohne/vor Besuch der DOS. H5: Die Lehrenden machen nach der DOS die Lernziele ihrer Lehrveranstaltungen deutlicher transparent als Lehrende ohne/vor Besuch der DOS.
Reaktion	H6: Die Lehrenden sind mit der DOS zufrieden.
Resultate:	
Organisatorische Veränderungen	H7: Die Lehrenden sind nach der DOS motiviert, an weiteren didaktischen Weiterbildungen teilzunehmen.

Anmerkung. DOS = Dozentenschulung

6.2 Methodik

6.2.1 Durchführung

Versuchs- und Vergleichsgruppen. In dieser Studie wurden die Teilnehmer drei aufeinanderfolgender DOSen zweimal schriftlich befragt: Einmal bevor sie an der DOS teilgenommen hatten, diese Lehrenden werden im Folgenden als Versuchsgruppe vor der Weiterbildung (VSG_{PRÄ}) bezeichnet. Und zum zweiten Mal nachdem sie die DOS vollständig besucht hatten, diese werden als Versuchsgruppe nach der Weiterbildung (VSG_{POST}) bezeichnet. Zudem wurden Lehrende untersucht, deren DOS mindestens ein Jahr zurücklag. Diese Lehrenden werden im Folgenden als die VSG_{ALT} bezeichnet. Die beiden Versuchsgruppen wurden mit Lehrenden verglichen, die (noch) nicht an der DOS teilgenommen hatten und im Folgenden als Vergleichsgruppe (VGG) bezeichnet werden. Es liegen somit eine Versuchsgruppe mit zwei Messzeitpunkten, eine Versuchsgruppe mit einem Messzeitpunkt und eine Vergleichsgruppe mit einem Messzeitpunkt vor (Abbildung 4).

Keine DOS/ Vor der DOS	Vor der DOS	Direkt nach der DOS	Mindestens ein Jahr nach der DOS
VGG	VSG _{PRÄ}	VSG _{POST}	VSG _{ALT}

VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Die meisten Lehrenden aus der VSG_{PRÄ} sind später in der VSG_{POST}.

Abbildung 4. Empirische Studie 1: Versuchsgruppen und Vergleichsgruppe

Stichprobengewinnung. Lehrende, die sich in dem Untersuchungszeitraum für die DOS angemeldet hatten, wurden etwa sieben Tage vor dem Beginn der DOS per E-Mail gebeten, an einer Befragung zu der Evaluation der DOS teilzunehmen (Anhang A1). Die Lehrende, welche die DOS vollständig besucht hatten, wurden nach Abschluss der zweiten Praxishospitation (vgl. Kapitel 5) erneut per E-Mail angeschrieben (Anhang A2). Um die vorliegenden Daten der VSG_{PRÄ} und VSG_{POST} aufeinander zu beziehen, und deren Anonymität zu gewährleisten, wurden die Lehrenden gebeten, einen vierstelligen persönlichen Code zu erstellen. Dieser setzte sich aus (a) dem ersten Buchstaben

des Vornamens der Mutter (wenn nicht bekannt: des Vaters), (b) dem ersten Buchstaben des Geburtsortes und (c) dem Geburtsmonat der Mutter zusammen.

Die Lehrenden, die zu einem früheren Zeitpunkt an einer DOS teilgenommen hatten (VSG_{ALT}) wurden ebenfalls per E-Mail angeschrieben, um sie für die Teilnahme zu gewinnen (Anhang A3). Für die VGG wurden Lehrende, die (noch) nicht an der DOS teilgenommen hatten, über einen E-Mail-Verteiler von Fachverantwortlichen angeschrieben (siehe Anhang A4). Ebenso sollten Lehrende für die VGG gewonnen werden, die auf der Warteliste für die DOS standen. Lehrende in der VGG und der VSG_{ALT} konnten nur zu einem Messzeitpunkt befragt werden.

Stichprobenbeschreibung. Eine Teststärkenanalyse im Vorfeld (mit GPOWER; Faul, Erdfelder, Lang & Buchner, 2007) ergab, dass bei einem β -Fehler Risiko von .20, die Aufdeckung mittlerer Effekte eine Gruppengröße von mindestens 51 Personen je Gruppe erfordert. Diese Größe wurde für alle Gruppen erreicht. Tabelle 11 beschreibt die Zusammensetzung der Gruppen und den jeweiligen Befragungszeitpunkt. Zwei Lehrende, die bei der Befragung angegeben hatten, noch keine eigenen Lehrveranstaltungen durchgeführt zu haben, wurden für die weitere Auswertung nicht herangezogen.

Die Befragung der Lehrenden erfolgte online über den Anbieter *Q-Set* (www.q-set.de). Der Rücklauf war bei den Lehrenden der VSG_{PRÄ} mit insgesamt 117 Personen höher (97 %), als bei Lehrenden der VSG_{POST} mit 72 Personen (69 %). Davon haben 42 Personen (58.3 %) einen Code hinterlegt, welcher den Daten der Befragung zum ersten Messzeitpunkt zugeordnet werden konnte. In der VGG waren mehrheitlich Lehrende vertreten, die zu dem Zeitpunkt der Befragung noch auf der Warteliste für die DOS standen und erst zu einem späteren Zeitpunkt an der DOS teilgenommen haben ($n = 49$). Diese wurden unmittelbar vor Besuch der DOS 1 mit der Bitte zur Studienteilnahme angeschrieben. Weiterhin konnten 24 Lehrende ohne Bezug zur Dozentenschulung über E-Mail-Verteiler innerhalb der medizinischen Fakultät durch Fachverantwortliche für die VGG gewonnen werden. Die Mehrheit der Lehrenden in der VSG_{ALT} absolvierte die DOS zwischen 2009 und 2011 ($n = 58$), drei Personen absolvierten die DOS 2008 oder früher. Von zwei Lehrenden in der VSG_{ALT} lagen keine Angaben vor.

Tabelle 11. Kohorten der Dozentenschulung (DOS) und Lehrende in den Versuchsgruppen sowie der Vergleichsgruppe.

Gruppe	Teilnehmer		Teilnehmer mit Code *	Befragungszeitraum
	der DOS	Rücklauf		
VSG_{PRÄ_POST}				
VSG _{PRÄ} (DOS 1.14)	42	41		10.03.2011 – 27.03.2011
VSG _{POST} (DOS 2.21)	39	26	17	01.03.2012 – 05.04.2012
VSG _{PRÄ} (DOS 1.15)	39	36		12.09.2011 – 21.09.2011
VSG _{POST} (DOS 2.22)	36	27	16	19.07.2012 – 13.09.2012
VSG _{PRÄ} (DOS 1.16)	41	40		28.02.2012 – 07.03.2012
VSG _{POST} (DOS 2.23)	30	19	9	28.03.2013 – 28.04.2013
Vergleichsgruppe				
DOS 1.17	42	26		12.10.2012 – 18.10.2012
DOS 1.18	38	23		02.03.2013 – 06.03.2013
Ohne DOS	-	24		16.06.2011 – 24.11.2011
VSG _{ALT}	-	63		09.11.2011 – 30.04.2012

Anmerkungen. VSG = Versuchsgruppe. Lehrende einer DOS-Kohorte wurden jeweils vor der DOS 1 (VSG_{PRÄ}) und nach der DOS 2 (VSG_{POST}) befragt. * = Teilnehmer mit Code, der sich eindeutig dem Code aus der ersten Befragung zuordnen ließ.

Die Mehrheit der Teilnehmer war in allen Gruppen männlich (63 % – 68 %). Tabelle 12 beschreibt die Zusammensetzung nach Geschlecht und Gruppenzugehörigkeit. Es lagen keine Unterschiede zwischen den Gruppen vor ($\chi^2(6; N = 324) = 1.65, p = .949$).

Tabelle 12. *Geschlechterverteilung nach Gruppenzugehörigkeit*

	Weiblich	Männlich	Keine Angabe	Gesamt
VGG	26	46	1	73
VSG _{PRÄ}	39	76	2	117
VSG _{POST}	24	48	0	72
VSG _{ALT}	19	43	1	63
Gesamt	108	213	4	325

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}).

6.2.2 Fragebogen

Aufgrund fehlender standardisierter und validierter Messinstrumente (vgl. Kapitel 4.3), wurden Fragen und Skalen gemäß der Empfehlung von Porst (2009) und Rost (2007) entwickelt. Um die Verständlichkeit aller verwendeten Items sicherzustellen, wurden diese bei einer Abschlussreflexion (im Rahmen einer vorherigen DOS) einer Gruppe von acht Lehrenden vorgelegt und besprochen. Daraufhin wurden kleinere sprachliche Anpassungen durchgeführt. Eine weitere empirische Prüfung der Befragung vor dem ersten Einsatz im März 2011 fand nicht statt. Fragen und Antwortoptionen die einen Rückschluss auf die Fachbereiche der Lehrenden ermöglichen, wurden vermieden. Die Beantwortung der Befragung sollte aus ökonomischen Gründen maximal zehn Minuten in Anspruch nehmen.

Skalen für die Hypothesentestung. Für die Überprüfung der Hypothesen wurden die folgenden Skalen entworfen, die sich den Zielebenen nach Steinert et al. (2010) zuordnen lassen: Die *Zufriedenheit mit der Weiterbildung* (Zielebene Reaktion), die *Einschätzung der eigenen Lehrkompetenz*, *Selbstwirksamkeitserwartungen der Lehre* (Ziel-

ebene Lernen) *Studierendenzentrierte Lehre* und *Lernziele in der Lehre* (Zielebene Verhalten). Zur Messung der *Weiterbildungsbereitschaft* (Zielebene Resultate) der Lehrenden wurde ein Item entworfen. Als Antwortformat wurde jeweils eine fünfstufige Likertskala gewählt. Die Skalen werden im Folgenden beschrieben:

Die Skala *Zufriedenheit mit der Weiterbildung* (5 Items, Beispielitem: „DOS hat mir bei meiner Lehrtätigkeit geholfen“, siehe Anhang B1) erfolgte erst drei bis vier Monate nach der zweiten Präsenzphase. So messen die Items, wie zufrieden die Teilnehmer mit der DOS waren, nachdem sie die Inhalte in der Praxis ausprobieren konnten. Damit sollte der Kritikpunkt, dass die Zielebene Reaktion nicht informativ für die Wirkung einer Weiterbildung sei (Stes et al., 2009), in dieser Studie entkräftet werden.

In der Humanmedizin an der Medizinischen Fakultät Heidelberg ist die Unterscheidung zwischen Vorlesungen und Seminaren nicht immer trennscharf und oft nicht an der Teilnehmerzahl zu erkennen. Da sich die Ausprägung studierendenzentrierter Lehre in verschiedenen Veranstaltungsformaten unterscheidet (vgl. Kapitel 3.4), wurde bei der Skala *Studierendenzentrierung in Vorlesungen* (6 Items, siehe Anhang B1) der Fokus auf Lehrveranstaltungen mit mindestens 40 Personen festgesetzt. Aufgrund der starken Lehrendenzentrierung von Lehrveranstaltungen mit größerer Teilnehmerzahl, sollten Unterschiede bei der Einstellung zu studierendenzentrierter Lehre besonders deutlich zutage treten. Da die Lehrenden in der DOS motiviert werden, traditionell eher lehrendenzentrierte Veranstaltungsformate stärker studierendenzentriert zu gestalten, erschien dieses Vorgehen besonders relevant für die Überprüfung der Wirksamkeit der DOS. Die Formulierung der Items zu studierendenzentrierter Lehre beinhaltete die Bereiche Einzel-, Partner- und Gruppenarbeit, sowie die möglichen Interaktionen in einer Lehrveranstaltung (Lehrende fragen Studierende, Studierende fragen Lehrende und Studierende diskutieren untereinander). Im Unterschied zu den Items des *ATI/ATI-R* bildet diese Skala so das Konstrukt Studierendenzentrierung in diesem Kontext vollständig ab (Johannes & Seidel, 2012; Lea et al., 2003; O'Neill & McMahon, 2005; vgl. Kapitel 3.4).

Die didaktische Kompetenz wurde mit drei Items (Beispielitem: „Meine didaktischen Fähigkeiten sind gut“, siehe Anhang B1) gemessen, die dem *absoluten* Selbstkonzept (vgl. Schöne, Dickhäuser, Spinath und Stiensmeier-Pelster, 2002) der Lehrenden zuzuordnen ist.

Bei der Skala zu *Selbstwirksamkeitserwartungen* in der Lehre (3 Items, siehe Anhang B1) sollte das wahrgenommene potentielle Leistungsvermögen von Lehrenden

abgebildet werden (Bandura, 2006). Die Items messen, ob die Lehrenden annehmen, dass die Gestaltung ihrer Lehre zu einer intensiven Beschäftigung mit den Inhalten führen, sowie den Lern- und Prüfungserfolg beeinflussen kann (Beispielitem: „Die Gestaltung meiner Lehre hat großen Einfluss auf den Lernerfolg der Studierenden“). Damit ist die hier gewählte Skala inhaltlich homogener als die Skala von Postareff et al. (2007) und von Schmitz und Schwarzer (2000).

Die Umsetzung *studierendenzentrierter Lehre* (7 Items, Beispielitem: „Ich kommuniziere intensiv mit den Studierenden“, siehe Anhang B1) wurde entsprechend den Definitionen bei Johannes und Seidel (2012), Lea et al. (2003) sowie O’Neill und McMahon formuliert (2005). Da sich die Items nicht auf ein bestimmtes Lehrformat bezogen, wie bei der Skala *Studierendenzentrierung in Vorlesungen*, konnten diese nur repräsentative Umsetzungen studierendenzentrierter Lehre abbilden, die in allen Lehrformaten an der Medizinischen Fakultät möglich sind. Die Skala *Lernziele in der Lehre* wurde mit zwei Items abgefragt (Beispielitem: „Ich teile den Studierenden die Lernziele mit“, „Ich richte meine Lehre an den Lernzielen aus“, siehe Anhang B1). So sollte die Facette Lernziele im Rahmen eines constructive alignment abgebildet werden (Biggs, 1996; Biggs & Tang, 2007; vgl. Kapitel 5.1). Auch wenn die Abstimmung von Lern- und Prüfzielen dabei als wichtig gilt, wird dieser Bezug hier aufgrund der Heterogenität der durchgeführten Lehrformate mit unterschiedlicher Prüfungsrelevanz nicht gemessen.

Die Skalen zu selbstberichtetem Lehrverhalten wurden bei der VSG_{POST} sowohl mit direkter als auch mit indirekter Veränderungsmessung erhoben. Voraussetzung für die indirekte Veränderungsmessung war, dass Lehrende sowohl in der VSG_{PRÄ} als auch in der VSG_{POST} einen gültigen Code angegeben haben. Dann wurden die Skalenwerte von beiden Messzeitpunkten miteinander in Beziehung gesetzt ($\text{Variablenwert}_{\text{VSG}_{\text{POST}}} - \text{Variablenwert}_{\text{VSG}_{\text{PRÄ}}}$), unter der Annahme, dass die Befragten zu beiden Messzeitpunkten die gleichen Vorstellungen von den Konstrukten haben (Kohlmann & Raspe, 1998). Positive Werte bedeuten, dass zu Messzeitpunkt 2 eine höhere Ausprägung vorlag. Zusätzlich wurden die Lehrenden in der VSG_{POST} bei der zweiten Erhebung gebeten, Veränderungen bei der Umsetzung der jeweiligen didaktischen Handlungen direkt zu bewerten (Instruktion: „Bitte schätzen Sie ein: Haben sich Ihre Lehrhandlungen in den letzten 12 Monaten verändert?“). So sollte die Möglichkeit ausgeschlossen werden, dass eventuelle Unterschiede darauf zurückgeführt werden, dass die Lehrenden im Rahmen der Weiterbildung ein verändertes Verständnis ihrer didaktischen Handlungen erworben

haben. Auch die Lehrenden in der VGG wurden gebeten, die Items mit direkter Veränderungsmessung zu bearbeiten, jedoch nur die Lehrenden, die über die DOS 1.18 und 1.19 für die Teilnahme motiviert wurden (vgl. Tabelle 11). Die Items zu Lehrhandlungen wurden gemeinsam mit den allgemeinen didaktischen Handlungen (werden im nächsten Abschnitt beschrieben) in einer Liste aufgeführt. Um mögliche Positionseffekte bei dieser Erhebung zu vermeiden, wurde hier bei der Erstellung des Fragebogens die Q-Set-Option "Zufällige Zeilenabfolge im Fragebogen" verwendet.

Kontrollvariablen. Zur Kontrolle möglicher Einflüsse auf die abhängigen Variablen wurden aus dem Modell zur Wirkung des Hochschulunterrichts nach Helmke und Schrader (2010, vgl. Kapitel 3.2) entsprechende Variablen abgeleitet. Individuelle Faktoren wie Vorwissen (*Lehrerfahrung in Semestern, bereits durchgeführte Lehrveranstaltungsformate, besuchte Weiterbildungen*), Einstellungen (*Präferenz von Lehrveranstaltungsformaten, dem Stellenwert der Lehre*), Lehrverhalten (*allgemeine didaktischen Handlungen*) sowie die Rahmenbedingungen der Lehre (*die wahrgenommener Akzeptanz studierendenzentrierter Lehre in Vorlesungen, der Stellenwert der Lehre in dem eigenen Fachbereich*) beeinflussen wahrscheinlich die Effekte einer Weiterbildung. Unterschiede auf den abhängigen Variablen lassen sich besser auf die DOS zurückführen, wenn die untersuchten Gruppen sich bezüglich der Kontrollvariablen nicht systematisch unterscheiden. Lediglich drei Variablen (Stellenwert der Lehre, Stellenwert der Lehre in dem Fachbereich, Akzeptanz aktivierender Lehr- und Lernmethoden in Vorlesungen) wurden als Skalen erfasst, die anderen Variablen bestehen jeweils aus einzelnen Items (Tabelle 13). Die allgemeinen Lehrhandlungen wurden nicht zu einer Skala zusammengefasst, da sie unterschiedliche Aspekte didaktischer Handlungen abdecken. Alle Items wurden für den Zweck dieser Untersuchung entworfen. Der Originalwortlaut aller Items und die Antwortoptionen sind in Anhang B1 abgebildet.

Tabelle 13. *Kontrollvariablen in der Lehrendenbefragung*

Skalen und Fragen	Beispielitem
- Lehrerfahrung in Semestern	Wie viele Semester haben Sie schon Lehrveranstaltungen durchgeführt?
- Bereits durchgeführte Lehrveranstaltungsformate	Vorlesung
- Besuch anderer didaktischer Weiterbildungen*	Haben Sie schon andere didaktische Weiterbildungen neben der Dozentenschulung besucht?
- Präferenz von Lehrveranstaltungsformaten*	Vorlesung
- Stellenwert der Lehre**	Lehre ist für mich eine unangenehme Pflicht
- Stellenwert der Lehre in dem Fachbereich**	Lehre hat insgesamt einen hohen Stellenwert in meiner Abteilung
- Akzeptanz aktivierender Lehr- und Lernmethoden in Vorlesungen**	Viele Studierende mögen aktivierende Lehr- und Lernmethoden in Lehrveranstaltungen
- Allgemeine didaktische Handlungen (insgesamt neun Items)	Ich fasse wichtige Inhalte zusammen

Anmerkungen. *Nicht alle Lehrenden wurden befragt. **Skalen mit jeweils drei Items.

Ein Teil der Skalen und Items wurde nicht in allen Befragungen eingesetzt. Die Frage zu anderen didaktischen Weiterbildungen wurde erst zu einem späteren Zeitpunkt in die laufende Studie integriert (23.11.2011), so dass nicht alle Teilnehmer befragt werden konnten. Es liegen daher keine Daten für die VSG_{PRÄ} vor. Die Präferenz bereits durchgeführter Lehrformate wurde nur bei der VGG und VSG_{PRÄ} abgefragt, ebenso die Frage zur Motivation für die eigene Lehre. Die Items zu allgemeinen didaktischen Handlungen (Veränderung) wurden nur der VSG_{POST} und der VGG ab dem 02.03.2013 vorgelegt, gemeinsam mit den Items zu studierendenzentrierter Lehre.

In der Befragung gab es keine Pflichtfragen, lediglich die Gruppenzugehörigkeit und der persönliche Code musste angegeben werden, um die Befragung abschließen zu können. So sollten die Lehrenden ermutigt werden, nur auf die Fragen zu antworten, die sie für sich auch zutreffend empfanden. Die Antworten sollten zudem Anonymität gewährleisten. Die Antwortoptionen bei der Frage Lehrerfahrung in Semestern („Wie viele Semester haben Sie schon Lehrveranstaltungen durchgeführt?“) wurde daher auf 1 bis

10 oder mehr Semester beschränkt, um die Anonymität älterer Lehrender zu ermöglichen. Außer dem Geschlecht wurden keine demographischen Angaben erhoben.

6.2.3 Auswertung

Für die statistischen Auswertungen wurde SPSS (Version 21.0.0.0) verwendet. Um die Hypothesen zu prüfen, wurden jeweils Gruppenvergleiche vorgenommen. Zum einen wurden die Lehrenden zu Beginn ($VSG_{PRÄ}$) sowie nach Abschluss der Weiterbildung (VSG_{POST}) miteinander verglichen. Zum anderen wurden die Lehrenden der VGG mit der VSG_{ALT} verglichen. Die Gruppen, welche die DOS abgeschlossen haben, sollten gemäß der Hypothesen signifikant höhere Werte ($p < .05$) auf den abhängigen Variablen aufweisen. Die Berechnung von Effektstärken sollte die praktische Bedeutsamkeit der Effekte aufzeigen, unter der Annahme, dass bei den hier untersuchten Variablen zumindest kleine Effekte zu beobachten sind. Entsprechend der Definition nach Cohen (1988) entsprechen Werte zwischen 0.20 – 0.50 kleinen, zwischen 0.51 – 0.80 mittleren und höher 0.80 hohen Effektstärken. Ein α -Fehler würde bei der Untersuchung bedeuten, die Wirksamkeit der DOS zu überschätzen, was zu unrealistisch positiven Annahmen über die Veränderbarkeit von Lehrverhalten und lehrbezogenen Einstellungen führen würde. Ein β -Fehler würde zu einer Unterschätzung der Wirksamkeit führen. Teststärkeanalysen (mit GPOWER; Faul et al., 2007) ergaben eine statistische Power von mindestens .84 für die Vergleiche von der VGG mit der VSG_{POST} und der VSG_{ALT} sowie von mindestens .93 für den Vergleich der $VSG_{PRÄ}$ mit der VSG_{POST} um mittlere Effekte aufzuzeigen. Da bei der VGG weniger Daten bei der direkten Veränderungsmessung vorlagen, ist die statistische Power dort für die Hypothesen mit .70 (Studierendenzentrierung) und .72 (Lernziele) etwas geringer. Das Signifikanzniveau wurde vorab mit $\alpha \leq .05$ festgesetzt und die Hypothesen einseitig getestet.

Gruppenunterschiede wurden entsprechend dem Skalenniveau ausgewertet: Chi²-Analysen für kategoriale Daten sowie t-Tests und Varianzanalysen für metrische Daten. Bei der Hypothesentestung wurde zunächst eine multifaktorielle Varianzanalyse berechnet, mit den fünf abhängigen Variablen, die in allen Gruppen untersucht wurden. Vorher wurde kontrolliert, dass die Voraussetzungen für die Varianzanalyse weitgehend erfüllt wurden (vgl. Leonhardt, 2013, S. 568). Die einzelnen Hypothesen wurden danach nur für die entsprechend formulierten Hypothesen mit multiplen t-Tests berechnet. Für die Untersuchung der Lehrenden in der Versuchsgruppe mit gültigem Code wurde jeweils ein t-Test für abhängige Stichproben berechnet, für die anderen Gruppen t-Tests für unabhängige Stichproben.

Die Überprüfung auf Normalverteilung beinhaltete Vergleiche von Mittelwert und Median, Analyse von Schiefe und Kurtosis sowie Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk. T-Tests wurden auch bei moderaten Abweichungen von der Normalverteilung verwendet, da diese als robust gelten (Bühner & Ziegler, 2009). Bei inhomogenen Varianzen (Levene-Test) wurden die p-Werte für die korrigierten Freiheitsgrade im t-Test herangezogen. Die Unterschiede zwischen den Gruppen bei den einzelnen Items zu allgemeinen Lehrhandlungen wurden mit dem Mann-Whitney-Test berechnet, da hier deutliche Abweichungen von der Normalverteilung vorlagen.

Wenn signifikante Ergebnissen bei den Varianzanalysen auftraten, wurde das partielle Eta-Quadrat (η^2) als Effektstärkemaß berechnet. Bei multivariaten Varianzanalysen wurde, entsprechend der Empfehlung bei Leonhart (2013), zur Signifikanzprüfung Pillai's Spur herangezogen. Bei Post-Hoc-Tests nach Varianzanalysen wurde, entsprechend der Empfehlung bei Field (2013), der Test nach Gabriel bei gleichen Varianzen und leicht unterschiedlichen Fallzahlen und der Test nach Games-Howell bei inhomogenen Varianzen (Levene-Test) und gleichen oder ungleichen Fallzahlen berechnet. Signifikante Unterschiede zwischen den einzelnen Gruppen werden im Text berichtet.

Die Berechnung der Effektstärken d erfolgte auf Grundlage der Differenzen der Mittelwerte in den Gruppen. Die Differenzen wurden an der gepoolten Standardabweichung standardisiert, da diese in der Studie am ehesten der Standardabweichung der zugrundeliegenden Population entspricht.

Alle Skalen wurden so kodiert, dass hohe Werte mit einer hohen Ausprägung auf der Variablen einhergehen. Skalenwerte wurden berechnet, wenn mindestens zwei Drittel aller Items bearbeitet wurden. Die interne Konsistenz wurde mit dem Alpha-Koeffizienten nach Cronbach erhoben (1951). Die Berechnung der Trennschärfe erfolgte mit der part-whole-Korrektur (Kelava & Moosbrugger, 2012).

6.3 Ergebnisse

6.3.1 Deskriptive Ergebnisse

Deskriptive Statistiken. Tabelle 14 zeigt die deskriptive Statistik, Trennschärfe und interne Konsistenz der abhängigen Variablen sowie der Kontrollvariablen (zu denen Skalen gebildet werden konnten) in den untersuchten Gruppen auf. Die Mittelwerte der Skalen der abhängigen Variablen variierten zwischen 2.82 und 3.93 (auf einer Skala von 1 – 5), der Median zwischen 2.83 und 4.00. Damit lagen keine Decken- oder Bodeneffekte vor. Die Ergebnisse auf Itemebene befinden sich in Anhang B2.

Tabelle 14. Deskriptive Statistiken der Skalen in der Lehrendenbefragung

Variablen (Anzahl der Items)	VGG, VSG _{PRÄ} und VSG _{ALT}								VSG _{POST}					
	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis	<i>r_{it}</i>	α	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis	<i>r_{it}</i>	α
Studierendenzentrierung-VL (6)	245	2.82	0.61	0.37	0.71	.61 – .69	.69	69	3.20	0.69	0.10	-0.66	.30 – .61	.77
Didaktische Kompetenz (3)	211	3.46	0.63	0.04	-0.04	.46 – .78	.68	68	3.72	0.55	-0.16	-0.35	.45 – .59	.68
Selbstwirksamkeitserwartungen (3)	218	3.71	0.73	-0.41	0.31	.58 – .79	.77	66	3.76	0.67	-1.05	3.44	.37 – .71	.63
Studierendenzentrierung (7)	241	3.35	0.69	-0.27	-0.12	.74 – .78	.78	69	3.53	0.63	-0.24	-0.50	.36 – .59	.76
Lernziele (2)	237	3.82	0.92	-0.79	0.37	-	.69	69	3.93	0.81	-0.56	-0.07	-	.72
Studierendenzentrierung DV (7)	26	3.08	0.50	0.15	5.23	.68 – .89	.93	71	3.62	0.49	0.00	-0.08	.36 – .61	.79
Lernziele DV (2)	28	3.20	0.67	-0.20	5.21	-	.89	69	3.72	0.72	0.37	-0.61	-	.76
Zufriedenheit* (5)	131	3.67	1.10	-0.82	-0.22	.82 – .92	.96	70	3.90	0.95	-0.42	-0.78	.82 – .92	.96
Weiterbildung* (1)	138	3.78	1.02	-0.52	-0.44	-	-	71	3.67	1.08	-0.96	0.32	-	-
Akzeptanz AM (3)	245	3.43	0.75	-0.36	-0.08	.29 – .64	.61	68	3.32	0.77	-0.77	0.16	.46 – .47	.65
Stellenwert der Lehre (3)	249	3.92	0.75	-0.95	1.18	.42 – .87	.69	70	3.00	0.87	0.12	-0.47	.45 – .66	.73
Stellenwert der Lehre im FB (3)	219	3.27	0.84	-0.28	-0.12	.42 – .69	.72	72	3.84	0.80	-0.04	-0.72	.63 – .79	.82

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Die VSG_{POST} ist getrennt aufgeführt, da die Lehrenden in der Gruppe eine Teilmenge der VSG_{PRÄ} darstellen. *r_{it}* = Trennschärfe. α = Alpha-Koeffizienten nach Cronbach. VL = in Vorlesungen. DV = Direkte Veränderungsmessung. AM = Aktivierender Methoden. FB = Fachbereich. *Nur Lehrende aus der VSG_{POST} und der VSG_{ALT} wurden befragt.

Trennschärfe und interne Konsistenz. Die Trennschärfen der Items lagen zwischen .30 und .92 bei den abhängigen Variablen sowie zwischen .29 und .87 bei den Kontrollvariablen. Nach Fisseni (1997) liegen somit akzeptable bis hohe Trennschärfen vor, nur bei einer Skala lag ein Item im niedrigen Bereich ($r_{it} < .30$). Die interne Konsistenz der Skalen der abhängigen Variablen war mit Werten zwischen .63 und .96. für den Vergleich von Gruppenmittelwerten akzeptabel, ebenso für die Kontrollvariablen mit Werten zwischen .61 und .82.

Prüfung der Normalverteilung. Bei den Vergleichen von Mittelwert und Median traten lediglich geringe Abweichungen bei den fünfstufigen Likertskalen auf (maximale Differenz = 0.33). Die Werte von Schiefe und Kurtosis lagen bei den meisten Variablen < 1 und können daher nach Miles und Shevlin (2001) als moderat abweichend von der Normalverteilung interpretiert werden. Die Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk zeigten für alle Skalen in mindestens einer der beiden Gruppenanalysen signifikante Abweichungen von der Normalverteilung.

Interkorrelationen. Etwa die Hälfte (22 von 42) der gefundenen Interkorrelationen der abhängigen Variablen unterschied sich signifikant von Null. Es zeigten sich kleine bis mittlere Zusammenhänge (Tabelle 15). Weiterhin wurde die Korrelation zwischen der direkten und indirekten Veränderungsmessung bestimmt. Dazu wurde bei den Lehrenden, von denen zu beiden Zeitpunkten gültige Codes vorlagen, die Differenz der indirekten Veränderung bei den Skalen *Studierendenzentrierung* und *Lernzielen in der Lehre* gebildet. Diese Werte wurden in Beziehung zu der Messung der direkten Veränderung gesetzt. Dabei zeigte sich bei der VSG_{POST} kein korrelativer Zusammenhang zwischen der direkten und indirekten Veränderungsmessung, weder bei der Studierendenzentrierung ($r = -.03$) noch bei den Lernzielen in der Lehre ($r = .01$). Die Interkorrelation der abhängigen Variablen mit Kontrollvariablen befindet sich in Anhang B3. Die Korrelation der Variablen der direkten Veränderungsmessung mit der Differenz der indirekten Veränderungsmessung war nicht signifikant (vgl. Anhang B4).

Tabelle 15. *Interkorrelationen der abhängigen Variablen*

	SZ-VL	DK	SWKE	SZ	LZ	Reaktion*	WB*
SZ-VL	-	.05	.18	.26	.07	.04	.30
DK	.16	-	.34	.31	.25	-.04	-.15
SWKE	.29	.49	-	.29	.23	.03	.02
SZ	.32	.25	.27	-	.28	.18	.16
LZ	-.06	.08	.07	.34	-	.10	.04
Reaktion	.14	.15	.00	.32	.15	-	.33
WB	.31	.44	.34	.32	.19	.31	-

Anmerkungen. Obere Diagonale: die Interkorrelationen der Vergleichsgruppe, der Versuchsgruppe vor der Dozentenschulung (VSG_{PRÄ}) und der Lehrenden nach Besuch der Dozentenschulung (VSG_{ALT}). Untere Diagonale: die Interkorrelationen der Versuchsgruppe zu Zeitpunkt 2 (VSG_{POST}). Alle Korrelationen nach Pearson, mit Ausnahme der Variablen WB (Korrelation nach Spearman). Signifikante Korrelationen ($p < .05$) sind fett markiert. SZ-VL = Studierendenzentrierung in Vorlesungen. DK = Didaktische Kompetenz. SWKE = Selbstwirksamkeitserwartungen. SZ = Studierendenzentrierung. LZ = Lernziele. WB = Weiterbildungsbereitschaft. *Hier sind nur die Werte der VSG_{ALT} einbezogen.

Analyse von fehlenden Werten. Aufgrund der Freiwilligkeit der Fragebeantwortung gab es fehlende Werte bei den Variablen (vgl. Tabelle 14). Bei den fünf abhängigen Variablen, die in allen Gruppen erhoben wurden, lagen insgesamt bei 72.6 % der Lehrenden alle Ergebnisse vor, bei 18.8 % die Ergebnisse von vier, und bei 2.4 % die Ergebnisse von weniger als drei Variablen.

Analyse von fehlenden Teilnehmern. Ein Teil der Lehrenden hat nur am ersten Messzeitpunkt an der Befragung teilgenommen. Lehrende, die zu beiden Messzeitpunkten einen gültigen Code hinterlegt haben, unterschieden sich in der VSG_{PRÄ} nur in einer abhängigen Variablen von Lehrenden, die keinen gültigen Code hinterlegt haben (Anhang B5). Sie wiesen eine geringere Ausprägung bezüglich der Orientierung an Lernzielen in ihrer Lehre auf, ansonsten zeigten sich keine Unterschiede.

6.3.2 Testung auf Gruppenunterschiede in den Kontrollvariablen

Um zu prüfen, ob sich die untersuchten Gruppen in relevanten Kontrollvariablen unterscheiden, wurde die Lehrerfahrung, der Stellenwert der Lehre, die wahrgenommene Akzeptanz aktivierender Methoden in Vorlesungen und das allgemeine Lehrverhalten zwischen den Gruppen verglichen.

Lehrerfahrung. Die Lehrerfahrung wurde über die Anzahl unterrichteter Semester, die Anzahl unterrichteter Lehrformate sowie die Anzahl an besuchten didaktischen Weiterbildungen erhoben. Tabelle 16 zeigt die Verteilung der Lehrerfahrung in den Gruppen.

Tabelle 16. *Vergleich der Lehrerfahrung: Anzahl der Semester, Anzahl der Lehrformate und Teilnahme an didaktischer Weiterbildung vor der DOS.*

Gruppe	Semester			Lehrformate			Weiterbildung	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	Ja	Nein
VGG	69	6.42	3.21	70	3.27	1.41	12	37
VSG _{PRÄ}	115	5.67	3.13	117	3.31	1.35	-	-
VSG _{POST}	71	7.27	2.61	72	3.78	1.51	19	53
VSG _{ALT}	63	8.10	2.01	63	3.98	1.55	15	25

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}).

Eine einfaktorielle Varianzanalyse zur Prüfung der Unterschiede von unterrichteten Semestern ergab insgesamt signifikante Unterschiede zwischen den Gruppen ($F(3, 314) = 11.24$; $\eta^2 = .097$; $p = .00$). Post-Hoc-Tests (nach Games-Howell) zeigten, dass Lehrende in der VSG_{POST} mehr Semester Lehrerfahrung haben, als die VGG ($p = .002$) und die VSG_{PRÄ} ($p = .000$). Erwartungsgemäß war die Anzahl unterrichteter Semester in der VSG_{POST} signifikant höher als vor Besuch der DOS ($p = .001$). Eine einfaktorielle Varianzanalyse zur Prüfung der Unterschiede bereits durchgeführter Lehrformate, zeigte ebenfalls signifikante Unterschiede zwischen den Gruppen ($F(3, 318) = 4.46$; $\eta^2 = .04$; $p = .004$). Post-Hoc-Tests (nach Gabriel) zeigten, dass Lehrende in der VSG_{ALT} bereits signifikant mehr Lehrformate durchgeführt haben, als die VGG ($p = 0.028$) und die VSG_{PRÄ} ($p = .016$). Die Gruppen unterscheiden sich nicht darin, ob sie bereits Weiterbildungen besucht haben ($\chi^2(2, N = 161) = 2.13$, $p = .35$). Insgesamt haben 40 % der

befragten Lehrenden bereits eine andere didaktische Weiterbildung als die DOS besucht. Die Gruppen der Lehrenden unterschieden sich nicht signifikant bezüglich der Durchführung von einzelnen Lehrformaten, mit Ausnahme von Seminaren (Tabelle 17).

Tabelle 17. *Lehrerfahrung in Lehrformaten*

	Durchgeführt	VGG	VSG _{PRÄ}	VSG _{POST}	VSG _{ALT}
Vorlesung	Ja	43	71	53	48
	Nie	27	46	19	15
$\chi^2 = 6.89, p = .075$					
Seminar	Ja	59	93	67	57
	Nie	11	24	5	6
$\chi^2 = 8.16, p = .043$					
POL	Ja	34	71	41	38
	Nie	36	46	31	25
$\chi^2 = 2.96, p = .398$					
Bedside Teaching	Ja	42	67	36	34
	Nie	28	50	36	29
$\chi^2 = 1.66, p = .645$					
Medi- KIT	Ja	16	24	17	18
	Nie	54	93	55	45
$\chi^2 = 1.50, p = .682$					
Skills-Lab	Ja	21	33	18	22
	Nie	49	84	54	41
$\chi^2 = 1.69, p = .639$					
Praktikum	Ja	-	-	38	30
	Nie	-	-	34	33
$\chi^2 = 0.36, p = .550$					

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). POL = Problemorientiertes Lernen. Medi- KIT = Medizinisches Kommunikations- und Interaktionstraining. N = 322, Freiheitsgrade (df) je = 3, mit Ausnahme bei Praktikum (N = 135, df = 1).

Lehrende der VGG und der VSG_{PRÄ} hatten insgesamt weniger Erfahrungen mit der Durchführung von Seminaren als Lehrende der VSG_{POST} und der VSG_{ALT}. Um mögliche Unterschiede bezüglich der Präferenz bereits durchgeführter Lehrformate aufzudecken, wurden die Lehrenden aus der VGG und die VSG_{PRÄ} verglichen, die entsprechende Lehrformate schon einmal durchgeführt haben. Tabelle 18 zeigt, dass sich die Gruppen nicht unterschieden. Die Lehrenden gaben an, die bereits durchgeführten Lehrformate eher gern durchzuführen, besonders das Bedside Teaching ($M_{VGG} = 1.45$ bzw. $M_{VSG_PRÄ} = 1.65$). Für das Lehrformat Praktikum wurde, aufgrund zu geringer Fallzahlen, keine Analyse durchgeführt.

Tabelle 18. *Präferenz von Veranstaltungsformaten*

	VGG			VSG _{PRÄ}			Vergleich		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>t</i>	<i>df</i>	<i>p</i>
Vorlesung	2.14	0.94	43	1.92	0.94	71	1.24	112	.22
Seminar	1.78	0.81	59	1.88	0.73	90	-0.77	143	.44
POL	2.29	1.17	34	2.23	0.90	70	0.31	102	.75
Bedside Teaching	1.45	0.71	42	1.65	0.64	66	-1.51	106	.13
Medi- KIT	2.73	0.88	15	2.25	1.07	20	1.42	33	.16
Skills Lab	1.76	0.94	21	2.13	0.89	31	-1.43	50	.16

Anmerkungen. VGG = Vergleichsgruppe, VSG_{PRÄ} = Versuchsgruppe vor Besuch der Dozentenschulung. POL = Problemorientiertes Lernen. Skala von 1 (Sehr gerne) bis 5 (Sehr ungern). Levene-Test zur Überprüfung der Varianzhomogenität für alle Gruppen nicht signifikant. Unkorrigierte p-Werte der t-Tests zur Berechnung von Mittelwertunterschieden.

Stellenwert der Lehre. Weiterhin wurde untersucht, welchen Stellenwert die Lehre für die Lehrenden in den Gruppen hat. Die Lehrenden wiesen der Lehre einen relativ hohen persönlichen Stellenwert zu (Tabelle 19), mit signifikanten Unterschieden zwischen den Gruppen ($F(3, 317) = 4.74$; $\eta^2 = .043$; $p = .003$). Post-Hoc-Tests (Gabriel) zeigten, dass die VSG_{ALT} der eigenen Lehre einen signifikant höheren Stellenwert einräumt, als die VSG_{PRÄ} ($p = .002$). Der Stellenwert der Lehre in dem Fachbereich der Lehrenden unterschied sich ebenfalls signifikant zwischen den Gruppen ($F(3, 305) = 2.73$; $\eta^2 = .026$; $p = .044$). Post-Hoc-Tests (Gabriel) zeigten, dass die VSG_{POST} von einem signifikant ($p = .048$) niedrigeren Stellenwert in dem Fachbereich berichtete, als die VSG_{PRÄ}.

Tabelle 19. *Stellenwert der Lehre*

Gruppe	Stellenwert der Lehre					
	Persönlich			Im Fachbereich		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
VGG	72	3.98	0.77	67	3.29	0.95
VSG _{PRÄ}	115	3.74	0.77	111	3.34	0.76
VSG _{POST}	72	3.84	0.80	70	3.00	0.87
VSG _{ALT}	62	4.17	0.62	61	3.12	0.85

Anmerkungen. VGG = Vergleichsgruppe, VSG = Versuchsgruppe. Skala von 1 (*trifft nicht zu*) bis 5 (*trifft voll zu*).

Wahrgenommene Akzeptanz aktivierender Methoden in Vorlesungen. Die wahrgenommene Akzeptanz aktivierender Methoden in Vorlesungen unterschied sich signifikant zwischen den Gruppen ($F(3, 309) = 3.18$; $\eta^2 = .03$; $p = .024$). Post-Hoc-Tests (Gabriel) ergaben, dass die VSG_{PRÄ} signifikant ($p = .048$) niedrigere Akzeptanz aktivierender Methoden in Vorlesungen wahrgenommen hat, als die VGG.

Tabelle 20. *Akzeptanz aktivierender Lehr-und Lernmethoden in Vorlesungen*

Gruppe	<i>N</i>	<i>M</i>	<i>SD</i>
VGG	70	3.65	0.71
VSG _{PRÄ}	113	3.33	0.77
VSG _{POST}	68	3.32	0.77
VSG _{ALT}	62	3.35	0.70
Gesamt	313	3.40	0.75

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Skala von 1 (*trifft nicht zu*) bis 5 (*trifft voll zu*).

Allgemeines Lehrverhalten. Das allgemeine Lehrverhalten wurde über neun Items erfasst (Tabelle 21). Die Lehrenden beschrieben ihre Lehre in allen Gruppen vor allem durch hohe Praxis- und Prüfungsrelevanz sowie dem Versuch, die Studierenden für ihr Fach begeistern zu wollen. Es zeigte sich im Mann-Whitney-Test bei keinem Item Unterschiede zwischen der VGG und der VSG_{PRÄ} (Anhang B6). Weiterhin wies die VGG

bei zwei Items signifikant niedrigere Werte auf als die VSG_{ALT} („Ich strukturiere die Inhalte klar“, „Ich unterbinde auftretende Störungen konsequent“). Die VSG_{POST} wies bei drei Items signifikant höhere Ausprägungen als die VSG_{PRÄ} auf (bei den Items: „Ich zeige die praktische Relevanz der Inhalte auf“, „Ich fördere die besten Studierenden“, „Ich versuche, die Studierenden für mein Fach zu begeistern“). Die VSG_{ALT} unterschied sich nicht von der VSG_{POST}.

Tabelle 21. *Allgemeines Lehrverhalten*

Ich...	VGG		VSG _{PRÄ}		VSG _{POST}		VSG _{ALT}	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
zeige die praktische Relevanz der Inhalte auf	4.31	0.79	4.05	1.09	4.45	0.77	4.40	0.84
fördere die besten Studierenden	2.83	1.03	2.75	1.10	3.17	1.00	2.89	0.95
strukturiere die Inhalte klar	3.96	0.82	3.87	1.05	4.21	0.74	4.24	0.69
unterbinde auftretende Störungen konsequent	2.95	0.79	2.88	0.99	3.13	0.70	3.33	0.82
fasse wichtige Inhalte zusammen	3.78	0.86	3.53	1.09	3.54	1.04	3.85	0.79
lehre mit einem ausgefeilten Präsentationsstil und sehr guter Rhetorik	3.34	1.05	3.13	1.07	3.27	1.02	3.39	1.08
bereite die Studierenden auf die Prüfung(en) vor	4.23	0.89	3.95	1.11	4.28	0.72	4.37	0.81
versuche, die Studierenden für mein Fach zu begeistern	4.32	0.90	4.05	1.16	4.46	0.71	4.48	0.84
helfe (schwächeren) Studierenden, die Unterstützung benötigen	3.26	1.07	3.19	1.23	3.29	0.96	3.43	0.92

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Skala von 1 (Das setze ich nicht um) bis 5 (Das setze ich um).

6.3.3 Hypothesentestung

Um zu prüfen, ob Unterschiede zwischen den beiden Versuchsgruppen (VSG_{POST} und VSG_{ALT}) und der VGG bestehen, wurde zunächst jeweils eine multivariate Varianzanalyse mit den fünf zentralen abhängigen Variablen berechnet. Aufgrund einzelner fehlender Werte umfasste die zugrundeliegende Stichprobe nur 164 von 208 Personen ($N_{VGG} = 50$, $N_{VSG_{POST}} = 61$, $N_{VSG_{ALT}} = 53$). Es zeigten sich signifikante Unterschiede zwischen der VGG und der VSG_{POST} ($F(5, 105) = 3.96$, $p = .002$; $\eta^2 = .16$) sowie zwischen der VGG und der VSG_{ALT} ($F(5, 97) = 3.48$, $p = .006$; $\eta^2 = .15$).

Zudem wurden die ergänzenden Annahmen zu den Gruppenunterschieden mit multivariaten Varianzanalysen geprüft. Die VSG_{POST} wies signifikant höhere Werte als die $VSG_{PRÄ}$ ($N_{VSG_{PRÄ}} = 72$) auf ($F(5, 127) = 9.35$, $p = .000$; $\eta^2 = .27$). Weitere multivariate Varianzanalysen zeigten eine Differenz zwischen der VGG und der $VSG_{PRÄ}$ ($F(5, 116) = 2.56$, $p = .031$; $\eta^2 = .10$) aber keine Unterschiede zwischen der VSG_{ALT} und der VSG_{POST} ($F(5, 108) = 1.27$, $p = .283$; $\eta^2 = .06$).

Die weitere Hypothesentestung erfolgte jeweils einzeln für die abhängigen Variablen. Es wurde angenommen, dass Lehrende nach der DOS (VSG_{POST} und VSG_{ALT}) höhere Ausprägungen auf den abhängigen Variablen aufweisen als die VGG. Es sollte zusätzlich gezeigt werden, dass die VSG_{POST} signifikant höhere Werte auf den abhängigen Variablen aufweist, als die $VSG_{PRÄ}$. Diese Hypothese wurde einmal für alle Lehrenden in den Gruppen berechnet und einmal nur für die Lehrenden, die zu beiden Zeitpunkten einen gültigen Code hinterlegt hatten. Weiterhin wurde angenommen, dass keine Unterschiede zwischen der VGG und der $VSG_{PRÄ}$ bestanden. Schließlich soll der Unterschied zwischen der VSG_{POST} und der VSG_{ALT} geprüft werden. Dabei wurde vorab keine Unterschiedshypothese formuliert. Tabelle 22 fasst die Hypothesen zusammen.

Tabelle 22. *Hypothesen und Fragestellungen in der Lehrendenbefragung.*

Gruppenvergleich	Hypothese/Fragestellung
VGG – VSG _{POST}	Die VSG _{POST} weist eine höhere Ausprägung auf als die VGG.
VGG- VSG _{ALT}	Die VSG _{ALT} weist eine höhere Ausprägung auf als die VGG.
Ergänzende Analysen	
VSG _{PRÄ} – VSG _{POST}	Die VSG _{POST} weist eine höhere Ausprägung auf als die VSG _{PRÄ} .
VSG _{PRÄ} – VSG _{POST} (Code)	Die VSG _{POST} weist eine höhere Ausprägung auf als die VGG (Nur Lehrende, bei denen zu beiden Messzeitpunkten gültige Codes vorlagen).
VGG – VSG _{PRÄ}	Es besteht kein Unterschied zwischen der VGG und der VSG _{PRÄ} .
VSG _{POST} - VSG _{ALT}	Besteht ein Unterschied zwischen der VSG _{POST} und der VSG _{ALT} ?

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}).

Zunächst werden die deskriptiven Ergebnisse abgebildet (Tabelle 23). Dann folgen nacheinander die Ergebnisse der Hypothesentestung.

Tabelle 23. Deskriptive Ergebnisse der abhängigen Variablen, getrennt nach Gruppen (Lehrendenbefragung)

	Alle Personen									Personen mit Code*								
	VGG			VSG _{PRÄ}			VSG _{POST}			VSG _{PRÄ}		VSG _{POST}			VSG _{ALT}			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Studierendenzentrierung	72	2.78	0.52	110	2.75	0.59	69	3.20	0.69	2.69	0.53	38	3.21	0.72	63	2.97	0.71	
Lehrkompetenz	56	3.45	0.65	96	3.30	0.60	68	3.72	0.55	3.35	0.57	33	3.73	0.56	59	3.73	0.57	
SWKE	60	3.76	0.79	100	3.68	0.76	66	3.76	0.67	3.66	0.57	34	3.66	0.60	58	3.71	0.58	
Studierendenzentrierte Lehre	68	3.29	0.60	110	3.34	0.78	69	3.53	0.63	3.24	0.74	38	3.53	0.55	63	3.39	0.68	
Lernziele in der Lehre	69	3.81	0.86	106	3.59	1.00	69	3.93	0.81	3.24	1.03	39	3.91	0.80	62	4.23	0.65	

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). SWKE = Selbstwirksamkeitserwartungen. *Die deskriptiven Ergebnisse der Personen mit Code umfassen nur die Lehrenden, von denen zu beiden Messzeitpunkten Daten vorlagen. Alle Skalen von 1 (niedrige Ausprägung) bis 5 (hohe Ausprägung).

Studierendenzentrierung in Vorlesungen. Die zu prüfende Hypothese 1 lautete, dass die Studierendenzentrierung bei Lehrenden nach der DOS höher ist, als bei Lehrenden ohne Besuch der DOS. Hypothesenkonform zeigten sich Unterschiede zwischen der VGG und der VSG_{POST} sowie der VSG_{ALT}. Ebenso hypothesenkonform wies die VSG_{POST} höhere Ausprägungen als die VSG_{PRÄ} auf (Tabelle 24). Die Effektstärken waren klein bis moderat ($d = 0.31 - 0.72$).

Tabelle 24. *Studierendenzentrierung in Vorlesungen.*

Gruppenvergleich	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Hypothesentestung				
VGG – VSG _{POST}	126.4	4.08	.000	0.69
VGG – VSG _{ALT}	112.9	1.80	.038	0.31
Ergänzende Analysen				
VSG _{PRÄ} – VSG _{POST}	177	4.57	.000	0.71
VSG _{PRÄ} – VSG _{POST} (Code)	37	5.05	.000	0.72
VGG – VSG _{PRÄ}	180	-0.26	.792	-
VSG _{POST} – VSG _{ALT}	130	-1.89	.061	-0.33

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Positive *t*-Werte und Effektstärken stehen für höhere Werte der jeweils zweitgenannten Gruppe.

Die berichtete Studierendenzentrierung in Großgruppenveranstaltungen unterschied sich nicht zwischen der VGG und der VSG_{PRÄ}. Ebenfalls zeigte sich kein signifikanter Unterschied zwischen der VSG_{POST} und der VSG_{ALT}.

Lehrkompetenz. Die zu prüfende Hypothese 2 lautete, dass Lehrende sich nach der DOS eine höhere Lehrkompetenz zuschreiben als Lehrende ohne Besuch der DOS. Hypothesenkonform zeigten sich signifikante Unterschiede zwischen der VGG und der VSG_{POST} sowie der VSG_{ALT}. Ebenso hypothesenkonform zeigten Lehrende der VSG_{POST} nach der DOS höhere Ausprägungen als davor (Tabelle 25). Die Effektstärken waren klein bis moderat ($d = 0.45 - 0.72$).

Tabelle 25. *Lehrkompetenz*

Gruppenvergleich	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Hypothesentestung				
VGG – VSG _{POST}	122	2.49	.007	0.45
VGG – VSG _{ALT}	113	2.46	.008	0.46
Ergänzende Analysen				
VSG _{PRÄ} – VSG _{POST}	177	4.57	.000	0.72
VSG _{PRÄ} – VSG _{POST} (Code)	33	2.55	.008	0.68
VGG – VSG _{PRÄ}	150	-1.49	.138	-
VSG _{POST} – VSG _{ALT}	125	0.11	.912	0.02

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Positive *t*-Werte und Effektstärken stehen für höhere Werte der jeweils zweitgenannten Gruppe.

Die selbst eingeschätzte Lehrkompetenz unterschied sich nicht zwischen der VGG und der VSG_{PRÄ}. Ebenfalls zeigte sich kein signifikanter Unterschied zwischen der VSG_{POST} und der VSG_{ALT}.

Selbstwirksamkeitserwartungen. Die zu prüfende Hypothese 3 lautete, dass Lehrende nach der DOS von höheren Selbstwirksamkeitserwartungen bezüglich ihrer Lehre berichten als Lehrende ohne Besuch der DOS. Die Ergebnisse stehen nicht in Einklang mit der Hypothese (Tabelle 26). Ebenfalls zeigte sich kein signifikanter Unterschied zwischen der VGG und der $VSG_{PRÄ}$ sowie der VSG_{POST} und der VSG_{ALT} .

Tabelle 26. *Selbstwirksamkeitserwartungen*

Gruppenvergleich	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Hypothesentestung				
VGG – VSG_{POST}	124	0.01	.498	0.06
VGG – VSG_{ALT}	108.23	0.402	.345	-0.07
Ergänzende Analysen				
$VSG_{PRÄ}$ – VSG_{POST}	164	-0.70	.242	0.17
$VSG_{PRÄ}$ – VSG_{POST} (Code)	32	-1.55	.065	0.25
VGG – $VSG_{PRÄ}$	158	-0.65	.519	-
VSG_{POST} – VSG_{ALT}	122	-0.45	.655	-0.08

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor ($VSG_{PRÄ}$) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). Positive *t*-Werte und Effektstärken stehen für höhere Werte der jeweils zweitgenannten Gruppe.

Umsetzung studierendenzentrierter Lehre. Die zu prüfende Hypothese 4 lautete, dass Lehrende nach der DOS häufiger studierendenzentriert lehren als Lehrende ohne Besuch der DOS. Hypothesenkonform zeigten sich Unterschiede zwischen der VGG und der VSG_{POST}, vor allem bei der direkten Veränderungsmessung (Tabelle 27). Keine Unterschiede zeigten sich zwischen der VGG und der VSG_{ALT}. Lehrende der VSG_{POST} zeigten nach der DOS hypothesenkonform höhere Ausprägungen gegenüber der VSG_{PRÄ}. Die Effektstärken waren insgesamt unbedeutend bis moderat (0.16 – 0.74).

Tabelle 27. *Umsetzung studierendenzentrierter Lehre*

Gruppenvergleich	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Hypothesentestung				
VGG – VSG _{POST}	135	2.24	.014	0.26
VGG – VSG _{POST} (DV)	95	4.80	.000	0.74
VGG – VSG _{ALT}	129	1.37	.086	0.16
Ergänzende Analysen				
VSG _{PRÄ} – VSG _{POST}	165.20	1.75	.041	0.26
VSG _{PRÄ} – VSG _{POST} (Code)	37	2.47	.009	0.44
VGG – VSG _{PRÄ}	167.44	0.45	.651	-
VSG _{POST} – VSG _{ALT}	130	-0.83	.410	-0.28

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). DV = Direkte Veränderungsmessung. Positive *t*-Werte und Effektstärken stehen für höhere Werte der jeweils zweitgenannten Gruppe.

Die Umsetzung studierendenzentrierter Lehre unterschied sich nicht zwischen der VGG und der VSG_{PRÄ}. Die VSG_{ALT} lehrten etwas weniger studierendenzentriert als die VSG_{POST}.

Lernziele. Die zu prüfende Hypothese 5 lautete, dass Lehrende nach der DOS ihre Lehre stärker an den Lernzielen der Lehrveranstaltung orientieren als Lehrende ohne Besuch der DOS. Hypothesenkonform zeigten sich Unterschiede zwischen der VGG und der VSG_{ALT}. Bei der indirekten Veränderungsmessung zeigten sich keine signifikanten Unterschiede zwischen der VGG und der VSG_{POST}, bei der direkten Veränderungsmessung zeigten sich signifikante Unterschiede. Ebenso hypothesenkonform zeigten Lehrende der VSG_{ALT} höhere Ausprägungen als die VGG (Tabelle 28). Die Effektstärken waren klein bis hoch (0.37 – 1.10).

Tabelle 28. *Lernziele*

Gruppenvergleich	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Hypothesentestung				
VGG – VSG _{POST}	136	0.87	.194	0.37
VGG – VSG _{POST} (DV)	95	-3.35	.001	1.10
VGG – VSG _{ALT}	125.38	3.12	.001	0.55
Ergänzende Analysen				
VSG _{PRÄ} – VSG _{POST}	173	2.40	.009	0.37
VSG _{PRÄ} – VSG _{POST} (Code)	38	3.59	.001	0.75
VGG – VSG _{PRÄ}	173	-1.51	.133	-
VSG _{POST} – VSG _{ALT}	129	2.53	.026	0.41

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). DV = Direkte Veränderungsmessung. Positive *t*-Werte und Effektstärken stehen für höhere Werte der jeweils zweitgenannten Gruppe.

Die Orientierung an Lernzielen in der Lehre unterschied sich nicht zwischen der VGG und der VSG_{PRÄ}. Weiterhin zeigte sich ein signifikanter Unterschied zugunsten der VSG_{ALT} gegenüber der VSG_{POST} mit mittlerer Effektstärke.

Reaktion. Die Hypothese 6 lautete, dass die Teilnehmer rückblickend zufrieden mit der DOS waren. Sowohl Lehrende in der VSG_{POST} ($M = 3.67, SD = 1.08$) als auch in der VSG_{ALT} ($M = 3.67, SD = 1.13$) berichteten insgesamt von eher hoher Zufriedenheit. Die Werte der VSG_{POST} ($t(70) = 5.21, p = .000$) und die der VSG ($t(59) = 4.59, p = .000$) waren jeweils signifikant höher als der numerische Mittelwert der Skala.

Weiterbildungsbereitschaft. Die Hypothese 7 lautete, dass die DOS Lehrende dazu motiviert, an weiteren didaktischen Weiterbildungen teilzunehmen. Sowohl Lehrende der VSG_{POST} ($M = 3.90, SD = 0.95$) als auch der VSG_{ALT} ($M = 3.69, SD = 1.06$) wiesen eine relativ hohe Weiterbildungsbereitschaft auf. Die Werte waren signifikant höher als der numerische Mittelwert der Skala, einmal bei der VSG_{POST} ($t(69) = 7.93, p = .000$) und einmal bei der VSG_{ALT} ($t(58) = 5.06, p = .000$). Die Weiterbildungsbereitschaft beider Gruppen unterschied sich nicht ($t(127) = 1.16, p = .248$).

6.4 Diskussion

Zuerst werden die Ergebnisse der Forschungshypothesen und der Kontrollvariablen kommentiert. Danach wird die Zusammensetzung der Stichprobe diskutiert, bevor auf den Untersuchungsplan, die statistische Auswertung, die psychometrischen Eigenschaften der Skalen, und die Kontrollvariablen eingegangen wird.

Die multivariaten Varianzanalysen ergaben signifikante Unterschiede zwischen der VGG und den Versuchsgruppen, mit hohen Effektstärken nach Cohen (1988). Weiterhin wies die VSG_{POST} höhere Werte auf als die VSG_{PRÄ}. Lehrende wiesen demnach insgesamt nach Besuch der DOS höhere Ausprägungen auf den abhängigen Variablen auf, als Lehrende vor, beziehungsweise ohne Besuch der DOS.

Studierendenzentrierung in Vorlesungen. Hypothesenkonform zeigten sich signifikante Unterschiede in der Studierendenzentrierung in Vorlesungen zwischen den Versuchsgruppen und der VGG, bei kleinen bis moderaten Effektstärken. Lehrende sind zudem nach der DOS eher davon überzeugt, dass eine studierendenzentrierte Lehre für Großgruppenveranstaltungen gut geeignet ist. Die Skalenwerte streuen jedoch um den mittleren Bereich der Skala, erreichen also keine hohen Ausprägungen.

Aufgrund der grundsätzlich starken Lehrendenzentrierung von Vorlesungen, sollten Unterschiede bei der Einstellung zu studierendenzentrierter Lehre besonders deutlich zutage treten. Der Nachteil bei der Verwendung der hier verwendeten Skala besteht darin, dass die

Lehrenden in dieser Untersuchung unterschiedliche Lehrerfahrungen mit Großgruppenveranstaltungen besitzen. Offen bleibt zudem, wie stark sich die hier berichteten Ansichten zur Studierendenzentrierung auf andere Lehrformate übertragen lassen. Die kleine bzw. moderate (bei der VSG_{POST}) Korrelation der Skalen zur Studierendenzentrierung steht zwar mit dieser Annahme in Einklang (Tabelle 15), in der Literatur fehlen dazu jedoch Belege (Lübeck, 2009). Darüber hinaus bleibt offen, inwiefern die Annahme zutrifft, dass studierendenzentrierte Lehrkonzeptionen der Lehrenden zu studierendenzentrierten Lehrhandlungen führen und diese wiederum zu besseren Lernergebnissen.

Didaktische Kompetenz. Angesichts der unklaren Konstruktdefinition (vgl. Kapitel 3.3) wurde eine Skala verwendet, die nur einen Teilbereich der Lehrkompetenz fokussiert. Hypothesenkonform zeigten sich signifikante Verbesserungen nach der Teilnahme an der DOS, bei kleinen bis moderaten Effektstärken. Möglicherweise erklärt sich die höhere Selbsteinschätzung durch das Feedback zu der Lehre, das den Teilnehmern der DOS im untersuchten Zeitraum gegen wurde: zweimal im Rahmen der kollegialen Hospitationen und möglicherweise unmittelbar durch die Studierenden in den durchgeführten Lehrveranstaltungen. Standardisierte studentische Evaluationen haben die Lehrenden nicht erhalten, da die Lehre an der Medizinischen Fakultät primär modular und nicht personenbezogen evaluiert wird. Da die Annahmen der Lehrenden über ihre eigenen Fähigkeiten mit Lehrerfolg in Zusammenhang stehen (Feldman, 1989; Roche & Marsh, 2000; Stehle, 2011), sollte das Konstrukt Didaktische Kompetenz in zukünftigen Studien näher erforscht werden.

Selbstwirksamkeitserwartungen bezüglich der Lehre. Die Zuversicht der Lehrenden, selbst eine positive Wirkung auf den Lernerfolg der Studierenden ausüben zu können, liegt in allen Gruppen über dem numerischen Mittelwert der Skala. Die Hypothese, dass die Teilnahme an der DOS mit höheren Selbstwirksamkeitserwartungen bezüglich der Lehre einhergeht, musste verworfen werden. Die Ergebnisse dieser Studie gleichen dem Befund bei Postareff et al. (2007). Dort wurde ebenfalls kein bedeutsamer Unterschied bei den Selbstwirksamkeitserwartungen bezüglich der Lehre festgestellt. Lindblom-Ylänne et al. (2006) nehmen an, dass sich lehrbezogene Selbstwirksamkeitserwartungen nur langsam verändern. Zudem zeigte sich hier, dass die Lehrenden schon vor Besuch der DOS die Annahme haben, dass sie mit ihrer Lehre den Lern- und Prüfungserfolg der Studierenden beeinflussen können. Lehrbezogene Selbstwirksamkeitserwartungen wurden in dieser und in der Studie von Postareff et al. (2007) untersucht, da sie möglicherweise in bedeutsamem Zusammenhang mit dem Erwerb studierendenzentrierter Lehransätze stehen. Das konnte jedoch nicht gezeigt werden

(Lübeck, 2009; Postareff et al., 2007) und auch in dieser Arbeit traten lediglich niedrige Korrelationen mit den Skalen zu studierendenzentrierter Lehre auf. Hohe Studierendenzentrierung ist damit kein Indikator für höhere Selbstwirksamkeitserwartungen in der Lehre.

Umsetzung studierendenzentrierter Lehre. Hypothesenkonform zeigten sich signifikante Unterschiede vor und nach der DOS, mit kleinen Effektstärken bei der indirekten Veränderungsmessung. Dies entspricht den berichteten Effekten in anderen Studien (Gibbs & Coffey, 2004; Postareff et al., 2007; Ulrich, 2013). Die höhere Effektstärke bei der direkten Veränderungsmessung lässt sich möglicherweise durch die hohe Veränderungssensitivität der direkten Veränderungsmessung begründen. Da die Vergleichsgruppe hier kleiner war, sind diese Daten jedoch etwas weniger repräsentativ. Dass die Ergebnisse der direkten und indirekten Veränderungsmessung nicht korrelieren, könnte zudem gegen die Validität der vorliegenden Messung sprechen.

Als Voraussetzung für den Abschluss der DOS müssen die Lehrenden zwei Lehrveranstaltungen im Zeitraum der Weiterbildung halten. Diese beiden Gelegenheiten sollten ausreichen, um Veränderungen umzusetzen. Nicht erfasst wurde, wie viele Lehrveranstaltungen die Lehrenden in der Zwischenzeit gehalten haben, dies sollte in zukünftigen Studien aufgegriffen werden.

Lernziele. Hypothesenkonform zeigten sich mehrheitlich signifikante Unterschiede, bei kleinen bis moderaten Effektstärken bei der indirekten Veränderungsmessung. Die Lehrenden der VSG_{POST} gaben an, sich deutlicher an den Lernzielen in ihrer Lehre zu orientieren und diese mitzuteilen als die Lehrenden der VSG_{PRÄ}. Allerdings unterschied sich lediglich die VSG_{ALT} signifikant von der VGG, nicht aber die VSG_{POST}. Die höhere Effektstärke bei der direkten Veränderungsmessung kann, wie schon bei der Studierendenzentrierung, möglicherweise durch die höhere Veränderungssensitivität der direkten Veränderungsmessung begründet sein. Dass indirekte und direkte Veränderungsmessung auch bei dieser Skala keine Übereinstimmung zeigte, spricht wieder gegen die Validität der vorliegenden Messung. Die relativ hohen Werte bei allen Lehrenden überraschen, da Lehrende erfahrungsgemäß eher selten die Lernziele transparent machen (vgl. Kapitel 7.1.3). Es ist möglich, dass Lehrende unter dem Begriff Lernziele weitere Handlungen verstehen, wie die bloße Nennung der Themen einer Lehrveranstaltung. In zukünftigen Studien sollten genauere Konstruktdefinitionen vorgelegt werden.

Reaktion. Sowohl Lehrende der VSG_{POST} als auch VSG_{ALT} der DOS berichteten von einer eher hohen Zufriedenheit. Diese Ergebnisse stehen in Einklang mit den unmittelbaren

Rückmeldungen, welche die Trainer innerhalb der DOS erhalten. Auch wenn die Reaktion auf eine Weiterbildung nicht zwingend mit Veränderungen auf den anderen Zielebenen einhergeht (Stes et al., 2009), und auch in dieser Studie nur geringe Korrelationen mit den anderen Variablen auftreten, ist eine hohe Zufriedenheit der Teilnehmer aus Sicht der Programmgestalter wichtig.

Weiterbildungsbereitschaft. Sowohl Lehrende in der VSG_{POST} als auch in der VSG_{ALT} berichteten von einer relativ hohen Weiterbildungsbereitschaft nach der DOS. Höhere Zufriedenheit mit der DOS ging mit höherer Weiterbildungsbereitschaft einher (Tabelle 15). Da das Interesse an Weiterbildungen bei Lehrenden mit geringer Lehrerfahrung höher ist (Ulrich et al., 2008; Schwaiger & Salzgeber, 2004), und die Lehrenden in dieser Untersuchung über Lehrerfahrung verfügen, stellt dies eine positive Wirkung der DOS dar. Auch dass die DOS eine verpflichtende Weiterbildung ist, spricht dafür, dass die Teilnehmer didaktische Weiterbildungen als effektiv erleben. Jedoch ist die Teilnahmeabsicht kein Garant für die tatsächliche Teilnahme an weiteren Weiterbildungen (Spiel & Fischer, 1998).

Da keine Vergleichsgruppe vorlag, wurde die Hypothese gegen den numerischen Mittelwert getestet. Besser wäre es, die Weiterbildungsbereitschaft einer vergleichbaren Gruppe ohne Weiterbildung als Maßstab heranzuziehen. Alternativ oder ergänzend, könnte die Frage zu der Weiterbildungsbereitschaft schon bei der Vorerhebung gestellt werden. Zudem wäre es interessant, die genauen Ziele zu erfahren, die die Lehrenden mit der Weiterbildung anstreben.

Kontrollvariablen. In dieser Studie wurden acht verschiedene Kontrollvariablen erhoben. Damit sollten die Mängel des Studiendesigns, wie die fehlende Randomisierung, das Vorliegen einer Gelegenheitsstichprobe und der Verlust von Teilnehmern, minimiert werden. Wenn Unterschiede auftraten, waren diese gering oder begünstigten die Hypothesentestung nicht. Eine Ausnahme stellte die Lehrerfahrung dar. Lehrende der VSG_{POST} und der VSG_{ALT} haben durchschnittlich ein bis zwei Semester mehr Lehrerfahrung als Lehrende in der VGG. Die Unterschiede bei der Lehrerfahrung wirkten sich vermutlich aber nur gering auf die abhängigen Variablen zur Studierendenorientierung aus. Bei Fendler und Gläser-Zikuda (2013) ging höheres Alter der Lehrenden (und damit wahrscheinlich höhere Lehrerfahrung) mit niedrigerer Studierendenorientierung einher, während Lübeck (2009) keinen Zusammenhang zwischen Lehrerfahrung und Studierendenorientierung fand.

Stichprobe. Die Stichprobengröße der Untersuchung war insgesamt höher als bei vergleichbaren Studien. So war, unter der Annahme mittlerer Effektstärken, die Wahrschein-

lichkeit eines β -Fehlers gering. Kritisch war dagegen, dass die Beteiligung in der VSG_{POST} geringer war als in der VSG_{PRÄ}. Die Anzahl gültiger Codes zu beiden Messzeitpunkten fiel hier zudem mit 58 % relativ gering aus. In anderen Untersuchungen mit personalisierten Codes (Grube, Morgan & Kearney, 1989; Kearney, Hopkins, Mauss & Weisheit, 1984; Pöge, 2005) zeigten sich höhere Werte übereinstimmender Codes. Vergleichbare Daten aus hochschuldidaktischen Studien liegen nicht vor. Für die fehlenden Angaben kommen neben Zeitmangel der Probanden, vor allem Bedenken bezüglich der Anonymität in Betracht. Lehrebezogene Merkmale können den Verlust von Versuchspersonen nicht erklären. Lehrende mit gültigem Code zu beiden Messzeitpunkten unterschieden sich zu Messzeitpunkt 1 (VSG_{PRÄ}) kaum von Lehrenden ohne Code (vgl. Tabelle 23). Da die Zufriedenheit mit der DOS erst zum zweiten Messzeitpunkt erhoben wurde, bleibt offen, ob sich mit der DOS unzufriedene Lehrende seltener an der zweiten Befragung beteiligt haben. Dies würde zu einer Überschätzung der Wirkung der DOS führen.

Zum Befragungszeitpunkt waren etwa 2000 Personen an der Medizinischen Fakultät beschäftigt, die gelegentlich Lehrveranstaltungen durchführen ohne die DOS besucht zu haben und etwa 500 – 700 Personen, welche die DOS (oder die frühere, kürzere Variante, vgl. Kapitel 5.1) absolviert hatten. Wie repräsentativ die Lehrenden in der VGG und in der VSG_{POST} für diese Grundgesamtheit sind, kann nicht beantwortet werden.

Statistische Auswertung. Die multivariate Varianzanalyse konnte eine generelle Wirkung der DOS auf die wichtigsten abhängigen Variablen aufzeigen. Der Vorteil einer simultanen Analyse der abhängigen Variablen bestand darin, dass keine Alpha-Fehler-Kumulierung auftritt. Aufgrund fehlender Werte wurden für diese Analyse allerdings relativ viele Lehrende von der Untersuchung ausgeschlossen. Die exakte Hypothesentestung erfolgte danach mit multiplen t-Tests, unter der Annahme, dass die Skalen intervallskaliert und die Abstände zwischen den Antwortkategorien gleich groß sind. Wenn ordinalskalierte Items als intervallskaliert betrachtet, und fälschlicherweise parametrische Verfahren eingesetzt werden, kann dies zu falschen Schlussfolgerungen führen (Jamieson, 2004). Da bei diesen Folgeanalysen auf eine Alpha-Korrektur verzichtet wurde, bestand bei den Signifikanztests die Gefahr, dass Hypothesen zu Unrecht beibehalten wurden. Weiterhin erfolgte die Hypothesentestung einseitig, da keine Unterschiede in die entgegengesetzte Richtung zu erwarten waren. Zur Einschätzung der praktischen Relevanz wurden hier Effektstärken berechnet, um die praktische Relevanz der Ergebnisse aufzuzeigen. Dies führte zu einer geringeren Teststärke. Aufgrund der fehlenden Codes, mussten die Unterschiede von VSG_{PRÄ} und VSG_{POST}

mit t-Tests für unabhängige Stichproben berechnet werden, wodurch die Teststärke bei diesen Vergleichen verringert wurde.

Untersuchungsplan. Die gefundenen Effekte sollten sich kausal durch die Teilnahme an der DOS erklären lassen. Es ist aber nicht auszuschließen, dass sich die Teilnehmer über die DOS hinaus mit der Weiterentwicklung ihrer Lehre beschäftigt haben. Wenn diese Entwicklung nicht durch die Inhalte der DOS angeregt wurde, würde dies zu einer Überschätzung der direkten Wirksamkeit der Weiterbildung führen (vgl. Kapitel 4.2). Möglicherweise haben sich auch Unterschiede durch sozial erwünschte Antworten ergeben, insbesondere bei Lehrenden in den Versuchsgruppen, denen die Ziele der DOS bewusst waren. Diese Limitation sollte zum einen durch die Messinstrumente und zum anderen durch die Bitte im Anschreiben, ein möglichst realistisches Bild der eigenen Lehre und der Rahmenbedingungen aufzuzeigen, minimiert werden.

Psychometrische Eigenschaften der Skalen. Die für diese Untersuchung neu entwickelten Skalen wiesen interne Konsistenzen von mindestens $\alpha = .61$ auf, was für Gruppenvergleiche als akzeptabel gilt (Rost, 2007). Wie in anderen Studien zu der Evaluation didaktischer Weiterbildungen kann die Validität der Messungen kritisiert werden (Steinert et al., 2006; Stes et al., 2010). Für den Einsatz selbst konstruierter Items und Skalen sprach der Mangel an geeigneten Alternativen. Die Entscheidung gegen den Einsatz des *ATI-R* (Trigwell et al., 2005) geschah angesichts dessen mangelnden Validität (Meyer & Eley, 2006; Lübeck, 2009; vgl. Kapitel 4.3). Die hier verwendete Skala zur Umsetzung studierendenzentrierter Lehre ist eher studierendenzentrierten Lehransätzen (Intentionen und Strategien) zuzuordnen, die Studierendenzentrierung in Vorlesungen eher den studierendenzentrierten Lehrkonzeptionen (Überzeugungen, vgl. Kapitel 3.4). Der lediglich kleine bzw. moderate (bei der VSG_{POST}) statistische Zusammenhang beider Skalen (vgl. Tabelle 15) lässt sich durch die asymmetrischen Konstruktoperationalisierungen erklären. Der Lehransatz der Lehrenden wurde hier allgemein gemessen, die Lehrkonzeption hingegen konkret für Lehrveranstaltungen mit mindestens 40 Studierenden. So konnte hier eine höhere Studierendenzentrierung anhand zwei verschiedener Indikatoren nachgewiesen werden. In zukünftigen Studien sollte die Studierendenzentrierung als Konstrukt weiter erforscht und unterschiedliche Messinstrumente entwickelt werden, die in verschiedenen Lehrkontexten eingesetzt werden können. Dazu müsste auch theoretisch und empirisch untersucht werden, wie sich lehrendenzentrierte Merkmale zu Skalen zusammenfassen lassen.

Die Entscheidung gegen die vorliegenden Skalen zur Selbstwirksamkeit (Postareff et al., 2007; Schmitz & Schwarzer, 2000) fiel aufgrund der zu breiten Konstruktoperationalisierung. Die hier verwendete Skala sollte ausschließlich messen, inwiefern Lehrende annehmen, eine positive Wirkung auf den Lernerfolg der Studierenden ausüben zu können. Sowohl die zufriedenstellende interne Konsistenz als auch die mittleren Korrelationen zu der selbst eingeschätzten didaktischen Kompetenz sprechen für eine zufriedenstellende Messung in dieser Arbeit.

Fazit. Die Ergebnisse stehen in Einklang mit der Annahme, dass die DOS eine positive Wirkung hat. Lehrende, welche die DOS besucht haben, berichteten über eine höhere Studierendenzentrierung, stärkere Orientierung an den Lernzielen und höhere didaktische Kompetenz als die VGG. Die Zufriedenheit mit der DOS war hoch und Lehrende gaben eine hohe Weiterbildungsbereitschaft an. Die fehlende Randomisierung und das Fehlen eines zweiten Messzeitpunktes bei der VGG mindern die interne Validität der Untersuchung. Zudem wurde auf nicht validierte Messinstrumente zurückgegriffen, die erst für diese Studie entwickelt wurden. Andererseits lassen sich die Ergebnisse nicht auf Kontrollvariablen zurückführen. Die Wirkung der DOS konnte sowohl unmittelbar danach, als auch bei Lehrenden nachgewiesen werden, deren DOS länger zurücklag. Wie in vergleichbaren Studien fielen die Ergebnisse eher gering bis moderat aus (vgl. Stes et al., 2009). Veränderungen von lehrebezogenen Kognitionen und Verhalten verändern sich nur eher langsam (Lindblom-Ylänne et al., 2006; Kember & Kwan, 2000; Knol, Vorst, van Driel & Mellenbergh, 2013). Auch die praktische Erfahrung im Rahmen der DOS zeigte, dass Lehrende eher kleine Veränderungen bei ihrer Lehre vornehmen, teilweise auch bedingt durch Zeitmangel und wahrgenommene Beschränkungen durch curriculare Vorgaben. Daher ist es unwahrscheinlich, dass eine zeitlich noch umfassendere Weiterbildung entsprechend höhere Effekte bewirkt. Offen bleibt zudem, inwiefern die Ergebnisse mit tatsächlichen Veränderungen in der Lehre einhergehen. Um den Transfer der Ergebnisse in die Praxis zu analysieren, benötigt es weitere Untersuchungen. Dazu wurden die folgenden Studien konzipiert.

7 Empirische Studie 2: Das Lehrverhalten

In diesem Kapitel wird dargestellt, wie an der Medizinischen Fakultät gelehrt wird und welchen Einfluss die DOS auf die Lehre hat. Die wichtigsten Ziele der DOS lauten, dass Lehrende studierendenzentrierte Lehre umsetzen und die Lernziele transparent machen sollen (vgl. Kapitel 5). Zuerst wurde in einer Vorstudie (Kapitel 7.1) ermittelt, wie häufig studierendenzentrierte Lehre in Lehrveranstaltungen generell umgesetzt wird. Für den Vergleich von Lehrenden mit und ohne Besuch der DOS wurde ein Beobachtungsbogen entwickelt (Kapitel 7.2), dessen Einsatz in Kapitel 7.3 beschrieben wird. In Kapitel 7.4 werden die Ergebnisse der kollegialen Hospitation beschrieben, die von den Teilnehmern der DOS durchgeführt, beobachtet und dokumentiert wurden.

7.1 Basisdaten Beobachtungsstudie

7.1.1 Fragestellung

Die DOS soll Lehrende an der Medizinischen Fakultät Heidelberg dazu motivieren, studierendenzentrierte Lehre in Seminaren und Vorlesungen umzusetzen und die genauen Lernziele zu nennen. Um festzustellen, wie häufig studierendenzentrierte Lehre umgesetzt wird und wie häufig die Lernziele schon ohne Besuch der DOS von Lehrenden genannt werden, wurde eine repräsentative Stichprobe von Lehrveranstaltungen empirisch untersucht.

7.1.2 Methodik

Vorgehen. Um eine Schätzung der Lehrhandlungen im Studium der Humanmedizin vorzunehmen, setzten zwei Studierende (beide weiblich und im vierten Studienjahr), einen Beobachtungsbogen in Lehrveranstaltungen ein. Beide waren bereits bei der Entwicklung des umfassenderen Beobachtungsbogens für Lehrveranstaltungen beteiligt (vgl. Kapitel 7.2.1) und dadurch für diese Aufgabe geschult. Sie besuchten zwischen Mai und November 2011 gemeinsam 119 Seminare und Vorlesungen im klinischen Abschnitt. Die Lehrveranstaltungen fanden im Rahmen verschiedener Fächer statt, die modular aufeinander folgten. Andere Lehrformate des Medizinstudiums (vgl. Kapitel 2.2) wurden nicht berücksichtigt, da der Anteil studierendenzentrierter Lehre dort mehr vom Veranstaltungsformat als von den Lehrenden vorgegeben wird. Die Seminare und Vorlesungen sind an der Medizinischen Fakultät Heidelberg traditionell eher lehrendenzentriert ausgerichtet. Um den Einfluss sozialer Erwünschtheit bei der Messung von Lehrhandlungen auszuschließen, wurden die Lehrenden

nicht über die Beobachtung der Lehre informiert. Daher wurden keine personenbezogenen Daten der Lehrenden erhoben.

Stichprobe. Es wurden 14 Seminare und 104 Vorlesungen ($n = 104$) beobachtet, dies stellt nahezu eine Vollerhebung der Seminare und Vorlesungen in den beiden untersuchten klinischen Semestern dar. Bei einer Lehrveranstaltung fehlte die Angabe zum Veranstaltungsformat. Nur fünf der untersuchten Lehrveranstaltungen hatten mehr als 40 Teilnehmer. Es ist davon auszugehen, dass nur wenige Lehrende mehrfach beobachtet wurden, da Lehrende in der Regel nur eine Lehrveranstaltung innerhalb eines Moduls durchführen. Da zudem 18 verschiedene Fächer in dem Untersuchungszeitraum beobachtet wurden, ist die Varianz der Lehrenden in dieser Untersuchung hoch.

Ein t-Test für unabhängige Stichproben zeigte insgesamt keinen Unterschied beider Formate bezüglich der Anzahl der Teilnehmer ($t(116) = 1.11, p = .263$) oder der Dauer der Lehrveranstaltungen ($t(116) = 0.94, p = .350$). Die Studierenden protokollierten die Teilnehmeranzahl und die Dauer der Lehrveranstaltungen weitgehend übereinstimmend, bei Abweichungen wurde der mittlere Wert für die Berechnung der Stichprobeneigenschaften herangezogen (Tabelle 29).

Tabelle 29. *Teilnehmeranzahl und Dauer der untersuchten Lehrveranstaltungen*

	<i>M</i>	<i>SD</i>	Minimum	Maximum
Teilnehmer	31.0	21.5	10	150
Dauer (in Minuten)	68.1	26.0	30	180

Anmerkung. $N = 119$ Lehrveranstaltungen.

Material. Da die Beobachtungen einfach durchführbar sein sollten, wurde ein kurzer Beobachtungsbogen mit dichotomer bzw. ordinalskaliertem Kategorisierung verwendet (Abbildung 5). Die fünf ausgewählten Items sollten mit *individuellen Lern- und Arbeitsphasen* und *studentischen Wortbeiträgen* die Kernelemente studierendenzentrierter Lehre erfassen (Johannes & Seidel, 2012; Lea et al., 2003; O'Neill & McMahon, 2005; vgl. Kapitel 3.4). Dazu wurden weitere Aspekte ausgewählt: die *Nennung der Lernziele*, ein ausführlicher *Einstig in die Lehrveranstaltung* und die *Zusammenfassung der Inhalte*. Die Ein- und Ausschlusskriterien der Items wurden mit den Beobachterinnen im Vorfeld besprochen und fortlaufend während der Erhebungsphase diskutiert.

Veranstaltungsformat:	
Anzahl der Teilnehmer:	
Dauer:	Datum:
Welche didaktischen Handlungen treten auf?	
Bitte <u>nur</u> ankreuzen, wenn sie eingesetzt werden!	
Führt zu Thema hin (Frage, Problem, Fall)	<input type="radio"/>
Stellt Lernziele ausdrücklich dar	<input type="radio"/>
Fasst Wichtiges zusammen	<input type="radio"/>
Wortmeldungen von Teilnehmern*:	
Keine <input type="checkbox"/>	1-5 <input type="checkbox"/>
6-10 <input type="checkbox"/>	> 10 <input type="checkbox"/>
Methodeneinsatz¹ (Individuelle Lernphasen): Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Anteil Dozentenzentrierter Lehre (Kollektive Lernphasen)	
0% <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 100%	
<small>¹ Methoden sind (mentale) Aktivierungen <u>aller</u> Studierenden (z.B. Übungsaufgaben, Partnergespräch, One-Minute Paper usw.)</small>	
<small>* Anzahl aller Beiträge z.B. Fragen, Antworten usw.</small>	

Abbildung 5. Beobachtungsbogen bei der Erhebung der Basisdaten

Bei einem Item (Methodeneinsatz) gab es einmal fehlende Werte. Um die Güte der Beobachtungen zu überprüfen, wurde der Kappa-Koeffizient nach Cohen (1960) als zufallskorrigiertes Übereinstimmungsmaß herangezogen. Beurteilerübereinstimmungen von $\kappa > .60$ gelten als gut (Fleiss & Cohen, 1973; Landis & Koch, 1977, Wirtz & Caspar, 2002). Es zeigten sich hier Beobachterübereinstimmungen von $\kappa > .70$ zwischen den beiden Beobachtungen (Tabelle 30).

Tabelle 30. Beobachtungskategorien und Beobachterübereinstimmung

Item	κ
Führt zu Thema hin (Frage, Problem, Fall) ^a	.84
Stellt Lernziele ausdrücklich dar ^a	.72
Fasst Wichtiges zusammen ^a	.78
Wortmeldungen von Teilnehmern ^b	.93
Methodeneinsatz (Individuelle Lernphasen) ^a	.81

Anmerkungen. $N = 119$ Lehrveranstaltungen. κ = Cohens Kappa. a = Dichotomes Beobachtungsformat (Handlung tritt auf oder tritt nicht auf). b = Ordinalskaliertes Beobachtungsformat (0; 1–5; 6–10; > 10.)

7.1.3 Ergebnisse

In beinahe allen Lehrveranstaltungen gab es mindestens einmal *Wortmeldungen der Studierenden*, *individuelle Lernphasen* traten dagegen selten auf (in 13.2 % der besuchten Lehrveranstaltungen, vgl. Tabelle 31). Nur in etwa jeder zehnten (10.1 %) Lehrveranstaltung wurden die *Lernziele* genannt.

Tabelle 31. *Didaktische Handlungen (Basisdaten Beobachtungsstudie)*

Item	Prozentsatz des Auftretens		
	B1	B2	M
Führt zu Thema hin (Frage, Problem, Fall)	35.3	37.8	36.6
Stellt Lernziele ausdrücklich dar	9.2	10.9	10.1
Fasst Wichtiges zusammen	28.6	29.4	29.0
Wortmeldungen von Teilnehmern gesamt	96.6	96.6	96.6
0	3.4	3.4	3.4
1–5	23.1	22.7	23.5
6–10	42.9	43.7	42.0
> 10	30.7	30.3	31.1
Methodeneinsatz (Individuelle Lernphasen)	12.7	13.6	13.2

Anmerkungen. N = 119 Lehrveranstaltungen. B = Beobachterin.

7.1.4 Diskussion

Ergebnisse. In beinahe allen Lehrveranstaltungen traten *Wortmeldungen der Studierenden* auf, *individuelle Lernphasen* dagegen kaum. Auch die *Lernziele* wurden nur selten genannt. Damit wurden die zentralen Inhalte der DOS in der Lehre kaum umgesetzt. Auch die anderen beiden Beobachtungskategorien (*Fasst Wichtiges zusammen*; *Führt zu Thema hin*) wurden relativ selten beobachtet. Offen blieb bei dieser Erhebung, wie ausgeprägt und zeitlich umfassend die beobachteten didaktischen Handlungen waren.

Vorgehen und Stichprobe. Da hier eine annähernde Vollerhebung der Seminare und Vorlesungen in zwei klinischen Semestern vorliegt, sind die Daten mit hoher Wahrscheinlichkeit repräsentativ für entsprechende Lehrveranstaltungen im Studium der Humanmedizin. Die Datenerhebung durch zwei Studierende schließt Effekte durch soziale Erwünschtheit weitgehend aus. Wie viele Lehrende in dem beobachteten Zeitraum bereits die DOS absolviert haben, bleibt ebenfalls offen. Da hier zwei klinische Abschnitte untersucht wurden

und jeder klinische Abschnitt jedes Jahr 12 Lehrende für die DOS anmeldet, ist davon auszugehen, dass auch Lehrende beobachtet wurden, welche bereits die DOS absolviert haben. Exakte Zahlen sind hier kaum zu bestimmen, gemessen an dem Anteil aller Lehrender sollte deren Anteil etwa zehn bis zwanzig Prozent betragen.

Material. Die Auswahl der Items leitete sich aus der Literatur (vgl. Kapitel 3.4) und den Inhalten der DOS (vgl. Kapitel 5.1) ab. Die Trainerbefragung (vgl. Kapitel 9) validierte die Auswahl der Items nachträglich. Der zugrundeliegende Beobachtungsbogen war inhaltlich sehr reduziert und die Beobachtungskategorien eingeschränkt. Die gute Beobachterübereinstimmung lässt sich auch durch das wenig differenzierte Antwortformat erklären. Wenn eine didaktische Handlung zweimal auftritt und Beobachterin 1 lediglich das erste Auftreten bemerkt und protokolliert hat und Beobachterin 2 nur das zweite oder ein weiteres, dann ist zwar eine absolute, aber keine exakte Übereinstimmung vorhanden. Mindernde Einflüsse auf die Beobachterübereinstimmung waren Undeutlichkeit (Lehrende sind schwierig zu verstehen), Inkongruenz (die Power Point Folien standen nicht in Einklang mit den Worten der Lehrenden) und Unklarheit (die operationalen Definitionen der Items waren nicht präzise genug).

Fazit. Die zentralen Inhalte der DOS werden in Vorlesungen und Seminaren kaum umgesetzt. Dieser Befund sollte als Vergleichsgrundlage für die folgenden Datenerhebungen dienen. Einmal, indem sie mit den Beobachtungsdaten der DOS-Trainer (vgl. Kapitel 7.1.3) und einmal mit den Beobachtungsdaten der Lehrenden, die sich gegenseitig hospitieren, verglichen werden (vgl. Kapitel 7.4.3). Die Erfahrungen aus dieser Studie haben zudem die Konstruktion des umfassenderen Beobachtungsbogens unterstützt, der im folgenden Kapitel dargestellt wird.

7.2 Entwurf eines Beobachtungsbogens

7.2.1 Konstruktion

Autoren von Reviews zu hochschuldidaktischen Interventionen (Levinson-Rose & Menges, 1981; Stes et al., 2010) fordern den Einsatz objektiver Beobachtungen. Da bei der Planung dieser Arbeit kein validiertes Beobachtungsinventar vorlag, welches für die Evaluation der DOS geeignet erschien (vgl. Kapitel 4.3), wurde ein solches entwickelt. Dazu sollte es in allen Lehrveranstaltungen der Medizinischen Fakultät einsetzbar sein und die Inhalte der DOS (vgl. Kapitel 5.1) abbilden. Entsprechend der Forderung bei Stes et al. (2010) sollte das

Instrument auch in vergleichbaren Studien eingesetzt werden können und so einen Beitrag zur Weiterentwicklung von Messinstrumenten leisten.

Um möglichst hohe Akzeptanz für die Beobachtungen zu schaffen, sollte die Datenerhebung von den Lehrenden als hilfreich und nicht als Kontrolle ihrer Lehrkompetenz betrachtet werden. Lehrende sehen in Lehrbesuchen zwar einen Nutzen, doch bereits die Vorstellung eines Lehrbesuchs ist für viele Lehrende unangenehm (Adshead, White & Stephenson, 2006; Siddiqui, Jones-Dwyer & Carr, 2007). Insbesondere die Annahme bewertet zu werden, kann zu Widerstand führen (Centra, 1979). Bewertungen der Lehre durch Beobachter können in ihrer Validität angezweifelt werden, da diese untereinander beträchtlich variieren (MET-Projekt, 2012; Strong, Gargani & Hacifazlıoğlu, 2011). Es ist anzunehmen, dass Bewertungen die Aufmerksamkeit der Beobachter zusätzlich beanspruchen (Hora & Ferrare, 2013) und von persönlichen Vorannahmen zu Lehre und Lernen abhängig sind (Cash, Hamre, Pianta & Myers, 2012). Beschreibende Beobachtungen sind in der Regel vorteilhafter als Bewertungen für Rückmeldegespräche nach Lehrhospitationen (Wehr, 2008). Die Datenerhebung in Lehrhospitationen ist am besten durch Personalentwicklungsmaßnahmen vertretbar (Hora & Ferrare, 2013). Beratungen nach Lehrhospitationen können positive Effekte, wie bessere Evaluationen durch Studierende, bewirken (Cohen, 1980; Marsh & Roche, 1993; Knol et al. 2013). Daher sollte in dieser Studie ein möglichst beschreibender, wenig wertender Beobachtungsbogen entwickelt werden, welcher informative Feedbackgespräche erleichtert. Um eventuelle Bedenken der Lehrenden gegen die Lehrhospitationen gering zu halten, sollten die Beobachtungen ohne Videoaufzeichnung möglich sein. Der Einsatz von Videoanalysen ist zudem nicht immer praktikabel, da sie das Einverständnis aller teilnehmenden Studierenden voraussetzen. Außerdem sollten die Ergebnisse quantitativ auswertbar sein und die Beobachterübereinstimmung empirisch bestimmt werden können.

Die wesentlichen Konstruktionsschritte beinhalteten die Auswahl der Beobachtungskriterien und des Kodierschemas. Zunächst sollten die Beobachtungskriterien bestimmt werden. Die Auswahl orientierte sich an den Inhalten der DOS (vgl. Kapitel 5.1), den Fragestellungen hochschuldidaktischer Evaluationen (Stes et al, 2010), an der Literatur zu menschlichen Lern- und Gedächtnisprozessen (siehe Anderson, 2007; Goldstein, 2011) sowie an pädagogisch-psychologischen Übersichtsarbeiten zur optimalen Gestaltung von Lehrveranstaltungen (Hattie, 2009; Helmke, 2012; Klauer & Leutner, 2007; Rost, 2010). Insbesondere alle Indikatoren studierendenzentrierter Lehre (vgl. Kapitel 3.4) sollten hier erfasst werden, dazu weitere lehrendenzentrierte Handlungen, denen eine positive Wirkung in der Hochschullehre

zugeschrieben wird. Die ursprüngliche Itemsammlung des so entwickelten Beobachtungsbogens für Lehrveranstaltungen (BEO-LV) umfasste 42 didaktische Handlungen zur Vortragsgestaltung, zur Kommunikation zwischen Lehrenden und Studierenden und zur methodischen Umsetzung von individuellen Lern- und Arbeitsphasen und wurde mehrfach überarbeitet. In einer Handreichung für die Beobachter wurden die Items kontinuierlich präziser definiert und mit Beispielen sowie Ein- und Ausschlusskriterien versehen. Anhand der Operationalisierung sollten die Beobachter entscheiden, ob die Kriterien für ein Item erfüllt sind (konzeptorientiertes Rating; Langer & Schulz von Thun, 2007). In der Konstruktionsphase wurden vier unterschiedliche Varianten eingesetzt, bevor die finale Version erprobt wurde. Die Vorabprüfung erfolgte durch insgesamt sieben Personen, darunter waren drei wissenschaftliche Hilfskräfte aus der Humanmedizin (zwei davon haben bereits in der Vorstudie in Kapitel 7.1 mitgewirkt), eine Diplom-Pädagogin und drei Personen (darunter der Autor), die hochschuldidaktisch an der Medizinischen Fakultät arbeiten. Jeweils zwei bis fünf Personen beobachteten insgesamt 34 Lehrveranstaltungen (zwischen Dezember 2010 und Mai 2011). Die Mehrheit der Lehrveranstaltungen fand an der Medizinischen Fakultät statt, die übrigen in Lehrveranstaltungen für Lehramtstudierende ($n = 3$) und in der Psychologie ($n = 3$). Die Beobachter protokollierten das Auftreten der Kriterien und machten dazu freie Notizen. Die besuchten Lehrveranstaltungen umfassten 30 – 150 Teilnehmer und dauerten zwischen 45 – 90 Minuten. Nach jeder Lehrveranstaltung verglichen die Beobachter ihre Ergebnisse auf Übereinstimmung und diskutierten über die Differenzen. Bei Nichtübereinstimmung wurden die Itembeschreibungen fortlaufend präzisiert oder Items aus dem Itempool entfernt. Tabelle 32 stellt die einzelnen Konstruktionsschritte dar, Version 1 ist exemplarisch im Anhang C1 aufgeführt.

Tabelle 32. *Konstruktion des Beobachtungsbogens*

Version (Zeitraum)	LV	Beob	Übergeordnete Beobachtungskategorien und Veränderungen
1 (12/2010)	3	2-3	Begrüßung/Persönliche Vorstellung, Einleitung, Präsentationsform, Lehrendenzentrierte Lehre, Methodeneinsatz, Kommunikation/Interaktion, Umgang mit Störungen, Ende der Veranstaltung.
2 (12/2010)	8	2	Wie Version 1, mit anderen Items bei Kommunikation/Interaktion
3 (01/2011)	14	2-5	Aufbau der Lehrveranstaltung, Methodeneinsatz, Präsentationsform, Kommunikation/Interaktion (erneut überarbeitete Items), Lehrendenzentrierte Lehre mit zwei Unterkategorien (Erleichtert Verständnis/Strukturiert Wissen; Zeigt Bedeutung der Inhalte auf) und Unterteilung des Auftretens in: Einleitung, Hauptteil, Schluss.
4 (04 – 05/2011)	7	2	Methodeneinsatz, Kommunikation/Interaktion (erneut überarbeitet), Lehrendenzentrierte Lehre. Zusammenfassende Einschätzung von Lehrveranstaltungsmerkmalen. Präsentationsform inkl. Bewertung von Blickkontakt und Rhetorik entfernt, Items zu lehrendenzentrierter Lehre anhand von Redundanzen und inhaltliche Überschneidungen gekürzt. Unterteilung des Auftretens: Einleitung, Hauptteil, Schluss entfernt.
5 BEO-LV (06/2011)	2	2	Zusammenfassende Einschätzung von Lehrveranstaltungsmerkmalen entfernt. Zwei Items aus Version 4 zusammengefasst, die ändern beibehalten. Aufteilung der Beobachtungen in 15-Minuten-Intervalle.

Anmerkungen. LV = Anzahl der besuchten Lehrveranstaltungen. Beob = Minimale und maximale Anzahl gleichzeitiger Beobachter in den beobachteten Lehrveranstaltungen.

Die Auswahl des Kodierschemas entstand in Abhängigkeit der untersuchten Konstrukte. Zunächst mussten die Beobachtungsintervalle für die Evaluation festgelegt werden. Bei den ersten Versionen des Beobachtungsbogens wurde das Auftreten der Handlungen in verschiedenen Phasen der Lehrveranstaltung erhoben (Einleitung, Hauptteil, Schluss). Dies erwies

sich als ungeeigneter Ansatz, da in vielen hospitierten Lehrveranstaltungen weder die Einleitung noch der Schlussteil der Lehrveranstaltungen als eigene Phasen erkennbar waren. Zudem hätte dieses Vorgehen die Frequenz der Handlungen wenig differenziert abgebildet. Daher sollten hier Zeitintervalle für die Beobachtung didaktischer Handlungen bestimmt werden. Die Länge der Intervalle wurde im Rahmen dieser Studie aus pragmatischen Gründen auf 15 Minuten festgelegt. Damit ließ sich die Beobachterübereinstimmung in jedem Intervall überprüfen (Rost, 2007) und es konnte eine ausreichende Differenzierung vorgenommen werden (bei 90-minütigen Lehrveranstaltungen bestehen sieben Abstufungen).

Studierendenzentrierte Lehre, also die Umsetzung individueller Lern- und Arbeitsphasen (Einzel, Partner- oder Gruppenarbeit) sowie Interaktion/Kommunikation, sollte möglichst präzise erfasst werden, da sowohl deren zeitliche Ausdehnung, als auch deren Positionierung in der Gesamtstruktur der Lehrveranstaltung für die Evaluation der DOS von Bedeutung ist (vgl. Kapitel 5.1). Eine likertskalierte Bewertung der Frequenz des Auftretens, wie sie in vielen anderen Beobachtungsinstrumenten vorgenommen wird (z. B. Murray, 1983; Wainwright et al., 2003) erschien für diesen Zweck unpassend. Die Fragen der Lehrenden, die Wortmeldungen der Studierenden sowie die Dauer individueller Lern- und Arbeitsphasen sollten gezählt werden und damit als kontinuierliche Daten vorliegen. Beobachtungen in der Konstruktionsphase zeigten, dass sich die Wechsel von studierendenzentrierter und lehrendenzentrierten Handlungen gut voneinander abgrenzen lassen. Das konnte auch in einer Studie von Johannes und Seidel (2012) durch gute Beurteilerübereinstimmung nachgewiesen werden. Fragen der Lehrenden ließen sich leicht zählen und individuelle Lern- und Arbeitsphasen traten als Methodenwechsel deutlich hervor.

Für lehrendenzentrierte Handlungen erschien eine exakte zeitliche Aufschlüsselung nicht notwendig, weder für die quantitative Analyse noch für die Rückmeldung nach Lehrhospitation. Daher sollte hier eine möglichst ökonomische Erhebungsweise Anwendung finden. Wenn ein Dozent beispielsweise innerhalb von 10 Minuten zwei konkrete Fallbeispiele abwechselnd aufgreift (Kategorie: Problem/Fall/Herausforderung) dann lässt sich diese didaktische Handlung auf drei verschiedene Arten quantifizieren. Einmal ließe sich der Zeitraum der Kategorie benennen (hier 10 Minuten), das Auftreten der Kategorie (hier zweimal) oder das Auftreten der Kategorie in dem Intervall (hier einmal). Die ersten beiden Alternativen sind mit erheblichem kognitivem Mehraufwand für die Beobachter verbunden, insbesondere wenn mehrere lehrendenzentrierte Handlungen intermittierend und parallel auftre-

ten. Daher wurde festgelegt, dass lediglich das Auftreten der Kategorie in dem Intervall protokolliert werden sollte.

Nach der Konstruktionsphase setzte sich der BEO-LV aus dreizehn Items zusammen, das Item *Fragen der Lehrenden* wurde noch einmal in fünf Unterkategorien unterteilt (Tabelle 33). Die vollständige Beschreibung aller Items, deren theoretische und empirische Fundierung, sowie die ein- und Ausschlusskriterien befinden sich in Anhang C2. Für die psychometrische Überprüfung des BEO-LV wurden zwei Vorstudien durchgeführt, die im Folgenden beschrieben werden.

Tabelle 33. *Items des BEO-LV*

Lehrendenzentrierte Lehre	Kurzbeschreibung
Problem/Fall/Herausforderung	Lehrende stellen zu lösende Fragestellungen vor (mit dem Ziel zu veranschaulichen bzw. eine Lösung/Antwort/Stellungnahme zu finden).
Visualisierung	Lehrende präsentieren Bilder, Filme oder Anschauungsmaterial.
Struktur/Systematisierung	Lehrende strukturieren Zusammenhänge oder Vergleiche mit Mind-Maps, Advance-Organizer, vergleichenden Tabellen, Diagrammen o. ä.
Lernziele	Lehrende stellen Lernziele der Lehrveranstaltung ausformuliert dar.
Persönlicher Bezug	Lehrende stellen persönliche Bezüge zu dem Thema der Lehrveranstaltung dar.
Handlungsrelevanz	Lehrende zeigen berufliche oder private Relevanz der Inhalte für die Studierenden auf.
Wichtige Inhalte	Lehrende weisen auf Relevanz der Inhalte für Prüfungen oder Abschlussarbeiten hin.
Interdisziplinarität	Lehrende stellen Bezüge her, die über das Thema der Veranstaltung hinausgehen.
Zusammenfassung	Lehrende wiederholen Inhalte.
Studierendenzentrierte Lehre	
Individuelle Lern- und Arbeitsphasen	Lehrende geben Arbeitsaufträge an alle Studierenden (Einzel-, Partner- oder Gruppenarbeit).
Fragen der Lehrenden	Lehrende richten Fragen an die Studierenden. Verschiedene Kategorien werden unterschieden: (a) Wissen (b) Vorwissen/Vorkenntnissen, (c) Interesse an den Inhalten, (d) Lernprozess/Verständnis, (e) Meinungen/Haltungen.
Wortmeldungen von Studierenden	Studierenden äußern sich verbal.
Studierende präsentieren	Referate, Rollenspiele, praktische Demonstrationen.

7.2.2 Vorstudie 1: Beurteilerübereinstimmung und Interraterreliabilität

In der Vorstudie 1 sollte die Beurteilerübereinstimmung und Interraterreliabilität von zwei Beobachtern überprüft werden. Es sollte die Annahme geprüft werden, dass sich die Beobachtungskriterien mit guter Übereinstimmung protokollieren lassen. Wie bei Wirtz und Caspar (2002) beschrieben, ist eine hinreichende Reliabilität bei differenzierteren Beobachtungen schwieriger zu erzielen. Es sollte daher auch geprüft werden, ob die Unterkategorien der *Fragen der Lehrenden* hinreichende Reliabilität aufweisen.

Stichprobe und Vorgehen. Zwei Beobachter (der Autor dieser Arbeit und eine Dipl. Pädagogin) setzten den BEO-LV unabhängig voneinander in denselben 20 Vorlesungen und Seminaren der Studiengänge Humanmedizin ($n = 10$), Lehramt Deutsch ($n = 5$), Psychologie ($n = 3$), Biologie ($n = 1$) und Jura ($n = 1$) ein. Die Auswahl der Lehrveranstaltungen ergab sich nach persönlicher Absprache mit den Lehrenden. Die Anzahl der Studierenden pro Lehrveranstaltung variierte zwischen 16 und 300 ($M = 76$, $SD = 74$), die Dauer der Lehrveranstaltungen zwischen 45 und 90 Minuten ($M = 79$, $SD = 18$). Der Beobachtungsumfang über alle besuchten Lehrveranstaltungen beinhaltete insgesamt 107 Beobachtungsintervalle von je 15 Minuten.

Statistische Analysen. Zur Beurteilung der Beobachterübereinstimmung bei dichotomen Variablen wurden, entsprechend der Empfehlung von Rost (2007) sowie Wirtz und Caspar (2002), folgende Koeffizienten berechnet: Die prozentuale Übereinstimmung ($P\ddot{U}$), die prozentuale Auftretensbeobachterübereinstimmung ($P\ddot{U}+$), die prozentuale Nichtauftretensbeobachterübereinstimmung ($P\ddot{U}-$) und Cohens Kappa (κ). Die $P\ddot{U}$ gibt an, wie häufig die Beobachtungen übereinstimmen, gute Werte liegen bei etwa 80 % vor (Wirtz & Caspar, 2002). Insbesondere bei seltenen oder sehr häufig auftretenden Merkmalen sind die Ergebnisse der $P\ddot{U}+$ und der $P\ddot{U}-$ aufschlussreich. Die Ergebnisse erläutern, in wie viel Prozent der Fälle beide Beobachter eine didaktische Handlung übereinstimmend wahrnehmen ($P\ddot{U}+$), oder übereinstimmend nicht wahrnehmen ($P\ddot{U}-$). Mit dem Kappa-Koeffizient nach Cohen (1960) wurde noch ein zufallskorrigiertes Übereinstimmungsmaß herangezogen. Hier gelten Beurteilerübereinstimmungen ab .40 als akzeptabel, und ab .60 als gut (Fleiss & Cohen, 1973; Landis & Koch, 1977, Wirtz & Caspar, 2002). Bei Beobachtungen von Lehrhandlungen konnten höhere Werte erzielt werden (Hora, 2013; Johannes & Seidel, 2012).

Die Beobachterübereinstimmung für die kontinuierlichen Variablen wurden mit der unjustierten Intraklassenkorrelation (ICC_{unjust}) analysiert. Hier gelten Werte ab .70 als akzeptabel (Wirtz & Caspar, 2002). Im zweifaktoriellen Modell der ICC_{unjust} werden

die Differenzen der Beurteiler und unterschiedliche Mittelwerte bei der Beobachtung abgebildet. Mittelwertunterschiede wirken sich mindernd auf die ICC_{unjust} aus. Als Voraussetzungsprüfung wurden die Varianzhomogenität mittels des Levene-Tests (Alpha = .25) und die Interaktion zwischen Beobachter und Beobachtungsintervall mit dem Tukey Additivitätstest (Alpha = .25) ermittelt.

Ergebnisse. Tabelle 34 zeigt Gemeinsamkeiten und Differenzen zwischen den beiden Beobachtern. Die $P\ddot{U}$ überstieg bei allen dichotomen Items den empfohlenen Mindestwert von 80 % (Rost, 2007). Die $P\ddot{U}$ - war höher als die $P\ddot{U}$ +, d. h. die Beobachter stimmten bei Nichtauftreten einer Kategorie häufiger überein, als bei dem Auftreten einer didaktischen Handlung. Cohens Kappa variierte zwischen .35 und .94 und betrug im Mittel $\kappa = .66$. Die Beobachter erzielten bei dem Item *Individuelle Lern- und Arbeitsphasen* die höchste Übereinstimmung ($\kappa = .94$), bei dem Item *Interdisziplinarität* die niedrigste ($\kappa = .35$). Für das Item *Studierende präsentieren* konnten keine Werte berichtet werden, da keine studentischen Vorträge im Rahmen der Beobachtungen stattfanden.

Die Ergebnisse der übergeordneten kontinuierlichen Items lagen über den anvisierten Zielwerten. *Fragen gesamt* ($ICC_{\text{unjust}} = .85$), die *Wortmeldungen von Studierenden* ($ICC_{\text{unjust}} = .90$) sowie die *Dauer individueller Lern- und Arbeitsphasen* ($ICC_{\text{unjust}} = .96$) ließen sich mit einer sehr guten Interraterreliabilität erheben. Die Werte der fünf Items zu den Fragen der Lehrenden waren dagegen geringer. Hier zeigte sich eine Interraterreliabilität zwischen .38 und .92. Die Items *Fragen zu Meinungen/Haltungen* sowie zu *Vorwissen/Vorkenntnissen* erzielten Koeffizienten $< .60$.

Tabelle 34. Statistische Kennwerte, Beurteilerübereinstimmung und Interraterreliabilität in 107 Beobachtungsintervallen.

Items	Prozentsatz des Auftretens		κ	Prozentuale Übereinstimmung ($P\ddot{U}$)		
	Beobachter 1	Beobachter 2		$P\ddot{U}_+$	$P\ddot{U}_-$	$P\ddot{U}$
Problem/Fall/Herausforderung	22	25	.84	78	93	94
Visualisierung	48	47	.83	83	86	92
Struktur/Systematisierung	37	38	.69	67	81	86
Lernziele	3	4	.85	75	99	99
Persönlicher Bezug	50	53	.76	78	79	88
Handlungsrelevanz	56	61	.66	72	67	82
Wichtige Inhalte ^a	33	43	.64	65	73	82
Interdisziplinarität	9	15	.35	26	86	87
Zusammenfassung	12	17	.64	53	91	92
Individuelle Lern- und Arbeitsphasen	9	8	.94	89	99	99
	$M (SD)$	$M (SD)$	ICC_{unjust}			
Dauer Individuelle Lern- und Arbeitsphasen	0.29 (1.30)	0.35 (1.65)	.96 ^{bd}			
Fragen insgesamt	2.00 (2.35)	1.75 (1.96)	.85 ^{bd}			
Wissensfragen	1.28 (1.84)	1.17 (1.69)	.87 ^b			
Vorwissen/Vorkenntnissen	0.16 (0.42)	0.16 (0.52)	.41 ^d			
Interesse	0.05 (0.25)	0.04 (0.24)	.92			
Lernprozess/Verständnis	0.28 (0.71)	0.29 (0.63)	.63 ^d			
Meinungen/Haltungen	0.23 (0.59)	0.09 (0.32)	.38 ^{bcd}			
Wortmeldungen von Studierenden	4.09 (3.63)	3.98 (3.60)	.90			

Anmerkungen. κ = Cohens Kappa. ICC_{unjust} = unjustierte Intraklassenkorrelation. $P\ddot{U}_+ = P\ddot{U}$ bei Auftreten eines Items; $P\ddot{U}_- = P\ddot{U}$ bei Nichtauftreten eines Items.

^aHier liegen nur 89 Beobachtungseinheiten vor, bei denen die Voraussetzung (Prüfungsrelevanz) der Items erfüllt war. ^bZweifaktorielle Varianzanalyse auf Mittelwertsunterschiede signifikant ($p < .25$). ^cLevene-Test signifikant ($p < .25$). ^dTukeys Additivitätstest signifikant ($p < .25$).

Diskussion. Die zentralen Items, die für die Evaluation der DOS von Bedeutung sind, ließen sich in dieser Vorstudie mit guter Übereinstimmung und Interraterreliabilität protokollieren. Insgesamt brachte der erste Einsatz des BEO-LV zufriedenstellende Ergebnisse.

Anhand der Koeffizienten und den Erfahrungen bei der Erhebung ließen sich Fehlerquellen identifizieren. Die signifikanten Ergebnisse der Voraussetzungstestungen zeigten, dass Mittelwertunterschiede, Varianzunterschiede und Abweichungen der Beobachter bei einzelnen Beobachtungsintervallen einen mindernden Einfluss auf die ICC_{unjust} hatten. Vor allem bei den einzelnen Fragen der Lehrenden zeigten sich hier niedrigere Reliabilitäten als bei der Verwendung der übergeordneten Kategorie. Nicht übereinstimmende Kategorisierungen führten dabei zu einer doppelten Minderung der Übereinstimmung bzw. der Interraterreliabilität. Wählt ein Beobachter Kategorie A anstelle von Kategorie B, und der andere Beobachter entscheidet sich gegensätzlich, so geht der Fehler zu Lasten der Übereinstimmungskoeffizienten beider Items. Daher soll hier, entsprechend der Empfehlung von Wirtz und Caspar (2002), bei der weiteren Verwendung des BEO-LV nur die übergeordnete Skala *Fragen gesamt* als abhängige Variable verwendet werden.

Bei den diskreten Variablen ergaben sich ebenfalls heterogene Ergebnisse. Die etwas niedrigere $P\ddot{U}^+$ zeigte, dass didaktische Handlungen übersehen wurden. Das könnte auf die Beobachter, unklare Darbietungen der didaktischen Handlungen seitens der Lehrenden oder auf unklare operationale Definitionen der Items in der Handreichung zurückgeführt werden. Damit verbunden konnten auch bei diesen Items nicht übereinstimmende Kategorisierungen zu einer doppelten Minderung der Übereinstimmung bzw. der Interraterreliabilität geführt haben. Der Einsatz des BEO-LV erfordert daher eine sehr konzentrierte Anwendung und setzt eine gute Einarbeitung in die Operationalisierungen der Konstrukte voraus.

Die Übereinstimmungen in dieser Studie sind insgesamt etwas geringer als in vergleichbaren Studien (Hora, 2013; Johannes & Seidel, 2012). Nach Murray (1997) lassen sich Variablen leichter übereinstimmend beobachten, wenn sie möglichst eindeutig sind und die Beobachter nur wenige Schlussfolgerungen treffen müssen. Das kann sich sowohl durch klar definierte übergeordnete Kategorien erzielen lassen (wie bei Johannes & Seidel, 2012) mit insgesamt sechs Kategorien, oder durch starke Aufschlüsselungen mit 47 (Hora, 2013) bis 60 Kategorien (Murray, 1983). Bei den lehrendenzentrierten Merkmalen war hier vor allem die Erfassung des Items *Interdisziplinarität*

für die Beobachter schwierig. Die zentrale operationale Definition in dem Beobachtungsmanual (vgl. Anhang C2, S. 197) lautet hierzu: „Wissenschaftliche Bezüge, die über das eigentliche Thema der Veranstaltung hinausgehen, werden (kurz) erklärt“. Die Entscheidung, wann genau die Grenzen eines Themas überschritten werden, erfordert fachspezifisches Wissen und hängt, bei näherliegenden Themen, eher von der Definition der Lehrenden ab, als von objektivierbaren Kriterien. Da die Beobachtungskategorie nicht reliabel erfasst werden konnte, stellte sich die Frage, ob die Kategorie ausgeschlossen werden muss.

Eine Limitation dieser Studie liegt in der geringen Beobachteranzahl von nur zwei Personen, zudem beide Beobachter in die Entwicklung des Instrumentes involviert waren. Es kann nur die Aussage getroffen werden, dass eine reliable und übereinstimmende Beobachtung möglich ist. Die Gütekriterien eines Beobachtungsinstrumentes müssen bei jeder Beobachtungsstudie neu ermittelt werden (Wirtz & Caspar, 2002). Die Auswahl besuchter Lehrveranstaltungen stellte nur eine Gelegenheitsstichprobe dar, daher bleibt offen, wie repräsentativ die erfassten didaktischen Handlungen waren. Es zeigte sich weiterhin, dass die zentralen Variablen bei der Untersuchung der DOS, *Lernziele* sowie *Individuelle Lern- und Arbeitsphasen*, hier kaum aufgetreten sind. Dies zeichnete sich bereits in der Konstruktionsphase (7.2.1) und bei der Erhebung der Basisdaten (7.1.3) ab. Auch die deutlich unterschiedlichen Auftretenswahrscheinlichkeiten der didaktischen Handlungen insgesamt (zwischen 3% – 61%) führen zu einer geringen Vergleichbarkeit der Koeffizienten, da Übereinstimmungswerte bei seltenen Beobachtungskategorien ungenauer sind (Wirtz & Caspar, 2002). Um die Tauglichkeit der BEO-LV zu bestimmen sollte daher eine größere Zahl von Beobachtern eingesetzt werden und dabei die Zahl der didaktischen Handlungen experimentell kontrolliert werden. Dies wird in der folgenden Vorstudie 2 dargestellt. Zusätzlich sollte eine weitere Limitation von Beobachtungsstudien überwunden werden. Da in Vorstudie 1 der wahre Wert der didaktischen Handlungen nicht bekannt war, drücken die Beobachterübereinstimmung und die Interraterreliabilität nicht aus, wie oft *richtige* Klassifikationen vorgenommen wurden. Auch dies wurde in der folgenden Vorstudie 2 geprüft.

7.2.3 Vorstudie 2: Beurteilerübereinstimmung und interne Validität

Ziel dieser Vorstudie war es, die Beurteilerübereinstimmung erneut mit einer größeren Beobachteranzahl zu prüfen. Als Indikator für die interne Validität sollte zusätzlich die Konzepttreue, also die Anzahl richtiger Kategorisierungen, untersucht werden (Langer & Schulz von Thun, 2007). Dabei wurde weiterhin der Einfluss didaktischer Expertise auf die Nutzung des BEO-LV analysiert. Wie Vorstudie 1 zeigte, setzt der Einsatz des BEO-LV ein gutes Verständnis der Kategorisierungen voraus. Erschwerend wirkt, dass die simultane Beobachtung verschiedener Merkmale Beobachter schnell überfordern kann (Wirtz & Caspar, 2002). Angenommen wurde daher, dass Experten im Gegensatz zu Novizen konzepttreuer protokollieren und dabei eine höhere Beurteilerübereinstimmung erzielen.

Stichprobe. Als Beobachter wurden sechs Personen rekrutiert, die mindestens ein Jahr in der hochschuldidaktischen Beratung tätig waren und nachfolgend als *Experten* bezeichnet werden. Es wurde erwartet, dass diese bereits über ein gutes Verständnis verschiedener didaktischer Handlungen verfügen. Als *Novizen* wurden sieben Studierende rekrutiert, vier Lehramtstudierende und drei in der Hochschuldidaktik arbeitende studentische Hilfskräfte (alle zwischen dem 4. – 9. Fachsemester).

Vorgehen. Vor Beginn der Testung sahen alle Teilnehmer ein Schulungsvideo (14 Minuten), in dem die einzelnen Kategorien vorgestellt wurden und jeweils anhand einer Situation geübt werden konnten. Die richtige Antwort wurde danach mit einer kurzen Begründung präsentiert. Zusätzlich wurden die Teilnehmenden aufgefordert, sich mit der Handreichung vertraut zu machen.

Um die Beobachtungsgenauigkeit zu prüfen, wurden in einer Onlinestudie danach zehn Videos präsentiert, die nachgestellte Lehrsituationen aus der Psychologie und Pädagogik wiedergaben (siehe Anhang D). Die einzelnen Videos dauerten durchschnittlich etwa neun Minuten. In diesen Videos führte jeweils ein Lehrender drei bis fünf systematisch variierte didaktische Handlungen durch (Tabelle 35). Die dargestellten Szenen wurden so entworfen, dass sie unterschiedliche Facetten der Kategorien abbilden, nicht den Beispielen des Schulungsvideos entsprachen und etwa einen mittleren Schwierigkeitsgrad aufwiesen. So konnte sowohl der Schwierigkeitsgrad als auch die Anzahl der dargebotenen didaktischen Handlungen kontrolliert werden. Die Rollen der Lehrenden nahmen diesmal die beiden Beobachter aus Vorstudie 1 ein. Da die Kategorien *Fragen der Lehrenden gesamt*, *Wortmeldungen der Studierenden* und die *Dauer individueller Lern- und Arbeitsphasen* in Vorstudie 1 sehr gute Werte erzielten, fiel die

Entscheidung, auf die Einbindung von Studierenden in den Videos zu verzichten. Daher wurde auch hier das Item *Studierende präsentieren* nicht berücksichtigt. Es lagen also jeweils zehn Items zur Bearbeitung vor (Tabelle 35).

Während die Teilnehmer die Videos ansahen, sollten sie alle didaktischen Handlungen online protokollieren, die sie beobachtet hatten. Jedes Video stellte in der Untersuchung ein eigenes Beobachtungsintervall dar. Den Beobachtern waren die Auftretenswahrscheinlichkeiten der didaktischen Handlungen (drei bis fünf je Video) vorab nicht bekannt. Alle Beobachter bearbeiteten die Videos vollständig. Es lagen so für jede Kategorie dichotome Werte vor (Handlung tritt auf vs. Handlung tritt nicht auf). So konnten die Teilnehmer bis zu 100 richtige Antworten erzielen.

Statistische Auswertung. Wie in Vorstudie 1 wurde Cohens Kappa (κ) als Übereinstimmungskoeffizient berechnet, (Fleiss & Cohen, 1973; Landis & Koch, 1977, Wirtz & Caspar, 2002). Zum einen wurde die Übereinstimmung eines Beurteilers mit allen anderen Beobachtern bestimmt (κ Gesamt). Zum anderen wurde die Übereinstimmung innerhalb der Expertengruppe (κ Experten) und innerhalb der Novizengruppe (κ Novizen) berechnet. Außerdem wurde die Schwierigkeit der Items ermittelt. Hier wurde erwartet, dass Experten mindestens 80 % der Items richtig lösen sollten. Zur Prüfung von Unterschieden zwischen der Experten- und Novizengruppe wurde der Mann-Whitney-Test verwendet (einseitige Testung, $\alpha = .05$).

Ergebnisse. Die Ergebnisse sind in Tabelle 36 dargestellt. Die Beobachter nahmen insgesamt zwischen 62 – 89 richtige Kategorisierungen vor ($M = 80.31$, $SD = 6.96$). Beide Gruppen übertrafen die Wahrscheinlichkeit der zufälligen Korrektheit einer Antwort von 50 %. Experten erzielten mehr richtige Kategorisierungen ($M = 84.67$; $SD = 2.50$) als die Novizen ($M = 76.57$; $SD = 7.50$). Dieser Unterschied war statistisch signifikant ($U = 4.50$, $p = .007$). Die Differenz entsprach einer hohen Effektstärke von $d = 1.40$ (Cohens d , standardisiert an der gepoolten Standardabweichung) zugunsten der Experten. Items mit höherer Schwierigkeit zeigten niedrigere Übereinstimmungskoeffizienten als Items mit niedrigerer Schwierigkeit. Die Werte von Cohens Kappa variierten zwischen .11 und 1.00 (κ Experten) sowie .13 und .75 (κ Novizen). Die Experten erreichten nur in sieben von zehn Kategorien sowohl 80 % richtige Antworten als auch ein $\kappa = .60$, die Novizen in nur einer von zehn. Die Kategorie *Individuelle Lern- und Arbeitsphasen* zeigte insgesamt die höchste Übereinstimmung, die Kategorien *Struktur/Systematisierung* und *Interdisziplinarität* die geringste.

Tabelle 35. *Didaktische Handlungen in den Übungsvideos*

	Video Nr.										Summe
	1	2	3	4	5	6	7	8	9	10	
Problem/Fall/Herausforderung	x	x						x	x		4
Visualisierung			x		x	x			x		4
Struktur/Systematisierung					x		x	x	x		4
Lernziele			x	x			x			x	4
Persönlicher Bezug	x				x		x		x	x	5
Handlungsrelevanz	x	x	x	x							4
Wichtige Inhalte	x	x			x		x			x	5
Interdisziplinarität	x		x	x							3
Zusammenfassung	x						x	x		x	4
Individuelle Lern- und Arbeitsphasen		x		x		x			x		4
Summe	6	4	4	4	4	2	5	3	5	4	41

Anmerkungen. x = didaktische Handlung tritt auf.

Tabelle 36. Beurteilerübereinstimmung und Anteil richtiger Antworten.

Didaktische Handlungen	Kappa			<i>p</i>		
	Experten	Novizen	Gesamt	<i>Experten</i>	Novizen	Gesamt
Problem/Fall/Herausforderung	.56	.36	.44	.85	.74	.79
Visualisierung	.62	.35	.48	.83	.76	.79
Struktur/Systematisierung	.11	.14	.11	.72	.64	.68
Lernziele	.79	.56	.73	.93	.91	.92
Persönlicher Bezug	.75	.36	.54	.93	.79	.85
Handlungsrelevanz	.69	.22	.40	.80	.76	.78
Wichtige Inhalte	.60	.24	.40	.83	.71	.77
Interdisziplinarität	.17	.13	.17	.75	.63	.68
Zusammenfassung	.62	.31	.41	.82	.76	.78
Individuelle Lern- und Arbeitsphasen	1	.75	.86	1	.94	.97
Mittelwert	.59	.35	.45	.85	.76	.80

Anmerkungen. Die Kappa-Werte sind durchschnittliche Werte. Die p-Werte geben die Schwierigkeit je Item an, definiert als Prozentsatz richtiger Antworten von 100.

Diskussion. Die hohen Werte der Kategorie *Individuelle Lern- und Arbeitsphasen* standen in Einklang mit Vorstudie 1 und vergleichbaren Analysen (Johannes und Seidel, 2012), können aber auch durch die Darbietung in den Videos erklärt werden. Da hier keine Studierenden in den Videos aufgetreten sind, wurde nur die Instruktion präsentiert und somit ist die Varianz der Darbietungen gegenüber der Realität eingeschränkt. Die niedrigeren Werte bei *Struktur/Systematisierung* sowie *Interdisziplinarität* entsprechen den Ergebnissen aus Vorstudie 1. Beide Kategorien sollten daher in der nun folgenden Hauptstudie dieser Arbeit mit besonderer Vorsicht interpretiert werden, da deren operationalen Definitionen vermutlich nicht genau genug sind. Alle weiteren Kategorien zeigten zumindest akzeptable Werte, so dass hier der Einsatz des BEO-LV in dieser Evaluationsstudie sinnvoll erschien.

Die Ergebnisse zeigen insgesamt, dass erfahrene Hochschuldidaktiker den BEO-LV erfolgreicher anwendeten als Novizen. Offen bleibt, inwiefern eine intensivere Schulung diese Differenz vermindern kann. Als einschränkende Faktoren der Repräsentativität müssen die kleine Stichprobe von Beobachtern und die Auswahl der Themen in den Videos angeführt werden (ausschließlich pädagogisch-psychologische Themen). Für den Einsatz von Videos sprach grundsätzlich, dass so der Schwierigkeitsgrad und die Anzahl didaktischer Handlungen experimentell kontrolliert werden konnte.

7.3 Beobachtung von Lehrverhalten durch Hochschuldidaktiker

Im folgenden Kapitel wird dargestellt, welchen Einfluss die DOS auf das Lehrverhalten der Lehrenden hat. Entsprechend den Zielen der DOS (vgl. Kapitel 5) wurde untersucht, wie häufig Teilnehmer der DOS studierendenzentrierte Lehre umsetzen. Zudem wurde geprüft, ob sie die Lehre anhand der Lernziele ausrichten. Wie in der Befragung der Lehrenden (Kapitel 6) wurden hier zwei Versuchsgruppen unterschieden. Lehrende, die sich zum Untersuchungszeitraum in der Praxisphase der DOS befanden (vgl. Kapitel 5.1) werden im Folgenden als VSG1 bezeichnet. Lehrende, deren Besuch der DOS schon länger zurücklag, werden als VSG2 bezeichnet. Die Untersuchung der VSG2 sollte Aufschluss über die Langzeitwirkungen der DOS geben, da diese bei didaktischen Weiterbildungen häufig offen bleiben (Stes et al., 2009). Vergleichend wurden Lehrende untersucht, die nicht an der DOS teilgenommen haben (Vergleichsgruppe, VGG)

7.3.1 Hypothesen

Die Hypothesen beziehen sich alle auf das unmittelbare Lehrendenverhalten und lauten folgendermaßen:

- H1: Die VSG1 und die VSG2 setzen häufiger individuelle Lern- und Arbeitsphasen um als die VGG.
- H2: Die VSG1 und die VSG2 setzen zeitlich länger individuelle Lern- und Arbeitsphasen um als die VGG.
- H3: Die VSG1 und die VSG2 ermöglichen mehr Wortmeldungen der Studierenden als die VGG.
- H4: Die VSG1 und die VSG2 stellen den Studierenden mehr Fragen als die VGG.
- H5: Die VSG1 und die VSG2 nennen häufiger die Lernziele ihrer Lehrveranstaltung als die VGG.

Weitere didaktische Handlungen, die lehrendenzentrierter Lehre zuzuordnen sind, wurden begleitend erhoben. Deren Untersuchung sollte ausschließen, dass systematische Unterschiede zwischen den Gruppen auftreten. Da studierendenzentrierte Lehre eine Ergänzung und keine Alternative zu lehrendenzentrierter Lehre darstellt (vgl. Kapitel 3.4 und 4.3), sollte die Einführung studierendenzentrierter Lehre nicht zu einer Reduzierung bewährter lehrendenzentrierter Handlungen führen.

7.3.2 Methodik

Beobachtungsinstrument. In dieser Beobachtungsstudie kam der BEO-LV zum Einsatz (Anhang C3). Die Vorstudie 1 zeigte, dass sich mit dem BEO-LV zufriedenstellende Beurteilerübereinstimmung und Interraterreliabilität in einer Felduntersuchung erzielen lassen. In Vorstudie 2 wurde nachgewiesen, dass didaktische Handlungen mit einer zufriedenstellenden Konzepttreue erfasst werden konnten. Voraussetzung dafür war eine gute Kenntnis der operationalen Definitionen. Die didaktischen Handlungen wurden unmittelbar in den Lehrveranstaltungen protokolliert. Die umfangreiche Beschreibung der didaktischen Handlungen, deren lerntheoretische Grundlagen, Wirksamkeitsbelege, sowie Ein- und Ausschlusskriterien sind im Anhang C2, und die Kurzversion für den Einsatz in Lehrveranstaltungen in Anhang C4, aufgeführt.

Die Studierendenzentrierung der Lehrenden wurde anhand der *Fragen der Lehrenden*, der *Wortmeldungen der Studierenden* und dem *Auftreten individueller Lern- und Arbeitsphasen* erhoben. Dabei sollte nicht nur die Dauer der individuellen Lern- und Arbeitsphasen erfasst werden, sondern auch, ob die Phasen intermittierend in verschie-

denen Intervallen der Lehrveranstaltungen auftreten, wie es bei Wahl (2006) empfohlen wird.

Die Fragen und die Wortmeldungen wurden jeweils als Striche (I) protokolliert. Diese Variablen sind daher als verhältnisskaliert zu betrachten. Die Summe des Auftretens wurde für die gesamte Lehrveranstaltung berechnet und anschließend durch die Anzahl der Intervalle geteilt. Wenn Lehrende beispielsweise 20 Fragen in einer 90-minütigen Lehrveranstaltung (6 Intervalle a 15 Minuten) gestellt haben, ergab sich ein Mittelwert von 3.33 für die entsprechende Lehrveranstaltung. So ließen sich Lehrveranstaltungen ungleicher Veranstaltungslänge miteinander vergleichen. Die Dauer der individuellen Lern- und Arbeitsphasen wurde durch die Dauer der Lehrveranstaltungen (in Minuten) dividiert, so dass der berechnete Wert den relativen zeitlichen Anteil angibt. Wenn Lehrende beispielsweise 15 Minuten individuelle Lern- und Arbeitsphasen in einer 90-minütigen Lehrveranstaltung ansetzen, ergibt sich ein Wert von .17.

Das Auftreten individueller Lern- und Arbeitsphasen wurde mit einem Kreuz (X) in dem entsprechenden Zeitintervall protokolliert, maximal war ein Kreuz (X) möglich. Die Summe des Auftretens wurde für die gesamte Lehrveranstaltung gezählt und anschließend durch die Anzahl der Intervalle geteilt. Der maximale Wert konnte dementsprechend 1 betragen. Wenn Lehrende beispielsweise in einer 90-minütigen Lehrveranstaltung (6 Intervalle a 15 Minuten) drei individuelle Lern- und Arbeitsphasen eingesetzt haben, ergab sich ein Mittelwert von .50 für die entsprechende Lehrveranstaltung. Die so für eine Lehrveranstaltung berechnete Variable ist dementsprechend ordinalskaliert. Die lehrendenzentrierten didaktischen Handlungen wurden ebenfalls mit maximal einem Kreuz (X) in dem entsprechenden Zeitintervall protokolliert.

Vorgehen. Teststärkenanalysen mit GPOWER (siehe Faul et al., 2007) zeigten, dass für den Nachweis von hohen Effektstärken 21 Personen, für den Nachweis von mittleren Effekten 51 Personen und für den Nachweis von kleinen Effekten 310 Personen je Gruppe notwendig sind (bei einer Teststärke von .80). Da in der Literatur bei vergleichbaren Studien maximal moderate Effekte gezeigt wurden, mussten viele Lehrende für die Untersuchung gewonnen werden. Um hohe Akzeptanz für die Beobachtungen zu schaffen, wurde der Schwerpunkt dieser Untersuchung auf Beobachtungen gelegt, auf den verpflichtenden Einsatz von Videoanalysen verzichtet und den Teilnehmern eine strukturierte Rückmeldung zu ihrer Lehre angeboten. Dennoch lehnten einige Lehrende eine Hospitation ab, mehrfach mit der Begründung, dass dies für sie eine unangenehme Situation darstellen würde. Zudem war es nur möglich, die Teilnehmer der

DOS zu hospitieren, die in dem entsprechenden Zeitraum gut zugängliche Lehrveranstaltungen durchgeführt haben. Bedside-Teaching und Vorträge vor Fachkollegen wurden von dieser Untersuchung ausgenommen. Dies schränkt einerseits die Generalisierbarkeit der Ergebnisse ein, erhöht aber andererseits die Vergleichbarkeit der Lehrveranstaltungen. Vorrangig wurden, wie bei der Erhebung der Basisdaten (vgl. Kapitel 7.1) Seminare und Vorlesungen hospitiert.

Stichprobe. Die Hospitationen für diese Untersuchung fanden zwischen Mai 2011 und Juli 2012 statt. Die Stichprobe umfasste drei Gruppen: Lehrende, die aktuell an DOS teilnehmen (VSG1, $n = 43$ mit insgesamt 215 Beobachtungsintervallen), Lehrende, bei denen die Teilnahme an DOS mindestens ein Jahr (maximal acht Jahre) zurücklag (VSG2, $n = 34$ mit 174 Intervallen) und Lehrende, die nicht an DOS teilgenommen haben oder auf der Warteliste stehen (VGG, $n = 32$ mit 153 Intervallen). Jede Lehrperson wurde für diese Studie nur einmal hospitiert. Die Lehrenden waren in erster Linie Human- und Zahnmediziner, aber auch Naturwissenschaftler, Psychologen, Soziologen und Mathematiker. Drei Hochschuldidaktiker der medizinischen Fakultät führten die Lehrhospitationen durch und protokollierten die aufgetretenen didaktischen Handlungen im BEO-LV. Die Mehrheit der Hospitationen (91 von 109) wurde durch den Autor selbst durchgeführt. Es traten keine fehlenden Werte auf.

Tabelle 37 beschreibt die Stichprobenzusammensetzung der Versuchs- und Vergleichsgruppen. Für den Vergleich der Randbedingungen in den Gruppen wurden Unterschiedstests entsprechend dem Datenniveau eingesetzt: χ^2 -Analysen für kategoriale Daten sowie t-Tests für metrische Daten. Bei inhomogenen Varianzen (Levene-Test) wurden die p-Werte für die korrigierten Freiheitsgrade im t-Test herangezogen. Die Profession und das Geschlecht der Lehrenden, sowie die Länge der Lehrveranstaltungen unterschieden sich nicht signifikant zwischen den Gruppen. Erwartungsgemäß waren männliche Lehrende und Lehrende aus der Medizin am stärksten vertreten. Die Anzahl der Studierenden in Lehrveranstaltungen der Vergleichsgruppe war höher als in den beiden Versuchsgruppen, da hier weniger Kleingruppenveranstaltungen beobachtet werden konnten. Da studierendenzentrierte Lehre eher im Kleingruppenunterricht umgesetzt wird (Lübeck, 2009; Trigwell & Prosser, 2004), begünstigt dies hypothesenkonforme Ergebnisse. Daher sollten die Hypothesen zusätzlich unter dem Ausschluss von Lehrveranstaltungen im Kleingruppenunterricht (mit weniger als 10 Studierenden) getestet werden.

Tabelle 37. *Stichprobenbeschreibung mit Unterschiedstests*

		VGG	VSG1	VSG2
Lehrende	Gesamt	32	43	34
	Human-/Zahnmedizin	28	32	28
	andere Fachrichtung	4	11	6
		$\chi^2 = 1.96$		$\chi^2 = 0.34$
	weiblich	4	11	10
	männlich	28	32	24
		$\chi^2 = 1.96$		$\chi^2 = 2.82$
Studierende in den LV	<i>M</i>	79	27.8	32.9
	<i>SD</i>	76.9	34.5	20.8
		$T(40) = 3.51^*$		$T(35) = 3.28^*$
Dauer (in Minuten)	<i>M</i>	69.2	73	75.1
	<i>SD</i>	18.8	20.6	22.7
		$T(70) = -.84$		$T(64) = -1.15$

Anmerkungen: VGG = Vergleichsgruppe ohne Dozentenschulung (DOS). VSG1 = Versuchsgruppe in der Praxisphase der DOS. VSG2 = Versuchsgruppe deren DOS mindestens ein Jahr zurücklag. LV = Lehrveranstaltung.

* $p \leq .01$ (χ^2 nach Pearson, t-Test für unabhängige Stichproben).

Statistische Analysen. Die Überprüfung auf Normalverteilung beinhaltete Vergleiche von Mittelwert und Median, Analyse von Schiefe und Kurtosis sowie Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk. Hier zeigten sich starke Abweichungen von der Normalverteilung (Kolmogorov-Smirnov-Test und Shapiro-Wilk mit $p < .05$; Schiefe und Kurtosis mehrheitlich > 1), daher wurden die Unterschiede zwischen den Gruppen mit dem Mann-Whitney-Test und die Interkorrelationen nach Spearman berechnet. Weiterhin zeigten sich im Box's M-Test keine homogenen Varianz-Kovarianzmatrizen in den Gruppen, so dass die Voraussetzungen für multivariate Varianzanalysen nicht erfüllt waren.

Das Signifikanzniveau wurde vorab mit $\alpha \leq .05$ festgesetzt und die Hypothesen einseitig getestet. Die Effektstärkenberechnung (Cohens d) erfolgte auf Grundlage der Differenzen der Mittelwerte von der Vergleichsgruppe mit den jeweiligen Versuchsgruppen, standardisiert an der gepoolten Standardabweichung.

7.3.3 Ergebnisse

Nach der Vorstellung der deskriptiven Ergebnisse aller Variablen werden die Interkorrelationen der abhängigen Variablen und der Teilnehmerzahl in den Lehrveranstaltungen berichtet. Danach werden die Unterschiedstestungen und der Vergleich der Gruppen unter Ausschluss aller Kleingruppenveranstaltungen berichtet. Dadurch sollte die Vergleichbarkeit der Ergebnisse erhöht werden, da in der VSG1 mehr Lehrveranstaltungen in Kleingruppen hospitiert wurden. Schließlich werden die Vergleiche der sonstigen lehrendenzentrierten Handlungen dargestellt.

Deskriptive Statistiken. Wie in Tabelle 38 dargestellt, unterschieden sich die didaktischen Handlungen in der Häufigkeit ihres Auftretens. *Wortmeldungen* und *Fragen der Lehrenden* traten bei den abhängigen Variablen in fast allen Lehrveranstaltungen auf, ebenso wie *Visualisierungen* bei den sonstigen Variablen. *Lernziele* (in 6 % der Intervalle) sowie *individuelle Lern- und Arbeitsphasen* (in 15 %) traten vergleichsweise selten auf. *Individuelle Lern- und Arbeitsphasen* wurden in 4 % der beobachteten Zeit beobachtet. Die meist eingesetzte individuelle Lern- und Arbeitsphase waren Partnergespräche, in denen sich zwei oder drei Studierende zu einem vorgegebenen Thema unterhalten sollten (vgl. z. B. Crouch & Mazur, 2001; Smith et al., 2009).

Interkorrelationen. Die Interkorrelationen der abhängigen Variablen und der Teilnehmerzahl in den Lehrveranstaltungen (Tabelle 39) zeigten, dass letztere einen signifikanten korrelativen Zusammenhang mit den *Wortmeldungen* und den *Fragen der Lehrenden* aufwies. In kleineren Gruppen traten mehr Wortmeldungen auf. Kein Unterschied ließ sich bei dem Einsatz und der Dauer individueller Lern- und Arbeitsphasen nachweisen. Erwartungskonform zeigte sich ein hoher Zusammenhang von *Fragen der Lehrenden* und *Wortmeldungen der Studierenden* sowie dem *Auftreten* und der *Dauer individueller Lern- und Arbeitsphasen*. Die übrigen Interkorrelationen waren gering oder niedriger.

Tabelle 38. Deskriptive Statistik der beobachteten Lehrveranstaltungen ($N = 109$)

Didaktische Handlung	<i>M</i>	<i>SD</i>	Min	Max	%
Abhängige Variablen					
ILAP (Auftreten)	.15	.24	.00	1.00	39.4
ILAP (Dauer)	.04	.09	.00	.53	39.4
Fragen der Lehrenden	3.30	2.86	.00	12.50	90.8
Wortmeldungen von Studierenden	4.09	2.93	.00	10.00	92.7
Lernziele	.06	.12	.00	.67	23.9
Sonstige didaktische Handlungen					
Problem/Fall/Herausforderung	.35	.35	.00	1.00	63.3
Visualisierung	.73	.33	.00	1.00	92.7
Struktur/Systematisierung	.43	.35	.00	1.00	76.1
Persönlicher Bezug	.35	.32	.00	1.00	67.9
Handlungsrelevanz	.43	.38	.00	1.00	69.7
Weist auf wichtige Inhalte hin	.24	.26	.00	1.00	63.0
Interdisziplinarität	.08	.17	.00	.75	24.8
Zusammenfassung	.18	.20	.00	.75	55.0

Anmerkungen. ILAP = Individuelle Lern- und Arbeitsphasen. *M* = mittleres Auftreten einer didaktischen Handlung in einem 15-minütigen Intervall. % = Prozentsatz des mindestens einmaligen Auftretens einer didaktischen Handlung in einer Lehrveranstaltung.

Tabelle 39. *Interkorrelationen der abhängigen Variablen und der Teilnehmerzahl in den Lehrveranstaltungen (N = 109).*

	Lernziele	ILAP (Auftreten)	ILAP (Dauer)	Fragen	Wort- meldungen
Anzahl der Teilnehmer	.13	-.14	-.16	-.36	-.49
Lernziele	-	.14	.12	.05	.03
ILAP (Auftreten)		-	.73	.24	.21
ILAP (Dauer)			-	.18	.17
Fragen				-	.85

Anmerkungen. ILAP = Individuelle Lern-und Arbeitsphasen. Korrelation nach Spearman. Fett markierte Werte zeigen signifikante Korrelationen ($p \leq .05$).

Sonstige didaktische Handlungen. In Tabelle 40 sind die Unterschiedstestungen der sonstigen didaktischen Handlungen dargestellt. Die Unterschiede zwischen den Gruppen waren insgesamt gering, nur bei jeweils zwei Variablen zeigten sich Unterschiede zwischen der VGG und einer VSG. Lehrende in der VGG haben signifikant häufiger Visualisierungen eingesetzt, als die VSG1 und VSG2, häufiger interdisziplinäre Bezüge dargestellt als die VSG1 und seltener Inhalte zusammengefasst als die VSG2.

Tabelle 40. *Ergebnisse der Beobachtungsdaten und Gruppenvergleiche (sonstige didaktische Handlungen)*

Didaktische Handlung	VGG (n = 32)		VSG1 (n = 43)		VSG2 (n = 34)		VGG - VSG1		VGG - VSG2	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>z</i>	<i>p</i>	<i>z</i>
Problem/Fall/Herausforderung	.39	.40	.31	.33	.35	.34	.492	-0.69	.692	-0.40
Visualisierung	.88	.21	.65	.36	.69	.33	.002	-3.14	.005	-2.83
Struktur/Systematisierung	.46	.37	.36	.33	.49	.33	.260	-1.13	.515	-0.65
Persönlicher Bezug	.40	.37	.30	.32	.36	.27	.341	-0.95	.770	-0.29
Handlungsrelevanz	.41	.38	.41	.39	.48	.35	1.00	0.00	.412	-0.82
Wichtige Inhalte	.30	.29	.23	.29	.21	.19	.192	-1.30	.276	-1.09
Interdisziplinarität	.11	.17	.04	.14	.10	.19	.010	-2.59	.545	-0.60
Zusammenfassung	.14	.20	.15	.17	.25	.23	.633	-4.80	.028	-2.20

Anmerkung: VGG = Vergleichsgruppe ohne Dozentenschulung (DOS). VSG1 = Versuchsgruppe in der Praxisphase der DOS. VSG2 = Versuchsgruppe deren DOS mindestens ein Jahr zurücklag. Die Mittelwerte (*M*) geben das durchschnittliche Auftreten der didaktischen Handlungen in einem 15-minütigen Intervall an. Unterschiedstestungen mit dem Mann-Whitney-U-Test. Da hier mit SPSS (Version 21.0.0.0) gerechnet wurde, sind die berichteten *z*-Werte immer negativ. Die Berechnung der *p*-Werte basiert auf zweiseitiger Testung.

Hypothesentestung. Alle studierendenzentrierten Merkmale traten bei der VSG1 und VSG2 signifikant häufiger auf, als bei der VGG (Tabelle 41): Die VSG1 und die VSG2 setzten häufiger individuelle Lern- und Arbeitsphasen um als die VGG (Hypothese 1), sie setzten zeitlich länger individuelle Lern- und Arbeitsphasen um als die VGG (Hypothese 2), sie ermöglichten mehr Wortmeldungen der Studierenden als die VGG (Hypothese 3) und sie stellten den Studierenden mehr Fragen als die VGG (Hypothese 4). Die VSG1 nannte häufiger die Lernziele ihrer Lehrveranstaltung als die VGG, allerdings konnte dieser Unterschied bei der VSG2 nicht nachgewiesen werden (Hypothese 5). Eine kleine Effektstärke zeigte sich zugunsten der VSG1 gegenüber der VGG bei der Nennung der Lernziele, während sich bei der VSG2 kein bedeutsamer Effekt feststellen ließ. Die Ergebnisse standen also in Einklang mit den Hypothesen 1 bis 4, aber nur teilweise mit der Hypothese 5. Die Effektstärken gegenüber der VGG waren bei der VSG1 generell höher als bei der VSG2.

Die Ergebnisse blieben stabil, wenn lediglich Seminare und Vorlesungen mit mindestens zehn teilnehmenden Studierenden betrachtet werden (Tabelle 42). Jedoch blieb auch hier die Anzahl der Studierenden in Lehrveranstaltungen der VGG signifikant höher als in den beiden Versuchsgruppen.

Tabelle 41. *Ergebnisse der Beobachtungsdaten und Hypothesenprüfung*

Didaktische Handlung	VGG		VSG1		VSG2		VGG - VSG1			VGG - VSG2		
	<i>(n = 32)</i>		<i>(n = 43)</i>		<i>(n = 34)</i>		<i>p</i>	<i>z</i>	<i>d</i>	<i>p</i>	<i>z</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>						
ILAP (Auftreten)	.02	.07	.27	.29	.13	.21	.000	-4.77	1.12	.002	-2.82	0.71
ILAP (Dauer)	.03	.08	.06	.10	.04	.08	.001	-3.47	0.37	.024	-1.98	0.21
Fragen der Lehrenden	2.78	3.36	3.51	2.58	3.55	2.69	.026	-1.95	0.25	.030	-1.88	0.25
Wortmeldungen von Studierenden	2.92	3.02	4.89	2.66	4.19	2.89	.002	-2.93	0.70	.023	-1.99	0.43
Lernziele	.03	.13	.08	.14	.05	.10	.022	-2.03	0.33	.052	-1.62	0.15

Anmerkung: VGG = Vergleichsgruppe ohne Dozentenschulung (DOS). VSG1 = Versuchsgruppe in der Praxisphase der DOS. VSG2 = Versuchsgruppe deren DOS mindestens ein Jahr zurücklag. ILAP = Individuelle Lern-und Arbeitsphasen. Die Mittelwerte (*M*) geben das durchschnittliche Auftreten der didaktischen Handlungen in einem 15-minütigen Intervall an. Unterschiedstestungen mit dem Mann-Whitney-U-Test. Da hier mit SPSS (Version 21.0.0.0) gerechnet wurde, sind die berichteten *z*-Werte immer negativ. Die Berechnung der *p*-Werte basiert auf einseitiger Testung.

Tabelle 42. *Ergebnisse der Beobachtungsdaten und Hypothesenprüfung (nur Seminare und Vorlesungen)*

Didaktische Handlung	VGG		VSG1		VSG2		VGG - VSG1			VGG - VSG2		
	(n = 30)		(n = 30)		(n = 31)		p	z	d	p	z	d
	M	SD	M	SD	M	SD						
ILAP (Auftreten)	.02	.07	.24	.25	.11	.18	.000	-4.37	1.20	.003	-2.72	0.66
ILAP (Dauer in Minuten)	.02	.07	.04	.05	.03	.05	.000	-3.57	0.33	.023	-2.00	0.21
Fragen der Lehrenden	2.57	3.28	3.40	2.84	3.45	2.58	.039	-1.76	0.27	.012	-2.25	0.30
Wortmeldungen von Studierenden	2.72	2.88	4.58	2.81	4.19	2.89	.005	-2.57	0.65	.011	-2.30	0.51
Lernziele	.04	.13	.10	.15	.06	.10	.011	-2.29	0.43	.047	-1.67	0.17

Anmerkung: VGG = Vergleichsgruppe ohne Dozentenschulung (DOS). VSG1 = Versuchsgruppe in der Praxisphase der DOS. VSG2 = Versuchsgruppe deren DOS mindestens ein Jahr zurücklag. ILAP = Individuelle Lern- und Arbeitsphasen. Die Mittelwerte (M) geben das durchschnittliche Auftreten der didaktischen Handlungen in einem 15-minütigen Intervall an. Unterschiedstestungen mit dem Mann-Whitney-U-Test. Da hier mit SPSS (Version 21.0.0.0) gerechnet wurde, sind die berichteten z-Werte immer negativ. Die Berechnung der p-Werte basiert auf einseitiger Testung.

7.3.4 Diskussion

Hypothesentestung. Die Ergebnisse der Beobachtungsstudie haben gezeigt, dass Lehrende nach der DOS häufiger studierendenzentrierte Lehre umsetzen als Lehrende ohne Besuch der DOS. Vier Indikatoren studierendenzentrierter Lehre wurden erfasst, dabei zeigten sich mittlere bis hohe Effektstärken bei dem Auftreten individueller Lern- und Arbeitsphasen sowie bei Wortmeldungen von Studierenden, und kleine Effektstärken bei der Dauer individueller Lern- und Arbeitsphasen sowie bei Fragen der Lehrenden. Diese Ergebnisse zeigten sich sowohl bei der VSG1 als auch bei der VSG2.

Da nur ein Messzeitpunkt vorlag, bleibt offen, ob die DOS kausal für die Unterschiede verantwortlich ist. In persönlichen Gesprächen unmittelbar nach den Lehrveranstaltungen gaben die Lehrende jedoch mehrheitlich an, dass sie erst durch die DOS studierendenzentrierte Lehre eingeführt oder stärker umgesetzt haben. Die beobachteten Unterschiede sind trotzdem teilweise eher gering, so ist beispielsweise die Dauer individueller Lern- und Arbeitsphasen in den Versuchsgruppen mit 6 % (VSG1), bzw. 4 % (VSG2), gegenüber 3 % der Zeit in der VGG nur unwesentlich höher. Es ist unwahrscheinlich, dass dieser Unterschied mit bedeutsamen Wirkungen einhergeht (vgl. Kapitel 3.5). Da die Daten in dieser Studie vorrangig Beobachtungen darstellen, kann zudem nicht daraus der Schluss gezogen werden, dass die Häufigkeit des Auftretens bestimmter didaktischer Handlungen tatsächlich mit besserer Gestaltung der Lehre einhergeht. Ohne Evaluationsdaten oder Leistungsindikatoren kann zu der Qualität der Lehre oder ihrer Ergebnisse keine Aussage gemacht werden.

Vorgehen und Stichprobe. Das querschnittliche und quasi-experimentelle Design der Studie schränkt die Generalisierbarkeit der Ergebnisse ein. Wie in anderen Studien zu hochschuldidaktischen Weiterbildungen (vgl. Stes et al., 2009) war hier keine randomisierte Zuordnung der Lehrenden zu einer der Gruppen möglich. Besonderheiten der Stichprobe haben möglicherweise die Ergebnisse beeinflusst. Dass die Lehrenden ihre Lehrveranstaltungen freiwillig hospitieren ließen, stellt eine Selektion dar. Lehrende, die an der DOS teilgenommen hatten, haben möglicherweise eher die Beobachtungskategorien durch die Beobachter antizipiert, so dass hier der Einfluss sozialer Erwünschtheit nicht auszuschließen ist. Zudem ist unbekannt, welche anderen Weiterbildungen die Lehrenden aller Gruppen bereits besucht haben.

Die Anzahl der Studierenden in Lehrveranstaltungen der VGG war höher als in der VSG1 und VSG2. Dadurch kann das Auftreten studierendenzentrierter Lehre in der VSG1 und VSG2 gegenüber der VGG begünstigt worden sein (vgl. Lübeck, 2009). Dies

betraff vor allem die Indikatoren *Fragen der Lehrenden* und *Wortmeldungen der Studierenden*, da diese korrelative Zusammenhänge mit der Teilnehmerzahl aufwiesen, weniger das Auftreten und die Dauer *individueller Lern- und Arbeitsphasen*.

Material. Für diese Untersuchung wurde der BEO-LV entworfen, da ein Mangel an objektiven Beobachtungsinstrumenten für die Evaluation hochschuldidaktischer Trainings vorlag (siehe Stes et al., 2010). Die Stärke des BEO-LV besteht in ausgearbeiteten operationalen Definitionen mit umfangreichen Ein- und Ausschlusskriterien, die es ermöglichen, die Ziele der DOS verhaltensbasiert zu erfassen. Zudem konnten alle studierendenzentrierten Handlungen der Lehrenden vollständig erfasst werden. Auch wenn die Beobachtungskategorien als Indikatoren erfolgreicher Lehre wissenschaftlich fundiert sind (vgl. Kapitel 7.2.1; Anhang C2), stehen Untersuchungen dazu aus, welche die Ergebnisse des BEO-LV mit Indikatoren von Lehrerfolg validieren. Zur Bestimmung der prädiktiven Validität des BEO-LV würden sich Leistungsergebnisse der Studierenden oder kompetenzorientierte Evaluationen der Studierenden eignen (Braun, Gusy, Leidner & Hannover, 2008). Auch Selbstreflexionen oder summative Bewertungen durch Studierende könnten herangezogen werden. Ton- und Videoaufnahmen sollten nach Möglichkeit ebenfalls genutzt werden, um zu prüfen, wie zuverlässig Daten sind, die unmittelbar bei der Beobachtung der Lehre protokolliert werden.

Da nur eine Person jeweils die Lehrveranstaltungen hospitieren konnte, müssen die in den Vorstudien (vgl. Kapitel 7.2.2 und 7.2.3) ermittelte Beobachterübereinstimmung, die Interraterreliabilität und die Validität als bester Schätzer der Beobachtungsgüte herangezogen werden. Dort zeigten sich akzeptable bis gute Ergebnisse für die abhängigen Variablen dieser Studie. Die Ergebnisse der Beobachtergüte sollten jedoch bei jeder Beobachtungsstudie neu ermittelt werden (Wirtz & Caspar, 2002), dies stellt hier also eine Limitation der Ergebnisse dar. Die Vorstudien haben gezeigt, dass die Beobachtungen mit dem BEO-LV eine gute Einarbeitung in die operationalen Definitionen erfordern, diese Voraussetzung war bei den eingesetzten Beobachtern erfüllt.

7.4 Beobachtung von Lehrverhalten durch Lehrende

Jeweils im Anschluss an die DOS 1 und 2 haben die Lehrenden Zeit, die gelernten Inhalte in der Lehrpraxis anzuwenden und sich wechselseitig mit einem Kollegen aus der DOS zu hospitieren (vgl. Kapitel 5.1). Die Beobachtungen werden schriftlich reflektiert. Die Untersuchung der kollegialen Hospitationen in diesem Kapitel beschreibt, wie Lehrende in der Praxisphase der DOS ihre Lehre gestalten.

7.4.1 Hypothesen

Die Ergebnisse der Basisdaten (vgl. Kapitel 7.1.3) haben gezeigt, dass die Mehrheit der Lehrenden Wortmeldungen der Studierenden ermöglicht, aber in weniger als 15 % der Lehrveranstaltungen die Lernziele genannt oder individuelle Lern- und Arbeitsphasen umgesetzt werden. Die Beobachtungsstudie (vgl. Kapitel 7.3.3) ergab, dass Lehrende nach der DOS häufiger Lernziele präsentieren und individuelle Lern- und Arbeitsphasen umsetzen als die Vergleichsgruppe. Dieser Befund sollte durch die Daten der kollegialen Hospitationen ergänzt werden. Die Ergebnisse werden mit den Werten der Basisdaten und der Beobachtungsstudie verglichen, da in dieser Studie keine Vergleichsgruppe vorlag. Die Hypothesen lauten:

- H1: Lehrende in der Praxisphase der DOS berichten von häufigerer Umsetzung individueller Lern- und Arbeitsphasen gegenüber dem Befund der Basisdaten
- H2: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen von Kommunikation mit den Studierenden (stellen Fragen oder regen Diskussionen an).

Diese Hypothese lässt sich weiter differenzieren, da Kommunikation in der Lehre verschiedene Facetten umfasst.

- H2a: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Wissensfragen stellen.
- H2b: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Fragen zum Vorwissen der Studierenden stellen.
- H2c: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Fragen zum Interesse der Studierenden stellen.
- H2d: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Fragen zum Verständnis der Studierenden stellen.

- H2e: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Meinungen/Haltungen der Studierenden aufgreifen.
- H2f: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Diskussionen unter den Studierenden entstehen lassen.
- H2g: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie den Studierenden Feedback geben.
- H2h: Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie Feedback von den Studierenden erhalten.
- H3: Lehrende in der Praxisphase der DOS berichten von häufigerer Präsentation der Lernziele gegenüber dem Befund der Basisdaten.

7.4.2 Methoden

Entwurf der Beobachtungsbögen. Für diese Studie wurde eine quantifizierbare Beschreibung didaktischer Handlungen angestrebt. Dazu wurden die bestehenden Feedbackinstrumente verändert, welche zuvor eine rein qualitative Bewertung der Lehre beinhalteten. Ein neu entwickelter Beobachtungsbogen, angelehnt an den BEO-LV wurde zwei Kohorten der DOS (zwischen März 2011 und April 2012) zur Bearbeitung gegeben. In den abschließenden vier Praxisberatungen, die jeweils durch den Autor dieser Arbeit moderiert wurden, wurden die Erfahrungen mit dem Beobachtungsbogen besprochen. Dabei zeigten sich kritische Punkte für die Zuverlässigkeit der Daten. Inhaltlich bestand die Schwierigkeit, dass das Verständnis der Teilnehmer von den Beobachtungskategorien stellenweise sehr heterogen war. Auch die Tendenz zu sozial erwünschten Antworten wurde offen angesprochen. Einige Lehrende haben alle didaktischen Handlungen in der Selbst- oder Fremdbeobachtung angekreuzt, was angesichts praktischer Erfahrungen aus den Vorstudien (vgl. Kapitel 7.1 und 7.2) unwahrscheinlich erschien. Zudem haben einzelne Lehrende bestätigt, die Bögen das erste Mal in der Lehrveranstaltung angesehen zu haben oder diese erst nachträglich ausgefüllt zu haben. Aus diesen Erfahrungen wurden mehrere Schlüsse gezogen: (a) Der Einsatz des BEO-LV für Lehrende in der gegenseitigen Hospitation ist nicht sinnvoll, eine reduzierte Version eignet sich besser. (b) Um den Einfluss der sozialen Erwünschtheit auf die Messergebnisse zu verringern, sollte nicht nur das Auftreten einer didaktischen Handlung erhoben

werden, sondern auch eine kurze Beschreibung wie die Handlung umgesetzt wurde. Gleichzeitig sollte so auch die Selbstreflexion der Lehrenden intensiviert werden. (c) Die Trainer der jeweiligen Praxisberatungsgruppen sollen im Vorfeld die Bearbeitung der Beobachtungsbögen mit den Lehrenden ausführlich besprechen und den Zweck der Untersuchung darstellen.

Material. Die überarbeiteten Beobachtungsbögen sind in Anhang E abgebildet. Es gab zwei identische Versionen: eine für die Person, die hospitiert und eine für die Person, die die Lehrveranstaltung durchführt. Die Beobachtungsbögen beinhalteten acht Fragen zu lehrendenzentrierten Handlungen und acht Fragen zur Kommunikation, zu denen die Personen jeweils eine kurze Beschreibung angeben sollten, in welcher Form sie die entsprechenden Handlungen beobachtet bzw. durchgeführt haben. Zudem beinhalteten die Beobachtungsbögen drei Fragen zu individuellen Lern- und Arbeitsphasen, vier allgemeine Fragen und drei offene Fragen. Die Auswahl der Beobachtungskategorien basierte anhand der in Kapitel 7.2.1 beschriebenen Kriterien und war so umfassend, dass alle möglichen Indikatoren studierendenzentrierter Lehre abgebildet werden konnten (vgl. Kapitel 3.4) Die Kategorien wurden so formuliert, dass sie in jedem Lehrformat einsetzbar sind.

Durchführung. Nach der wechselseitigen Hospitation fanden jeweils Abschlussreflexionen statt, in denen die schriftlich protokollierten Beobachtungen besprochen wurden (vgl. Kapitel 5.1). Für die statistische Untersuchung der Beobachtungen wurden die Hospitationsbögen der Lehrenden analysiert, die in dem Untersuchungszeitraum (März 2012 bis März 2013) an Abschlussreflexionen teilgenommen haben. Der Zweck der Untersuchung wurde den Lehrenden mitgeteilt. Die Bögen wurden durch die Trainer an den Autor dieser Arbeit weitergeleitet und standen so für eine Analyse zur Verfügung. Für die Evaluation der DOS werden im Folgenden nur die beschreibenden Daten, nicht die gegenseitigen Bewertungen herangezogen.

Auswertung. Um eine möglichst hohe Datenqualität zu erzielen, wurde folgendes Vorgehen gewählt: (a) Es wurden nur die Hospitationen Lehrender berücksichtigt, bei denen sowohl die Bögen der Selbst- als auch der Fremdbeobachtungen durch den Hospitationspartner vorlagen. (b) Bei Lehrenden, die in dem entsprechenden Zeitraum in beiden Praxisberatungen (der DOS 1 und der DOS 2) vollständig bearbeitete Bögen abgegeben haben, wurde ausschließlich auf die Daten von der DOS 2 zugegriffen, um den Einfluss der DOS bestmöglich zu schätzen. (c) Bei offensichtlichem Missverstehen des Beobachtungsbogens (beispielsweise bei Lehrenden mit geringen Deutschkenntnis-

sen) oder bei falscher Nutzung der Antwortkategorien (beispielsweise wenn nur Freitexte beantwortet wurden) wurden die Angaben nicht berücksichtigt. Die Hypothesen wurden geprüft, indem die Anzahl von positiven Werten (wenn eine didaktische Handlung mindestens einmal angegeben wurde) gezählt wurde. Für die Analyse von Hypothese 2 wurde die Kategorie *Dialog gesamt* gebildet, deren Wert 1 ergibt, wenn mindestens eine Kategorie zu Kommunikation als aufgetreten protokolliert wurde.

Als Übereinstimmungsmaß von Selbst- und Fremdbeobachtung wurde hier zur besseren Veranschaulichung die prozentuale Übereinstimmung vorliegender Daten verwendet. Da die Lehrenden nach der Hospitation die Ergebnisse gemeinsam besprechen und die Daten der Selbsthospitation oft erst im Anschluss daran eingetragen wurden, kann nicht von einer Unabhängigkeit bei der Datenerfassung ausgegangen werden.

Stichprobe. Den Teilnehmern der DOS stand es frei, wann sie an einer Abschlussreflexionsgruppe teilnehmen, in der Regel nehmen sie jedoch in dem Semester unmittelbar nach dem Workshop teil. Insgesamt konnten die Selbst- und Fremdbeobachtungsbögen von 103 Personen für diese Untersuchung berücksichtigt werden.

Tabelle 43 zeigt die Verteilung der verwendeten Hospitationsbögen nach Zugehörigkeit zu den jeweiligen DOS-Kohorten und nach Geschlecht. Weitere demographische Angaben wurden aus Datenschutzgründen hier nicht erfasst.

Tabelle 43. *Vorliegende Beobachtungsbögen im Bearbeitungszeitraum*

Kohorte der Dozentenschulung	<i>N</i>
1.16	23
1.17	23
2.21	6
2.22	35
2.23	16
Geschlecht	<i>N</i>
weiblich	35
männlich	68

Die Teilnehmer konnten frei wählen, welche Lehrveranstaltungen sie im Rahmen der DOS für die kollegiale Hospitation nutzen. Tabelle 44 zeigt die hier gewählten Lehr-

formate. Die Darstellung in der Tabelle beschränkt sich auf die Angaben aus den Selbstbeobachtungsbögen, da die Angaben bei den Fremdbeobachtungen weitgehend übereinstimmen. Erwartungsgemäß war die Zahl der Studierenden in den Vorlesungen deutlich höher als bei den anderen Lehrformaten.

Tabelle 44. *Deskriptive Darstellung der Lehrformate*

Lehrformat	N	Anzahl Studierende		Dauer	
		M	SD	M	SD
Seminar/PJ-Unterricht	43	20.9	12.5	81.5	35.4
Vorlesung	29	46.3	42.4	79.0	55.5
Bedside-Teaching	9	3.5 ^a	0.9	73.3	20.5
Problemorientiertes Lernen	7	10.6	2.2	85.7	11.3
Praktikum	6	10.2	1.3	267.5	116.4
Skills-Lab	4	6.5	1.3	90.0	0.0
Workshop	3	8.3	5.8	206.7	57.7
Abteilungsfortbildung	1	5.0		90.0	
Medi-KIT.	1	5.0		45.0	

Anmerkungen. PJ = Praktisches Jahr. Medi-KIT = Medizinisches Kommunikations- und Interaktionstraining. a = Bei einem Fall fehlte die Angabe.

7.4.3 Ergebnisse

In wie vielen Lehrveranstaltungen die didaktischen Handlungen aus Sicht der Teilnehmer der DOS umgesetzt wurden, ist in Tabelle 45 abgebildet. Jede didaktische Handlung wurde mit mindestens 66.3 % bei der Selbst- und mit 58.6 % bei der Fremdbeobachtung protokolliert. Die Kategorie *Einleitung/Hinführung zum Thema* wurde mit etwa 98 % am häufigsten protokolliert.

Tabelle 45. *Deskriptive Darstellung aller beobachteten Lehrhandlungen in der kollegialen Hospitation*

	Beobachtung				PÜ
	Selbst		Fremd		
Lehrendenzentrierte Merkmale	N	%	N	%	
Einleitung/Hinführung zum Thema	95	97.9	89	97.8	96
Visualisierung	93	91.4	89	92.1	86
Strukturierung des Themas	91	91.2	88	92.0	94
Nennung der Lernziele	100	88.0	99	88.9	92
Zusammenfassung einzelner Abschnitte	93	84.9	89	84.3	84
Darstellung der Praxisrelevanz der Inhalte	92	96.7	89	95.5	93
Darstellung der Prüfungsrelevanz der Inhalte	93	68.8	88	67.0	81
Abschließende Zusammenfassung	91	84.6	90	92.2	87
Studierendenzentrierte Merkmale					
Wissensfragen gestellt	98	96.8	94	96.8	87
Nach Vorwissen/Vorkenntnissen gefragt	97	89.7	93	88.2	87
Nach Interesse an den Inhalten gefragt	96	63.5	92	70.7	67
Fragen zum Lernprozess/Verständnis gestellt	94	86.6	92	81.5	87
Meinungen/Haltungen aufgegriffen	95	77.0	91	81.3	76
Diskussionen entstehen (lassen)	95	83.2	92	83.7	81
Den Studierenden Feedback gegeben	95	75.8	90	77.8	74
Feedback von den Studierenden erhalten	95	66.3	87	58.6	66
Dialog gesamt*	96	100.0	93	100.0	100
Individuelle Lern- und Arbeitsphasen	100	90.0	98	85.7	86

Anmerkungen. PÜ = Prozentuale Übereinstimmung (N zwischen 84 und 99). * Voraussetzung, dass mindestens ein Item zu Kommunikation aufgetreten ist.

Die Werte von *Nennung der Lernziele* und *Individuelle Lern- und Arbeitsphasen* sind höher als die Ergebnisse der Basisdaten (vgl. Kapitel 7.1.3) und der Beobachtungsstudie (vgl. Kapitel 7.3.3). Lehrende in der Praxisphase der DOS berichten in 90 % (Selbstbeobachtung) bzw. 86 % (Fremdbeobachtung) von der Umsetzung individueller Lern- und Arbeitsphasen. In allen Lehrveranstaltungen haben die Lehrenden zudem Wortmeldungen ermöglicht. Lehrende in der Praxisphase der DOS berichten in der Mehrheit der Lehrveranstaltungen, dass sie *Wissensfragen stellen* (Hypothese 2a), dass sie *Fragen zum Vorwissen* der Studierenden stellen (Hypothese 2b), dass sie *Fragen zum Interesse* der Studierenden stellen (Hypothese 2c), dass sie *Fragen zum Verständnis* der Studierenden stellen (Hypothese 2d), dass sie *Meinungen/Haltungen* der Studierenden *aufgreifen* (Hypothese 2e), dass sie *Diskussionen* unter den Studierenden entstehen lassen (Hypothese 2f), dass sie sie den Studierenden *Feedback geben* (Hypothese 2g) und sie *Feedback von den Studierenden erhalten* (Hypothese 2h). Lehrende in der Praxisphase der DOS berichten in 88 % (Selbstbeobachtung) bzw. 89 % (Fremdbeobachtung) der Lehrveranstaltungen davon, die Lernziele genannt zu haben.

Die Ergebnisse stehen somit in Einklang mit den Hypothesen 1 bis 3. Angesichts der hohen Werte der abhängigen Variablen wurde auf eine differenzierte und vertiefende Analyse nach Lehrformat, Geschlecht oder DOS-Kurs verzichtet, da hier keine bedeutsamen Unterschiede erwartet wurden.

7.4.4 Diskussion

Ergebnisse. Die Auswertung der Beobachtungsbögen ergab, dass alle Beobachtungskategorien in mindestens 58 % der Lehrveranstaltungen protokolliert wurden. Die Hypothesen, dass Lehrende nach der DOS von häufigerer Umsetzung individueller Lern- und Arbeitsphasen und von häufigerer Präsentation der Lernziele gegenüber dem Befund der Basisdaten berichten, stand in Einklang mit den vorliegenden Daten. Ebenso wurde in der Mehrheit der Lehrveranstaltungen von intensiver Kommunikation mit den Studierenden berichtet.

Die hohen Werte sprechen zwar einerseits für eine hohe Akzeptanz und Umsetzung der didaktischen Prinzipien, die in der DOS vermittelt werden, stellen aber wahrscheinlich eine Überschätzung der wahren Werte dar. Die Annahme, dass Lehrende bei der Anwesenheit eines Trainers (wie in der Beobachtungsstudie, vgl. Kapitel 7.3) weniger intensiv die Inhalte der DOS umsetzen, als bei einer kollegialen Hospitation, ist eher unwahrscheinlich. Andererseits kann die Übereinstimmung von Lehrenden und Hospitati-

onspartnern als Beleg für die Validität der Daten sprechen. Allerdings ist anzunehmen, dass die Lehrenden die Beobachtungsbögen nicht unabhängig voneinander bearbeitet haben. Selbstberichte der Lehre sind anfällig für erwartungskonforme Verzerrungen (Desimone, 2009; Kane et al., 2002). Da sowohl die Beobachterübereinstimmung als auch die Höhe der Ergebnisse höher als erwartet ausfallen, sind die Daten wahrscheinlich positiv durch soziale Erwünschtheit verzerrt. Da die Beobachtungsbögen zudem mit Namen versehen sind, um den Abschluss der DOS zu zertifizieren, die Lehrenden die Inhalte der Beobachtungsbögen bei der Abschlussreflexion besprechen und schließlich persönlich einem Trainer der DOS übergeben, waren die Ergebnisse nicht anonym. Den Lehrenden war bewusst, dass die Trainer der DOS den Einsatz individueller Lern- und Arbeitsphasen und die Nennung der Lernziele befürworten. Es wurde den Teilnehmern der DOS zwar schriftlich bestätigt, dass diese Ergebnisse anonymisiert werden, aber dies reichte wohl nicht aus, um die Gefahr sozial erwünschter Antworten zu minimieren. Auch die Aufforderung, eine kurze Beschreibung zu geben, wie die Handlung umgesetzt wurde, hat dazu nicht ausgereicht.

Eine andere Erklärung lautet, dass die Lehrenden weniger strenge operationale Definitionen vorgenommen haben, als die trainierten Beobachter in den Beobachtungsstudien. Das ist beispielsweise bei einem Konstrukt wie *Nennung der Lernziele* denkbar, wenn die Lehrenden zwar die Themen der Lehrveranstaltung nennen, aber nicht die genauen Kompetenzen, die in Bezug zu diesem Thema erreicht werden sollen. Die können im weiteren Verlauf einer Lehrveranstaltung zwar häufig erschlossen werden, sollen im BEO-LV aber nur protokolliert werden, wenn die Lernziele ausdrücklich verbal formuliert werden. Auch bei anderen Kategorien wie den *individuellen Lern- und Arbeitsphasen* werden teilweise strengere Kategorien herangezogen, als Lehrende dies möglicherweise intuitiv annehmen. Jedoch lassen sich auch so nicht die hohen Unterschiede zwischen den Ergebnissen hier und in der Beobachtungsstudie (vgl. Kapitel 7.3.3) erklären.

Die Ergebnisse dieser Untersuchung zeigen, dass die Selbstberichte der Lehrenden alleine wenig geeignet sind, um den Erfolg einer didaktischen Weiterbildung zu bemessen. Darauf verwiesen bereits mehrere Autoren (Levinson-Rose & Menges, 1981; Stes et al, 2010). Insgesamt ist hier davon auszugehen, dass die Ergebnisse der Beobachtungsstudie (vgl. Kapitel 7.3.3) eine bessere Schätzung der Effektivität der Weiterbildung darstellen.

Vorgehen und Stichprobe. Da hier alle Beobachtungsbögen des untersuchten Zeitraums vorlagen, ist von einer hohen Repräsentativität der Ergebnisse auszugehen. Mit insgesamt 103 Personen, die sowohl ihre eigene Lehrveranstaltung protokolliert haben, als auch die eines Kollegen, liegen hier im Vergleich zu anderen Studien umfangreiche Beobachtungsdaten vor. Die Entscheidung, Lehrende auszuschließen, denen die Handhabung mit dem Beobachtungsinstrument offensichtlich schwer fiel, hat dabei die Wahrscheinlichkeit begünstigt, hypothesenkonforme Ergebnisse vorzufinden.

Material. Die hier verwendeten Beobachtungsinstrumente wurden für den Zweck dieser Studie entworfen. Sie sollten quantifizierbare Ergebnisse ermöglichen und gleichzeitig die Selbstreflexion im Rahmen der kollegialen Hospitation unterstützen. Die geringe Differenzierung der Beobachtungskategorien war notwendig, da sich die Anwendung eines differenzierten Schemas als zu komplex für den hier vorliegenden Rahmen herausgestellt hatte. Offen blieb bei dem hier gewählten Verfahren, wie intensiv sich die Teilnehmer der DOS im Vorfeld mit dem Instrument auseinandergesetzt hatten und wie umfassend die vorherige Besprechung in den Kleingruppen bei der DOS erfolgte. Die relativ hohe Anzahl von fehlenden Werten weist darauf hin, dass die Bearbeitung des Beobachtungsbogens nicht immer eindeutig für die Lehrenden war. Vor allem die Verbindung von quantitativer und qualitativer Erhebung hat womöglich zu unterschiedlichen Schwerpunkten bei der Bearbeitung geführt. Dass der Beobachtungsbogen zwei Funktionen erfüllen sollte, also zum einen Daten für die Evaluation der DOS bereitstellen und zum anderen die kollegiale Reflexion nach einer Lehrhospitation anregen, hat wahrscheinlich dazu geführt, dass beide Ziele nicht optimal erreicht wurden.

8 Empirische Studie 3: Die Lehre aus Sicht der Studierenden

In diesem Kapitel wird die Studierendenbefragung vorgestellt. Die übergeordnete Forschungsfrage lautete, ob Studierende in der Humanmedizin studierendenzentrierter Lehre positiv gegenüberstehen und diese als hilfreich für ihr Studium betrachten. Auch wenn aktivierende Lehrmethoden als effektiv gelten, ist deren Wirkung und Akzeptanz von der konkreten Umsetzung abhängig (vgl. Kapitel 3.5). Diese Studie sollte die Wahrnehmungen der Studierenden differenziert untersuchen und daraus praktische Schlüsse für die optimale Umsetzung aktivierender Methoden ableiten. Dazu wurde ein Fragebogen entwickelt. In Kapitel 8.1 werden die Konstruktionsschritte der Befragung vorgestellt. In Kapitel 8.1.2 werden die Ergebnisse des Pretest dargestellt und in Kapitel 8.1.3 die daraus resultierenden Überarbeitungen. In Kapitel 8.2 werden die Hypothesen, die eingesetzten Skalen und die Stichprobe dargestellt. In Kapitel 8.2.3 werden schließlich die zentralen Ergebnisse der Studie beschrieben.

8.1 Vorstudie

8.1.1 Methodik

Die Umsetzung der Studierendenbefragung umfasste drei wesentliche Schritte: (a) Konstruktion einer Skala zur Akzeptanz studierendenzentrierter Lehre, (b) Auswahl weiterer Skalen für die Untersuchung und (c) zwei Pretests an größeren Stichproben.

Die Akzeptanz studierendenzentrierter Lehre kann sich in verschiedenen Veranstaltungsformaten unterscheiden (vgl. Kapitel 3.5). In der DOS werden die Lehrenden motiviert, traditionell eher lehrendenzentrierte Veranstaltungsformate stärker studierendenzentriert zu gestalten. Daher sollte der Fokus auf Vorlesungen liegen, da Unterschiede bei der Akzeptanz studierendenzentrierter Lehre hier besonders deutlich zutage treten sollten. In einer Vorlesung können vier verschiedene Formen studentischer Aktivierung auftreten: Einzelarbeit, Partnerarbeit (zu zweit), Gruppenarbeit (mehr als zwei Personen) und durch den Lehrenden gelenkte Diskussionen sowie Fragen an die Studierenden. Alle Facetten sollten durch die Items abgebildet werden. Zwei Items sollten noch den subjektiven Nutzen der aktivierenden Methoden erheben (Beispielitem: „Gelegentliche Aktivierungen (Fragen, Aufgaben...) erhöhen in Vorlesungen meinen Lernerfolg“). Die Antwortmöglichkeiten reichten von 1 (*stimme nicht zu*) bis 5 (*stimme zu*). Da Menschen die Tendenz haben, bei Items eher zuzustimmen (Bühner, 2011, S. 134), wurde darauf geachtet, Items mit unterschiedlicher Polung einzusetzen. Zwölf Items wurden nacheinander fünf Studierenden (drei aus der Humanmedizin, zwei aus der Psy-

chologie) vorgelegt. Die Studierenden wurden gebeten, die Items zu bearbeiten und ihre Gedankengänge bei der Beantwortung jeweils rückblickend zu artikulieren, angelehnt an die Think Aloud-Methode (Prüfer & Rexroth, 2005). Nach jeder Befragung wurden die Formulierungen präzisiert. Der Fachbegriff *studierendenzentrierte Lehre* wurde durch den Begriff *aktivierende Methode* ersetzt, da dieser für die Studierenden besser verständlich war. Die überarbeitete Skala *Aktivierende Methoden in Vorlesungen* (AM-V) umfasste schließlich zehn Items. Weitere Skalen sollten Hinweise auf die konvergente und diskriminante Validität geben. Da die Effektivität studentischer Aktivierung von der Qualität der Instruktion abhängig ist (vgl. Kapitel 3.5), sollte dies auch auf die Akzeptanz aktivierender Methoden zutreffen. Eine Skala *Probleme bei der Umsetzung aktivierender Methoden* wurde erstellt, um die subjektive Wahrnehmung der Instruktionqualität zu quantifizieren. Sieben Items wurden so in den Fragebogen aufgenommen und ein vier-stufiges Antwortformat (von 4 = *Ja, habe ich oft erlebt* bis 1 = *Nein, habe ich nie erlebt*) verwendet. Es wurde erwartet, dass Studierende, die eine gute Umsetzung aktivierender Methoden erleben, diesen positiver gegenüberstehen.

Testlauf für die Vorstudie. Die Items wurden in einem Onlinefragebogen über den Anbieter Q-Set (www.q-set.de) einer Gelegenheitsstichprobe von 56 Studierenden vorgelegt. Die Studierenden wurden über drei Studierendenportale (www.studivz.de, www.medi-learn.de, www.uni-turm.de) rekrutiert und stammen aus 24 verschiedenen Studiengängen. Häufigste Nennungen waren Humanmedizin ($n = 14$) und Betriebswirtschaftslehre ($n = 4$). Die erste Testung der Skala *Aktivierende Methoden in Vorlesungen* zeigte gute Ergebnisse. Der Mittelwert der Skala lag bei 3.07 ($SD = 0.85$). Laut der Überprüfung mit den Tests nach Kolmogorov-Smirnov und Shapiro-Wilk lag bei der Skala eine Normalverteilung vor. Die interne Konsistenz der Skala war gut (Cronbachs Alpha = .89). Angesichts der guten psychometrischen Eigenschaften der Skala, sollte das Konstrukt inhaltlich homogener abgebildet werden. Die Akzeptanz aktivierender Methoden und deren wahrgenommene Wirkung sollten theoretisch miteinander in Zusammenhang stehen (vgl. Kapitel 3.5), sie sind aber keine austauschbaren Indikatoren eines Konstrukts. Daher wurden die beiden Items zum subjektiven Nutzen aktivierender Methoden aus der Skala entfernt, die Skala sollte jetzt ausschließlich die Akzeptanz aktivierender Methoden abbilden.

Die Skala *Probleme bei der Umsetzung aktivierender Methoden* sollte explorativ geprüft werden. Die Rückmeldungen der Studierenden bestätigten, dass die durch die Items repräsentierten Probleme den Studierenden bekannt waren. Zudem nannten sie in

den ergänzenden Freitexten drei weitere Probleme, von denen neue Items abgeleitet werden konnten. Inhaltlich umfasste die Skala dann Probleme bei der Instruktion (Beispielitem: „Die Aufgabe war unklar instruiert“), der Umsetzung aktivierender Methoden („Die Aufgabe war zu einfach“), und situativ-motivationale Hindernisse (Beispielitem: „Ich selber hatte keine Lust mitzuarbeiten“).

Neben den beiden Skalen zu der Umsetzung aktivierender Methoden sollten weitere Items in die Befragung aufgenommen werden. Motivationale Voraussetzungen der Studierenden beeinflussen möglicherweise die Wirkung studierendenzentrierter Lehre (Baeten et al., 2010, vgl. Kapitel, 3.5). Der Einsatz tiefer Lernansätze bei Studierenden steht in Zusammenhang mit höherer Leistungszielorientierung (Cano & Berben, 2009) und intrinsischer Lernmotivation (Harris, 2004). Welcher Zusammenhang zwischen Lernmotivation und der Akzeptanz aktivierender Methoden besteht, wurde bislang kaum untersucht. Erfahrungen in Leistungskontexten können als Herausforderung oder als selbstwertbedrohlich erlebt werden (Köller & Schiefele, 2010). Aktivierende Lehrmethoden bewirken, dass Studierende offene Handlungen vollziehen, unter den Augen der Lehrenden und denen anderer Studierenden. Möglicherweise bevorzugen Studierende mit hoher *Lernzielorientierung*, sowie Studierende mit hoher *Annäherungs-Leistungszielorientierung*, diese offene Auseinandersetzung mit den Lerninhalten. Studierende mit einer hohen *Vermeidungs-Leistungszielorientierung* lehnen studierendenzentrierte Lehre dagegen möglicherweise ab, da sie in einer solchen Situation ihre (vermeintlich) unterlegene Kompetenz schlechter verbergen können. Daher sollte in dieser Studie geprüft werden, in welchem Zusammenhang die Konstrukte stehen. Für die Erfassung von Zielorientierungen wurden die *Skalen zur Erfassung der Lern- und Leistungsmotivation* für Studierende (*SELLMO-ST*; Spinath, Stiensmeier-Pelster, Schöne & Dickhäuser, 2002) gewählt. Die Skalen wurden im Rahmen der Vorabtestung leicht modifiziert und gekürzt. Zusätzlich dazu wurden die Kurzskalen *Arbeitsvermeidung* (*SELLMO-ST*), sowie das *studienbezogene Interesse* (Westermann, Heise, Spies und Trautwein, 1996) mit jeweils drei Items erhoben. Weiterhin wurde untersucht, ob Korrelationen mit einzelnen Items auf eine unerwünschte Konfundierung bei der Messung hinweisen. Hier wurde die *Einschätzung der Studienbedingungen* (Item: „Mit den Studienbedingungen an meiner Fakultät bin ich sehr zufrieden“) und die wahrgenommene *Qualität der Vorlesungen* herangezogen („Mit der Qualität der Vorlesungen bin ich insgesamt zufrieden“). Im Anhang F1 sind alle Skalen, die entsprechenden Items und die Antwortformate aufgelistet.

Vorstudie. Die Skalen wurden im August 2011 zur Vorabtestung an einer größeren Stichprobe getestet. Die Befragung erfolgte erneut online über den Anbieter *Q-Set*. Die Rekrutierung der Teilnehmer erfolgte über sieben Onlineportale (www.uni-turm.de, www.medi-learn.de, www.unicum.de, www.thieme.de, www.psychologieheute.de, www.studivz.de). Den Teilnehmern wurde eine Zusammenfassung der Ergebnisse in Aussicht gestellt. Die gewonnene Gelegenheitsstichprobe umfasste 199 Personen, aus 34 verschiedenen Studiengängen. Die häufigste Nennungen waren Lehramt ($n = 61$), Informatik ($n = 14$) und Maschinenbau ($n = 10$). Mehr Frauen (56.8 %) als Männer (41.2 %) nahmen an der Befragung teil (2 % der Befragten gaben kein Geschlecht an).

Da es in der laufenden Studie eine Rückmeldung gab, dass eine Person keine Erfahrungen mit aktivierenden Methoden in Vorlesungen hatte, wurde nachträglich eine entsprechende Frage eingefügt. Nur vier von 99 Befragten antworteten hier, dass sie diese „*nie*“ erlebt haben. Daher wird im Folgenden davon ausgegangen, dass Studierende Fragen zu der Akzeptanz aktivierender Methoden zuverlässig beantworten können.

Statistische Auswertung. Alle Skalen wurden so kodiert, dass hohe Werte mit einer hohen Ausprägung einhergehen. Skalenwerte wurden nur berechnet, wenn mindestens zwei Drittel aller Items bearbeitet wurden. Der Skalenwert einer Person wird aus der Summe der gegebenen Antworten berechnet, dividiert durch die Zahl der beantworteten Fragen. Die interne Konsistenz wurde mit dem Alpha-Koeffizienten (α) nach Cronbach erhoben (1951). Die Berechnung der Trennschärfe erfolgte mit der *part-whole-correction* (Kelava & Moosbrugger, 2012).

8.1.2 Ergebnisse

Deskriptive Ergebnisse. In der Stichprobe zeigte sich eine Tendenz zugunsten aktivierender Methoden in Vorlesungen (Tabelle 46). Die statistische Überprüfung mit einem Ein-Stichproben t-Test zeigte eine signifikante Abweichung vom numerischen Mittelwert der Skala, $t(195) = 3.28$, $p < .01$. Probleme bei der Umsetzung aktivierender Methoden wurden von vielen Studierenden wahrgenommen. Die Ergebnisse der weiteren Skalen zeigten, dass die Studierenden sich als sehr lernzielorientiert und sehr interessiert an ihrem Studium beschrieben.

Tabelle 46. Deskriptive Statistik Studierendenbefragung (Vorstudie)

Skala (Anzahl der Items)	<i>N</i>	<i>M</i>	<i>SD</i>	<i>r_{it}</i> Bereich	α	<i>r</i> <i>AM-V</i>
AM-V (8)	196	3.22	.92	.39 – .73	.86	-
AM-P (10)	185	2.76	.42	-.03 – .46	.62	-.36
LO (3)	195	4.48	.53	.35 – .39	.56	.13
AL (3)	195	3.45	.83	.26 – .61	.65	.01
VL (3)	195	2.27	.93	.59 – .69	.77	-.16*
AV (3)	195	2.24	.88	.79 – .82	.90	-.02
Interesse (3)	195	4.12	.84	.69 – .77	.84	.01

Anmerkungen. r_{it} = korrigierte Trennschärfe. α = Cronbachs Alpha. r_{AM-V} = Korrelation nach Pearson mit der Skala AM-V. AM-V = Aktivierende Methoden in Vorlesungen, AM-P = Probleme bei der Umsetzung aktivierender Methoden, LO = Lernzielorientierung, AL = Annäherungs-Leistungszielorientierung, VL = Vermeidungs-Leistungszielorientierung, AV = Arbeitsvermeidung. * $p < .05$.

Stichprobenverteilung. Die Überprüfung auf Normalverteilung mit den Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk zeigte für alle Skalen und Items signifikante Abweichungen von der Normalverteilung ($p < .05$). Die Werte von Schiefe und Kurtosis sind kleiner 1 und können daher nach Miles und Shevlin (2001) nur als moderat abweichend von der Normalverteilung interpretiert werden.

Interne Konsistenz der Skalen. Die korrigierten Trennschärfen der Items pro Skala lagen mehrheitlich über dem empfohlenen Wert von .30 (Backhaus, Erichson, Plinke & Weiber, 2008). Eine Ausnahme stellte die Skala *Probleme bei der Umsetzung aktivierender Methoden* dar, bei der vier von zehn Items in der Vortestung Werte $< .30$ erzielten. Die Skalen *Aktivierende Methoden in Vorlesungen*, *Arbeitsvermeidung* und *Interesse* zeigten eine akzeptable interne Konsistenz, während die internen Konsistenzen der Zielorientierungen relativ gering waren.

Interkorrelationen. Die Korrelation mit der Skala *Probleme bei aktivierenden Methoden* ergaben einen erwartungskonformen moderat negativen Zusammenhang ($r = .36$). Studierende, die bei der Umsetzung aktivierender Methoden bereits häufiger Probleme erlebt haben, sind aktivierender Lehre gegenüber tendenziell kritischer eingestellt. Es zeigten sich nur geringe Zusammenhänge mit den übrigen Skalen, ein statistisch signifikanter Zusammenhang fand sich nur zu der *Vermeidungs-*

Leistungszielorientierung. Es zeigte sich weiterhin kein Zusammenhang zu der Einschätzung der *Studienbedingungen* ($r = .01$), oder zu der Einschätzung der *Qualität der Vorlesungen* insgesamt ($r = .08$).

8.1.3 Diskussion und Skalenüberarbeitung

Die ermittelten Werte sprachen dafür, alle Items der Skala *Aktivierende Methoden in Vorlesungen* beizubehalten. Es wurde davon ausgegangen, dass die Skala gut geeignet ist, um die Akzeptanz aktivierender Methoden durch die Studierenden der medizinischen Fakultät abzubilden. Die nicht signifikanten Korrelationen mit den inhaltlich weiter entfernten Konstrukten können als Beleg für die diskriminante Validität der Skala gewertet werden. Die Skala erfasst die Akzeptanz aktivierender Methoden unabhängig von lernmotivational bedeutsamen Variablen. Vor dem Hintergrund der Befunde zum Zusammenhang von studentischen Lernansätzen und lernmotivationalen Konstrukten (Baeten et al., 2010) sind diese Ergebnisse jedoch überraschend. Als mögliche Erklärung können Eigenschaften der Stichprobe herangezogen werden. Deckeneffekte (insbesondere bei den Skalen *Lernzielorientierung* und *Interesse*), Abweichungen von der Normalverteilung sowie die Zusammensetzung der Gelegenheitsstichprobe schränken die Varianz der Ergebnisse ein. Studierende mit höherer Lernmotivation und Interesse an dem Thema der Befragung waren wahrscheinlich überrepräsentiert. Die Interkorrelationen sind möglicherweise geringer als in der Population. In vergleichbaren Erhebungen zeigten sich unter Studierenden ebenfalls hohe Werte bei der *Lernzielorientierung* und niedrige bei der *Arbeitsvermeidung* (Spinath et al., 2002; Wolgast, Stiensmeier-Pelster & von Aufschnaiter, 2014). Da für die Skalen des *SELLMO-ST* keine Normwerte vorliegen, kann dies nicht abschließend beurteilt werden.

Es bestand die Annahme, dass Studierende, die aktivierende Methoden in Vorlesungen befürworten, diese auch in anderen Lehrformaten befürworten. Um dies zu überprüfen, wurden zwei zusätzliche Skalen entwickelt. Einmal wurde das Antwortformat verändert und der Fokus auch auf Seminare mit höherer Teilnehmerzahl erweitert: „Wenn Sie die Didaktik in den `theoretischen` Lehrveranstaltungen (Vorlesungen und Seminare) mit mindestens 30 Teilnehmern Ihres Studiums für sich selbst bestimmen könnten, welche Verteilung würden Sie wählen?“ Die Studierenden sollten bei diesem Item hundert Prozentpunkte auf die von ihnen präferierten didaktischen Inhalte verteilen. Zur Auswahl standen neben dem Vortrag der Lehrenden vier verschiedene Merkmale aktivierender Methoden (Fragen und Diskussionen, Einzelarbeit, Partnerarbeit und

Gruppenarbeit). Die Teilnehmerzahl größer 30 stellt für Wildt (2003) die Grenze zu einer Massenveranstaltung dar, da hier für Lehrende keine echte Chance mehr für individuelle Interaktionen mit allen Teilnehmern besteht. Da relativ viele Seminare in der Humanmedizin mit 30 Studierenden oder mehr stattfinden, sind die Antworten für eine breitere Anzahl von Lehrveranstaltungen gültig.

In einer weiteren Skala wurde der Fokus von einzelnen Lehrveranstaltungen auf eine modulare Lehrveranstaltungsreihe hin gerichtet. In der Humanmedizin beinhalten diese, je nach Fach, eine Vielzahl von verschiedenen Lehrformaten. Die Instruktion lautete „Stellen Sie sich bitte eine zweiwöchige Lehrveranstaltungsreihe vor, an deren Ende eine Prüfung steht. Unter welchen Umständen würden Sie einen eher frontalen Unterricht bevorzugen und unter welchen Umständen einen eher aktivierenden Unterricht mit vielen individuellen Lernphasen?“ Als Antwortoptionen stehen hier vier Kombinationen als Items zur Bearbeitung („Interessantes Fach - Schwierige Prüfung“, „Interessantes Fach - Einfache Prüfung“, „Uninteressantes Fach - Schwierige Prüfung“, „Uninteressantes Fach - Einfache Prüfung“). Mit diesen Fragen lässt sich unterscheiden, inwiefern die Akzeptanz aktivierender Methoden von den Rahmenbedingungen abhängig ist.

Die Skalen zur Lernmotivation wurden durch eine Skala zur *sozialen Eingebundenheit* ergänzt. Damit sollte die Möglichkeit ausgeschlossen werden, dass eine geringe Akzeptanz aktivierender Methoden durch eine geringe soziale Eingebundenheit erklärt werden kann. Die hier eingesetzte Skala von drei Items wurde ohne empirische Vorabprüfung eingesetzt.

Die Skala *Probleme bei der Umsetzung aktivierender Methoden* wurde ebenfalls beibehalten, auch wenn nicht alle zehn Items gute Trennschärfen erzielten. Diese Entscheidung fiel aufgrund der Annahme, dass die umfassende Aufzählung möglicher Probleme die Konstruktvalidität der Skala erhöht.

8.2 Hauptstudie

Ziel der Hauptstudie war es, zu überprüfen, ob die Studierenden der Humanmedizin aktivierenden Methoden positiv gegenüberstehen und diese als hilfreich für ihr Studium betrachten. Diese Studie sollte die Wahrnehmungen der Studierenden differenziert untersuchen und daraus praktische Schlüsse für die optimale Umsetzung aktivierender Methoden abgeleitet werden.

8.2.1 Hypothesen

Die genauen Hypothesen der Untersuchung lauteten folgendermaßen:

H1: Die Studierenden weisen auf der Skala *Akzeptanz aktivierender Methoden in Vorlesungen* Werte signifikant größer 3.0 auf.

H2: Die Studierenden weisen auf der Skala *Aktivierende Methoden in Modulen* Werte signifikant größer 3.0 auf.

Die Hypothesen 1 und 2 wurden gegen den numerischen Mittelwert der Skalen getestet, da hier keine Referenzwerte aus der Literatur vorliegen. Das Signifikanzniveau wurde vorab mit $p \leq .05$ festgesetzt. Vertiefend sollte explorativ untersucht werden, ob in Abhängigkeiten der Fächerkombinationen Unterschiede bei der Skala *Aktivierende Methoden in Modulen* auftreten.

H3: Die Studierenden weisen auf der Skala *Aktivierende Methoden in Vorlesungen und Seminaren* Werte signifikant größer 25 % auf.

Akzeptanz aktivierender Methoden liegt also hier vor, wenn entsprechende didaktische Handlungen in mehr als einem Viertel der Lehrveranstaltungszeit gewünscht werden. Dieser Testwert ist weniger streng formuliert als bei den Hypothesen 1 und 2. Auch wenn Wahl (2006) einen etwa gleichwertigen Einsatz aktivierender Methoden fordert, sind effektive Aktivierungen in der Regel kürzer (vgl. Kapitel 3.5). Das Signifikanzniveau wurde vorab mit $p \leq .05$ festgesetzt.

Weiterhin sollten die Skalen zur Akzeptanz aktivierender Methoden untereinander moderat positiv korrelieren. Wahrgenommene Probleme bei der Umsetzung sollten negativ mit der Akzeptanz korrelieren. Die Hypothese 4 lautete daher:

H4: Die Skalen zur Akzeptanz aktivierender Methoden weisen untereinander mittlere bis hohe Korrelationen auf.

Explorativ sollte zudem untersucht werden, wie die Korrelationen der Skalen zu der Akzeptanz aktivierender Methoden mit den lernmotivationalen Variablen ausfallen. Die Ergebnisse zu wahrgenommenen Problemen bei der Umsetzung aktivierender Me-

thoden sollten weiterhin Hinweise geben, wie die Effektivität aktivierender Methoden an der Medizinischen Fakultät gesteigert werden kann.

8.2.2 Methodik

Material. Tabelle 47 zeigt die verwendeten Skalen im Überblick. Die vollständige Auflistung aller Items findet sich in Anhang F1. Zusätzlich wurde die Akzeptanz des Fragebogens überprüft („Wie ist Ihre Meinung zu diesem Fragebogen?“). Die Befragung erfolgte online über den Anbieter *Q-Set*. In der Befragung gab es keine Pflichtfragen.

Tabelle 47. *Skalen in der Studierendenbefragung*

Skalen
Akzeptanz aktivierender Methoden
Akzeptanz aktivierender Methoden in Vorlesungen ^a
Aktivierende Methoden in Vorlesungen und Seminaren ^b
Aktivierende Methoden in Modulen ^a
Probleme bei der Umsetzung aktivierender Methoden ^c
Lernmotivation
Zielorientierungen und Arbeitsvermeidung ^a (vgl. SELLMO-ST; Spinath, et al, 2002)
Interesse ^a (Fragebogen zur Studienzufriedenheit; Westermann, Heise, Spies und Trautwein; 1996)
Soziale Eingebundenheit im Studium ^a (angelehnt an Müller, 2001)
Soziodemographische Fragen
Studienabschnitt
Noten im Studium
Geschlecht

Anmerkungen. Abstufung der Antwortskalen: a = fünffach, b = Prozentzahlen zwischen 0 und 100, c = vierfach.

Der Hintergrund der Befragung wurde im Anschreiben kurz erläutert (Anhang A5). Damit Studierende auch an der Befragung teilnehmen konnten, ohne ihre Daten einer wissenschaftlichen Verwertung in dieser Arbeit zu ermöglichen, konnten sie ihr Einverständnis zur wissenschaftlichen Nutzung der Daten am Ende der Befragung widerrufen. Weiterhin wurde zur Steigerung der Rücklaufquoten eine spätere Erinnerung zur Teil-

nahme versendet, der Fragebogen über einen längeren Zeitraum geöffnet, die Anonymität gewährleistet und die Befragung kurz gehalten. Die Befragung enthielt auch offene Fragen, um individuelle Meinungen der Studierenden zu ermöglichen. Diese sind nicht Gegenstand dieser Arbeit.

Stichprobenbeschreibung. Die Befragung der Studierenden erfolgte zwischen Mai und Juli 2012. Dazu wurden alle eingeschriebenen Studierenden der Humanmedizin durch das Studiendekanat angeschrieben. Insgesamt wurde der Link zur Befragung 686-mal geöffnet. Davon haben 56 Studierende die Teilnahme nach der ersten Seite abgebrochen. Die Studierenden ($n = 14$), die sich nicht mit der wissenschaftlichen Verwertung der Daten einverstanden erklärten, wurden von der weiteren Auswertung ausgeschlossen. Die Stichprobe umfasste schließlich 375 Studierende, der Rücklauf beträgt damit 17 %. Um zu überprüfen, wie die Repräsentativität der Gelegenheitsstichprobe einzuschätzen ist, wurden drei Merkmale erhoben: Geschlecht, Studienabschnitt und Noten im Studium (Tabelle 48). Die drei vorgegebenen Antwortalternativen zu der Akzeptanz des Fragebogen (*Gut*, *Mäßig*, *Schlecht*) zeigten bei 359 Antworten folgende Verteilung auf: *Gut* mit 67.4 %, *Mäßig* mit 31.5 % und *Schlecht* mit 1.1 %.

Tabelle 48. *Demographische Angaben der Studierenden*

Noten	Geschlecht		Gesamt
	Weiblich (<i>n</i> = 203)	Männlich (<i>n</i> = 168)	
1.0 – 1.4	18	19	37
1.5 – 1.9	62	61	123
2.0 – 2.4	49	40	89
2.5 – 2,9	24	14	38
3.0 – 3.4	9	3	12
3.4 oder höher	1	1	2
Gesamt	163	138	301 ^a
Vorklinik	70	75	145
Propädeutik	4	5	9
Klinik	98	72	170
PJ	28	16	44
Gesamt	200	168	368 ^b

Anmerkungen. Noten = Selbst eingeschätzter Notenschnitt im Studium.

PJ = Praktisches Jahr. Fehlende Werte: a = 74, b = 7.

Der Anteil weiblicher Studierender von 54 % entsprach der zugrundeliegenden Verteilung im Sommersemester 2012, nach der 53 % der Studierenden weiblich waren (Ruprecht-Karls-Universität Heidelberg, 2012). Die Mehrheit der Studierenden gaben gute Noten im Studium an. Hier liegen keine exakten Vergleichswerte für Noten im Studium vor. Vor dem Hintergrund, dass Studierende der medizinischen Fakultät Heidelberg bei zentralen Abschlussprüfungen durchschnittlich gute Ergebnisse erzielen (Institut für medizinische und pharmazeutische Prüfungsfragen, 2014), werden gute Leistungen im Studium hier als Indikator für Repräsentativität gewertet. Die Verteilung der Studierenden auf die einzelnen Studienabschnitte entspricht etwa deren Umfang im Studium. Dementsprechend ist der klinische Abschnitt mit fünf Fachsemestern am stärksten vertreten. Insgesamt kann hier also von einer guten Repräsentativität ausgegangen werden.

8.2.3 Deskriptive Ergebnisse

Deskriptive Statistiken. Die Werte der drei Skalen zur Akzeptanz aktivierender Methoden liegen unter den numerischen Mittelwerten der Skalen (Tabelle 49). Die Skala *Aktivierende Methoden in Vorlesungen und Seminaren* setzt sich aus der Summe von vier verschiedenen Aspekten studierendenzentrierter Lehre zusammen. Den höchsten Anteil an der Summe hat das Item „Die Lehrenden stellen Fragen und diskutieren mit den Studierenden“ mit 22.7 % (siehe Anhang F2).

Tabelle 49. *Deskriptive Statistik Studierendenbefragung*

Skala (Anzahl der Items)	<i>N</i>	<i>M</i>	<i>Md</i>	<i>SD</i>	Schiefe	Kurtosis	<i>r_{it}</i> Bereich	α
AM-VS (4)	374	46.59	50.00	22.70	0.18	-0.70	-	-
AM-M (4)	352	2.75	2.75	0.83	-0.08	-0.19	.36 – .50	.64
AM-V (8)	371	2.63	2.63	0.86	0.27	-0.37	.37 – .68	.83
AM-P (10)	365	2.81	2.80	0.50	-0.47	0.37	.10 – .58	.74
LO (4)	375	4.43	4.50	0.52	-0.93	0.58	.40 – .48	.64
AL (3)	373	3.29	3.33	0.89	-0.06	-0.61	.26 – .61	.65
VL (3)	373	2.28	2.00	0.95	0.49	-0.60	.59 – .69	.77
AV (3)	373	2.26	2.33	0.90	0.62	0.28	.79 – .82	.90
Interesse (3)	374	4.23	4.33	0.70	-0.75	0.07	.69 – .77	.84
SE (3)	374	4.06	4.33	0.95	-0.99	0.52	.80 – .84	.90

Anmerkungen. r_{it} = korrigierte Trennschärfe, α = Cronbachs Alpha. AM-VS = Aktivierende Methoden in Vorlesungen und Seminaren, AM-M = Aktivierende Methoden in Modulen, AM-V = Aktivierende Methoden in Vorlesungen, AM-P = Probleme bei der Umsetzung aktivierender Methoden, LO = Lernzielorientierung, AL = Annäherungs-Leistungszielorientierung, VL = Vermeidungs-Leistungszielorientierung, AV = Arbeitsvermeidung, SE = Soziale Eingebundenheit.

Trennschärfe und interne Konsistenz. Die Trennschärfen der Items liegen alle im akzeptablen Bereich. Auch die interne Konsistenz der Skalen ist für den Zweck dieser Studie akzeptabel. Die Maßnahme, der Skala zur *Lernzielorientierung* ein Item hinzuzufügen, hat zu einer etwas höheren internen Konsistenz im Vergleich zu der Vorstudie geführt.

Prüfung der Normalverteilung. Die Überprüfung auf Normalverteilung beinhaltete Vergleiche von Mittelwert und Median, Analyse von Schiefe und Kurtosis sowie Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk. Bei den Vergleichen von Mittelwert und Median (Tabelle 49) ergaben sich Abweichungen bei den Skalen *Vermeidungs-Leistungszielorientierung* und *Soziale Eingebundenheit*. Die Werte von Schiefe und Kurtosis sind kleiner 1 und können daher nach Miles und Shevlin (2001) als nur moderat abweichend von der Normalverteilung interpretiert werden. Die Tests auf Normalverteilung nach Kolmogorov-Smirnov und Shapiro-Wilk zeigten für alle Skalen signifikante Abweichungen von der Normalverteilung auf. Insgesamt traten, insbesondere bei den motivationalen Variablen, unsymmetrische Werteverteilungen auf. Die Skalen zu der *Akzeptanz aktivierender Methoden* wiesen moderate Abweichungen auf. Deckeneffekte traten bei der *Lernzielorientierung*, *Interesse* und *sozialer Eingebundenheit*, Bodeneffekte bei *Vermeidungs-Leistungszielorientierung* und *Arbeitsvermeidung* auf. Da parametrische Tests bei größeren Fallzahlen robust gegenüber moderaten Abweichungen von der Normalverteilung sind (Bühner & Ziegler, 2009), fanden sie hier Anwendung.

Interkorrelationen. Hypothesenkonform zeigten sich positiv signifikante Korrelationen unter den Skalen, welche die Akzeptanz aktivierender Methoden messen (Tabelle 50). Diese fallen nach Cohen (1988) hoch aus. Ebenso hypothesenkonform sind die negativen Korrelationen zu der Intensität von erlebten *Problemen bei dem Einsatz aktivierender Methoden*. Weiterhin zeigen sich keine bedeutsamen Korrelationen zwischen der *Akzeptanz aktivierender Methoden* und lernmotivational bedeutsamen Variablen. Die Interkorrelationen von *Zielorientierungen*, *Arbeitsvermeidung*, *Interesse* und *sozialer Eingebundenheit* sind nach Cohen (1988) klein bis moderat.

Tabelle 50. *Interkorrelationen der Skalen in der Studierendenbefragung*

	AM-M	AM-V	AM-P	LO	VL	AL	AV	INT	SE
AM-VS	.60	.55	-.33	.02	.01	.09	.04	.04	.13
AM-M	-	.56	-.28	.05	.04	.07	.06	-.03	.07
AM-V		-	-.45	.05	.05	.12	.00	-.02	.09
AM-P			-	-.05	.10	-.08	.06	-.11	-.20
LO				-	-.06	.38	-.36	.27	.16
VL					-	.13	.24	-.16	-.18
AL						-	-.16	.17	.12
AV							-	-.30	-.12
INT								-	.31
SE									-

Anmerkungen. Fett markierte Werte zeigen signifikante Korrelationen ($p \leq .05$). AM-VS = Aktivierende Methoden in Vorlesungen und Seminaren, AM-M = Aktivierende Methoden in Modulen, AM-V = Aktivierende Methoden in Vorlesungen, AM-P = Probleme bei der Umsetzung aktivierender Methoden, LO = Lernzielorientierung, VL = Vermeidungs-Leistungszielorientierung, AL = Annäherungs-Leistungszielorientierung, AV = Arbeitsvermeidung, INT = Interesse, SE = Soziale Eingebundenheit.

8.2.4 Hypothesentestung

Die deskriptiven Ergebnisse (Tabelle 49) ließen bereits den Schluss zu, dass die Hypothese für zwei der drei Skalen verworfen werden muss. Tabelle 51 zeigt die Ergebnisse des Einstichproben-t-Tests. Die entgegengesetzten Abweichungen der Skalen *Aktivierende Methoden in Vorlesungen* sowie *Aktivierende Methoden in Modulen* unterscheiden sich signifikant zum numerischen Mittelwert. Studierende lehnen den Einsatz aktivierender Methoden in Vorlesungen durchschnittlich eher ab und wünschen sich in Modulen tendenziell eher Frontalunterricht als aktivierende Methoden. Die Prüfung der Skala *Aktivierende Methoden in Vorlesungen und Seminaren* gegen den Testwert zeigt dagegen ein hypothesenkonform signifikantes Ergebnis. In Vorlesungen und Seminaren sind aktivierende Methoden zu einem Anteil von mindestens 25 % akzeptiert.

Tabelle 51. *Ergebnisse der Signifikanztestungen (Studierendenbefragung)*

	<i>df</i>	<i>t</i>	<i>p</i>
AM in Vorlesungen	370	-8.24	.000
AM in Modulen	351	-5.65	.000
AM in Vorlesungen und Seminaren	373	18.40	.000

Anmerkungen. AM = Aktivierende Methoden.

Die Hypothesen 1 und 2 mussten daher verworfen werden, was inhaltlich bedeutet, dass die Studierenden sich nicht mehrheitlich aktivierende Methoden in der Lehre wünschen, weder in ganzen Modulen noch in Vorlesungen und Seminaren. Lediglich Hypothese 3 stand in Einklang mit den Daten: Studierende wünschen sich einen bedeutsamen Anteil aktivierender Methoden in Lehrveranstaltungen mit mehr als 30 Personen.

Als Vertiefung zur Hypothese 2 wurde geprüft, unter welchen Umständen die Studierenden eher aktivierende Lehrmethoden bevorzugen. Dazu wurden zunächst die deskriptiven Ergebnisse berechnet (Tabelle 52).

Tabelle 52. *Akzeptanz aktivierender Methoden nach Fachmerkmalen*

Items	<i>M</i>	<i>SD</i>	<i>N</i>
Interessantes Fach - Einfache Prüfung	3.16	1.19	356
Uninteressantes Fach - Einfache Prüfung	2.71	1.33	344
Interessantes Fach - Schwierige Prüfung	2.58	1.07	353
Uninteressantes Fach - Schwierige Prüfung	2.55	1.19	349

Um die Bedeutsamkeit der Unterschiede zu überprüfen, wurde eine einfaktorielle, univariate Varianzanalyse mit Messwiederholung berechnet. Die Faktorstufe Zeit wurde hier durch die vier verschiedenen Merkmalskombinationen definiert. Als Voraussetzungsprüfung für die Varianzanalyse, wurde der Mauchly-Test auf Sphärizität durchgeführt. Da hier inhomogene Varianzen der Differenzen zwischen den Faktorstufen vorliegen ($\chi^2(5) = 133.32, p = .000$) wurden zur Analyse der Haupteffekte die korrigierten Werte nach Greenhouse-Geisser herangezogen (Rasch, Frieze, Hofmann & Naumann, 2006). Die Varianzanalyse mit Messwiederholung zeigte ein signifikantes Ergebnis ($F(2.5) = 26.95, p = .000$) für die Merkmalskombinationen. Somit unterscheidet sich die Akzeptanz aktivierender Methoden je nach wahrgenommenem Schwierigkeitsgrad der Prüfung und dem subjektiven Interesse. Zur Einschätzung der praktischen Bedeutsamkeit der aufgetretenen Unterschiede wurde das partielle Eta-Quadrat (η^2) berechnet. Der hier gefundene Wert $\eta^2 = .07$ entspricht nach Cohen (1988) einem mittleren Effekt. Um zu analysieren, welche der Merkmalskombinationen signifikante Unterschiede aufzeigen, wurden paarweise Vergleiche berechnet (korrigiert nach Bonferroni). Die Testung erfolgte ungerichtet, da hier keine Hypothesen vorlagen. Die Kombination *Interessantes Fach - Einfache Prüfung* unterschied sich signifikant ($p = .000$) von den anderen Kombinationen, untereinander unterschieden sich die anderen Kombinationen nicht signifikant. Die Antwort auf die Fragestellung lautet demnach, dass Studierende dann aktivierende Methoden bevorzugen, wenn sie ein Fach interessant finden und die abschließende Prüfung als einfach eingeschätzt wird.

8.2.5 Diskussion

Hypothesentestung. Die Annahme, dass Studierende der Humanmedizin aktivierenden Methoden in Veranstaltungen mit höherer Teilnehmerzahl eher positiv gegenüberstehen, musste für zwei der drei eingesetzten Skalen verworfen werden. Die Ergebnisse der Skalen *Aktivierende Methoden in Vorlesungen* sowie *Aktivierende Methoden in Modulen* lagen unterhalb des numerischen Mittelwertes. Die Annahme, dass die Mehrheit der Studierenden sich mehrheitlich aktivierende Methoden in der Lehre wünscht, wurde dadurch widerlegt. Bei der Prozentsummenskala *Aktivierende Methoden in Vorlesungen und Seminaren* zeigte sich dagegen, dass die Studierenden im Mittel etwa 47 % der Zeit in Großgruppenveranstaltungen für aktivierende Methoden befürworten. Damit wünschen sich die Studierenden zumindest anteilig den Einsatz aktivierender Lehrmethoden in Großgruppenveranstaltungen. Offen bleibt, wie dieses Ergebnis bewertet werden kann, da keine empirischen Referenzwerte in der Literatur zu dem Anteil aktivierender Methoden vorliegen (vgl. Kapitel 3.5). Möglicherweise waren die hier herangezogenen Testwerte zu streng gewählt. So bedeutet bei der Frage zu aktivierenden Methoden in Modulen ein Mittelwert von 3,0, dass die Studierenden sich durchschnittlich einen genau so hohen Anteil von Frontalunterricht und aktivierenden Methoden wünschen. Es ließe sich einwenden, dass auch schon etwas geringere Mittelwerte Ausdruck einer Akzeptanz aktivierender Methoden darstellen.

Weiterhin zeigte sich, dass die Akzeptanz aktivierender Methoden von dem subjektiven Interesse und der wahrgenommenen Schwierigkeit der Prüfung abhängig sein kann. Von den dargebotenen Fachkombinationen wies das Item „Interessantes Fach - Einfache Prüfung“ einen Wert oberhalb des Mittelwertes und eine signifikante Abweichung zu den anderen Items auf. Jedoch kann eine solche Fachbeschreibung subjektiv für ganz unterschiedliche Fächer innerhalb der Humanmedizin zutreffen. Diese Form der Erhebung wurde gewählt, da sie eine zeitökonomische Erhebung ermöglichte. Eine genauere Vorgehensweise wäre es gewesen, verschiedene Fächer anhand verschiedener Merkmale seitens der Studierenden bewerten zu lassen und darauf basierend die entsprechende Akzeptanz aktivierender Methoden aller Fächer festzustellen. Unklar bleibt also, inwiefern dieser Befund generalisiert werden kann. Wenn es generell so wäre, dass Studierende aktivierenden Methoden unter günstigen Randbedingungen offener gegenüberstehen, hätte dieses Ergebnis Relevanz für die formative Evaluation von der DOS. Lehrende aus entsprechenden Fächern könnten sich ermutigt fühlen, aktivierende Methoden einzusetzen. Welchen Schluss Lehrende aus Fächern mit ungünstigeren Randbe-

dingungen ziehen könnten, hängt von der Interpretation der Ergebnisse ab. Wenn Studierende aktivierende Methoden als ineffektiv für die Vorbereitung schwieriger Prüfungen betrachten, dann müsste auf diese verzichtet oder die Prüfungsrelevanz der Aktivierungen erhöht werden. Das geschieht dann, wenn Lehrende die Aktivierungen in der Lehre in Einklang mit Lernzielen und Prüfungsinhalten umsetzen und entspricht der Idee des Constructive Alignment (vgl. Kapitel 5.1). Wenn Studierende einem Fach mit eher geringem Interesse gegenüberstehen, könnten Lehrende Aktivierungen einsetzen, die die persönliche oder praktische Relevanz für die Studierenden aufzeigen. Generell sollten Aktivierungen möglichst professionell, effizient und interessant gestaltet werden, um mögliche Vorbehalte zu minimieren. Die negativen Korrelationen mit der Skala *Probleme bei der Umsetzung aktivierender Methoden* legen den Schluss nahe, dass die Ablehnung aktivierender Methoden auf Probleme bei der Umsetzung zurückzuführen ist. Dann würde sich die Akzeptanz aktivierender Methoden in Abhängigkeit der erlebten Umsetzungen im Studium entwickeln. Da hier nur eine Querschnittstudie vorlag, ist diese Annahme aber nicht kausal überprüfbar.

Stichprobe. Die hier untersuchte Stichprobe konnte anhand der erhobenen Merkmale als repräsentativ für die Studierenden in HeiCuMed betrachtet werden. Der Rücklauf der Studierendenbefragung war relativ gering, entsprach aber dem vergleichbarer Erhebungen (Ramm, 2014; Woisch, Willige & Grützmacher, 2014). Die Bereitschaft an Studierendenbefragungen teilzunehmen, ist in den vergangenen 20 Jahren kontinuierlich gesunken, Rücklaufquoten bei Online-Befragungen liegen selten höher als 20 % (Ramm, 2014). Offen bleibt, inwiefern die Einstellung zu aktivierender Lehre zu einer Selektivität bei der Teilnahme beigetragen hat. Da die Studie als Maßnahme zur Qualitätssicherung der Lehre beworben wurde, könnten sich insbesondere Studierende für die Teilnahme motiviert haben, die aktivierenden Methoden eher ablehnend gegenüberstehen.

Es liegen in der Literatur keine Hinweise dafür vor, dass Studierende der Humanmedizin aktivierenden Methoden kritischer gegenüberstehen als Studierende anderer Studiengänge. Sander et al (2010) fanden beispielsweise keine bedeutenden Unterschiede in der Präferenz aktivierender Methoden von Medizinstudierenden gegenüber Studierenden der Psychologie und der Wirtschaftswissenschaften. Es fällt in dieser Arbeit allerdings auf, dass die Ergebnisse der Skala *Aktivierende Methoden in Vorlesungen* sich deutlich zwischen den Studierenden der Humanmedizin und den Studierenden

in der Vorstudie (v.a. Lehramtstudierende) unterscheiden. Für weitere Vergleiche wäre der Einsatz der Skala an anderen Standorten und Fachbereichen aufschlussreich.

Psychometrische Eigenschaften der Skalen. Ein Ziel dieser Arbeit bestand darin, eine Skala zur Akzeptanz aktivierender Methoden zu entwerfen. Deren Einsatz in einer Studierendenbefragung sollte eine Beschreibung der Rahmenbedingungen der Lehre ermöglichen. Die in Kapitel 8.2.3 berichteten Ergebnisse zeigen, dass die Akzeptanz aktivierender Methoden mit der vorgelegten Skala reliabel gemessen werden kann und dass positive Hinweise auf Konstruktvalidität vorliegen. Die Trennschärfen und internen Konsistenzen der Skalen *Aktivierende Methoden in Vorlesungen* sowie *Aktivierende Methoden in Modulen*, können als zufriedenstellend betrachtet werden. Die Skala *Aktivierende Methoden in Vorlesungen und Seminaren* konnte, wegen der voneinander abhängigen Werte in dem Prozentsummenformat, nur anhand der Interkorrelation zu anderen Skalen überprüft werden. Die Interkorrelationen der drei Skalen untereinander waren hoch und standen somit in Einklang mit der formulierten Hypothese. Moderate negative Korrelationen traten erwartungsgemäß mit der Skala *Probleme bei der Umsetzung aktivierender Methoden* auf. Die Skala eignet sich somit für Untersuchungsfragen wie in dieser Arbeit, in denen Gruppenvergleiche oder Zusammenhangsanalysen angestrebt werden. Die Validität der Fragen wurde hier lediglich durch Korrelationen mit anderen Skalen festgestellt. Wünschenswert wäre eine Validierung durch Verhaltensbeobachtungen, beispielsweise müsste geprüft werden, ob Studierenden sich tatsächlich mehr an aktivierenden Methoden beteiligen, wenn sie eine hohe *Akzeptanz aktivierender Methoden* aufweisen.

Wie schon in der Vorstudie (Kapitel 8.1.2) zeigten sich keine bedeutsamen und konsistenten Korrelationen zwischen der Akzeptanz aktivierender Methoden und lernmotivational bedeutsamen Variablen. Dies kann ebenfalls als Beleg für die diskriminante Validität der Skala gewertet werden. Doch auch in dieser Befragung schränken Deckeneffekte (erneut bei den Skalen Lernzielorientierung und Interesse), Abweichungen von der Normalverteilung sowie die Zusammensetzung der Gelegenheitsstichprobe, die Varianz der Ergebnisse ein. Wenn jedoch von einer annähernden Repräsentativität der Stichprobe ausgegangen werden kann, zeigen die hier ermittelten Ergebnisse, dass die Akzeptanz aktivierender Methoden deutlich stärker mit der wahrgenommenen Qualität der Umsetzung und situativen Faktoren einhergeht, als mit dispositionalen Eigenschaften der Studierenden.

Angemessenheit des Fragebogens. Die insgesamt eher gute Bewertung des Fragebogens kann als Indikator für eine gute Akzeptanz und Verständlichkeit des Fragebogens gewertet werden. Kritisch muss der in einigen Freitexten erwähnte Einwand von Teilnehmern aufgegriffen werden, dass aktivierende Lehrmethoden und Frontalunterricht sehr heterogene Umsetzungen in der Praxis aufweisen. Studierende können unterschiedliche Vorstellungen von beiden Konzepten haben, die einen negativen Einfluss auf die Konstruktvalidität ausüben. Um die Generalisierbarkeit der Ergebnisse besser bewerten zu können, wären ergänzende Messungen denkbar. In zukünftigen Studien könnten konkrete Beispiele für verschiedene Umsetzungen aktivierender Methoden von den Studierenden auf eine mögliche Akzeptanz hin bewertet werden. Offen bleibt dabei, welcher Zusammenhang zwischen der Bewertung aktivierender Methoden, der tatsächlichen Teilnahme an entsprechenden Aktivitäten im Kontext Lehre, und dem tatsächlichen Lernerfolg besteht. Auch dies müsste genauer untersucht werden.

9 **Zusätzliche Erhebung: Ziele der Dozentenschulung aus Sicht der Trainer**

Die Struktur der DOS mit verschiedenen Inhalten, verschiedenen Trainern und der individuellen Arbeit der Lehrenden in unterschiedlichen Kleingruppen (siehe Kapitel 5.1) könnte mit unterschiedlichen Schwerpunktsetzungen seitens der Trainer einhergegangen sein. Daher sollte hier geprüft werden, inwiefern die durch den Autor identifizierten Ziele der DOS von den Trainern ebenfalls als solche betrachtet werden. Die Trainerbefragung fand statt, als die Lehrendenbefragung (vgl. Kapitel 6) und die Beobachtungsstudie (vgl. Kapitel 7.3) bereits durchgeführt wurden.

9.1 Fragestellung

Die Hypothesen umfassten zwei Bereiche. Für die Selbsteinschätzung der Lehrenden bestand die Annahme, dass die Trainer eine größere Wirkung auf die studierendenzentrierte Lehre und Lernziele in der Lehre erwarten, als auf allgemeine didaktische Handlungen. Für die Beobachtung der Lehrenden bei der Lehre bestand die Annahme, dass die Trainer eine größere Wirkung auf individuelle Lern- und Arbeitsphasen, Lernziele und Fragen der Lehrenden erwarten, als auf allgemeine didaktische Handlungen.

9.2 Methodik

Es wurden die Trainer befragt, die in den Jahren 2011 und 2012 bei der DOS mindestens zwei verschiedene Themen betreut hatten. Der Autor und die Betreuerin dieser Arbeit (selbst Trainer bei der DOS) nahmen nicht an der Befragung teil. Insgesamt haben elf von insgesamt dreizehn Trainern der DOS teilgenommen (sieben waren weiblich, vier männlich). Die schriftliche Befragung erfolgte im Oktober 2012.

Grundlage der Befragung waren die Items der Lehrendenbefragung zum eigenen Lehrverhalten (vgl. Kapitel 6) und der Verhaltensbeobachtung (vgl. Kapitel 7). Den Trainern wurden die Listen der Items vorgelegt (Anhang G1). Sie sollten die Items einzeln auf einer Skala von -3 bis +3 dahingehend bewerten, inwiefern ein Verhalten nach der DOS häufiger (positive Werte) oder seltener (negative Werte) auftritt. Hohe Werte sollten einen stärkeren Effekt kennzeichnen, der mittlere Skalenwert Null sollte keinem Effekt entsprechen. Alle Trainer haben die Listen vollständig bearbeitet. Die vorliegenden Items wurden danach zu Skalen zusammengefasst (siehe Anhang G2). Die allgemeinen didaktischen Handlungen umfassten heterogene Items wie „Ich unterbinde auftretende Störungen konsequent“ und „Ich strukturiere die Inhalte klar“.

9.3 Ergebnisse

Wie in Tabelle 53 dargestellt, entsprachen die Ergebnisse weitgehend den Annahmen. Die Trainer erwarteten sowohl bei der Selbsteinschätzung als auch bei der Beobachtung der Lehrenden durchschnittlich eine größere Wirkung auf die studierendenzentrierte Lehre und Lernziele als auf allgemeine didaktische Handlungen. Insgesamt zeigten sich konsistent positive Werte, was bedeutet, dass die Trainer bei allen Variablen positive Effekte nach der DOS erwarteten. In Anhang G2 sind die Ergebnisse der einzelnen Items aufgelistet.

Tabelle 53. *Annahmen der Trainer zur Effektivität der Dozentenschulung (DOS).*

Verhalten aus Sicht der Lehrenden	<i>M</i>	<i>SD</i>	Min	Max
Studierendenzentrierte Lehre (7)	1.35	0.27	0.71	1.71
Lernziele in der Lehre (2)	1.96	0.69	0.50	2.50
Allgemeine didaktische Handlungen (9)	0.93	0.29	0.44	1.44
Verhalten aus Sicht von Beobachtern	<i>MW</i>	<i>SD</i>	Min	Max
Individuelle Lern- und Arbeitsphasen (1)	2.18	0.60	1	3
Lernziele (1)	2.27	0.65	1	3
Fragen der Lehrenden (5)	1.13	0.45	0.40	1.60
Allgemeine didaktische Handlungen (8)	1.14	0.33	0.50	1.75

Anmerkung. In Klammern steht die Anzahl der zugrundeliegenden Items. Skala von -3 (Verhalten tritt nach der DOS seltener auf) bis +3 (Verhalten tritt nach der DOS häufiger auf). Min = Maximaler durchschnittlicher Wert eines Trainers auf der Skala/dem Item. Max = Minimaler durchschnittlicher Wert eines Trainers auf der Skala/dem Item.

9.4 Diskussion

Die Trainer erwarteten nach der Dozentenschulung positive Effekte. Sie nahmen an, dass die Lehrenden häufiger berichten, studierendenzentrierte Lehre umzusetzen und ihre Lehre stärker an den Lernzielen auszurichten. Sie erwarteten weiterhin, dass der Einsatz individueller Lern- und Arbeitsphasen und die Nennung von Lernzielen häufiger nach der DOS zu beobachten sein würde, ebenso, dass Lehrende häufiger Fragen stellen. Diese Ergebnisse stehen in Einklang mit den Inhalten der Dozentenschulung. Gleichzeitig wurde deutlich, dass die Trainer auch eine Wirkung der DOS auf lehrendenzentrierte Handlungen wie „Ich strukturiere die Inhalte klar“ oder „Ich fasse wichtige Inhalte zusammen“ erwarten. Übungseinheiten in der DOS wie das Micro-Teaching, mit den Schwerpunkten Strukturierung und Rhetorik, stärken auch Aspekte lehrendenzentrierter Lehre (Hattie, 2009). Die tendenziell positive Bewertung aller möglichen Erfolgskriterien zeigt, dass die DOS aus Sicht der Trainer eine breite Wirkung zu erzielen soll. Das unterscheidet sich von den Annahmen anderer Weiterbildungs-evaluatio-nen. So wird in vergleichbaren Studien eine Reduzierung lehrendenzentrierter Lehran-sätze erwartet (z. B. Gibbs & Coffey, 2004; Postareff, et al., 2007), wenn diese mit dem *ATI* bzw. *ATI-R* gemessen werden. Offen bleibt hier, inwiefern sich die Inhalte der an-deren Programme unterscheiden, oder ob unterschiedliche Operationalisierungen von Lehransätzen zu anderen Erwartungen bezüglich der Wirksamkeit führen (zu den Ope-rationalisierungen in dieser Arbeit siehe Kapitel 6.2 und 7.2.1).

Die vorgelegten Items beinhalteten ausschließlich verhaltensbasierte Erfolgssin-dikatoren. Allgemeine Erfolgskriterien (wie didaktische Kompetenz oder Selbstwirk-samkeitserwartungen der Lehre), didaktisches Wissen und Wirkungen auf organisatori-scher Ebene wurden hier nicht berücksichtigt. Eine Limitation dieser Erhebung war zu-dem die Fokussierung auf den zu erwartenden Effekt nach der DOS, ohne Annahmen dazu, welche Ausprägung bei den Lehrenden vor der DOS bestand. Die hier berichteten Ergebnisse lassen sich daher nicht exakt in zu erwartende Effektstärken übertragen. Die positiven Werte legen aber den Schluss nahe, dass die Trainer mindestens kleine Effekt-stärken erwartet haben. Die Annahmen der Trainer werden in der nun folgenden, ab-schließenden Diskussion in Beziehung zu den vorliegenden Ergebnissen gesetzt.

10 Gesamtdiskussion

Zusammenfassung. In diesem Kapitel werden die wichtigsten Ergebnisse zusammengefasst und die Ziele, Wirkungen und Rahmenbedingen zueinander in Beziehung gesetzt. Hochschuldidaktische Weiterbildungen gelten als wirksam darin, die Qualität der Lehre zu steigern (Steinert et al. 2006; Stes et al., 2010). Bisher lagen jedoch kaum Befunde dazu vor, wie Weiterbildungen das Lehrverhalten verändern und welche Langzeitwirkungen Weiterbildungen haben. Die DOS der Medizinischen Fakultät Heidelberg ist mit 120 Unterrichtseinheiten, 13 Präsenzterminen und 13 verschiedenen Themengebieten eine inhaltlich und zeitlich umfassende Weiterbildung. Für die vorliegende Programmevaluation wurden der Transfer der Inhalte in die Praxis und die Langzeitwirkungen der DOS untersucht. Die DOS sollte primär dazu führen, dass die Lehrenden studierendenzentrierte Lehre umsetzen und die Lernziele in ihren Lehrveranstaltungen transparent machen.

Es wurden verschiedene Studien durchgeführt: Eine Fragebogenstudie zur Selbsteinschätzung der Lehrenden (Kapitel 6), eine Beobachtungsstudie zur Erfassung des Lehrverhaltens (Kapitel 7.3), eine Studierendenbefragung zur Wahrnehmung der Lehre (Kapitel 8) und eine Trainerbefragung zur Identifikation der wichtigsten Ziele der DOS (Kapitel 9). In der Lehrendenbefragung und der Beobachtungsstudie wurden zwei Versuchsgruppen untersucht, einmal Lehrende, die sich in der DOS befinden (Beobachtungsstudie) bzw. Lehrende, die die DOS unmittelbar vorher abgeschlossen haben (Lehrendenbefragung) und andererseits Lehrende, deren Besuch der DOS länger zurücklag. Beide Versuchsgruppen wurden jeweils mit einer Gruppe von Lehrenden verglichen, welche (noch) nicht an der DOS teilgenommen haben (Vergleichsgruppe). Die Trainerbefragung (Kapitel 9) hat gezeigt, dass die aus den Inhalten der DOS abgeleiteten Hypothesen in Einklang mit den Erwartungen der Trainer standen. Die Vorstudie der Beobachtungsstudie (vgl. Kapitel 7.1.3) zeigte, dass die zentralen Verhaltensweisen, die durch die DOS gestärkt werden sollten, in Vorlesungen und Seminaren generell nur selten umgesetzt werden.

In der Beobachtungsstudie (vgl. Kapitel 7.3.3) zeigten sich kleine bis hohe Effektstärken für alle Indikatoren studierendenzentrierter Lehre. Die Transparenz der Lernziele wies nur bei Lehrenden in der DOS einen praktisch bedeutsamen Effekt (mit einer kleinen Effektstärke) auf. Die Lehrendenbefragung (vgl. Kapitel 6.3.3) zeigte eine höhere Studierendenzentrierung bei beiden Versuchsgruppen. Lehrende, deren Besuch der DOS länger zurücklag, gaben häufiger an, ihre Lehre an den Lernzielen zu orientie-

ren, während sich bei Lehrenden unmittelbar nach Besuch der DOS kein Effekt feststellen ließ. Darüber hinaus ging die Teilnahme an der DOS mit einer höheren Einschätzung der Lehrkompetenz einher. Die Annahme, dass die DOS zu einer Steigerung der Selbstwirksamkeitserwartungen bezüglich der Lehre führt, musste jedoch verworfen werden. Die Teilnehmer waren mit der DOS zufrieden und äußerten hohe Weiterbildungsbereitschaft. Insgesamt hat die vorliegende Evaluationsstudie die Wirksamkeit der DOS nachgewiesen. Auf allen Zielebenen zur Evaluation einer hochschuldidaktischen Weiterbildung (vgl. Kapitel 4.1) wurden in dieser Arbeit hypothesenkonforme Ergebnisse festgestellt (Tabelle 54).

Tabelle 54. Zusammenfassung der Ergebnisse der Lehrendenbefragung (LB) und der Beobachtungsstudie (BS).

Abhängige Variable	Zielebene	Studie	Ergebnisse	
			VSG1	VSG2
Zufriedenheit mit der DOS	Reaktion	LB	+	+
Studierendenzentrierung	Lernen (Einst.)	LB	+	+
Lehrkompetenz	Lernen (Einst.)	LB	+	+
Selbstwirksamkeitserwartungen	Lernen (Einst.)	LB	-	-
ILAP (Auftreten)	Verhalten	BS	+	+
ILAP (Dauer)	Verhalten	BS	+	+
Wortmeldungen von Studierenden	Verhalten	BS	+	+
Fragen der Lehrenden	Verhalten	BS	+	+
Studierendenzentrierung	Verhalten	LB	+	-
Transparenz der Lernziele	Verhalten	LB	-	+
Nennung der Lernziele	Verhalten	BS	+	-
Weiterbildungsbereitschaft	Resultate	LB	+	+

Anmerkungen. Zielebenen nach Steinert et al. (2006). Ergebnisse: Hypothese in Einklang mit den Daten (+) oder Hypothese widerlegt (-). VSG1 = Versuchsgruppe 1 (Lehrende während oder unmittelbar nach Besuch der DOS) und VSG2 (Lehrende, dessen Besuch der DOS mindestens ein Jahr zurückliegt) gegenüber der Vergleichsgruppe. ILAP = Individuelle Lern- und Arbeitsphasen. Einst. = Einstellungen.

Studierendenzentrierte Lehre wurde in dieser Arbeit, in Übereinstimmung mit anderen Autoren (Lübeck, 2009; Pratt, 2002; Reinmann & Mandl, 2006), als Ergänzung zu leh-

rendenzzentrierter Lehre aufgefasst, nicht als Widerspruch. Weder in der Lehrendenbefragung (vgl. Kapitel 6.3) noch in der Beobachtungsstudie (vgl. Kapitel 7.3.3) zeigten sich Unterschiede bei den untersuchten lehrendenzentrierten didaktischen Handlungen. Die intendierten Wirkungen der DOS wurden also erzielt, ohne dass bewährte lehrendenzentrierte Elemente an Bedeutung verloren haben.

Diskussion. Insgesamt entsprechen die gefundenen moderaten Effektstärken den Ergebnissen aus der Literatur (Steinert et al., 2006; Stes et al., 2010). In dieser Arbeit konnten Effekte allerdings an einer größeren Stichprobe anhand von Beobachtungsdaten nachgewiesen werden. Zu den Langzeiteffekten hochschuldidaktischer Weiterbildungen lagen bislang lediglich qualitative Analysen vor (Stes et al., 2007; Stewart, 2014). Die Effekte, die auf die DOS zurückgeführt wurden, haben zwar praktische Relevanz, sind aber eher gering. Der überwiegende Anteil der Lehrenden in den Versuchsgruppen setzte in der Beobachtungsstudie keine individuellen Lern- und Arbeitsphasen um. Wenn doch, war deren Dauer gering (insgesamt 6 % bzw. 4 % der beobachteten Zeit). Damit liegen die Ergebnisse deutlich unter den von Wahl (2013) mindestens empfohlenen 50 % und unter dem Anteil von 10-20 % in Studien zur Wirksamkeit aktivierender Methoden (vgl. Kapitel 3.5). Dafür kommen verschiedene Gründe in Betracht: Lehrende der Medizinischen Fakultät Heidelberg greifen regelmäßig auf vorgegebene Lehrveranstaltungsinhalte zurück und betrachten ihre Freiheit bei der Gestaltung der Lehre dann als eingeschränkt. Des Weiteren berichteten Lehrende, dass eine erhöhte Vorbereitungszeit ein Hindernis für die Umsetzung individueller Lern- und Arbeitsphasen darstellt. Die Zeit für die Lehre wird stark durch die weiteren Aufgaben der Lehrenden in der Humanmedizin eingeschränkt (vgl. Kapitel 2.2). Dass die Teilnahme an der DOS für die meisten Teilnehmer verpflichtend war, hat sich eventuell auch negativ auf die Umsetzung studierendenzentrierter Lehre ausgewirkt. Als weiteres Hindernis für den Einsatz studierendenzentrierter Lehre wurde angeführt, dass diese einen Zeitverlust für die Vermittlung von weiteren Inhalten bedeutet. Vereinzelt wurde auch eine theoretische Grundlage der DOS in Frage gestellt. Manche Lehrende sind der Meinung, dass Lehrveranstaltungen wie Vorlesungen und Seminare klassisch unterrichtet werden sollten, mit Frontalunterricht und gelegentlichen Fragen durch die Lehrenden. Diese Ansicht ist verständlich vor dem Hintergrund, dass Lehrende dazu neigen, die Lehre fortzuführen, die sie selbst erlebt haben (vgl. Kapitel 2.1) und die Lehrenden in dieser Studie mehrheitlich vor der Einführung neuer Lehr- und Lernformate studiert haben. Wahl (2013) hat sogar bei „der Mehrzahl der Lehrenden“ (S.108) eine Ablehnung gegenüber aktivie-

renden Lehrmethoden festgestellt, häufig mit der Begründung, diese seien nicht angemessen für erwachsene Studierende.

Die Studierenden der Humanmedizin in Heidelberg bevorzugen tatsächlich lehrendenzentrierte Lehre, wünschen sich aber auch studierendenzentrierte Anteile, sowohl in Vorlesungen als auch in ganzen Unterrichtsmodulen (vgl. Kapitel 8.2). Aus Sicht der Studierenden treten bei der Umsetzung aktivierender Methoden allerdings Probleme auf: Uninteressante Aufgaben, Unlust der Studierenden, unklare Instruktionen oder unpassende Aufgabenschwierigkeiten wurden von vielen Studierenden beobachtet. Möglicherweise geht dies mit einer geringeren Akzeptanz und Wirkung studierendenzentrierter Lehre einher. Andrews et al. (2011) haben in ihrer Studie zu der Wirksamkeit aktivierender Methoden festgestellt, dass viele Lehrende aktivierende Methoden nicht effektiv umsetzen und vermutlich aus diesem Grund positive Effekte ausgeblieben sind. Die Ergebnisse der Studierendenbefragung weisen darauf hin, wie aktivierende Methoden umgesetzt werden müssen, um eine ähnlich hohe Wirkung zu erzielen, wie in der Literatur berichtet wurde (Freeman et al., 2014; vgl. Kapitel 3.5).

Es kann weitere Gründe dafür geben, dass viele Studierende der Humanmedizin eher lehrendenzentrierten Unterricht in Seminaren und Vorlesungen bevorzugen. Die Studierenden beschrieben sich in der Studierendenbefragung als sehr lernzielorientiert, mit hohem Interesse an dem Studium und sozial gut eingebunden. Die Beurteilung der Lehre in studentischen Evaluationen ist überwiegend positiv (Porsche, 2013) und die Heidelberger Absolventen der Humanmedizin erzielen in der Regel überdurchschnittliche Prüfungsergebnisse (Institut für medizinische und pharmazeutische Prüfungsfragen, 2014). HeiCuMed (vgl. Kapitel 1) bietet aus Sicht der Studierenden eventuell bereits genügend alternative Aktivierungen für ein erfolgreiches Studium. Der Einsatz individueller Lern- und Arbeitsphasen in Vorlesungen und Seminaren wird möglicherweise als wirksam, aber wenig effizient betrachtet. Dann hätten die Lehrenden in diesen Studien die didaktischen Ideen aus der DOS nur so weit angepasst, wie sie es für ihre Zielgruppe für angemessen halten.

In dieser Arbeit wurden unterschiedliche Lehrende in verschiedenen Studien untersucht, ohne dass die Beobachtungsdaten mit den Selbstberichten verknüpft werden konnten. Dadurch konnten sich die Vor- und Nachteile der Methoden nicht gegenseitig ausgleichen. Somit konnte diese Arbeit keinen Beitrag zu der Frage leisten, wie hoch der individuelle Zusammenhang zwischen Lehrkonzeptionen, Lehransätzen und beobachtetem Lehrverhalten ist. Auch gelang es nicht, alle bekannten Limitationen bei der

Evaluation hochschuldidaktischer Weiterbildungen (vgl. Kapitel 4.4) zu überwinden. Es konnten nur Gelegenheitsstichproben untersucht werden, so dass sich nicht eindeutig feststellen lässt, dass nur die DOS für die Unterschiede verantwortlich ist (vgl. Kapitel 4.2). Die Personen konnten den Gruppen nicht randomisiert zugewiesen und in der Beobachtungsstudie konnte kein Pre-Test umgesetzt werden. Gleichzeitig wurden andere Limitationen überwunden. Die Stichprobengröße war für eine Evaluation hochschuldidaktischer Weiterbildungen hoch. Es wurden nicht nur Selbstberichte, sondern auch Beobachtungsdaten erhoben. Es lagen Vergleichsgruppen vor. Bei der Lehrendenbefragung fand eine Vortestung statt. Die Kontrollvariablen zeigten kaum Unterschiede zwischen Vergleichs- und Versuchsgruppen auf. Die Hypothesen standen bei auch bei vertiefenden Analysen in Einklang mit den Daten.

Studierendenzentrierung und die Transparenz der Lernziele standen in dieser Evaluationsstudie als abhängige Variablen im Vordergrund. Die Zahl zu testender Hypothesen wurde möglichst gering gehalten, um das Risiko einer Alpha-Fehler-Inflation zu minimieren. Mehr abhängige Variablen hätten die vielfältigen Inhalte der DOS besser repräsentiert. Angesichts der zu erwartenden Stichproben- und Effektgrößen hätte eine dann notwendigerweise durchzuführende Alpha-Fehler-Korrektur (Rost, 2007) zu einem deutlichen Verlust der Teststärke geführt.

Der Mangel an objektiven Beobachtungsinstrumenten für die Evaluation hochschuldidaktischer Trainings hat bislang dazu geführt, dass die Verhaltensebene nicht angemessen berücksichtigt wurde (Stes et al., 2010). Mit dem BEO-LV liegt nun ein Beobachtungsbogen vor, der in hochschuldidaktischen Evaluationen auf der Verhaltensebene eingesetzt werden kann, um studierendenzentrierte und lehrendenzentrierte Merkmale der Lehre zu messen. Die Stärke des BEO-LV liegt in den umfangreichen und geprüften operationalen Definitionen der Lehrhandlungen. Ohne die Beobachtungsstudie hätten hier, wie in vergleichbaren Arbeiten (vgl. Stes et al., 2010) lediglich die Selbstberichtete der Lehrenden für die Evaluation herangezogen werden können, was zu einer Überschätzung der Wirkung geführt hätte. Daraus könnten unrealistische Erwartungen an die Wirksamkeit hochschuldidaktischer Weiterbildungen entstehen. Der Nachweis, dass die hier beobachteten Veränderungen im Lehrverhalten zu praktisch bedeutsamen Verbesserungen der Lernleistungen Studierender führen, steht jedoch noch aus.

Bislang liegen keine Metaanalysen zur Evaluation hochschuldidaktischer Weiterbildungen vor. Stes et al. (2010) haben in ihrem Review verschiedene Variationen

von Weiterbildungen (Dauer, Art und Zielgruppe) bezüglich der Wirkung auf verschiedenen Zielebenen verglichen. Sie schlagen vor, zukünftig die theoretischen Grundlagen, die Ziele und die Inhalte zum Gegenstand der Evaluation zu erheben. Diese Forderung ist sehr wichtig, um die Gestaltung zukünftiger Weiterbildungen auf eine bessere empirische Basis zu stellen. Aus den unbeantworteten Fragen und Limitationen dieser Arbeit lassen sich weitere Forschungsperspektiven ableiten. Es müssen bessere Untersuchungspläne verwirklicht werden, um die Wirkung von hochschuldidaktischen Weiterbildungen besser zu verstehen (vgl. Desimone, 2009; Steinert et al., 2006; Stes et al., 2010). Da die Randomisierung von Lehrenden nur selten möglich ist, müssen hier andere Lösungen gefunden werden. Cook und Beckman (2010) empfehlen, anstelle einer Vergleichsgruppe ohne Intervention, verschiedene Variationen einer Weiterbildung zu untersuchen. Unterschiede zwischen den Gruppen lassen sich dann eher auf die Variation zurückführen, als auf die Zusammensetzung der Gruppen. Bei Ullrich (2013) kam ein experimentelles Warte-Kontrollgruppendesign zum Einsatz. Eine zufällig ausgewählte Hälfte der Teilnehmer nahm erst an der Intervention teil, nachdem die andere Hälfte diese abgeschlossen hatte. Beide Ideen haben das Potenzial, die Untersuchungspläne von Weiterbildungen zu verbessern. Damit verbunden muss die Entwicklung und Messung von Lehrkompetenz (vgl. Kane et al., 2002) sowie Lehransätzen (vgl. Lübeck, 2009) genauer untersucht werden. Dieser Aspekt erscheint besonders dringlich, denn erst danach kann der Forderung von Steinert et al. (2006) und Stes et al. (2010) nachgekommen werden, möglichst identische Instrumente in verschiedenen Evaluationsstudien zu verwenden. Auch die optimale Umsetzung studierendenzentrierter Lehre muss besser verstanden werden. Lehrende sollten diese so umsetzen, dass sie lernpsychologisch wirksam ist und gleichzeitig eine hohe Akzeptanz erfährt. Die Besonderheiten verschiedener Lehrformate sind dabei zu berücksichtigen.

Abschließend stellt sich die Frage, wie die hochschuldidaktische Aus- und Weiterbildung von Lehrenden geregelt sein sollte, um effektiv zu sein. Empfehlungen zu einer bestimmten Anzahl von Unterrichtseinheiten sind bislang kaum durch den aktuellen Forschungsstand zu rechtfertigen (Desimone, 2009). Um die Empfehlungen nach Spencer für wirksame Weiterbildungen (2014, vgl. Kapitel 4.1) angemessen zu berücksichtigen, sind allerdings wie in der DOS etwa 120 Unterrichtseinheiten notwendig. Damit sich die Ergebnisse dieser Programmevaluation auf vergleichbare Angebote generalisieren lassen, sollte dieser Umfang nicht unterschritten werden. Gleichzeitig sollten die hochschuldidaktischen Angebote so professionell gestaltet sein, dass eine ver-

pflichtende Teilnahme mit hoher Akzeptanz seitens der Lehrenden einhergeht. Welche Verbesserungen aus dieser Arbeit abgeleitet wurden, beschreibt der folgende Abschnitt.

Formative Evaluation der DOS. Nach Scriven (1972) sollte eine Evaluation Hinweise zur Optimierung einer Weiterbildung geben. Diese werden nun abschließend dargestellt. Die DOS der medizinischen Fakultät befindet sich, wie das Curriculum, in einer ständigen Weiterentwicklung. Mit der Einführung von HeiCuMed sollten Lehrende auf neue praktische Lehr- und Prüfformate vorbereitet und ermutigt werden, in allen Lehrveranstaltungen einen hohen Anteil studierendenzentrierter Lehre umzusetzen. Der Begriff *Dozentenschulung* verdeutlicht, dass diese Ziele in der Vergangenheit eher direktiv vermittelt wurden. Rückmeldungen von Lehrenden, Studierenden und auch die Ergebnisse dieser Arbeit haben gezeigt, dass die flächendeckende Umsetzung studierendenzentrierter Lehre nicht von allen Personen als Ziel geteilt wird. Bei der individuellen Beratung der Lehrenden wird dies während der DOS nun stärker berücksichtigt. Wenn Lehrende einen lehrendenzentrierten Ansatz in Seminaren und Vorlesungen beibehalten wollen, werden sie dabei unterstützt, entsprechende didaktische Ideen möglichst wirksam umzusetzen. Das ändert nichts daran, dass gestiegene Studierendenzentrierung ein Erfolgsindikator der DOS bleiben wird. Den Lehrenden werden die Ergebnisse der Studierendenbefragung präsentiert und die typischen Probleme bei der Umsetzung werden gemeinsam diskutiert. Der Fokus liegt nun stärker darauf, Aktivierungen lernförderlich umzusetzen. Lehrende können zudem wählen, in welchem Lehrformat sie ihr Wissen zu studierendenzentrierter Lehre vertiefen wollen: (a) im Bedside-Teaching, (b) in Seminaren mit maximal 30-40 Personen oder (c) in Seminaren und Vorlesungen mit mehr als 30 Personen. Die Studierenden wünschen sich Lehrende, die sowohl studierendenzentrierte als auch lehrendenzentrierte Lehre in möglichst hoher Qualität umsetzen, beides soll die DOS unterstützen.

Während Lehrende selbst entscheiden sollen, ob sie studierendenzentrierte Lehre umsetzen, bleibt die explizite Nennung der Lernziele ein wichtiges Ziel zur Verbesserung der Lehr- und Studienbedingungen (vgl. Wissenschaftsrat, 2008). Um die Lernzieltransparenz in der Lehre zu stärken, wurde ein entsprechendes Item in die studentischen Evaluationen an der Medizinischen Fakultät Heidelberg seit 2013 neu aufgenommen. Dies soll ein Anreiz sein, die Lernziele transparent zu machen.

Die Erfahrungen mit den kollegialen Hospitationen (vgl. Kapitel 7.4) haben gezeigt, dass die Reflexionsfähigkeit der Teilnehmer besser unterstützt werden kann. Dazu wurden die Beobachtungsbögen für die kollegiale Hospitation überarbeitet. Diese konn-

ten nicht die Doppelfunktion erfüllen, sowohl eine Reflexion zu ermöglichen als auch objektive Daten zu erheben. Lehrende wählen jetzt selbst die Beobachtungsschwerpunkte aus, zu denen sie nach der Hospitation ein ausführliches, schriftliches Feedback von den Kollegen bekommen. Ebenso geben sie nach deren Lehrveranstaltung Feedback. Beobachtungsdaten wie diese können Diskrepanzen zwischen den Vorstellungen der Lehrenden und den tatsächlichen Lehrhandlungen aufzeigen und so zur Professionalisierung von Hochschullehrenden beitragen (Johannes & Seidel, 2012).

Einige Entwicklungen vollzogen sich unabhängig von den Ergebnissen dieser Arbeit. So beinhaltet die DOS mittlerweile zehn Workshop- Tage und findet wieder in Heidelberg statt. Neue Inhalte, wie die *Betreuung von Promotionen* und *Kollegiale Beratung* sind dazu gekommen. Andere, wie das Thema Lernstile und die Darstellung bestimmter Methoden, die in der Humanmedizin kaum umgesetzt werden (z. B. *Lerntempo-Duett*; vgl. Strittmatter-Haubold & Ehlail, 2012) wurden reduziert.

Da die DOS in ein System der Qualitätssicherung der Lehre an der medizinischen Fakultät eingebunden ist, haben die Ergebnisse dieser Arbeit praktische Relevanz. Die Erfahrungen aus etwa 150 Lehrhospitationen waren sehr hilfreich, um Lehrende und Fachbereiche in der Weiterentwicklung ihrer Lehre zu unterstützen. Die Konzeption von Workshops für andere Zielgruppen an der Medizinischen Fakultät konnte davon profitieren. Durch die Teilnahme an der wissenschaftlichen Untersuchung der DOS haben auch viele Lehrende erstmals Kontakt mit den Angeboten zur Qualitätssicherung der Lehre bekommen. Das Leitbild der Universitätsmedizin Heidelberg gibt „Exzellenz in Forschung und Lehre“ als Ziel vor. Die in dieser Arbeit gewonnenen Erfahrungen sollen Lehrende auf ihrem Weg dorthin motivieren und unterstützen.

Literaturverzeichnis

- Adshead, L., White, P. T. & Stephenson, A. (2006). Introducing peer observation of teaching to GP teachers: a questionnaire study. *Medical Teacher*, 28, 68-73.
- Al-Kabbani, D., Trautwein, C. & Schaper, N. (2012). Modelle hochschuldidaktischer Lehrkompetenz – Stand der Forschung. In B. Szczyrba & S. Gotzen (Hrsg.), *Das Lehrportfolio – Darstellung, Entwicklung und Nachweis von Lehrkompetenz an Hochschulen* (S. 29 -50). Münster: LIT-Verlag.
- Alliger, G.M., Tannenbaum, S.I., Bennett, W. Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341-358.
- Amundsen, C. & Wilson, M. (2012). Are we asking the right questions? A conceptual review of the educational development literature in higher education. *Review of Educational Research*, 82, 90-126.
- Anderson, J. R. (2007). *Kognitive Psychologie* (6. Aufl.). Heidelberg: Spektrum, Akademischer Verlag.
- Andrews, T. M., Leonard, M. J., Colgrove, C. A & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE-Life Sciences Education*, 10, 394-405.
- Argyris, C. & Schön, D. A. (1996). *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley.
- Attard, A., Di Iorio, E., Geven, K. & Santa, R. (2010). *Student-Centred Learning: Toolkit for students, staff and higher education institutions*. Zugriff am 04.05.2015 unter http://www.aic.lv/bologna/2010/Reports/SCL_toolkit_ESU_EI.pdf
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden. Eine anwendungsbezogene Einführung* (12., vollständig überarbeitete Aufl.). Heidelberg: Springer.
- Baeten, M., Dochy, F. & Struyven, K. (2008). Students' approaches to learning and assessment preferences in a portfolio-based learning environment. *Instructional Science*, 36, 359-374.
- Baeten, M., Kyndt, E., Struyven, K. & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, 5, 243-260.

- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Hrsg.). *Self-efficacy beliefs of adolescents*, (S. 307-337). Greenwich, CT: Information Age Publishing.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122.
- Barczyk, C., Buckenmeyer, J., Feldman, L & Hixon, E. (2011). Assessment of a University-Based Distance Education Mentoring Program from a Quality Management Perspective. *Mentoring & Tutoring: Partnership in Learning*, 19, 5-24.
- Barr, R. B. & Tagg, J. (1995). From teaching to learning-A new paradigm for undergraduate education. *Change: The Magazine of Higher Learning*, 27, 12-26.
- Battaglia, S. (2010). Quo vadis, hochschuldidaktische (Hochschul)Forschung. *Journal Hochschuldidaktik*, 1, 28-32.
- Benor, D. E. (2000). Faculty development, teacher training and teacher accreditation in medical education: twenty years from now. *Medical Teacher*, 22, 503-512.
- Benz, C. (2005). *Das Kompetenzprofil des Hochschullehrers. Zur Bestimmung der Kompetenzanforderungen mittels Conjoint-Analyse*. Aachen: Shaker.
- Biggs, J. B. & Tang, C. (2007). *Teaching for quality learning at university* (3. Aufl.). Maidenhead: Open University Press.
- Biggs, J.B. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Biggs, J.B. (1999). What the student does: teaching for enhanced learning. *Higher Education Research & Development*, 18, 57-75.
- Böttger, H.; Gien, G. & Müller, T. (2011). Eine Befragung zur Studiensituation der Katholischen Universität Eichstätt-Ingolstadt. In H. Böttger, Heiner & G. Gien (Hrsg.), *Aspekte einer exzellenten universitären Lehre* (S. 18-45). Bad Heilbrunn: Klinkhardt.
- Booth, K. M & James, B. W. (2001). Interactive learning in a higher education Level 1 mechanics module. *International Journal of Science Education*, 23, 955-967.
- Borchard, C. (2002). *Hochschuldidaktische Weiterbildung-Akzeptanz und Wirkung: eine Analyse am Beispiel des Bausteinprogramms WindH-Weiterbildung in der Hochschullehre, seiner Konzeption und Evaluation* (Vol. 2). Münster: LIT Verlag.
- Boyce, T.E. & Hinline, P.N. (2002). Interteaching: A strategy for enhancing the userfriendliness of behavioral arrangements in the college classroom. *The Behavior Analyst*, 25, 215 - 226.

- Brauchle, P. E. & Jerich, K. F. (1998). Improving instructional development of industrial technology graduate teaching assistants. *Journal of Industrial Teacher Education*, 35, 67-92.
- Braun, E. & Hannover, B. (2008). Zum Zusammenhang zwischen Lehr-Orientierung und Lehrgestaltung von Hochschuldozierenden und subjektivem Kompetenzzuwachs bei Studierenden. *Zeitschrift für Erziehungswissenschaft*, Sonderheft 9, 277–291.
- Braun, E. & Hannover, B. (2011). Gelegenheiten zum Kompetenzerwerb in der universitären Lehre: Zusammenhänge zwischen den Einschätzungen Studierender und unabhängigen Beobachtungen relevanter Merkmale universitärer Lehrveranstaltungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 22–28.
- Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Kompetenzorientierte Lehr evaluation - Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54, 30-42.
- Brophy, J. & Good, T. L. (1986). Teacher behavior and student achievement. *Handbook of research on teaching*, 328-375.
- Brunner-Schwaiger, A. & Salzgeber, G. (2004). „Qualität der Lehre ist kein Zufall!“ In S. Laske, T. Scheytt & C. Meister-Scheytt (Hrsg.), *Personalentwicklung und universitärer Wandel. Programm, Aufgaben, Gestaltung* (S. 331-353). München: Hampp.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. München: Pearson.
- Bundesministerium der Justiz und für Verbraucherschutz. *Ärztliche Approbationsordnung (ÄApprO) in der Fassung vom 27.6.2002*. Bundesgesetzblatt 2002. Zugriff am 04.05.2015 unter http://www.gesetze-im-internet.de/bundesrecht/_appro_2002/gesamt.pdf
- Cano, F. & Berbén, A. B. G. (2009). University students' achievement goals and approaches to learning in mathematics. *British Journal of Educational Psychology*, 79, 131-153.
- Cash, A. H., Hamre, B. K., Pianta, R. C. & Myers, S. S. (2012). Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly*, 27, 529-542.

- Cavanagh, M. (2011). Students' experiences of active engagement through cooperative learning activities in lectures. *Active Learning in Higher Education*, 12, 23-33.
- Centra, J. A. (1979). *Determining Faculty Effectiveness. Assessing Teaching, Research, and Service for Personnel Decisions and Improvement*. San Francisco, CA: Jossey-Bass.
- Chur, D. (2005). Didaktische Qualifizierung im Rahmen des Heidelberger Modells der (Aus-) Bildungsqualität. Strategien und Konzepte hochschuldidaktischer Qualifizierung im internationalen Vergleich. In S. Brendel, K. Kaiser & G. Macke (Hrsg.), *Hochschuldidaktische Qualifizierung. Strategien und Konzepte im internationalen Vergleich* (179-195). Bielefeld: Bertelsmann.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.Aufl.). New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112, 155.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 13, 321-341.
- Cohen, R., Murnaghan, L., Collins, J. & Pratt, D. (2005). An update on master's degrees in medical education. *Medical Teacher*, 27, 686-692.
- Cook, D. A. & Beckman, T. J. (2010). Reflections on experimental research in medical education. *Advances in Health Sciences Education*, 15, 455-464.
- Craik, F. I. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Crouch, C. H. & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69, 970-977.
- Dany, S. (2007). *Start in die Lehre: Qualifizierung von Lehrenden für den Hochschulalltag*. Münster: LIT Verlag.
- Deutsche Gesellschaft für Hochschuldidaktik (2014). *Hochschuldidaktische Institutionen. Institutionelle Mitglieder der dghd*. Zugriff am 04.09.2014 unter <http://www.dghd.de/hochschuldidaktische-institutionen.html>

- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38, 181-199.
- Deslauriers, L., Schelew, E. & Wieman, C. (2011). Improved Learning in a Large-Enrollment Physics Class. *Science*, 332, 862-864.
- Dewey, J. (2002). *Wie wir denken*. Mit einem Nachwort neu herausgegeben von Rebekka Horlacher und Jürgen Oelkers. Zürich: Pestalozzianum.
- Diez, A., Fischer, M., Rühmann, M. & Weiß, S. (1999). Bessere Lehre durch kollegiale Supervision - ein neuer Weg in der hochschuldidaktischen Weiterbildung. In B. Berendt, H.-P. Voss & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre* (L. 3.1, S.1-18). Berlin: Raabe.
- Dubs, R. (2009). *Lehrerverhalten. Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht* (2. Aufl.). Stuttgart: Steiner.
- Dunkin, M. J. & Barnes, J. (1986). Research on teaching in higher education. *Handbook of research on teaching*, 3, 754-777.
- Elen, J., Clarebout, G., Léonard, R. & Lowyck, J. (2007). Student-centred and teacher-centred learning environments: what students think. *Teaching in Higher Education*, 12, 105-117.
- Entwistle N. & Walker, P. (2000). Strategic alertness and expanded awareness within sophisticated conceptions of teaching. *Instructional Science*, 28, 335-361.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. G*POWER 3.1.7. Download vom 23.07.2013 unter <http://www.gpower.hhu.de/>
- Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 583-645.
- Fendler, J., Seidel, T. & Johannes, C. (2013). Wie wirksam sind hochschuldidaktische Workshops? Auswirkungen auf die Lehrkompetenz von Hochschullehrenden. *Beiträge zur Hochschulforschung*, 35, 28-48.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.
- Fisseni, H. J. (1997). *Lehrbuch der psychologischen Diagnostik* (2. Aufl.). Göttingen [u. a.]: Hogrefe.

- Fleiss, J. L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111, 8410-8415.
- Futter, K. (2011). Fachspezifisch-didaktisches Lehrcoaching: Ein wirkungsvolles Modell zur Förderung der Lehrexpertise? In M. Weil, M. Schiefner, B. Eugster & K. Futter (Hrsg.), *Aktionsfelder der Hochschuldidaktik. Von der Weiterbildung zum Diskurs* (S. 205-220). Münster [u. a.]: Waxmann..
- Gallos, M. R., Berg, E. V. D & Treagust, D. F. (2005). The effect of integrated course and faculty development: Experiences of a university chemistry department in the Philippines. *International Journal of Science Education*, 27, 985-1006.
- Garner, I. (2000). Problems and inconsistencies with Kolb's learning styles. *Educational Psychology*, 20, 341-348.
- Gerbig-Calcagni, I. (2010). *Wie aufmerksam sind Studierende in Vorlesungen und wie viel können sie behalten?* Unveröffentlichte Dissertation, Pädagogische Hochschule Weingarten.
- Gibbs, G. & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active learning in Higher Education*, 5, 87-100.
- Gijbels, D. & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference? *Educational Studies*, 32, 399-409.
- Gijbels, D., Dochy, F., Van den Bossche, P & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, 75, 27-61.
- Goldstein, H. (2011). *Multilevel statistical models*. Chichester: John Wiley & Sons.
- Good, T. & Brophy J. (1986). School effects. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching. A project of the American Educational Research Association* (3. Aufl.). New York: Macmillan.
- Groves, M. (2005). Problem-based learning and learning approach: Is there a relationship? *Advances in Health Sciences Education*, 10, 315-326.
- Grube, J. W., Morgan, M. & Kearney, K. A. (1989). Using self-generated identification codes to match questionnaires in panel studies of adolescent substance use. *Addictive Behaviors*, 14, 159-171.

- Hanbury, A., Prosser, M. & Rickinson, M. (2008). The differential impact of UK accredited teaching development programmes on academics' approaches to teaching. *Studies in Higher Education*, 33, 469-483.
- Harden R.M.; Crosby J. (2000). AMEE Guide No 20: The good teacher is more than a lecturer – the twelve roles of the teacher. *Medical Teacher*, 22, 334-347.
- Harris, C. (2004). *Understanding the role of epistemological beliefs in post-graduate studies: Motivation and conceptions of learning in first-year law students*. Unveröffentlichte Dissertation, University of Texas.
- Hativa, N. & Goodyear, P. (2002). Research on teacher thinking, beliefs, and knowledge in higher education: Foundations, status and prospects. In N. Hativa & P. Goodyear (Hrsg.), *Teacher thinking, beliefs and knowledge in higher education* (S. 335-359). Dordrecht [u. a.]: Springer.
- Hattie, J. & Marsh, H.W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, 66, 507-542.
- Hattie, J. (2009). *Visible learning. A synthesis of meta-analyses relating to achievement*. London: Routledge.
- Heger, M. (1985). *Verbesserung der Lehr-und Lernsituation in ingenieurwissenschaftlichen Studiengängen: eine Evaluation hochschuldidaktischer Aus- und Fortbildung*. Unveröffentlichte Dissertation, Technische Universität Berlin.
- Helmke, A. & Schrader, F.-W. (2010). Hochschuldidaktik. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. erweiterte Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (Unterricht verbessern – Schule entwickeln, 4. erweiterte Aufl.). Seelze: Klett-Kallmeyer.
- Hochschulrektorenkonferenz (HRK, 2008). *Für eine Reform der Lehre in Hochschulen* (3. Mitgliederversammlung der HRK am 22.4.2008). Bonn. Zugriff am 04.05.2015 unter <http://www.hrk.de/positionen/gesamtlistebeschluesse/position/convention/fuer-eine-reform-der-lehre-in-den-hochschulen/>
- Hofer, M., Jansen, M. & Soboll, S. (2005). Effektive Didaktiktrainings für Dozenten der Medizin. *GMS Zeitschrift für Medizinische Ausbildung*, 22, 2005-22.

- Hora, M. T. (2013). *Exploring the Use of the Teaching Dimensions Observation Protocol to Develop Fine-grained Measures of Interactive Teaching in Undergraduate Science Classrooms*. Working Paper 2013-6. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research. Zugriff am 04.05.2015 unter http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2013_06.pdf
- Hora, M. T. & Ferrare, J. J. (2013). *A review of classroom observation techniques used in postsecondary settings*. Working Paper 2013-1. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research. Zugriff am 04.05.2015 unter http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2013_01.pdf
- Hora, M. T. & Ferrare, J. J. (2014). *The Teaching Dimensions Observation Protocol (TDOP) 2.0*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Huwendiek S., Dern P., Hahn E.G., Pediaditakis D., Tönshoff B. & Nikendei C. (2008). Training requirements, expertise and basic conditions of committed medical educators in Germany. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 102, 613-617.
- Huwendiek S., Mennin S., Dern P., Ben-David M.F., van der Vleuten C., Tönshoff B. et al. (2010). Expertise, needs and challenges of medical educators: Results of an international web survey. *Medical Teacher*, 32, 912-918.
- Huwendiek S., Hahn E. G., Tönshoff B. & Nikendei C. (2013). Herausforderungen Lehrender in der Medizin: Ergebnisse einer Umfrage unter Mitgliedern der Gesellschaft für Medizinische Ausbildung. *GMS Zeitschrift für Medizinische Ausbildung*, 30, 3.
- Institut für medizinische und pharmazeutische Prüfungsfragen (2014). *Ergebnisse vorangegangener Examina*. Zugriff am 04.05.2015 unter <https://www.impp.de/internet/de/archiv.html>
- Irby, D. M. (1994). What clinical teachers in medicine need to know. *Academic Medicine*, 69, 333-342.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Jaudzims, S. & Schnitzer, K. (2007). *Deutschland und die Schaffung eines europäischen Hochschul- und Forschungsraumes*. In F. Bretschneider & J. Wildt, (Hrsg.), *Handbuch Akkreditierung von Studiengängen* (2. Auflage, S. 21-34). Bielefeld: Bertelsmann.
- Johannes, C. & Seidel, T. (2012). Professionalisierung von Hochschullehrenden. *Zeitschrift für Erziehungswissenschaft*, 15, 233-251.

- Kane, R., Sandretto, S. & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, 72, 177-228.
- Kane, T. J. & Staiger, D. O. (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*. Research Paper. MET Project. Bill & Melinda Gates Foundation.
- Kearney, K. A., Hopkins, R. H., Mauss, A. L. & Weisheit, R. A. (1984). Self-generated identification codes for anonymous collection of longitudinal questionnaire data. *Public Opinion Quarterly*, 48, 370-378.
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg), *Testtheorie und Fragebogenkonstruktion* (S. 75-102). Heidelberg: Springer.
- Kember, D. & Kwan, K. P. (2000). Lecturers' approaches to teaching and their relationship to conceptions of good teaching. *Instructional Science*, 28, 469-490.
- Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7, 255-275.
- Kirkpatrick, D. & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler Publishers.
- Kirkpatrick, D. (1967). Evaluation of training, in: R. Craig & I. Mittel (Hrsg), *Training and Development Handbook* (S. 87-112). New York, NY, USA: McGraw-Hill.
- Kirkpatrick, D. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler Publishers.
- Klauer, K. J. & Leutner, D. (2007). *Lehren und Lernen. Einführung in die Instruktionspsychologie* (Studentexte). Weinheim [u. a.]: Beltz, PVU.
- Knol, M. H., Vorst, H. C., van Driel, J. H. & Mellenbergh, G. J. (2013). Experimental effects of student evaluations coupled with collaborative consultation on college professors' instructional skills. *Research in Higher Education*, 54, 825-850.
- Kohlmann, T. & Raspe, H. (1998). Zur Messung patientennaher Erfolgskriterien in der medizinischen Rehabilitation: Wie gut stimmen „indirekte“ und „direkte“ Methoden der Veränderungsmessung überein. *Rehabilitation*, 37, 30-37.
- Kolb, D. A. (1985). *Learning Style Inventory*. Boston, MA: Hay Group.
- Köller, O. & Schiefele, U. (2010). Zielorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarbeitete und erweiterte Aufl.) (S. 880-886). Weinheim [u. a.]: Beltz, PVU.

- Kröber, E. (2010). *Wirksamkeit hochschuldidaktischer Weiterbildung am Beispiel der Veränderung von Lehrkonzeptionen: eine Evaluationsstudie*. Unveröffentlichte Dissertation, Technische Universität Dortmund.
- Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education*, 18, 315-328.
- Kurtz, S. M., Silverman, D. J., Draper, J., van Dalen, J. & Platt, F. W. (2005). *Teaching and learning communication skills in medicine*. Oxford: Radcliffe Pub.
- Lammerding-Köppel, M., Fabry, G., Hofer, M., Ochsendorf, F. & Schirlo, C. (2006). Hochschuldidaktische Qualifizierung in der Medizin: I. Bestandsaufnahme. *GMS Zeitschrift für Medizinische Ausbildung*, 23, 73.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
- Langer, I. & Schulz von Thun, F. (2007). *Messung komplexer Merkmale in Psychologie und Pädagogik. Ratingverfahren*. Münster [u. a.]: Waxmann.
- Lea, S. J., Stephenson, D. & Troy, J. (2003). Higher education students' attitudes to student-centred learning: beyond 'educational bulimia'? *Studies in Higher Education*, 28, 321-334.
- Leitner, E. (1984). *Hochschul-Pädagogik. Zur Genese und Funktion der Hochschul-Pädagogik im Rahmen der Entwicklung der deutschen Universität 1800-1968*. *Schriften zur Hochschulbildung*. Bd. 4. Frankfurt/M. [u. a.]: Peter Lang.
- Leonhart, R. (2013). *Lehrbuch Statistik. Einstieg und Vertiefung* (3. überarbeitete und erweiterte Auflage). Bern: Verlag Hans Huber.
- Levin, A. (2005). *Lernen durch Fragen*. Münster [u. a.]: Waxmann.
- Levinson-Rose, J. & Menges, R. J. (1981). Improving college teaching: A critical review of research. *Review of Educational Research*, 51, 403-434.
- Lindblom-Ylänne, S., Trigwell, K., Nevgi, A. & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. *Studies in Higher Education*, 31, 285-298.
- Looß, M. (2001). Lerntypen. *Die Deutsche Schule*, 93, 186-198.
- Lortie, D. C. (1975). *School teacher: A sociological inquiry*. Chicago: University of Chicago Press.
- Lübeck, D. (2009). *Lehransätze in der Hochschullehre*. Unveröffentlichte Dissertation, Freie Universität Berlin.
- Macke, G., Hanke, U. & Viehmann, P. (2008). *Hochschuldidaktik. Lehren, vortragen, prüfen*. Weinheim [u. a.]: Beltz, PVU.

- Mahler, S. & Benor, D. E. (1984). Short and long term effects of a teacher-training workshop in medical school. *Higher Education*, 13, 265-273.
- Marsh, H. W. & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30, 217-251.
- Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77-95.
- Marton, F. & Säljö, R. (1976). On Qualitative Differences in Learning: I-Outcome and process. *British Journal of Educational Psychology*, 46, 4-11.
- McAlpine, L. (2003). Het belang van onderwijskundige vorming voor studentgecentreerd onderwijs: de praktijk geëvalueerd. [The importance of instructional development for student centered teaching: An examination of practice.] In N. Druine, M. Clement, M & K. Waeytens (Hrsg.), *Dynamiek in het hoger onderwijs. Uitdagingen voor onderwijsondersteuning* [Dynamics in higher education: Challenges for teaching support] (S. 57-71). Leuven: Universitaire Pers.
- McGaghie, W. C., Issenberg, S. B., Petrusa, E. R & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003-2009. *Medical Education*, 44, 50-63.
- McLean, M., Cilliers, F. & Van Wyk, J. M. (2008). Faculty development: Yesterday, today and tomorrow. *Medical Teacher*, 30, 555-584.
- McShannon, J. & Hynes, P. (2005). Student achievement and retention: Can professional development programs help faculty GRASP it? *The Journal of Faculty Development*, 20, 87-93.
- Medsker, K. L. (1992). NETWORK for Excellent Teaching; A Case Study in University Instructional Development. *Performance Improvement Quarterly*, 5, 35-48.
- Meyer, J. H. & Eley, M. G. (2006). The approaches to teaching inventory: A critique of its development and applicability. *British Journal of Educational Psychology*, 76, 633-649.
- Michael, J. (2007). Faculty perceptions about barriers to active learning. *College Teaching*, 55, 42-47.
- Miles, J. & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage.
- Milner, R. J., Gusic, M. E. & Thorndyke, L. E. (2011). Perspective: Toward a competency framework for faculty. *Academic Medicine*, 86, 1204-1210.

- Moniz, R., Fine, J. & Bliss, L. (2008). The effectiveness of direct-instruction and student-centered teaching methods on students' functional understanding of plagiarism. *College & Undergraduate Libraries*, 15, 255-279.
- Müller, F. H. (2001). *Studium und Interesse*. Münster [u. a.]: Waxmann.
- Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology*, 75, 138.
- Murray, H. G. (1997). Effective teaching behaviors in the college classroom. Effective teaching in higher education: *Research and Practice*, 171-204.
- Murray, K. & Macdonald, R. (1997). The disjunction between lecturers' conceptions of teaching and their claimed educational practice. *Higher Education*, 33, 331-349.
- Neville, A. J. (2009). Problem-based learning and medical education forty years on. *Medical Principles and Practice*, 18, 1-9.
- Norton, L., Richardson, T. E., Hartley, J., Newstead, S. & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education*, 50, 537-571.
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, 33, 301-314.
- Nurrenbern, S. C., Mickiewicz, J. A. & Francisco, J. S. (1999). The impact of continuous instructional development on graduate and undergraduate students. *Journal of Chemical Education*, 76, 114-117.
- O'Neill, G. & McMahon, T. (2005). Student-centred learning: What does it mean for students and lecturers. *Emerging issues in the practice of university learning and teaching*, 1, 27-36.
- O'Sullivan, P. S. & Irby, D. M. (2014). Promoting Scholarship in Faculty Development: Relevant Research Paradigms and Methodologies. In Y. Steinert (Hrsg), *Faculty Development in the Health Professions* (S. 375-398). Dordrecht: Springer.
- Paetz, N. V., Ceylan, F., Fiehn, J., Schworm, S. & Harteis, C. (2011). *Kompetenz in der Hochschuldidaktik: Ergebnisse einer Delphi-Studie über die Zukunft der Hochschullehre*. Springer-Verlag.
- Parsons, D., Hill, I., Holland, J. & Willis, D. (2012). *Impact of teaching development programmes in higher education*. York: Higher Education Academy.
- Päuler, L. & Jucks, R. (2013). Direkt erfragt: die Messung von Lehrorientierungen per Fragebogen mit offenem Antwortformat. *Zeitschrift für Hochschulentwicklung*, 8, 95-109.

- Piaget, J. (1978). *Das Weltbild des Kindes*. Stuttgart: Klett-Cotta.
- Pleschová, G., Simon, E., Quinlan, K., Murphy, J., Roxa, T. & Szabó, M. (2012). *The professionalisation of academics as teachers in higher education*. Science Position paper. European Science Foundation. Zugriff am 04.05.2015 unter http://www.esf.org/fileadmin/Public_documents/Publications/professionalisation_academics.pdf
- Pöge, A. (2005). Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht. *Zentralarchiv für Empirische Sozialforschung*, 56, 50-69.
- Porsche, M. (2013). *Evaluationsbericht für das Studienjahr 2012/13*. Medizinische Fakultät, Ruprecht-Karls-Universität Heidelberg.
- Porst, R. (2009). *Fragebogen. Ein Arbeitsbuch. Studienskripte zur Soziologie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Postareff, L. & Lindblom-Ylänne, S. (2008). Variation in teachers' descriptions of teaching: Broadening the understanding of teaching in higher education. *Learning and Instruction*, 18, 109-120.
- Postareff, L., Lindblom-Ylänne, S. & Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teaching and Teacher Education*, 23, 557-571.
- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42 (4), 203-220.
- Pratt, D. D. (2002). Good teaching: One size fits all? *New Directions for Adult and Continuing Education*, 2002, 5-16.
- Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G. & Zepke, N. (2004). *Impact of student support services and academic development programs on student outcomes in undergraduate tertiary study: A synthesis of the research*. Report to the Ministry of Education, Massey University College of Education. Zugriff am 02.06.2015 unter http://www.educationcounts.govt.nz/__data/assets/pdf_file/0013/7321/ugradstudentoutcomes.pdf
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93, 223-232.
- Prosser, M. & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham: Open University Press.
- Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the Approaches to Teaching Inventory. *British Journal of Educational Psychology*, 76, 405-419.

- Prosser, M., Trigwell, K. & Taylor, P. (1994). A phenomenographic study of academics' conceptions of science learning and teaching. *Learning and Instruction*, 4, 217-231.
- Prüfer, P. & Rexroth, M. (2005). Kognitive Interviews. *ZUMA How-to-Reihe*, 15, 95-116.
- Ramm, M. (2014). *Response, Stichprobe und Repräsentativität. Hefte zur Bildungs- und Hochschulforschung*, 72. Zugriff am 02. 04.2015 unter <http://cms.uni-konstanz.de/ag-hochschulforschung/publikationen/hefte-zur-bildungs-und-hochschulforschung/>
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006). *Quantitative Methoden. Einführung in die Statistik. Band 1.* (2., erweiterte Aufl.). Heidelberg: Springer.
- Reiber, K. (2005). Die Lehrhospitation mit teilnehmender Beobachtung – Expertengestützte Beratung als Beitrag zur Kompetenzentwicklung. In B. Berendt, H.-P. Voss & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre* (L 3.3/S. 1-12). Berlin: Raabe.
- Reichmann, G. (2008). Welche Kompetenzen sollten gute Universitätslehrer aus der Sicht von Studierenden aufweisen? Ergebnisse einer Conjointanalyse. *Das Hochschulwesen*, 56, 52-57.
- Reid, W. A., Duvall, E. & Evans, P. (2005). Can we influence medical students' approaches to learning? *Medical Teacher*, 27, 401-407.
- Reinmann, G. & Mandl, H. (2006): Unterrichten und Lernumgebungen gestalten. In: A. Krapp; B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (5. vollständig überarbeitete Aufl.) (S. 613-658). Weinheim [u. a.]: Beltz, PVU.
- Renkl, A. (2005). Fünf Dogmen in der Diskussion zum Lernen und Lehren. Provokante Thesen zu dysfunktionalen Voreingenommenheiten. In S. R. Schilling, J. R. Sparfeldt & C. Pruisken (Hrsg.), *Aktuelle Aspekte pädagogisch-psychologischer Forschung. Detlef H. Rost zum 60. Geburtstag* (S. 11-23). Münster [u. a.]: Waxmann..
- Roche, L. A. & Marsh, H. W. (2000). Multiple dimensions of university teacher self-concept. *Instructional Science*, 28, 439-468.
- Rodgers, R., Christie, J. & Wideman, M. (2014). *The effects of a required faculty development program on novice faculty self-efficacy and teaching*. Toronto: Higher Education Quality Council of Ontario.
- Roos, M., Kadmon, M., Kirschfink, M., Koch, E., Jünger, J., Strittmatter-Haubold, V., & Steiner, T. (2014). Developing medical educators – a mixed method evaluation of a teaching education program. *Medical Education Online*, 19. Zugriff am 02.04.2015 unter <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969510/>

- Rost, D. H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien: eine Einführung*. Weinheim [u. a.]: Beltz, PVU.
- Rowe, M. B. (1986). Wait time: slowing down may be a way of speeding up! *Journal of Teacher Education*, 37, 43-50.
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R. & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, 331, 1269-1270.
- Ruprecht-Karls-Universität Heidelberg. Zentrale Universitätsverwaltung (2012). *Studierendenstatistik Sommersemester 2012*. Zugriff am 02.04.2015 unter: http://www.uni-heidelberg.de/md/studium/download/ss2012_www.pdf
- Samuelowicz, K. & Bain, J. D. (1992). Conceptions of teaching held by academic teachers. *Higher Education*, 24, 93-111.
- Sander, P., Stevenson, K., King, M. & Coates, D. (2000). University students' expectations of teaching. *Studies in Higher Education*, 25, 309-323.
- Saville, B. K., Zinn, T. E. & Elliott, M. P. (2005). Interteaching versus traditional methods of instruction: A preliminary analysis. *Teaching of Psychology*, 32, 161-163.
- Saville, B. K., Zinn, T. E., Neef, N. A., Norman, R. V. & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of applied Behavior Analysis*, 39, 49-61.
- Sawada, D., Piburn, M., Falconer, K., Turley, J., Benford, R. & Bloom, I. (2000). *Reformed Teaching Observation Protocol: Technical Report No. IN00-1*. Tempe, AZ: Arizona State University.
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie*, 14, 12-25.
- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 60-91). München: Piper.
- Seidel, T & Hoppert, A. (2011). Merkmale von Lehre an der Hochschule. Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalysen. *Unterrichtswissenschaft*, 39, 154-172.
- Siddiqui, Z. S., Jonas-Dwyer, D. & Carr, S. E. (2007). Twelve tips for peer observation of teaching. *Medical Teacher*, 29, 297-300.
- Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592.

- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N. et al. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323, 122-124.
- Smith, C. V. & Cardaciotto, L. (2012). Is active learning like broccoli? Student perceptions of active learning in large lecture classes. *Journal of the Scholarship of Teaching and Learning*, 11, 53-61.
- Snyder, S. (2001). A program to teach curriculum development to junior faculty. *Family Medicine*, 33, 382-387.
- Sonntag, K. (2002). Personalentwicklung und Training. Stand der psychologischen Forschung und Gestaltung. *Zeitschrift für Personalpsychologie*, 2, 59-79.
- Spencer, J. (2014). Faculty Development Research: The 'State of the Art' and Future Trends. In Y. Steinert (Hrsg), *Faculty Development in the Health Professions* (S. 353-374). Dordrecht: Springer.
- Spiel, C. & Fischer, U. (1998). Evaluierung eines Weiterbildungsangebots für Hochschullehrende. *Zeitschrift für Hochschuldidaktik*, 22, 83-99.
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2002). *Skalen zur Erfassung der Lern- und Leistungsmotivation. SELLMO*. Göttingen: Hogrefe.
- Stehle, S. (2011). *Quality Development of University Teaching: The Role of Student Evaluations*. Unveröffentlichte Dissertation, Ruprecht-Karls-Universität Heidelberg.
- Steinert, Y. (2014). Faculty development: core concepts and principles. In Y. Steinert (Hrsg), *Faculty Development in the Health Professions* (S. 3-25). Dordrecht: Springer.
- Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M et al. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*, 28, 497-526.
- Steinert, Y., Naismith, L. & Mann, K. (2012). Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. *Medical Teacher*, 34, 483-503.
- Stes, A., Clement, M. & Van Petegem, P. (2007). The effectiveness of a faculty training programme: Long-term and institutional impact. *International Journal for Academic Development*, 12, 99-109.
- Stes, A., Min-Leliveld, M., Gijbels, D. & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, 5, 25-49.

- Stewart, M. (2014). Making sense of a teaching programme for university academics: Exploring the longer-term effects. *Teaching and Teacher Education*, 38, 89-98.
- Strittmatter-Haubold, V. & Ehlail, F. (2012). *Lernen im Aufwind. Methodenreader zur Gestaltung von Lernprozessen* (8., völlig überarbeitete und erweiterte Aufl.). Heidelberg: Inst. für Weiterbildung.
- Strong, M., Gargani, J. & Hacifazlıoğlu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367-382.
- Sydow, D. L. (1998). *Outcomes of the VCCS Professional Development Initiative: 1993-1998*. Big Stone Gap, VA: Mountain Empire Community College. Zugriff am 02.04.2015 unter <http://files.eric.ed.gov/fulltext/ED423016.pdf>
- Thumser-Dauth, K. (2007). *Evaluation hochschuldidaktischer Weiterbildung: Entwicklung, Bewertung und Umsetzung des 3P-Modells*. Hamburg: Kovač.
- Trautwein, C. & Merkt, M. (2012). Lehrportfolios für die Darstellung und Entwicklung akademischer Lehrkompetenz. In B.Szczyrba & S.Gotzen (Hrsg.), *Das Lehrportfolio. Entwicklung, Dokumentation und Nachweis von Lehrkompetenz an Hochschulen*, 75-97.
- Trautwein, C. (2013). Lehrebezogene Überzeugungen und Konzeptionen- eine konzeptuelle Landkarte. *Zeitschrift für Hochschulentwicklung*, 8 (3), 1-14. Zugriff am 02.04.2015 unter <http://www.zfhe.at/index.php/zfhe/article/view/557/562>
- Trigwell, K. & Prosser, M. (1996). Changing approaches to teaching: A relational perspective. *Studies in Higher Education*, 21, 275-284.
- Trigwell, K. & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16, 409-424.
- Trigwell, K., Prosser, M. & Ginns, P. (2005). Phenomenographic pedagogy and a revised approaches to teaching inventory. *Higher Education Research & Development*, 24, 349-360.
- Trigwell, K., Prosser, M. & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education*, 27, 75-84.
- Trigwell, K., Prosser, M. & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57-70.
- Turpen, C. & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction. *Physical Review Special Topics-Physics Education Research*, 5, 020101.

- Ulrich, I. (2013). *Strategisches Qualitätsmanagement in der Hochschullehre*. Wiesbaden: Springer Fachmedien.
- van der Wende, M. C. (2000). The Bologna Declaration: Enhancing the transparency and competitiveness of European higher education. *Higher Education in Europe*, 25, 305-310.
- Wahl, D. (2006). *Lernumgebungen erfolgreich gestalten. Vom trägen Wissen zum kompetenten Handeln* (2., erweiterte Aufl.). Bad Heilbrunn: Klinkhardt.
- Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten Vom trägen Wissen zum kompetenten Handeln* (3., erweiterte Aufl.). Bad Heilbrunn: Klinkhardt.
- Wainwright, C. L., Flick, L. B. & Morrell, P. D. (2003). Development of instruments for assessment of instructional practices in standards-based teaching. *Journal of Mathematics and Science: Collaborative Explorations*, 6, 21-46.
- Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L. et al. (2012). *Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program* (UTeach Technical Report 2011/ 01). Texas: Universität, UTeach Natural Sciences.
- Webler, W. D. (2000). Weiterbildung der Hochschullehrer als Mittel der Qualitätssicherung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualitätssicherung im Bildungsbereich: Schule, Sozialpolitik, Hochschule* (S. 225-247.). Weinheim [u. a.]: Beltz, PVU.
- Wegner, E. & Nückles, M. (2011). Die Wirkung hochschuldidaktischer Weiterbildung auf den Umgang mit widersprüchlichen Handlungsanforderungen. *Zeitschrift für Hochschulentwicklung*, 6, 171-188.
- Wehr, S. (2008). Unterrichtshospitationen lernwirksam gestalten. In: S. Wehr & H. Ertel (Hrsg.), *Lernprozesse fördern an der Hochschule* (S. 71-85). Bern: Haupt.
- Weimer, M. & Lenze, L. F. (1997). Instructional interventions: A review of the literature on efforts to improve instruction. *Effective Teaching in Higher Education: Research and Practice*, 205-240.
- Westermann, R., Heise, E., Spies, K. & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie in Erziehung und Unterricht*, 43, 1-22.
- Wildt, J. (2003). Reflexives Lernen in der Lehrerbildung: ein Mehrebenenmodell in hochschuldidaktischer Perspektive. In A. Obolenski & H. Meyer (Hrsg.), *Forschendes Lernen. Theorie und Praxis einer professionellen LehrerInnenbildung* (71-84). Bad Heilbrunn: Klinkhardt.

- Wildt, J. (2004). Welche Schlüsselkompetenzen braucht ein Hochschullehrer? Einige Anmerkungen aus hochschuldidaktischer Sicht. In Stifterverband für die Deutsche Wissenschaft e. V. (Hrsg.), *Schlüsselkompetenzen und Beschäftigungsfähigkeit. Konzepte für die Vermittlung überfachlicher Qualifikationen an Hochschulen* (S. 22–24). Essen.
- Wilson, K. & Korn, J. H. (2007). Attention during lectures: Beyond ten minutes. *Teaching of Psychology*, 34, 85-89.
- Wilson, M. E. (2012). *What is Known about the Relationship between Instructional Development Approaches and Effective Teaching Outcomes? A Meta-study of the Instructional Development Research Literature*. Unveröffentlichte Dissertation, Simon Fraser University, Burnaby.
- Winteler, A. (2006). Lehrende an Hochschulen. In A. Krapp & B. Weidenmann (Hrsg.), *Lehrbuch Pädagogische Psychologie* (5., vollständig überarbeitete Aufl.) (S. 334-347). Weinheim [u. a.]: Beltz, PVU.
- Winteler, A. (2008). *Professionell lehren und lernen: ein Praxisbuch*. Darmstadt: WBG.
- Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen [u. a.]: Hogrefe.
- Wissenschaftsrat (2005). *Empfehlungen zur Ausgestaltung von Berufungsverfahren*. Zugriff am 04.05.2015 unter <http://www.wissenschaftsrat.de/download/archiv/berufungsverfahren.pdf>
- Wissenschaftsrat (2008). *Empfehlungen zur Qualitätsverbesserung von Lehre und Studium*. Zugriff am 04.05.2015 unter <http://www.wissenschaftsrat.de/download/archiv/8639-08.pdf>
- Wittmann, W. W. (2009). Evaluationsmodelle. In H. Holling (Hrsg.), *Grundlagen und statistische Methoden der Evaluationsforschung* (Enzyklopädie der Psychologie, Themenbereich B, Methodologie und Methoden, Serie IV, Evaluation, Band 1) (S. 59-98). Göttingen: Hogrefe.
- Woisch, A., Ortenburger, A. & Multrus, F. (2013). *Studienqualitätsmonitor 2012. Studienqualität und Studienbedingungen an deutschen Hochschulen*. HIS Projektbericht, Universität Konstanz. Zugriff am 04.05.2015 unter <http://www.dzhw.eu/pdf/24/sqm2012.pdf>
- Wolgast, A., Stiensmeier-Pelster, J. & Aufschnaiter, C. v. (2014). Papierbasierte oder internetbasierte Skalen zur Erfassung von Motivation (SELLMO) und Selbstkonzept (SESSKO)? *Diagnostica*, 60, 46–58.

Tabellenverzeichnis

Tabelle 1. <i>Beispiele für die Umsetzung studierendenzentrierter Lehre</i>	22
Tabelle 2. <i>Zielebenen, Forschungsmethodik und Ergebnisse in Studien zur Wirksamkeit didaktischer Weiterbildungen (Angaben nach Stes et al., 2010)</i>	29
Tabelle 3. <i>Schlüsselmerkmale effektiver Weiterbildungen</i>	30
Tabelle 4. <i>Untersuchungspläne nach Rost (2007)</i>	31
Tabelle 5. <i>Anteil der Studien bei Stes et al.(2010) mit Vortestung und Kontrollgruppe</i>	33
Tabelle 6. <i>Implikationen für die hochschuldidaktische Forschung</i>	41
Tabelle 7. <i>Zeitlicher Rahmen der Dozentenschulung (DOS) 1 und 2</i>	43
Tabelle 8. <i>Themen der Dozentenschulung (DOS)</i>	44
Tabelle 9. <i>Hypothesen zur Evaluation der Dozentenschulung (DOS)</i>	48
Tabelle 10. <i>Hypothesen der Lehrendenbefragung</i>	50
Tabelle 11. <i>Kohorten der Dozentenschulung (DOS) und Lehrende in den Versuchsgruppen sowie der Vergleichsgruppe.</i>	53
Tabelle 12. <i>Geschlechterverteilung nach Gruppenzugehörigkeit</i>	54
Tabelle 13. <i>Kontrollvariablen in der Lehrendenbefragung</i>	58
Tabelle 14. <i>Deskriptive Statistiken der Skalen in der Lehrendenbefragung</i>	62
Tabelle 15. <i>Interkorrelationen der abhängigen Variablen</i>	64
Tabelle 16. <i>Lehrerfahrung: Anzahl der unterrichteten Semester, Anzahl der Lehrformate und bereits besuchte didaktische Weiterbildungen.</i>	65
Tabelle 17. <i>Lehrerfahrung in Lehrformaten</i>	66
Tabelle 18. <i>Präferenz von Veranstaltungsformaten</i>	67
Tabelle 19. <i>Stellenwert der Lehre</i>	68
Tabelle 20. <i>Akzeptanz aktivierender Lehr-und Lernmethoden in Vorlesungen</i>	68
Tabelle 21. <i>Allgemeine didaktische Handlungen in der Lehre (Lehrendenbefragung)</i> ..	69
Tabelle 22. <i>Überblick über die Hypothesen und Fragestellungen in der Lehrendenbefragung.</i>	71
Tabelle 23. <i>Deskriptive Ergebnisse der abhängigen Variablen, getrennt nach Gruppen (Lehrendenbefragung)</i>	72
Tabelle 24. <i>Studierendenzentrierung in Vorlesungen.</i>	73
Tabelle 25. <i>Lehrkompetenz</i>	74
Tabelle 26. <i>Selbstwirksamkeitserwartungen</i>	75
Tabelle 27. <i>Umsetzung studierendenzentrierter Lehre</i>	76

Tabelle 28. <i>Lernziele</i>	77
Tabelle 29. <i>Teilnehmeranzahl und Dauer der untersuchten Lehrveranstaltungen</i>	86
Tabelle 30. <i>Beobachtungskategorien und Beobachterübereinstimmung</i>	87
Tabelle 31. <i>Didaktische Handlungen (Basisdaten Beobachtungsstudie)</i>	88
Tabelle 32. <i>Konstruktion des Beobachtungsbogens</i>	92
Tabelle 33. <i>Items des BEO-LV</i>	95
Tabelle 34. <i>Statistische Kennwerte, Beurteilerübereinstimmung und Interraterreliabilität in 107 Beobachtungsintervallen.</i>	98
Tabelle 35. <i>Didaktische Handlungen in den Übungsvideos</i>	103
Tabelle 36. <i>Beurteilerübereinstimmung und Anteil richtiger Antworten.</i>	104
Tabelle 37. <i>Stichprobenbeschreibung mit Unterschiedstests</i>	109
Tabelle 38. <i>Deskriptive Statistik der beobachteten Lehrveranstaltungen (N = 109)</i> ...	111
Tabelle 39. <i>Interkorrelationen der abhängigen Variablen und der Teilnehmerzahl in den Lehrveranstaltungen (N = 109).</i>	112
Tabelle 40. <i>Ergebnisse der Beobachtungsdaten und Hypothesenprüfung</i>	113
Tabelle 41. <i>Ergebnisse der Beobachtungsdaten und Hypothesenprüfung (nur Seminare und Vorlesungen)</i>	116
Tabelle 42. <i>Ergebnisse der Beobachtungsdaten und Gruppenvergleiche (sonstige didaktische Handlungen)</i>	113
Tabelle 43. <i>Vorliegende Beobachtungsbögen im Bearbeitungszeitraum</i>	122
Tabelle 44. <i>Deskriptive Darstellung der Lehrformate</i>	123
Tabelle 45. <i>Deskriptive Darstellung aller beobachteten Lehrhandlungen in der kollegialen Hospitation</i>	124
Tabelle 46. <i>Deskriptive Statistik Studierendenbefragung (Vorstudie)</i>	132
Tabelle 47. <i>Skalen in der Studierendenbefragung</i>	136
Tabelle 48. <i>Demographische Angaben der Studierenden</i>	138
Tabelle 49. <i>Deskriptive Statistik Studierendenbefragung</i>	139
Tabelle 50. <i>Interkorrelationen der Skalen in der Studierendenbefragung</i>	141
Tabelle 51. <i>Ergebnisse der Signifikanztestungen (Studierendenbefragung)</i>	142
Tabelle 52. <i>Akzeptanz aktivierender Methoden nach Fachmerkmalen</i>	143
Tabelle 53. <i>Annahmen der Trainer zur Effektivität der Dozentenschulung</i>	149
Tabelle 54. <i>Zusammenfassung der Ergebnisse der Lehrendenbefragung (LB) und der Beobachtungsstudie (BS).</i>	152

Abbildungsverzeichnis

<i>Abbildung 1.</i> Rahmenmodell zur Wirkungsanalyse des Hochschulunterrichts nach Helmke und Schrader (2010, S. 275).	17
<i>Abbildung 2.</i> Zusammenhang von Lehrkonzeption, Lehransatz und Lehrverhalten (nach Lübeck, 2009, S. 40).	19
<i>Abbildung 3.</i> Komponenten und Dimensionen von Lehransätzen (Kember & Kwan, 2000, S. 476)	20
<i>Abbildung 4.</i> Empirische Studie 1: Versuchsgruppen und Vergleichsgruppe	51
<i>Abbildung 5.</i> Beobachtungsbogen bei der Erhebung der Basisdaten	87

Abkürzungsverzeichnis

Abb.	Abbildung
α	Cronbachs Alpha
ÄApprO	Approbationsordnung für Ärzte
ATI	Approaches to Teaching Inventory
ATI-R	Approaches to Teaching Inventory-revised
AV	Abhängige Variable
Bsp.	Beispiel
BEO-LV	Beobachtungsbogen für Lehrveranstaltungen
Dipl.	Diplom
DOS	Dozentenschulung
d	Effektstärke d
HeiCuMed	Heidelberger Curriculum Medicinale
H	Hypothese
ICC _{unjust}	unjustierte Intraklassenkorrelation
κ	Cohens Kappa
M	Arithmetischer Mittelwert
Medi-KIT	Medizinisches Kommunikations- und Interaktionstraining
N	Anzahl der Studienteilnehmer
p	Irrtumswahrscheinlichkeit
POL	Problemorientierten Lernen
PÜ ₊	prozentuale Auftretensbeobachterübereinstimmung
PÜ ₋	prozentuale Nichtauftretensbeobachterübereinstimmung
PÜ	Prozentuale Übereinstimmung
r	Korrelationskoeffizient
S.	Seite
SD	Standardabweichung
SELLMO-ST	Skalen zur Erfassung der Lern- und Leistungsmotivation für Studierende
vgl.	vergleiche
VGG	Vergleichsgruppe
VSG	Versuchsgruppe
VSG1	Versuchsgruppe in der Praxisphase der DOS
VSG2	Versuchsgruppe deren Besuch der DOS mindestens ein Jahr zurücklag

VSG_{ALT}	Versuchsgruppe deren DOS mindestens ein Jahr zurücklag
VSG_{POST}	Versuchsgruppe nach Besuch der DOS (2. Messzeitpunkt)
$VSG_{PPÄ}$	Versuchsgruppe vor Besuch der DOS (1. Messzeitpunkt)
η^2	Effektstärke Partielles Eta Quadrat
%	Prozent
&	und

Erklärung gemäß § 8 (1) c) und d) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Erklärung gemäß § 8 Abs. (1) c) und d) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

Vorname Nachname _____

Datum, Unterschrift _____

Anhang

Anhang A1. Anschreiben an DOS Teilnehmer

Liebe/r ... (Anrede)

vor kurzem haben Sie an der Dozentenschulung teilgenommen, nun stehen Ihre überarbeiteten Lehrveranstaltungen bevor.

Wir untersuchen, welchen Einfluss die didaktischen Weiterbildungen auf die medizinische Lehre haben. Mit einer Teilnahme an dieser Evaluation ermöglichen Sie uns, Ihnen und den zukünftigen Teilnehmern die Wirkung der Dozentenschulung auf die Lehre aufzuzeigen.

Folgendes Vorgehen ist geplant:

Ich (oder meine Kollegin Janine Kahmann) besuche eine Ihrer Lehrveranstaltungen und schaue (gemeinsam mit Ihrem Tandempartner), welche didaktischen Elemente Sie einsetzen. Grundlage für die Hospitation ist ein wissenschaftlich geprüfter Beobachtungsbogen (keine Bewertung!). Für die Teilnahme ist es unwichtig, welchem Fachbereich Sie angehören, wie Sie Ihre Lehre gestalten oder wie gut Sie Ihre eigene Lehrkompetenz einschätzen! Ihr Vorteil: Auf Wunsch erhalten Sie eine ergänzende Rückmeldung zu Ihrer Lehre, gerne mit von Ihnen selbst ausgewählten Beobachtungsschwerpunkten und (auf Wunsch und wenn möglich) mit zusätzlicher Videoanalyse.

Wir würden uns sehr freuen, wenn Sie das Projekt unterstützen! Sollten Sie weitere Fragen haben, kontaktieren Sie mich bitte einfach!

Senden Sie für Ihre Teilnahme bitte eine Antwortemail zurück und tragen Sie hier ein, wann und wo Sie Ihre Lehrveranstaltung halten:

ORT:

ZEIT:

Gewünschter Beobachtungsschwerpunkt (optional):

Vielen Dank und schöne Grüße!

Gerald Wibbecke und das DOS-Team

P.S. Es liegt selbstverständlich eine zustimmende Bewertung der Ethikkommission vor. Alle Daten werden garantiert ohne die Nennung von Namen, Fachbereichen und Themen gespeichert. Die Teilnahme an der Untersuchung kann jederzeit ohne Angabe von Gründen beendet werden. Ebenso kann jederzeit die Erlaubnis zur wissenschaftlichen Verwertung der Daten zurückgezogen und eine Löschung der Daten verlangt werden.

P.P.S. Falls Sie später das Gesamtzertifikat Hochschuldidaktik Baden-Württemberg anstreben, können Sie die Ergebnisse in die schriftliche Abschlussreflexion einfließen lassen.

Anhang A2. Anschreiben mit der Bitte zur Teilnahme an der Lehrendenbefragung

Liebe Kollegen,

wir bitten Sie an der Abschlussevaluation der Dozentenschulung (Nr. der DOS-Kohorte) teilzunehmen!

<http://www.heile2012.de/DOS2Absolventen/>

Vielleicht erinnern Sie sich daran, dass Sie vor der ersten Dozentenschulung bereits einen ähnlichen Fragebogen ausgefüllt haben.
Für aussagekräftige Ergebnisse ist es wichtig, dass möglichst viele DOS-2 Absolventen teilnehmen.

Die Befragung ist anonym. Die Daten werden ohne Namensnennung oder der Nennung eines Fachbereichs gespeichert und analysiert.
Die Codes werden nach Abschluss der Befragung umgehend gelöscht.

Sie benötigen maximal 10 Minuten für diese Befragung. Gerne senden wir Ihnen später die Gesamtergebnisse der Evaluation zu!

Vielen Dank!

Ihr DOS-Team

Anhang A3. Anschreiben an Absolventen der DOS

Sehr geehrte Damen und Herren,

vor einiger Zeit haben Sie an der Dozentenschulung der Medizinischen Fakultät teilgenommen. Nun würden wir von Ihnen gern eine ehrliche Rückmeldung zur langfristigen Wirkung von DOS auf die medizinische Lehre erhalten. Mit Ihrer möglichen Teilnahme an dieser Studie ermöglichen Sie es uns, die Weiterentwicklung von DOS effektiv zu planen.

Die Studie beinhaltet zwei Aspekte der Lehre:

1. Der Fragebogen „HeiLe“. Hier geht es um Ihre Einstellungen zu der Lehre und zu DOS. Die Teilnahme nimmt etwa 10 Minuten in Anspruch. Die Beantwortung ist vollkommen anonym.

Hier geht es zum Fragebogen: <http://www.heile2011.de/Absolventen/>

2. Der Besuch einer beliebigen Lehrveranstaltung von Ihnen. Grundlage für die Hospitation ist ein wissenschaftlich geprüfter Beobachtungsbogen.

Wichtig: Eine Bewertung Ihrer Lehre o.ä. findet hier nicht statt. Es geht nur um eine formelle Analyse der Lehrhandlungen. Es ist dafür unwichtig, welchem Fachbereich Sie angehören, wie Sie Ihre Lehre gestalten oder wie gut Sie Ihre eigene Lehrkompetenz einschätzen!

Ihr Vorteil: Auf Wunsch erhalten Sie eine ergänzende Rückmeldung zu Ihrer Lehre, gerne mit von Ihnen selbst ausgewählten Beobachtungsschwerpunkten.

Senden Sie für Ihre Teilnahme bitte eine Antwort-Email zurück an: gerald.wibbecke@med.uni-heidelberg.de und tragen Sie dafür bitte hier ein, wann und wo Sie Ihre Lehrveranstaltung halten:

ORT:

ZEIT:

Beobachtungsschwerpunkt (optional):

Sie erhalten dann eine Terminbestätigung. Darüber hinaus entsteht kein weiterer Zeitaufwand für Sie! Wir würden uns sehr freuen, wenn Sie das Projekt unterstützen! Sie können auch nur an einer der beiden Optionen teilnehmen! Sollten Sie weitere Fragen haben, kontaktieren Sie bitte gerald.wibbecke@med.uni-heidelberg.de!

Vielen Dank und schöne Grüße!

das DOS-Team

P.S. Es liegt selbstverständlich eine zustimmende Bewertung der Ethikkommission vor. Alle Daten werden ohne die Nennung von Namen, Fachbereichen und Themen gespeichert. Die Teilnahme an der Studie kann jederzeit ohne Angabe von Gründen beendet werden.

Anhang A4. Anschreiben an Lehrende ohne DOS

Sehr geehrte Damen und Herren,

im Rahmen eines Forschungsprojektes der Integrativen Lehrberatung untersuche ich aktuell, wie die Lehre an der medizinischen Fakultät gestaltet wird und welchen Einfluss die didaktischen Weiterbildungen auf diese haben (können).

Daher frage ich hiermit an, ob ich eine Ihrer Lehrveranstaltung hospitieren darf und bitte Sie einen Fragebogen zu bearbeiten:

Wenn Sie nie an der Dozentenschulung der medizinischen Fakultät (DOS) teilgenommen haben: <http://www.heile2011.de/>

Wenn Sie bereits die Dozentenschulung der medizinischen Fakultät (DOS) absolviert haben: <http://www.heile2011.de/Absolventen/>

Der praktische Teil der Studie:

Ich besuche eine Ihrer Lehrveranstaltungen an der medizinischen Fakultät. Grundlage für die Hospitation ist ein wissenschaftlich geprüfter Beobachtungsbogen (keine Bewertung!). Für die Teilnahme ist es unwichtig, welchem Fachbereich oder welcher Klinik Sie angehören, wie Sie Ihre Lehre gestalten oder wie gut Sie Ihre eigene Lehrkompetenz einschätzen!

Senden Sie für Ihre Teilnahme bitte eine Antwort-E-Mail zurück und tragen Sie hier ein, wann und wo Sie Ihre Lehrveranstaltung halten:

Ort:

Zeit:

Beobachtungsschwerpunkt (optional):

Ihr Vorteil: Auf Wunsch (!) erhalten Sie eine kostenlose, individuelle und professionelle Rückmeldung zu Ihrer Lehre. Sie können aber auch ohne Feedback an der Studie teilnehmen.

Unser Ziel ist es eine große und repräsentative Studie zu erstellen. Ich würde mich daher sehr freuen, wenn Sie das Projekt unterstützen! Sollten Sie weitere Fragen zu dem Forschungsprojekt haben, kontaktieren Sie mich bitte einfach!

Vielen Dank!

Gerald Wibbecke

P.S. Es liegt selbstverständlich eine zustimmende Bewertung der Ethikkommission vor. Alle Daten werden garantiert ohne die Nennung von Namen, Fachbereichen und Themen gespeichert. Die Teilnahme an der Untersuchung kann jederzeit ohne Angabe von Gründen beendet werden.

Anhang A5. Anschreiben an die Studierenden

Liebe Studierende,

in dieser Online-Befragung geht es um die Gestaltung der Lehre an der Medizinischen Fakultät Heidelberg.

Die zentrale Frage dieser Studie lautet:

Bevorzugen Sie eher Lehrveranstaltungen, in denen ausschließlich die Lehrenden Wissen präsentieren oder bevorzugen Sie Lehrveranstaltungen in denen aktivierende Lehrmethoden eingesetzt werden?

Im Rahmen der hochschuldidaktischen Beratung von Lehrenden arbeiten wir daran, die Lehre zu verbessern. Mit Ihrer möglichen Teilnahme helfen Sie uns dabei, die didaktischen Programme für die Lehrenden stärker am Bedarf der Studierenden auszurichten.

Die Befragung dauert etwa 10-15 Minuten! Sie können zwischenspeichern und zu einem späteren Zeitpunkt vorfahren. Alle Angaben sind freiwillig. Die Ergebnisse stellen wir Ihnen selbstverständlich später zur Verfügung.

Vielen herzlichen Dank!

Gerald Wibbecke (Dipl. Psych.)
Integrative Lehrberatung Heidelberg
Medizinische Fakultät

Anhang B1. Lehrendenbefragung: Items in der Hauptstudie

Skala	Item im Wortlaut (in Klammern die Abkürzung)
Studierendenzentrierung in Vorlesungen (SZ-VL) ^a	Einzelarbeit (an Übungsaufgaben/Fällen usw.) (SZ-VL1)
	Partnerarbeit (an Übungsaufgaben/Fällen usw.) (SZ-VL2)
	Gruppenarbeit (an Übungsaufgaben/Fällen usw.) (SZ-VL3)
	Den Studierenden viele Fragen stellen. (SZ-VL4)
	Studierende viele Fragen stellen lassen. (SZ-VL5)
	Studierende miteinander diskutieren lassen. (SZ-VL6)
Didaktische Kompetenz (DK) ^b	Meine Lehrveranstaltungen gelingen mir in der Regel sehr gut. (DK1)
	Meine didaktischen Fähigkeiten sind gut. (DK2)
	Mit meinen Lehrveranstaltungen bin ich meistens sehr zufrieden. (DK3)
Selbstwirksamkeitserwartungen (SWKE) ^b	Die Gestaltung meiner Lehre hat großen Einfluss auf den späteren Prüfungserfolg der Studierenden (in meinem Thema). (SWKE1)
	Die Gestaltung meiner Lehre hat großen Einfluss auf den Lernerfolg der Studierenden. (SWKE2)
	Die Gestaltung meiner Lehre hat großen Einfluss darauf, wie intensiv sich die Studierenden mit den Inhalten beschäftigen. (SWKE3)

Anmerkungen. Antwortformat: a = 1 (In Vorlesungen überhaupt nicht geeignet), 2 (Eher selten geeignet), 3 (Gelegentlich geeignet), 4 (Häufig geeignet), 5 (In fast jeder Vorlesung geeignet), Weiß nicht. b = 1 (trifft nicht zu) bis 5 (trifft voll zu), Kann oder will ich nicht beurteilen.

Fortsetzung Anhang B1. Lehrendenbefragung: Items in der Hauptstudie

Studierendenzentrierung der Lehre (SZ) ^c	<p>Ich ermutige Studierende dazu, eigene Lösungen zu finden (z. B. in Gruppenarbeit). (SZ1)</p> <p>Ich binde alle Studierenden aktiv ein (Übungsaufgaben, Gruppenarbeit etc.). (SZ2)</p> <p>Ich gestalte die Lehre methodisch sehr abwechslungsreich. (SZ3)</p> <p>Ich kommuniziere intensiv mit den Studierenden. (SZ4)</p> <p>Ich fördere die anwendungsorientierte Auseinandersetzung der Studierenden mit den Inhalten. (SZ-L5)</p> <p>Ich lasse die Studierenden wichtige Inhalte selbst zusammenfassen. (SZ-L6)</p> <p>Ich reflektiere mit den Studierenden gemeinsam deren Lernprozess. (SZ-L7)</p>
Durchgeführte Lehrveranstaltungen (DLV) ^d	<p>Vorlesung (DLV1)</p> <p>Seminar (DLV2)</p> <p>POL (DLV3)</p> <p>Bedside Teaching (DLV4)</p> <p>Medi-Kit (DLV5)</p> <p>Skills Lab (DLV6)</p> <p>Praktikum (DLV7)</p> <p>Sonstiges</p>

Anmerkungen. Antwortformat: c = 1 (Das setze ich nicht um) bis 5 (Das setze ich um) bei indirekter Veränderungsmessung bzw. 1 (Das setze ich nun weniger um als vor einem Jahr), 3 (Das setze ich genau so viel um wie vor einem Jahr) bis 5 (Das setze ich nun mehr um als vor einem Jahr) bei direkter Veränderungsmessung, Kann ich (noch) nicht beurteilen. d = Habe ich schon mindestens einmal durchgeführt (Ja oder nein).

Fortsetzung Anhang B1. Lehrendenbefragung: Items in der Hauptstudie

Präferenz durchgeführter Lehrveranstaltungsformate (PLV) ^e	Vorlesung (PLV1) Seminar (PLV2) POL (PLV3) Bedside Teaching (PLV4) Medi-Kit (PLV5) Skills Lab (PLV6) Praktikum (PLV7)
Akzeptanz aktivieren der Methoden in Vorlesungen ^b (AAM-VL)	Die meisten Studierenden bevorzugen vorwiegend dozenten-zentrierten (frontalen) Unterricht.* (AAM-VL1) In einer Vorlesung sollte sich der Dozentenvortrag regelmäßig mit aktivierenden Lehr- und Lernmethoden abwechseln. (AAM-VL2) Viele Studierende mögen aktivierende Lehr- und Lernmethoden in Lehrveranstaltungen. (AAM-VL3)
Allgemeine didaktische Handlungen ^c (ADH)	Ich zeige die praktische Relevanz der Inhalte auf. (ADH1) Ich fördere die besten Studierenden. (ADH2) Ich strukturiere die Inhalte klar. (ADH3) Ich lehre mit einem ausgefeilten Präsentationsstil und sehr guter Rhetorik. (ADH4) Ich bereite die Studierenden auf die Prüfung(en) vor. (ADH5) Ich unterbinde auftretende Störungen konsequent. (ADH6) Ich fasse wichtige Inhalte zusammen. (ADH7) Ich versuche, die Studierenden für mein Fach zu begeistern. (ADH8) Ich helfe (schwächeren) Studierenden, die Unterstützung benötigen. (ADH9)

Anmerkungen. Antwortformat: b = 1 (trifft nicht zu) bis 5 (trifft voll zu), Kann oder will ich nicht beurteilen. c = 1 (Das setze ich nicht um) bis 5 (Das setze ich um) bei indirekter Veränderungsmessung bzw. 1 (Das setze ich nun weniger um als vor einem Jahr), 3 (Das setze ich genau so viel um wie vor einem Jahr) bis 5 (Das setze ich nun mehr um als vor einem Jahr) bei direkter Veränderungsmessung, Kann ich (noch) nicht beurteilen. e = 1 (Sehr ungerne), 2 (Eher ungerne), 3 (Weder gern noch ungern), 4 (Eher gern), 5 (Sehr gerne), Weiß nicht. * Items wurden rekodiert.

Fortsetzung Anhang B1. Lehrendenbefragung: Items in der Hauptstudie

Stellenwert der Lehre (STL) ^b	Meinen Lehrauftrag erfülle ich sehr gerne. (STL1)
	Lehre ist für mich eine unangenehme Pflicht. (STL2)*
	Meine anderen beruflichen Verpflichtungen sind mir viel wichtiger als die Lehre.(ST-L3)*
Stellenwert der Lehre im Fachbereich (STL-FB) ^b	Meine Kollegen/Kolleginnen zeigen einen hohen Einsatz für die Lehre. (STL-FB1)
	Mein Einsatz für die Lehre wird von meinen Vorgesetzten gewürdigt. (STL-FB2)
	Lehre hat insgesamt einen hohen Stellenwert in meiner Abteilung. (STL-FB3)
Zufriedenheit mit der Weiterbildung (ZWB) ^f	Die DOS hat meine Lehre insgesamt verbessert. (ZWB1)
	Bei der DOS habe ich wichtiges für meine Lehre gelernt. (ZWB2)
	Die DOS hat mir bei meiner Lehrtätigkeit geholfen. (ZWB3)
	Bei der Planung neuer Lehrveranstaltungen greife ich noch heute auf mein Wissen aus DOS zurück. (ZWB4)
	DOS hat mir insgesamt gut gefallen. (ZWB5)
Weiterbildung ^g	Haben Sie schon andere didaktische Weiterbildungen neben DOS besucht?
Motivation für die Lehre ^h	Ich glaube, dass meine Motivation für die Lehre...
Semester Lehrerfahrung	Wie viele Semester haben Sie schon Lehrveranstaltungen durchgeführt?
Weiterhin didaktische Weiterbildung ^b	Ich denke, ich werde auch in Zukunft didaktische Weiterbildungen besuchen.
Geschlecht	Ihr Geschlecht

Anmerkungen. DOS = Dozentenschulung. Antwortformat: b = 1 (trifft nicht zu) bis 5 (trifft voll zu), Kann oder will ich nicht beurteilen. f = 1 (niedrige Zustimmung) bis 5 (hohe Zustimmung). g = ja /nein. h = ...kleiner geworden ist/ ...etwa gleich geblieben ist/ ...größer geworden ist.

* Items wurden rekodiert.

Anhang B2. Lehrendenbefragung: Deskriptive Statistiken der Items

Item	VGG			VSG _{PRÄ}			VSG _{POST}			VSG _{ALT}		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
SZ-VL1	2.03	0.88	71	2.05	0.90	110	2.87	1.00	69	2.53	0.94	62
SZ-VL2	2.27	0.84	71	2.15	0.83	110	3.09	0.94	69	2.90	1.12	63
SZ-VL3	2.25	0.85	72	2.27	0.94	110	2.75	1.01	69	2.37	1.01	63
SZ-VL4	3.69	0.92	71	3.55	0.87	112	3.54	1.09	69	3.56	0.95	63
SZ-VL5	3.92	0.90	72	3.76	0.94	112	3.74	0.92	69	3.71	0.97	63
SZ-VL6	2.50	1.14	70	2.73	1.13	110	3.20	1.13	69	2.73	1.22	63
DK1	3.60	0.71	55	3.56	0.79	97	3.81	0.58	69	3.81	0.66	58
DK2	3.32	0.74	56	3.00	0.92	108	3.55	0.78	69	3.66	0.72	56
DK3	3.43	0.73	44	3.24	0.86	62	3.78	0.75	69	3.75	0.65	60
SWKE1	3.45	0.95	56	3.51	0.95	90	3.46	1.01	63	3.54	0.86	54
SWKE2	3.85	0.91	59	3.76	0.92	98	3.88	0.73	66	3.78	0.70	58
SWKE3	3.95	0.90	66	3.76	0.93	101	3.97	0.84	69	3.77	0.68	57
SZ1	3.44	0.99	66	3.33	1.16	111	3.55	1.02	69	3.37	1.10	63
SZ2	3.45	0.97	67	3.42	1.06	106	3.56	1.00	70	3.41	1.09	63
SZ3	3.00	0.92	65	3.04	1.11	104	3.35	0.84	69	3.29	0.80	62
SZ4	3.72	0.90	71	3.84	1.03	116	4.04	0.84	71	4.10	0.82	63
SZ5	3.79	0.89	66	3.79	0.96	109	4.00	0.76	67	3.90	0.90	62
SZ6	2.86	1.13	71	3.06	1.22	111	3.06	1.17	69	3.11	1.15	62
SZ7	2.84	1.04	69	2.84	1.13	109	3.09	1.13	67	2.89	1.18	63
LZ1	3.78	1.08	69	3.53	1.17	113	3.99	0.99	69	4.16	0.97	63
LZ2	3.82	0.93	71	3.64	1.10	107	3.88	0.83	69	4.26	0.68	62

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). SZ-VL = Studierendenzentrierung in Vorlesungen. DK = Didaktische Kompetenz. SWKE = Selbstwirksamkeitserwartungen in der Lehre. SZ-L = Studierendenzentrierung der Lehre. LZ = Lernziele.

Fortsetzung Anhang B2. Lehrendenbefragung: Deskriptive Statistiken der Items												
Item	VGG			VSG _{PRÄ}			VSG _{POST}			VSG _{ALT}		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
SZ1-DV	3.04	0.65	27				3.79	0.74	71			
SZ2-DV	3.19	0.57	26				3.69	0.77	71			
SZ3-DV	3.17	0.48	24				3.90	0.73	70			
SZ4-DV	3.29	0.54	28				3.46	0.67	70			
SZ5-DV	3.08	0.64	25				3.65	0.74	71			
SZ6-DV	2.96	0.58	28				3.34	0.72	70			
SZ7-DV	2.85	0.73	26				3.48	0.82	67			
LZ1-DV	3.22	0.75	27				3.90	0.77	69			
LZ2-DV	3.15	0.68	26				3.53	0.80	68			
AAM_VL1	3.14	0.96	69	2.92	1.05	110	2.77	1.00	66	2.89	0.94	62
AAM_VL2	4.07	0.97	70	3.65	1.07	115	3.91	1.02	70	3.92	0.95	62
AAM_VL3	3.71	0.93	69	3.36	0.99	111	3.26	0.87	68	3.26	0.83	62
ADH1	4.31	0.79	70	4.05	1.09	116	4.45	0.77	71	4.40	0.84	62
ADH2	2.83	1.03	66	2.75	1.10	104	3.17	1.00	66	2.89	0.95	61
ADH3	3.96	0.82	68	3.87	1.05	111	4.21	0.74	68	4.24	0.69	63
ADH4	2.95	0.79	60	2.88	0.99	93	3.13	0.70	64	3.33	0.82	54
ADH5	3.78	0.86	68	3.53	1.09	105	3.54	1.04	68	3.85	0.79	62
ADH6	3.34	1.05	67	3.13	1.07	107	3.27	1.02	70	3.39	1.08	61
ADH7	4.23	0.89	70	3.95	1.11	114	4.28	0.72	71	4.37	0.81	63
ADH8	4.32	0.90	69	4.05	1.16	116	4.46	0.71	71	4.48	0.84	63
ADH9	3.26	1.07	66	3.19	1.23	106	3.29	0.96	65	3.43	0.92	61

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). SZ-L-DV = Studierendenzentrierung der Lehre (direkte Veränderungsmessung). LZ-L-DV = Lernziele in der Lehre (direkte Veränderungsmessung). AAM-VL = Akzeptanz aktivierender Methoden in Vorlesungen. ADH = Allgemeine didaktische Handlungen.

Fortsetzung Anhang B2. Lehrendenbefragung: Deskriptive Statistiken der Items

Item	VGG			VSG _{PRÄ}			VSG _{POST}			VSG _{POST}		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
STL1	4.27	0.88	70	4.04	1.00	114	4.22	0.88	72	4.56	0.62	62
STL2	1.60	0.83	72	1.78	1.03	116	1.71	1.01	72	1.35	0.70	62
STL3	2.77	1.01	70	3.03	1.06	115	3.00	1.08	71	2.71	1.05	62
STL-FB1	3.76	0.89	63	3.57	0.89	112	3.46	1.02	67	3.69	0.90	59
STL-FB2	3.03	1.31	62	3.00	1.10	109	2.59	1.10	68	2.68	1.33	59
STL-FB3	3.33	1.17	69	3.45	1.15	114	2.97	1.16	72	2.95	1.20	61
ZWB1							3.66	1.08	71	3.63	1.22	60
ZWB2							3.73	1.18	71	3.75	1.27	60
ZWB3							3.63	1.13	70	3.60	1.18	60
ZWB4							3.58	1.27	71	3.58	1.18	60
ZWB5							3.73	1.16	71	3.78	1.25	60

Anmerkungen. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). STL = Stellenwert der Lehre. STL-FB = Stellenwert der Lehre im Fachbereich. ZWB = Zufriedenheit mit der Weiterbildung.

Anhang B3. Lehrendenbefragung: Interkorrelationen von abhängigen Variablen mit Kontrollvariablen

	VGG, VSG _{PRÄ} und VSG _{ALT}				VSG _{POST}			
	LE	STL	STL-FB	AAM-VL	LE	STL	STL-FB	AAM-VL
SZ-VL	.04	.14	.08	.19	.05	-.01	-.02	.58
DK	.24	.32	.25	.01	.17	.32	.30	.11
SWKE	-.14	.24	.22	.08	.28	.39	.19	.13
SZ	.09	.35	.26	.21	-.03	.28	.21	.25
LZ	.05	.26	.18	.03	.02	.23	.12	-.08
Reaktion	-.09*	.19*	.12*	.21*	-.15	.27	.05	.28
WB	-.03*	.37*	.23*	.38*	-.01	.40	.23	.13

Anmerkungen. Alle Korrelationen nach Pearson, mit Ausnahme der Variablen WB (Korrelation nach Spearman). * Hier sind nur die Werte der VSG einbezogen. Signifikante Korrelationen ($p \leq .05$) sind fett markiert. VGG = Vergleichsgruppe. Versuchsgruppen: Lehrende vor (VSG_{PRÄ}) und nach (VSG_{POST}) Besuch der Dozentenschulung (DOS) sowie Lehrende, deren Besuch der DOS mindestens ein Jahr zurückliegt (VSG_{ALT}). LE = Lehrerfahrung (in Semestern). STL = Stellenwert der Lehre. STL- FB = Stellenwert der Lehre im Fachbereich. AAM-VL = Akzeptanz aktivierender Methoden in Vorlesungen. SZ-VL = Studierendenzentrierung in Vorlesungen. DK-L = Didaktische Kompetenz. SWKE = Selbstwirksamkeitserwartungen in der Lehre. SZ = Studierendenzentrierung. LZ = Lernziele. WB = Weiterbildungsbereitschaft.

Anhang B4. Lehrendenbefragung: Interkorrelationen von direkter und indirekter Veränderungsmessung (VM) bei Studierendenzentrierung (SZ) und Lernzielen (LZ).

	SZ-direkte VM	SZ 2.MZP	LZ-direkte VM	LZ-2.MZP
SZ-Differenz Indirekte VM	-.03	.33		
SZ-direkte VM		.56		
LZ-Differenz Indirekte VM			.01	.50
LZ-direkte VM				.39

Anmerkungen. MZP = Messzeitpunkt. Alle Korrelationen nach Pearson. Signifikante Korrelationen ($p \leq .05$) sind fett markiert.

Anhang B5. Lehrendenbefragung: Unterschiedstestungen von Lehrenden mit und ohne gültigen Code (in der Versuchsgruppe vor Besuch der Dozentenschulung).

	Mit Code			Ohne Code			Vergleich		
	M	SD	N	M	SD	N	t-Wert	df	p
SZ-VL	2.76	0.61	40	2.75	0.59	70	-0.43	108	.965
DK	3.32	0.59	35	3.28	0.62	61	-0.27	94	.787
SWKE	3.67	0.72	36	3.68	0.79	64	0.10	98	.922
SZ	3.26	0.73	40	3.38	0.80	70	0.79	108	.432
Lernziele	3.24	1.03	39	3.79	0.94	67	2.80	104	.006
LE	5.76	3.09	42	5.62	3.17	73	-0.24	113	.812
Erfahrung mit LF	3.55	1.35	42	3.17	1.35	75	-1.44	115	.153
Stellenwert Lehre	3.85	0.68	42	3.68	0.81	73	-1.14	113	.258
Stellenwert Lehre Team	3.32	0.84	40	3.36	0.72	71	0.22	109	.829
AAM-VL	3.38	0.72	41	3.31	0.80	72	-0.48	111	.632

Anmerkungen. Nur Versuchsgruppe vor der Dozentenschulung (VSG_{PRÄ}). DK = Didaktische Kompetenz . SZ-VL = Studierendenzentrierung in Vorlesungen. SWKE = Selbstwirksamkeitserwartungen in der Lehre. SZ = Studierendenzentrierung. LE = Lehrerfahrung (in Semestern). LF = Lehrveranstaltungsformate (Anzahl). AAM-VL = Akzeptanz aktivierender Methoden in Vorlesungen.

Anhang B6. Lehrendenbefragung: Unterschiedstestungen allgemeine didaktische Handlungen

	VGG - VSG _{PRÄ}		VSG _{PRÄ} - VSG _{POST}		VGG - VSG _{ALT}		VSG _{POST} - VSG _{ALT}	
	z	p	z	p	z	p	z	p
Ich zeige die praktische Relevanz der Inhalte auf	-1.30	.195	-2.56	.011	-.98	.325	-.19	.852
Ich fördere die besten Studierenden	-.81	.419	-2.74	.006	-.32	.747	-1.56	.120
Ich strukturiere die Inhalte klar	-.01	.990	-1.95	.051	-2.01	.044	-.14	.893
Ich unterbinde auftretende Störungen konsequent	-.27	.786	-1.50	.134	-2.32	.020	-1.23	.218
Ich fasse wichtige Inhalte zusammen	-1.28	.199	-.08	.932	-.37	.715	-1.61	.108
Ich lehre mit einem ausgefeilten Präsentationsstil und sehr guter Rhetorik	-1.07	.286	-.73	.466	-.31	.757	-.66	.508
Ich bereite die Studierenden auf die Prüfung(en) vor	-1.55	.120	-1.59	.111	-.92	.355	-1.07	.284
Ich versuche, die Studierenden für mein Fach zu begeistern	-1.35	.177	-2.13	.033	-1.23	.220	-.64	.521
Ich helfe (schwächeren) Studierenden, die Unterstützung benötigen	-.21	.833	-.34	.731	-.93	.355	-.83	.404

Anmerkungen. VGG = Vergleichsgruppe. VSG = Versuchsgruppe. Unterschiedstestungen mit dem Mann-Whitney-U-Test (Zweiseitige Testung. Signifikante Unterschiede ($p \leq .05$) sind fett markiert. Da hier mit SPSS (Version 21.0.0.0) gerechnet wurde, sind die z-Werte immer negativ.

Anhang C1. Version 1 des Beobachtungsbogens

Beobachtungsbogen

Thema: _____ **Lehrende/r:** _____
Datum: _____ **Teilnehmer:** _____ (Anzahl)

Begrüßung ¹ Beginnt pünktlich
Persönliche Vorstellung² Keine Nennt/Zeigt Namen Nennt Namen und persönlichen Bezug zum Thema

Einleitung (Wenn keine, dann siehe **Vorträge**)

Themenhinführung (Frage, Problemstellung, Zitat...) Visualisiert zum Einstieg (Bild, Film, Demonstration...)
 Stellt Lernziele ausdrücklich dar Zeigt Struktur des Themas auf
 Stellt Relevanz (für Teilnehmer, Praxis...) dar Fragt Vorwissen/Erwartungen/Interesse usw. ab
 Sonstiges wie? *Methoden bitte bei **Methodeneinsatz** vermerken!*

Präsentation Verwendet PowerPoint Verwendet Flip-Chart/Tafel Sonstige Medien

Hält Blickkontakt zu Teilnehmern: Ja Teils teils Nein, oft abgewandt

Spricht frei ? Ja Teils teils Nein, liest oft ab

Spricht deutlich ? Ja Teils teils Nein, häufig undeutlich

Wie wird Wissenserwerb in kollektiven Lernphasen erleichtert?

Gibt Strukturierungshilfen (Überblick/Zusammenhänge...) Gibt Zwischenzusammenfassungen

Stellt Relevanz der Inhalte (für Prüfungen, Praxis...) dar Visualisiert (Bild, Film, Demonstration...)

Verweist ausdrücklich auf Lernziele Verweist auf Vorwissen

Sonstiges wie?

Gesamte Zeit kollektiver Lernphasen: _____

Setzt neben kollektiven Lernphasen auch individuelle Lernphasen ein (dann siehe **Methodeneinsatz**)

Methodeneinsatz / Individuelle Lernphasen

#	Methode/Sozialform ³	Instruktion ⁴	Synchronisation ⁵	
(1)		was <input type="radio"/> wie <input type="radio"/> mit wem <input type="radio"/> wie lange <input type="radio"/>	Ja <input type="radio"/>	Gesamtzeit: _____
(2)		was <input type="radio"/> wie <input type="radio"/> mit wem <input type="radio"/> wie lange <input type="radio"/>	Ja <input type="radio"/>	Gesamtzeit: _____
(3)		was <input type="radio"/> wie <input type="radio"/> mit wem <input type="radio"/> wie lange <input type="radio"/>	Ja <input type="radio"/>	Gesamtzeit: _____

Kommunikation/Interaktion

Fragt nach: Wissen Interesse Meinung Ergänzungen Unklarheiten Sonstiges

Anzahl der Antworten von Teilnehmern: Mehr (als fünf) Antworten

Teilnehmer stellen von sich aus Fragen: Mehr (als fünf) Fragen

Teilnehmer halten Vorträge, Referate usw. Es ergeben sich Diskussionen

Umgang mit Störungen durch Teilnehmer

Treten Störungen auf? Nein Ja Wenn ja, welche?

Beseitigt Störungen aktiv Ja Teils teils Nein Bleibt dabei freundlich Ja Teils teils Nein

Ausstieg

Fasst wichtigste Ergebnisse zusammen (Resümee...) Gibt einen Ausblick

Gibt Hinweise zur Vertiefung Verabschiedet Teilnehmer

Sonstiges wie?

*Methoden bitte bei **Methodeneinsatz** vermerken!*

Anmerkungen: ¹ Kreise nur ankreuzen, wenn Verhalten gezeigt wird. ² Bei unterstrichenen Fragen nur eine Antwortoption wählen! ³ Kurzbeschreibung der Methode (z.B. One-Minute-Paper in Einzelarbeit...) ⁴ Ankreuzen, ob Arbeitsauftrag ausführlich instruiert wird ⁵ Werden Ergebnisse individueller Lernphasen präsentiert/abgefragt...?

**Beobachtungsbogen für
Lehrveranstaltungen
BEO-LV**

Einleitung

Lehrhospitationen sind ein wesentlicher Bestandteil didaktischer Weiterbildungen. In der Hochschuldidaktik gibt es ganz unterschiedliche Vorgehensweisen bei Lehrhospitationen. Einige Beobachter verwenden eine Checkliste mit Beobachtungsschwerpunkten, andere beobachten frei, und wieder andere erstellen ein Ablaufprotokoll. Diese Herangehensweisen können hilfreich für die unmittelbare Rückmeldung an die Lehrenden sein, sie haben aber einen Nachteil: Sie generieren keine quantitativen Daten, die wir für Studien nutzen können. Der BEO-LV soll diese beiden wichtigen Funktionen gleichzeitig erfüllen: Sie erhalten eine Grundlage für eine unmittelbare Rückmeldung nach der Lehrveranstaltung und können Ihre Beobachtung quantitativ auswerten.

Mit dem BEO-LV protokollieren Sie das Auftreten didaktischer Handlungen in Lehrveranstaltungen. Dazu markieren Sie mit einem Kreuz (X) oder einem Strich (|) die entsprechende Kategorie in dem entsprechenden Zeitintervall. Am Ende einer Lehrveranstaltung erhalten Sie so eine Übersicht, welche didaktischen Handlungen aufgetreten sind. Die folgende Beschreibung erläutert Ihnen die zugrunde liegenden Konzepte.

Mit dem BEO-LV beobachten wir einen Teilaspekt des Lehr- und Lernprozesses, die Handlungen, die durch Lehrende initiiert werden, bzw. die „direkte Instruktion“ (siehe z. B. Weinert, 1996; Wellenreuther, 2004). Direkte Instruktion beinhaltet eine Vielzahl von möglichen Lehrmethoden: Frontalunterricht, fragend-entwickelnder Unterricht, Diskussionen, sowie Stillarbeit und kooperatives Lernen. Wir wissen, dass die Lehrenden die akademischen Leistungen niemals vollständig determinieren (siehe z. B. das Rahmenmodell zur Wirkungsanalyse des Hochschulunterrichts, Helmke & Schrader, 2010). Lehre ist ein Angebot, welches von den Studierenden unterschiedlich wahrgenommen und genutzt wird. Auch die Rahmenbedingungen, in denen Lehre stattfindet, sind hier von Bedeutung. Dennoch üben Lehrende einen bedeutsamen Einfluss aus, sowohl auf die kognitiven Fähigkeiten der Lernenden, als auch auf emotionalen und motivationalen Aspekte des Lernens. Die hier dargestellten didaktischen Handlungen sind allesamt praktikable Wege, Lehre an Hochschulen zu gestalten. Situativ angemessen eingesetzt, stellen sie wichtige Komponenten gelungener Lehre dar. Sie sind dabei weitestgehend unabhängig vom Veranstaltungsformat und von der Anzahl der Studierenden.

Die didaktischen Handlungen des BEO-LV werden auf den folgenden Seiten ausführlich beschrieben:

- Die **inhaltliche Beschreibung** der Beobachtungskategorien beinhaltet die Arbeitsdefinition der didaktischen Handlungen.
- Die **Lerntheoretischen Grundlagen** stellen die didaktische Handlung in den Kontext bestehender Theorien und Modelle zum Lernen und Lehren.
- Die Belege für die **didaktische Relevanz** beziehen sich primär auf empirische Studien, in denen das Merkmal positive Wirkungen auf kognitive, emotionale oder motivationale Variablen gezeigt hat.
- **Positivbeispiele** zeigen exemplarisch auf, welche Verhaltensweisen und Aussagen die Voraussetzungen der Arbeitsdefinition erfüllen. Diese Auflistung kann nicht vollständig sein. Im Zweifelsfall sollten Sie den Wortlaut aufschreiben und später entscheiden, wie Sie es einordnen.
- **Nicht gemeint** – Beispiele zeigen auf, welche didaktischen Handlungen ausdrücklich nicht gemeint sind. Wenn möglich, wird eine besser passende Kategorie vorgeschlagen. Diese Auflistung kann nicht vollständig sein. Im Zweifelsfall sollten Sie den Wortlaut aufschreiben und später entscheiden, wie Sie es einordnen.
- Unter **Sonstiges** stehen gegebenenfalls ergänzende Informationen.

1. Problem, Fall, Herausforderung

Beschreibung	<p>Lehrende veranschaulichen und erklären hier einen Denkvorgang oder Lösungsweg. Die Suche nach einer Antwort wird hier exemplarisch dargestellt. Sie können Situationen (Probleme, Fälle, Herausforderungen, Fragen usw.) aufzeigen, für die es eine Reaktion (eine Lösung, Antwort oder Stellungnahme) zu finden bzw. zu erarbeiten gilt. Das Spektrum ist groß: hier sind sowohl lebensnahe Situationen aus dem Anwendungskontext, kritische Fragestellungen, als auch abstrakte Aufgaben denkbar. Barrows (1986) unterscheidet: Vorlesungen mit Fallbeispielen, Fallbasierte Vorlesungen, Fallmethoden und Problembasiertes Lernen. Diese Vorgehensweisen können variiert werden.</p>
Lerntheoretische Grundlagen	<p>Klauer und Leutner (2007) sehen in der Fähigkeit ein interessantes Problem aufzuwerfen, eine Möglichkeit der Motivierung Lernender. Zudem kann es den Transfer des Wissens fördern und abstraktes Wissen in konkrete Zusammenhänge übertragen. Wissen kann „träge“ sein, nämlich dann, wenn es nicht vom Kontext Hochschule auf andere Bereiche des Lebens übertragen werden kann (siehe Renkl, 2010). Positiver Transfer ist an verschiedene Voraussetzungen gebunden und stellt sich daher nicht automatisch ein (siehe Mähler & Stern, 2010). Klauer und Leutner (2007) schreiben, was Lernende konkret tun können: „Es empfiehlt sich, diese Schritte der Übertragungsleistung an neuen Problemen auch einzuüben. Ohne Einübung des Transfers kann man nicht mit einer erfolgreichen Übertragung rechnen“ (S.149). Wenn Lehrende diesen Prozess exemplarisch anhand verschiedener Situationen simulieren, erleben die Lernenden wie sie Wissen einsetzen können und wie Lösungen gefunden werden. Die Überwindung <i>kognitiver Konflikte</i> kann die Relevanz der Lerninhalte eindrücklich hervorheben. Aber auch einfache Falldarstellungen oder Beispiele einer Lösungsstrategie können das Verständnis erleichtern.</p>
Belege für didaktische Relevanz	<p>Die Mehrheit der Studierenden erwartet im Studium die Vermittlung praktischer Fähigkeiten in Lehrveranstaltungen (in Form von regelmäßigen/kontinuierliches Einbringen von Beispielen aus der Praxis, siehe HIS-Studierendenbefragung 2010). Zumbach, Haider und Mandl (2008) merken an, dass Fallbeispiele durch ihren hohen Realitätsbezug gut verstanden werden.</p> <p>Hattie (2009) berichtet von einer bedeutsamen Effektstärke von $d = 0,61$ für “Problem-solving Teaching” (welches in den dort für die Meta-Analysen zugrundeliegenden Studien als umfassende Lehrmethode eingesetzt wurde).</p>
Positivbeispiele	<ul style="list-style-type: none"> - Fallvorstellung (z. B. Krankengeschichte) - Präsentation einer Zeitungsschlagzeile oder kontroversen Frage/Aussage - Mathematische Aufgabe wird exemplarisch gelöst - Problemlösungsprozesse wie z. B. Darstellungen von durchgeführten Studien (Aufbau und Ergebnisse)
Nicht gemeint	<ul style="list-style-type: none"> - Stellt Titel oder Struktur des Themas vor (dann evtl. 3) - Erlebnis des Lehrenden (dann 5) - Ein in Zukunft oder aktuell auftretendes Problem oder ein Fall der Lernenden (dann 6) - Überfachlicher Bezug (dann 8) - Übungsaufgabe für Lernende (dann Methode) - Beispiele, die nur als Belege für einen Sachverhalt gelten.
Sonstiges	<p>Es muss ein fachlicher Bezug bei den Fallbeispielen bestehen.</p>

2. Visualisierung

Beschreibung	Visualisierungen umfassen alle Methoden, bei denen die Lerninhalte durch visuelle Informationen angereichert werden. Möglich ist hier auch die Kombination mit anderen sensorischen Eindrücken (Auditive oder haptische Eindrücke). Gemeint sind hier realistische Bilder im weiteren Sinne.
Lerntheoretische Grundlagen	Wenn Visualisierungen das Arbeitsgedächtnis entlasten, können sie das Verständnis und Behalten von Sachinformationen unterstützen. Das ist v.a. der Fall, wenn: <ol style="list-style-type: none"> 1. der begleitende Text auditiv und nicht ebenfalls visuell präsentiert wird 2. der relevante Informationsgehalt des Bildes hoch ist (siehe z. B. Klauer & Leutner, 2007) und 3. wenn die Text-Bild-Kombination optimal auf das Verständnis ausgerichtet ist (Weidenmann, 1994).
Belege für didaktische Relevanz	Die Metaanalysen von Levie und Lenz (1982) sowie von Levin, Anglin und Carney (1987) zeigen, dass Bilder (im weitesten Sinne) die Lernleistung von Textinformation steigern können, unter der Voraussetzung, dass die Bilder nicht nur dekorative Zwecke erfüllen. Allgemeingültige Empfehlungen zu einer optimalen Bildgestaltung sind auf Basis der Literatur allerdings nicht möglich (siehe Lewalter, 1997).
Positivbeispiele	<ul style="list-style-type: none"> - Bild - Film - Grafiken, - Lehrende führen etwas vor/Demonstration - Landkarte - Etwas wird herumgegeben... - Auch Kombination mit anderen Reizen sind hier möglich: Haptisch: „Tasten Sie mal...“ oder olfaktorisch „riechen Sie mal“
Nicht gemeint	<ul style="list-style-type: none"> - Bilder, die keinen thematischen Informationsgehalt haben (z. B. dekorative Bilder oder Muster) - Logische Bilder wie Diagramme und Abbildungen oder Tabellen
Sonstiges	Bei Bildern, die eine längere Zeit gezeigt werden, ohne dass zwischenzeitlich auf sie eingegangen wird, sollten Sie diese nur bei der aktiven Bezugnahme als gezeigt notieren.

3. Struktur, Systematisierung

Beschreibung	<p>Beinhaltet drei verschiedene Komponenten:</p> <ol style="list-style-type: none"> 1. Die Darstellung der Struktur eines Themas (mit der Darstellung der Unterthemen): Zeigt Struktur des Themas auf. Bekannte Methoden: Advance Organizer und Concept Maps. "Advance Organizer" sind Organisationshilfen (Ausubel, 1974, S.159), welche früh im Lernprozess von den Lehrenden präsentiert werden. Das Thema ist dort allgemein und abstrakt dargestellt, so dass der inhaltliche Zusammenhang der Teilthemen deutlich wird. Concept-Maps (oder Mind-Maps) sind noch konkreter und zeigen ebenfalls eine Vernetzung und Gliederung von Inhalten auf (diese werden aber besser von den Lernenden selbst erarbeitet). 2. Die Darstellung von Lerninhalten durch logische Bilder, also Schemata und Diagramme. Diese bewirken eine Systematisierung und Gliederung von konkreten Informationen. Sie sind im Gegensatz zu realistischen Bildern (siehe Visualisierung) wesentlich abstrakter. 3. Tabellen. Zahlen und/oder Worte in tabellarischer Form.
Lerntheoretische Grundlagen	<p>Wahl (2006, S.141f) beschreibt positive kognitive und motivationale Auswirkungen von Advance Organizer, die sich auch auf andere Strukturierungshilfen übertragen lassen. Durch Strukturierung werden komplizierte Sachverhalte so aufbereitet, dass sie leichter verstanden werden können. Das Arbeitsgedächtnis wird entlastet, wenn bestimmte Voraussetzungen erfüllt sind. Logische Bilder sollen die Verarbeitung der verbalen Informationen und die Integration der Darstellung zu einem „mental Modell“ ebenfalls erleichtern (siehe Schnotz, 2009). Schnotz führt die syntaktische und semantische Klarheit, die implizite Ordnung und das Prinzip der Sparsamkeit als vorteilhaft für die Gestaltung logischer Bilder an.</p>
Belege für didaktische Relevanz	<p>Schnotz (2009) führt eine Reihe von Belegen auf, dass logische Bilder einen positiven Einfluss auf Wissenserwerb und Verständnis haben können. Preiss und Gayle (2006) zeigen in einer Metaanalyse auf, dass auch schon mündlich präsentierte Advance Organizer einen Effekt auf den Lernerfolg haben können, wobei jüngere Lerner stärker von der Gliederungshilfe präsentieren.</p>
Positivbeispiele	<ol style="list-style-type: none"> 1. Organisationshilfen: <ul style="list-style-type: none"> - Advance Organizer - Mind-Maps - Concept-Maps 2. Logische Bilder: <ul style="list-style-type: none"> - Schemata - Diagramme 3. Tabellen: <ul style="list-style-type: none"> - Vergleichende Tabellen - Statistische Tabellen
Nicht gemeint	<ul style="list-style-type: none"> - Darstellung der Gliederung eines Vortrags „Als erstes stelle ich XY dar, dann gehe ich auf YZ ein, und dann ...“ - Beiläufige Zusammenhänge oder Vergleiche - Aufzählungen /Auflistungen unter Überschriften - Zusammenfassung wichtiger Punkte (dann 9) - Sagt eine Regel auf / gibt Merkhilfen (dann evtl. 7 oder 9)
Sonstiges	<p>Logische Bilder und Tabellen werden relativ häufig präsentiert, Organisationshilfen dagegen selten. Letztere können daher bei der Protokollierung hervorgehoben werden.</p>

4. Lernziele

Beschreibung	<p>Lehrende benennen konkret, was die Lernenden innerhalb einer Lehrveranstaltung lernen sollen. Das beinhaltet die Nennung eines Themas und die damit zu erwerbenden Fertigkeiten. Dies umfasst kognitive, affektive, soziale und motivationale Lernziele.</p> <p>Für Lernziele gibt es in der Literatur verschiedene Bezeichnungen. Hinter der Bezeichnung „Lernziel“ steht die Annahme, dass die Lernenden die „Lehrziele“ für sich selbst als Lernziel annehmen. Es gibt Autoren, die daher die Formulierung „Lernziel“ bevorzugen (siehe z. B. Klauer & Leutner, 2007 oder Brezinka, 1974). Biggs und Tang sprechen von “Intended Learning Outcome“: “Intended Learning Outcome“ clarifies what the student should be able to perform after teaching that couldn’t be performed previously... (Biggs & Tang, 2007). Hattie spricht von “Learning intentions“: “Learning intentions describe what it is we want students to learn in terms of the skills, knowledge, attitudes, and values within any particular unit or lesson. Learning intentions should be clear, and provide guidance to the teacher about what to teach, help learners be aware of what they should learn from the lesson, and form the basis of assessing what the students have learnt and for assessing what the teachers have taught well to each student” (Hattie, 2009, S. 162). Weiterhin treten im englischsprachigen Raum die Begriffe “ Learning Objectives“ und „learning outcomes“ auf. Allen hier genannten Begriffen ist gemeinsam, dass sie einen Zustand beschreiben, den die Lernenden erreichen sollen.</p> <p>Aus der behavioristischen Tradition stammte der Vorschlag, den Inhalten eine Verhaltenskomponente hinzuzufügen (Mager, 1994). Damit entstand der Fokus auf klar beobachtbare Verben, welche den Lerngewinn in einem Thema präzisieren (Verben wie <i>erklären</i> und <i>aufzählen</i> und im Gegensatz zu unklaren Verben wie <i>wissen</i> und <i>verstehen</i>). Der Präzisionsgrad kann sich dennoch stark unterscheiden. Lernziele können konkrete, aber auch allgemeine, übergeordnete Kompetenzen ansprechen. Letztere haben im Zuge der Bologna-Reform an didaktischer Bedeutung gewonnen.</p>
Lerntheoretische Grundlagen	<p>Die wechselseitige Abstimmung von Lernzielen, Lehre und Prüfungen bezeichnen Biggs und Tang (2007) als „Constructive Alignment“. Klarheit und Transparenz von Lernzielen erleichtert den Studierenden das Lernen und den Lehrenden die Unterrichtsplanung. Klauer und Leutner (2007) heben ebenfalls deren Bedeutung für die spätere Erfolgskontrolle hervor: „Unterricht ohne klare Lehrziele hat es schon immer gegeben, und er kann auch sinnvoll sein. Soll jedoch eine bestimmte Kompetenz oder Qualifikation vermittelt werden, so ist ein klar definiertes Lehrziel unerlässlich. In diesem Fall hat das Lehrziel eine doppelte Funktion: Steuerung des Lehr-Lern-Prozesses und Kontrolle des Lehrerfolgs“ (S. 39). Die Mitteilung der Lernziele kann mehrfach erfolgen, beispielsweise am Anfang einer Lehrveranstaltung (Steuerungsfunktion) und am Ende (Kontrolle des Lehrerfolgs).</p>
Belege für didaktische Relevanz	<p>Hattie (2009) errechnet in seiner Meta-Analyse einen mittleren Effekt ($d = 0.56$) von Lernzielen auf akademische Leistungen. Effekte treten v.a. dann auf, wenn Lehrende anspruchsvolle Ziele setzen, Lehrsituationen auf die Zielerreichung hin strukturieren, die Lernenden ermutigen diese Ziele anzustreben und Rückmeldung zur Zielerreichung geben. Auch bei Feldman (2007) zeigen sich Lernziele als Einflussfaktor auf die Leistung der Studierenden.</p> <p>Die Metaanalyse von Klauer (1984) zeigte, dass vorangestellte Zielanga-</p>

	ben in schriftlichen Lehrtexten zu besseren Leistungen bei nachfolgenden Wissensfragen führten.
Positivbeispiele	<ul style="list-style-type: none"> - „Am Ende der Lehrveranstaltung können Sie die wichtigsten Nebenwirkungen von XY beschreiben.“ - „Sie definieren Bluthochdruck und nennen die wichtigsten Behandlungsmethoden.“ - „Sie entwerfen ein eigenes Diagnoseschema für XY.“ - Auch ungenaue Aussagen gelten: Verben wie „wissen/kennen/verstehen“, dann <u>wenn</u> sie auf ein konkretes Thema bezogen sind und in einem verständlichen Kontext sind. Z. B. „Diesen Stoffwechsel Vorgang sollten Sie jetzt genau verstehen“
Nicht gemeint	<ul style="list-style-type: none"> - Stellt nur zu behandelnde Themen dar (ohne Kompetenz). - Unklare Verben wie „wichtige Krankheitsbilder verstehen“ oder „das alles werden Sie wissen...“. - Unklare Themen wie: „Das Lehrbuch müssen Sie gelesen haben“ oder „Sie sollen hier soziale Kompetenzen erwerben“. - Lehrender sagt nur was wichtig ist, ohne konkretes Lernziel (dann 7). - Lernziele, welche von den Lehrenden verfolgt werden (Lehrziele), die den Lernenden aber nicht mitgeteilt werden.
Sonstiges	Natürlich kann Unterricht auch Ergebnisse bewirken, die nicht direkt intendiert wurden. Dies gilt dann als Lernziel, wenn die Lehrenden darauf ausdrücklich hinweisen (beispielsweise am Ende einer Lehrveranstaltung).

5. Persönlicher Bezug

Beschreibung	Der objektiv beobachtbare Bestandteil des Interesses an einem Thema, ist, ob die Lehrperson einen Bezug zu dem Thema offen äußert. Lehrende bringen sich selbst als Person ein und zeigen damit, dass sie sich selbst mit einem Thema intensiv auseinandergesetzt haben. Dies beinhaltet den beruflichen Bezug, persönliche Erlebnisse und v.a. das eigene Interesse an dem Thema der Lehrveranstaltung.
Lerntheoretische Grundlagen	Das Engagement der Lehrenden wird häufig als ein Faktor guter Lehre beschrieben (siehe Rindermann, 2001; Wolf, Spiel & Pellert, 2001; Helmke & Schrader 2010). Es besteht die Annahme, dass sich bestehendes Interesse der Lehrenden positiv auf die Lernenden auswirken kann. Im besten Fall können die Lehrenden das Interesse der Lernenden an einem Thema nachhaltig erhöhen. Darüber hinaus sollte sich ein hohes Interesse der Lehrenden auf weitere didaktische Handlungen positiv auswirken. Lehrende mit hohem Engagement werden sich möglicherweise mehr bemühen, ihre Lehre optimal zu gestalten.
Belege für didaktische Relevanz	Rindermann (1996) fand einen mittleren Zusammenhang in Evaluationsdaten zwischen dem wahrgenommenen, subjektiven Lerngewinn der Teilnehmer und dem wahrgenommenen Engagement der Lehrenden an Hochschulen.
Positivbeispiele	<ul style="list-style-type: none"> - „Ich habe dazu promoviert“ - „Die Frage beschäftigt mich selbst sehr intensiv...“ - „Mit dem Thema habe ich mir damals selbst schwergetan und habe es erst verstanden, als...“
Nicht gemeint	<ul style="list-style-type: none"> - Verneinungen von Interesse und Bezug zu dem Thema: „Das Thema halte ich selbst für völlig irrelevant“ „Das wenden wir heute überhaupt nicht mehr an“ - Allgemeine Bezüge, die nicht nah an der Person oder dem Thema sind: „Ich habe das natürlich auch lernen müssen...“ „Ich arbeite ja als Arzt und da könnte es evtl. vorkommen, dass irgendwann mal...“
Sonstiges:	Es gibt auch nonverbale und paraverbale Signale, die als Ausdruck hohen Interesses und hohes Engagements gelten. Diese sind hier aber nicht Teil der Beobachtungskategorie und sollten daher als freie Bemerkung festgehalten werden.

6. Handlungsrelevanz

Beschreibung	Lehrende weisen auf (potentielle) praktische Handlungsrelevanz für die Teilnehmer hin. Dies beinhaltet sowohl berufliche als auch private Handlungsrelevanz.
Lerntheoretische Grundlagen	Das gelehrte Wissen an Hochschulen hat höhere Relevanz für die Lernenden, wenn die Anwendbarkeit des Wissens deutlich wird. Prenzel (1996) sieht in der Darstellung anwendungsbezogener Verknüpfungen ein Beispiel, wie sich die wahrgenommene inhaltliche Relevanz des Lernstoffes erhöhen lässt. Dadurch soll selbstbestimmt motiviertes und interessiertes Lernen unterstützt werden.
Belege für didaktische Relevanz	Die absolute Mehrheit der Studierenden erwartet im Studium die Vermittlung praktischer Fähigkeiten in Lehrveranstaltungen (HIS-Studierendenbefragung 2011).
Positivbeispiele	<ul style="list-style-type: none"> - „Das ist wichtig für Ihre spätere Berufstätigkeit“ - „Wenn Sie später arbeiten, dann könnten Sie XY machen“ - „In Ihrem Praxiseinsatz haben Sie das schon angewendet...“ - „Wenn Sie mal in die Situation XY kommen, könnte Ihnen dieses Wissen nützlich sein“
Nicht gemeint	<ul style="list-style-type: none"> - Ausschließlich persönlicher beruflicher Bezug des Lehrenden (dann 5) - Allgemeinaussagen: z. B. „Das muss man behandeln“ - Allgemeine Falldarstellung (dann 1) - Ausformulierte Lernziele (dann 4)

7. Wichtige Inhalte (im Kontext Hochschule)

Beschreibung	Nicht alle Inhalte einer Lehrveranstaltung sind gleich wichtig. Manche Inhalte sind für das Verständnis bestimmter Sachverhalte besonders wichtig, andere für das erfolgreiche Bestehen späterer Prüfungen. Lehrende können <i>besonders</i> wichtige Informationen ausdrücklich hervorheben. Da hier die Betonung auf die Relevanz innerhalb des Kontextes Hochschule liegt, beinhaltet dies auch die Relevanz für sämtliche Prüfungen und auch für wissenschaftliche Arbeiten im weiteren Sinne (also Bachelorarbeiten, Promotionen, usw.)
Lerntheoretische Grundlagen	In Lehrveranstaltungen ist es den Teilnehmern nicht immer möglich, die ganze Zeit mit voller Konzentration und Aufmerksamkeit zu erleben. Doch es gilt: „Je aufmerksamer ein Mensch ist, desto besser wird er bestimmte Inhalte behalten“ (Spitzer, 2007). Mit der Betonung besonders wichtiger und relevanter Inhalte können Lehrende die Aufmerksamkeit der Teilnehmer zumindest einen zu einem gewissen Grad steuern. Lehrende können die Betonung wichtiger Inhalte als Teil der ausdrücklichen Prüfungsvorbereitung (auf Basis der Lernziele) einsetzen.
Belege für didaktische Relevanz	In einer Befragung zu Aspekten guter Lehre (Ledic, Rafajac & Kovac; 1999): wird die Betonung wichtiger Inhalte von Studierenden und Lehrenden als sehr wichtiger Aspekt guter Lehre benannt.
Positivbeispiele	<ul style="list-style-type: none"> - Betonung <i>besonders</i> wichtiger Informationen „Jetzt müssen Sie gut aufpassen, das ist wichtig“ - Beinhaltet Relevanz der Inhalte in Bezug auf Prüfungen/Promotion: „Das sind typische Prüfungsfragen: ...“ „So können Sie die Prüfungsfrage leicht beantworten...“ „Wenn Sie später promovieren, dann müssen Sie diesen Aspekt unbedingt berücksichtigen“
Nicht gemeint	<ul style="list-style-type: none"> - Ausformulierte Lernziele (4) - Praktische Handlungsrelevanz (6) - Allgemeine Aussagen zur Wichtigkeit von XY - Verneinung von Wichtigkeit („...müssen Sie nicht wissen“) - Allgemeine Hinweise zu Prüfungen (Termine, Prozedur...)

8. Interdisziplinarität

Beschreibung	Wissenschaftliche Bezüge, die über das eigentliche Thema der Veranstaltung hinausgehen, werden (kurz) erklärt. Dies kann den Bezug zu Nebenthemen, Erklärungsmodelle/Ansätze anderer Disziplinen oder Ergebnisse zu ähnlichen Themen anderer Fächer beinhalten. Im weiteren Sinne beinhaltet diese Kategorie interdisziplinäre Bezüge.
Lerntheoretische Grundlagen	Prenzel (1996) sieht in der Darstellung fachübergreifender Verknüpfungen ein Beispiel, wie sich die wahrgenommene inhaltliche Relevanz des Lernstoffes erhöhen lässt. Dadurch soll selbstbestimmt motiviertes und interessiertes Lernen unterstützt werden. Die didaktische Handlung, ein Thema in einen breiteren Kontext zu stellen, lässt sich als Beispiel für erfolgreiche „Erklärung“ bezeichnen, nach Helmke und Schrader (2010) ein Schlüsselmerkmal erfolgreicher Lehre. Wenn Lehrende ihr Thema in Zusammenhang mit anderen Themen stellen, erleichtern sie den Lernenden die Konstruktion positiven Transfers, welcher sich nicht automatisch einstellt (siehe Mähler & Stern, 2010). Lehrende können so dass Erkennen von Zusammenhängen anbahnen.
Belege für didaktische Relevanz	Von den Lernenden wird in der Regel erwartet, dass sie eigenständig in der Lage sind, die einzelnen Themen zueinander in Beziehung setzen zu können. Untersuchungen, welche die Wirksamkeit des Einsatzes fachübergreifender und interdisziplinärer Bezüge auf kognitive Leistungen untersucht haben, stehen bislang aus. Prenzel (1996) findet einen mittleren Zusammenhang der wahrgenommenen Relevanz mit motivationalen Variablen. Fachübergreifendes Denken sowie die Fähigkeit, vorhandenes Wissen auf neue Fragen und Probleme anzuwenden, wird von der Mehrheit der Studierenden im Studium erwartet. (siehe HIS-Studierendenbefragung 2011).
Positivbeispiele	<ul style="list-style-type: none"> - „Wie erklärt die Psychologie das Thema Stress?“ - „Ich stelle das Thema jetzt in den Zusammenhang mit Ihrem Wissen aus dem Grundstudium...“
Nicht gemeint	<ul style="list-style-type: none"> - Advance-Organizer, welche andere Themen mit einbeziehen, ohne dass der Bezug ausdrücklich hergestellt wird - Bezüge zu nicht fachlichen Themen - Bezüge sehr nahe an dem Hauptthema - Praktische Handlungsrelevanz (6) - Praktische Beispiele - Bezüge ohne Erklärungen

9. Zusammenfassung

Beschreibung	Beinhaltet die deutliche Wiederholung zuvor genannter und zunächst abgeschlossener Inhalte (auch bei Bezug auf vorherige Lehrveranstaltungen). Hier gelten auch Wiederholungen, welche in Kombination mit neuen Aspekten des Themas dargeboten werden. In Präsentationen häufig am Abschluss der Lehrveranstaltung mit der Bezeichnung „Zusammenfassung“ oder als „Take-Home-Message“.
Lerntheoretische Grundlagen	Die Funktionsweise des menschlichen Arbeitsgedächtnisses ist begrenzt (siehe Anderson, 2007). Wiederholungen und Zusammenfassungen erleichtern die Fokussierung auf die relevanten Informationen. Dadurch, dass sie das Tempo einer Lehrveranstaltung etwas verringern, können sie die tiefe Verarbeitung der Inhalte unterstützen. Die Zusammenfassung wichtiger Inhalte wird in der Regel der übergeordneten Kategorie Strukturierung zugeordnet (siehe Good & Brophy, 1986). Wiederholungen können seitens der Lernenden mit tiefen Verarbeitungsstrategien kombiniert werden. Steiner (2006) sieht in Wiederholungsstrategien eine wichtige Voraussetzung für erfolgreiches Lernen.
Belege für didaktische Relevanz	Das wiederholte Üben von Inhalten steigert die nachhaltige Gedächtnisleistung (siehe Anderson, 2007; Goldstein, 2011). Zusammenfassungen die ausschließlich seitens der Lehrenden vorgenommen werden können diesen Prozess anbahnen. Butler und Roediger (2007) zeigten auf, dass die Präsentation der wichtigsten Fakten einen positiven Effekt auf die Lernleistung hat, der aber etwas geringer ist als die Wiederholung von Fakten in einem kurzen Test mit offenem Antwortformat.
Positivbeispiele	<ul style="list-style-type: none"> - „Hier noch einmal die wichtigsten Fakten...“ - „Das sind die Kernaussagen der heutigen Lehrveranstaltung“ - „Jetzt fasse ich das wichtigste noch einmal zusammen...“
Nicht gemeint	<ul style="list-style-type: none"> - Erstmalige Aufzählungen/Auflistungen von Inhalten - Aspekte des Themas werden in Beziehung zu anderen Themen gesetzt (dann 8) - Aspekte des Themas werden in Beziehung zu handlungsrelevanten Aspekten gesetzt (dann 6) - Wichtiges steht ohne Erklärung an der Tafel (oder auf der Folie) - Weist auf wichtige Inhalte hin (dann 7)

Fragen in Lehrveranstaltungen

Mit der stärkeren Kompetenzorientierung des Studiums (siehe Qualifikationsrahmen für Deutsche Hochschulabschlüsse, 2005) stellt sich die Frage, wie Lehrende den Kompetenzerwerb der Studierenden unterstützen können. Insbesondere der Erwerb kommunikativer Kompetenzen setzt voraus, dass Studierende die Gelegenheit erhalten, eigene Wortbeiträge zu formulieren. Daher kann es von Bedeutung sein, zu analysieren, in welchem Ausmaß Lehrveranstaltungen dazu die Gelegenheit bieten.

Fragen der Lehrenden sind eine häufig eingesetzte Methode in Lehrveranstaltungen, an Hochschulen und an Schulen. Helmke (2010) und Dubs (2009) berichten von durchschnittlich 50-80 Fragen in 45-minütigen Lehrveranstaltungen. Lehrer können durch Fragen verschiedene Zwecke verfolgen: Vorwissen und Wissen abrufen oder Meinungen, Haltungen, Stimmungen, Interessen und Verständnisschwierigkeiten zum Vorschein bringen. Fragen, häufig die des Lehrers, setzen im Unterricht Interaktionen in Gang und dienen letztlich auch der Leistungsüberprüfung. In Schulen beträgt die durchschnittliche Wartezeit nach der Fragestellung unter einer Sekunde (Bromme, 1997). Dies wird als zu gering erachtet (siehe Helmke, 2010). Good und Brophy (1986) kommen in einem Überblicksartikel zu dem Schluss, dass Lehrer nach der Fragestellung mindestens drei Sekunden warten sollen, bis Sie jemanden bitten, eine Frage zu beantworten. So können Fragen optimale Antworten hervorrufen.

In der pädagogisch-psychologischen Literatur gibt es aufgrund des hohen Stellenwertes von Fragen für die gesamte Lehrer-Schüler-Interaktion viele Verweise auf den sinnvollen Einsatz von Lehrerfragen. Dubs (2009) betont, dass gut gestellte Fragen zu weiteren Denkprozessen anregen, schlecht gestellte Fragen dagegen produzieren Unklarheit. Erfolgreiches Antworten führt dagegen zu Erfolgserlebnissen. Offene, anspruchsvolle Fragen sind lernförderlicher als geschlossene Fragen (Redfield und Rousseau 1981). Hattie (2009, 2012) weist zudem darauf hin, dass eine funktionierende Kommunikation über das Lernen zu besseren Lernerfolgen führt. Lehrende unterscheiden sich jedoch darin, wie gut Ihnen dies gelingt.

Die Wortmeldungen der Teilnehmer sind objektive Indikatoren für die Beteiligung der Teilnehmer. Darunter fallen die Antworten auf die Fragen der Lehrenden, aber auch die eigenen Fragen der Teilnehmer. In diesem Beobachtungsbogen wird ausschließlich die Frequenz von verschiedenen Fragetypen und studentischen Äußerungen protokolliert. Vertiefende Analysen der Interaktionen in Lehrveranstaltungen müssen durch die Hochschuldidaktiker mit anderen Methoden und Ansätzen durchgeführt werden.

Für alle Fragen gilt:

- Der Begriff *Frage* beinhaltet hier auch Aufforderungen, die eine Antwort hervorrufen sollen. Die Aufforderung muss also nicht zwingend als Frage formuliert sein! Beispielsweise:
 - „Bitte erläutern Sie XY!“
 - „Nennen Sie bitte die Definition von YZ“
 - „Bitte stellen Sie jetzt die Fragen, die noch offen sind“
- Es ist keine Voraussetzung, dass jemand auf die Frage antwortet. Entscheidend ist, dass die Frage so gestellt wird, dass jemand der Teilnehmer darauf antworten kann (Bedingung: mehr als eine Sekunde warten).
- Nicht gemeint sind dagegen Fragen, die von den Lehrenden sofort selbst beantwortet werden.
- Die Frage kann auch gezielt an bestimmte Studierende gestellt werden.
- Fragen an einzelne Teilnehmer während individueller Lernphasen werden hier dagegen nicht berücksichtigt.

Zu der praktischen Umsetzung:

- Die Fragen und Wortmeldungen werden im BEO-LV als Striche (|) protokolliert.
- Erfahrungsgemäß können Sie etwa zehn Striche in einem Intervall protokollieren.
- Sollten die Zahl der Fragen in einem Intervall größer werden, müssen Sie sich freie Notizen machen. Anschließend können Sie einfach die gesamte Zahl eintragen.

10. Wissensfragen

Beschreibung	<p>Stellt Fragen, auf die es eine richtige Antwort geben kann/soll: Definitionen, Fakten, Regeln, Erklärungen, Übersetzungen, Zusammenfassen in eigenen Worten, Analysen, Hypothesen aufstellen, Abwägen, Beurteilen u.v.m. Die Kategorisierung ist hier unabhängig von der Schwere oder Komplexität der Fragen.</p> <p>Diese Kategorie kann in Anlehnung an Modelle kognitiver Anspruchsniveaus (in Anlehnung an Anderson & Krathwohl, 2001) breiter gefasst werden und damit differenziert werden:</p> <p>A) Fokus auf Fakten, Definitionen, Regeln B) Verständnis (Erklären, Übersetzen, Zusammenfassen in eigenen Worten...) C) Analyse (Gliedern, Kriterien ermitteln, Fehler suchen...) D) Synthese: Divergente Ideen, Hypothesen E) Bewertung: Abwägen, Beurteilen, Entscheiden...</p> <p>Für eine genauere Darstellung siehe Dubs (2009, S.129). Diese differenzierten Kategorien stehen in der erweiterten Version (Kommunikation) zur Verfügung.</p>
Lerntheoretische Grundlagen	<p>Wissensfragen sind die meist verbreitete Form von Fragen im Kontext Schule und Hochschule. Wissensfragen erfüllen verschiedene Funktionen: Sie weisen auf relevantes Wissen hin und erhöhen gleichzeitig die kognitive Aktivität der Teilnehmer. Klauer und Leutner (2007) sehen in Wissensfragen eine gute, praktikable Möglichkeit, die Abrufbarkeit gelerntes Wissens zu erhöhen.</p> <p>Dubs (2009) weist darauf hin, dass ein höheres kognitives Anspruchsniveau nicht zwangsläufig mit anspruchsvolleren Kognitiven Leistungen verbunden sein muss.</p>
Belege für didaktische Relevanz	<p>Studien an Schulen zeigen, dass Fragen in der Regel eine moderat positive Wirkung auf den Lernzuwachs haben, ohne dass sich dabei ein bestimmtes Anspruchsniveau als grundsätzlich vorteilhaft erwiesen hat (siehe Good & Brophy, 1986; Hattie, 2009; Dubs, 2009). Das bedeutet, dass sowohl einfache, kognitiv wenig fordernde Fragen, als auch anspruchsvolle Fragen ihren Platz in der Lehre haben, solange sie passend eingesetzt werden.</p>
Positivbeispiele	<ul style="list-style-type: none"> - „Bitte definieren Sie Bluthochdruck!“ - „Wie würden Sie dies mit eigenen Worten erklären?“ - „Was können Sie aus diesem Befund ableiten?“ - „Welche Hypothesen leiten Sie aus den theoretischen Annahmen ab?“ - „Bitte beurteilen Sie den Test anhand wissenschaftlicher Gütekriterien“
Nicht gemeint	<ul style="list-style-type: none"> - Fragen, die unter 11-14 oder 18 fallen - Fragen, die von Lehrenden selbst beantwortet werden - Arbeitsauftrag von individuellen Lernphasen
Sonstiges	<p>Es kann sein, dass Studierende die Frage anspruchsvoller beantworten, als sie vermeintlich intendiert war. In dem Fall kann die höhere Kategorie angekreuzt werden (und dies als Rückmeldung thematisiert werden).</p>

11. Fragen zu Vorwissen/Vorkenntnissen

Beschreibung	Fragen, welche darauf abzielen, herauszufinden was die Lernenden schon wissen, erlebt oder gelernt haben. Es gibt hier daher keine richtige oder falsche Antwort.
Lerntheoretische Grundlagen	Die Aktivierung von Vorwissen hilft Wissen anschlussfähig zu machen. Für Lehrende ist es hilfreich zu erfahren, welches Vorwissen besteht. Roth (2004) erklärt es so: „Dinge, die für den Lernenden neu, d.h. nicht anschlussfähig sind, fallen durch die Gedächtnisnetze hindurch, weil sie nirgendwo Brücken zu bereits vorhandenem Wissen bilden können“ (S. 504). Das Wissen um das Vorwissen der Lernenden ist also wichtig, um didaktische Handlungen angemessen zu planen und steuern.
Belege für didaktische Relevanz	Bereichsspezifisches Vorwissen gilt als wichtiger Prädiktor für spätere Leistungen (siehe z. B. Helmke & Schrader, 2010). Chinn und Brewer (1993) zeigen, wie Vorwissen den Wissenserwerb hemmen kann. Widersprüchliches Wissen wird nicht selten zurückgewiesen oder uminterpretiert. Lehrende, welche die Annahmen der Lernenden kennen, können dagegen Maßnahmen ergreifen, die dem entgegenwirken.
Positivbeispiele	<ul style="list-style-type: none"> - „Wer hat schon in der Neurologie ein Praktikum gemacht?“ - „Wer hatte in der Schule Biologie-Leistungskurs?“ - „Was wissen Sie schon über XY?“ - „Was haben Sie in der Schule über XY gelernt?“
Nicht gemeint	<ul style="list-style-type: none"> - Arbeitsauftrag von individuellen Lernphasen - Fragen ohne inhaltlichen Bezug - Reine Wissensfragen ohne persönlichen Bezug - Fragen danach, wie die Lernenden, früher gelernt haben (dann 13)
Sonstiges	Der Unterschied zu Wissensfragen besteht darin, dass es hier nur darum geht, herauszufinden ob die Studierenden etwas wissen. Hier können Sie auch biographische Fragen protokollieren, wenn diese einen inhaltlichen Bezug haben.

12. Fragen zu Interesse an den Inhalten

Beschreibung	Lehrender stellt Fragen danach, was die Lernenden besonders viel oder wenig interessiert.
Lerntheoretische Grundlagen	Es ist anzunehmen, dass Interessen eine wichtige Rolle für die Lernmotivation spielen (siehe Deci & Ryan, 1993; Krapp, 2011). Die Frage „Ist der Lerner motiviert“ ist nach Klauer & Leutner (2007) die erste Frage, die im Lehr-Lernprozess von Bedeutung ist. Durch die Kommunikation über das Interesse erfahren Lehrende die Haltung der Lernenden. Sie erhalten so die Möglichkeit, dieses Wissen didaktisch zu nutzen.
Belege für didaktische Relevanz	Eine Metaanalyse zeigte einen Zusammenhang von Interesse und Leistung mit einer durchschnittlichen Korrelation von $r = .30$ (Schiefele, Krapp & Schreyer, 1993). Offen ist dagegen noch, ob das individuelle Interesse tatsächlich eine vertiefende Auseinandersetzung mit den Lerninhalten bewirkt (siehe Krapp, 2011) und ob es Lehrenden messbar gelingen kann, durch Kommunikation über Interessen, ihre Lehre unmittelbar daran anzupassen.
Positivbeispiele	<ul style="list-style-type: none"> - „Wer möchte sich mit diesem Thema in Zukunft intensiver beschäftigen?“ - „Wer möchte später bei XY arbeiten?“ - „Wer möchte später als YZ arbeiten?“ - „Haben Sie sich bereits selbstständig mit dem Thema auseinandergesetzt?“
Nicht gemeint	<ul style="list-style-type: none"> - Arbeitsauftrag von individuellen Lernphasen - Fragen, die eher Vorwissen abfragen: „Wer hat das schon mal woanders gelernt?“

13. Fragen zum Lernprozess/Verständnis

Beschreibung	Die Kategorie erfasst, ob Lehrende den Lernprozess auf einer Metaebene thematisieren. Dies beinhaltet zwei verschiedene Komponenten: 1. Fragen danach, <u>wie</u> die Studierenden (in der Lehrveranstaltung) lernen und Verständnis konstruieren. 2. Fragen, <u>ob</u> die Studierenden die Inhalte verstanden haben und welche Fragen noch offen sind
Lerntheoretische Grundlagen	Die Kenntnisse über die eigenen Kognitionen, kann man als <i>Metakognitive Kompetenz</i> bezeichnen. Lehrende können durch Fragen Informationen über die metakognitiven Prozesse erhalten. Lehrende erhalten so Feedback zu ihrer Erklärung und die Lernenden erhalten die Möglichkeit, Denkfehler und Unklarheiten zu korrigieren. Bestenfalls können Lehrende durch Fragen positiven Einfluss auf die Metakognitionen der Lernenden nehmen.
Belege für didaktische Relevanz	Bruning, Schraw, Norby & Ronning, (2004) verweisen auf einen positiven Zusammenhang der metakognitiven Kompetenz mit der Leistungsstärke bei Schülern. Hattie (2009) berechnet eine deutliche positive Wirkung von Programmen, die metakognitive Fähigkeiten fördern.
Positivbeispiele	<ul style="list-style-type: none"> - „Haben Sie noch Fragen?“ - „Ist Ihnen klar geworden, was ein XY ist?“ - „Fällt es Ihnen leicht, diesen Zusammenhang zu verstehen?“ - „Ist Ihnen klar, <i>warum</i> dies so ist?“ - „Können Sie XY erklären?“ - „Bitte stellen Sie jetzt die Fragen, die noch offen sind“
Nicht gemeint	<ul style="list-style-type: none"> - Dozent referiert über typische Missverständnisse/Fragen anderer Personen. - Das Angebot, Verständnisfragen im Anschluss oder in der Sprechstunde zu klären. - Die Frage: „Haben Sie noch Fragen?“ als rhetorische Frage, nur 1 Sekunde (häufig am Ende einer Lerneinheit) - Arbeitsauftrag von individuellen Lernphasen

14. Fragen zu Meinungen/Haltungen

Beschreibung	Beinhaltet Fragen, auf die es keine eindeutig richtige Antwort gibt, da sie eine subjektive Einschätzung erfragen. Dies kann eine Meinung zu einem bestimmten Thema beinhalten oder die Darstellung einer umfassenderen Haltung.
Lerntheoretische Grundlagen	Wenn Teilnehmer ihre eigene Meinung und Haltung zu einem Thema reflektieren, dann stellen sie damit den persönlichen Bezug zu den behandelten Themen her. Gedächtnisinhalte, welche für die Vorstellung vom eigenen Selbst, für das eigene Selbstkonzept, besonders relevant sind, werden besser erinnert. Dieses Phänomen ist bekannt als „Selbstreferenzeffekt“ (siehe z. B. Symons & Johnson, 1997).
Belege für didaktische Relevanz	Es kann sein, dass Studierende unterschiedlich stark von der Maßnahme profitieren, persönlichen Bezug zu dem Unterrichtsmaterial herzustellen. Hulleman und Harackiewicz (2009) zeigten, dass Schülerinnen und Schüler mit geringeren Erfolgserwartungen davon stärker profitieren, als solche mit höheren Erfolgserwartungen. Braun, Gusy, Leidner und Hannover (2008) sehen in der Fähigkeit, die eigene Meinung in Lehrveranstaltungen äußern zu können, einen Indikator für die kommunikative Kompetenz. Diese zu fördern ist im Rahmen der Bologna-Reform ein zu erreichendes Kompetenzziel für alle Studierenden.
Positivbeispiele	<ul style="list-style-type: none"> - „Wie stehen Sie zur Stammzellentherapie?“ - „Welche Meinung haben Sie zu...?“ - „Wie würden Sie entscheiden?“
Nicht gemeint	<ul style="list-style-type: none"> - Fragen zu Positionen, die nicht die persönliche Meinung widerspiegeln z. B. „Nennen Sie drei typische Pro-Argumente für Stammzellentherapie“ - Arbeitsauftrag von individuellen Lernphasen

Die Handlungen der Lernenden

Teilnehmer in Lehrveranstaltungen können auf ganz unterschiedliche Arten und Weisen aktiv sein. Die wohl geläufigste Form ist die Beteiligung durch Wortbeiträge. Davon abzugrenzen sind Aktivitäten in individuellen Lern- und Arbeitsphasen und in Präsentationen durch die Teilnehmer. Diese drei Kategorien sind hier nun genauer beschrieben.

15. Wortmeldungen von Teilnehmern

Beschreibung	Alle Wortmeldungen von den Lernenden. Beinhaltet Fragen, Beiträge zu Diskussionen und Antworten auf Fragen der Lehrenden.
Lerntheoretische Grundlagen	Fragen der Lehrenden bewirken eine Aktivierung der Teilnehmer und sollen verschiedene Wirkungen erzielen (s.o.). Die Wortmeldungen der Teilnehmer sind dann von hoher Bedeutung für die Interaktion. Die Möglichkeit, eigene Fragen zu stellen, ist ebenfalls wichtig für die erfolgreiche Kommunikation in Lehrveranstaltungen.
Belege für didaktische Relevanz	Hattie (2009, 2012) betont, dass erfolgreiche Kommunikation zu besseren Lernerfolgen führt. Die Anzahl der Wortmeldungen der Teilnehmer gibt zumindest Auskunft darüber, wie viele Personen sich mit Wortbeiträgen beteiligen.
Positivbeispiele	<ul style="list-style-type: none"> - Antworten auf Fragen der Lehrenden - Verständnisfragen - Aussagen zu dem eigenen (Nicht-)Verständnis: „Ich weiß nicht“. - Wissensfragen - Inhaltliche Fragen, welche sich auf die Prüfung beziehen: „Ist das Thema prüfungsrelevant?“
Nicht gemeint	<ul style="list-style-type: none"> - Gespräche unter den Teilnehmern, die nicht in die ganze Gruppe getragen werden. - Fragen nach Beendigung der Lehrveranstaltung - Aussagen und Fragen, welche sich nicht die Inhalte der Lehrveranstaltung beziehen: „Ich öffne kurz das Fenster“ „Können Sie die Folie noch mal zeigen?“ „Wie viel Uhr findet die Prüfung statt?“
Sonstiges	Hattie (2009) merkt an, dass die Lehrforschung die Fragen der Lernenden genauer untersuchen sollte. Zu diesem Zweck könnte diese Kategorie weiter differenziert werden.

16. Individuelle Lern-und Arbeitsphasen

Beschreibung	Individuelle Lern-und Arbeitsphasen bezeichnen Phasen, in denen <u>alle</u> Lernenden den Auftrag erhalten haben, selbst aktiv zu sein. Unabhängig davon, ob alle Personen dann tatsächlich aktiv sind. Individuelle Lern- und Arbeitsphasen stehen im Gegensatz zu kollektiven Lernphasen (Phasen, in denen einzelne präsentieren oder eine Kommunikation stattfindet und dabei nicht <u>alle</u> Teilnehmer aktiv beteiligt sind).
Lerntheoretische Grundlagen	<p>Die Empfehlung, Zeit für individuelle Lern- und Arbeitsphasen einzuräumen, basiert auf der Annahme, dass die Lernenden auch schon innerhalb einer Lehrveranstaltung Zeit benötigen, um das neue Wissen mit ihrem Vorwissen zu verknüpfen. Damit Lerninhalte im Anschluss an Lehrveranstaltungen erinnert werden können, müssen Informationen aus dem Arbeitsgedächtnis in das Langzeitgedächtnis übertragen werden. Kognitionspsychologische Befunde zeigen, dass hier eine aktive und tiefe Verarbeitung (Elaboration) von neuen Informationen zu besserem Behalten führt. Auch das wiederholte Üben von Inhalten steigert die nachhaltige Gedächtnisleistung (siehe Anderson, 2007; Goldstein, 2011). Individuelle Lern-und Arbeitsphasen sind erwartungsgemäß dann besonders wirksam, wenn sie die aktive Auseinandersetzung mit den Lerninhalten erhöhen.</p> <p>Es wird oft argumentiert, dass individuelle Lern-und Arbeitsphasen mehr Kompetenzerleben innerhalb von Lehrveranstaltungen ermöglichen. Dies hat möglicherweise günstige motivationale Auswirkungen. Ebenso wird häufig die soziale Komponente bestimmter Handlungsformen betont, beispielsweise von Diskussionen oder Arbeiten innerhalb von Kleingruppen. Mit der stärkeren Kompetenzorientierung des Studiums (siehe HRK/KMK/BMBF, 2005) sind diese Faktoren heute von höherer Relevanz.</p>
Belege für didaktische Relevanz	Individuelle Lern- und Arbeitsphasen können in verschiedensten Ausprägungen auftreten, denen nur gemeinsam ist, dass alle Lernenden selbst aktiv sind. Es kann beispielsweise vorteilhaft sein, Bilder zu gestalten (Seufert, Zander & Brünken, 2007), Zusammenfassungen zu entwerfen (Slotte & Lonka, 1999) oder Partnergespräche mit Übungsaufgaben zu verbinden (Deslauriers, Schelew & Wieman, 2011). Auch Kleingruppenarbeit kann vorteilhafte Ergebnisse auf verschiedenen Zieldimensionen zeigen (Springer, Stanne & Donovan; 1999). Seidel und Shavelson (2007) sehen in der Anleitung zu fachspezifischen Lernaktivitäten, welche die tiefe Verarbeitung fördern, die effektivste Komponente erfolgreicher Lehrhandlungen. Ruiz-Primo, Briggs, Iverson, Talbot und Shepard (2011) berechneten insgesamt mittlere Effektstärken auf die Lernleistung für verschiedene Formen aktivierender Methoden. Freeman et al. (2014) berechneten in der aktuell umfassendsten Metaanalyse ebenfalls eine mittlere Effektstärke auf Prüfungsleistungen.
Positivbeispiele	<ul style="list-style-type: none"> - Aufgaben in Einzelarbeit (Eine Person bearbeitet eine Aufgabe) - Aufgaben in Partnerarbeit (Zwei Personen arbeiten an einer Aufgabe oder besprechen eine Fragestellung...) - Aufgaben in Gruppenarbeit (Mindestens drei Personen) - Verschiedene Methoden, die in Lehrveranstaltungen eingesetzt werden können (z. B. <i>Bildergalerie</i>, <i>Brainwalking</i>, <i>Strukturlegetechnik</i>, <i>x-minute-paper</i> u.v.m siehe z. B. Strittmatter-Haubold & Ehlail, 2012). - Wissensfragen mit Abstimmungen (per Handheben oder TED-Systemen)

Nicht gemeint	<ul style="list-style-type: none"> - Lehrende/r stellt Fragen (dann 10-15) an die ganze Gruppe - Teilnehmer schreiben etwas ab - Abstimmungen ohne direkten Lernbezug: z. B. Vorwissen („wer hat schon einmal...“) oder zu Interesse („wer möchte später in die Forschung?“) - Einzelne Lernende sind aktiviert (dann 15 oder 16) - Teilnehmer sehen Lehrfilme ohne konkreten Arbeitsauftrag - Etwas wird herumgegeben/ausgeteilt (evtl. 2) - Methoden wie <i>Aquarium</i> (bzw. <i>Fish bowl</i>), da hier auch nicht alle Personen gleichzeitig eingebunden sind.
---------------	--

Die Differenzierung der individuellen Lern-und Arbeitsphasen

Die Umsetzung individueller Lern-und Arbeitsphasen (ILAP) können Sie mit dem BE-O-LV differenziert protokollieren. Dazu die folgende Erläuterung:

Methode/Sozialform	<p>Hier können Sie die Bezeichnung der Methode und/oder die Sozialform eintragen: Beispiele für Methoden lauten: <i>Strukturlegetechnik</i>, <i>x-minute-paper</i>, <i>Partnerggespräch</i>, <i>Übungsaufgabe</i>...</p> <p>Die Sozialformen lauten: Einzelarbeit, Partnerarbeit oder Gruppenarbeit.</p>
Dauer	<p>Hier können Sie die Zeit eintragen, welche für die ILAP aufgewendet wurde. Es empfiehlt sich, die Zeit selbst zu stoppen, da diese nicht zwingend identisch mit der Zeit ist, die Lehrende vorher einräumen.</p>
vollständige Instruktion?	<p>Diese Kategorie erfasst, ob die Instruktion der Lehrenden klar und eindeutig war. Im besten Fall geben die Lehrenden vor einer ILAP wichtige Informationen.</p> <ul style="list-style-type: none"> ▪ „Wie lange“: Lehrende sagen, welche Zeit sie einräumen ▪ „Was“: Lehrende erklären den Inhalt genau ▪ „Wie“: Lehrende nennen die Form der Umsetzung (z. B. schriftlich) ▪ „Mit wem“: Lehrende benennen die Sozialform (z. B. zu zweit) <p>Kreuzen Sie den entsprechenden Kasten an, wenn die Lehrenden die Instruktion ausdrücklich geben.</p>
Werden die Ergebnisse eingeholt und zusammengefasst?	<p>Wenn die Lehrenden die Ergebnisse einer ILAP aufgreifen und besprechen, nennt man dies <i>Synchronisation</i>. Die Synchronisation der Ergebnisse ist nicht immer notwendig, sie kann aber hilfreich sein.</p> <ul style="list-style-type: none"> ▪ Kreuzen Sie „Ja“ an, wenn alle Inhalte zusammengetragen werden. Dabei ist es nicht notwendig, dass alle Studierenden zu Wort kommen. ▪ Kreuzen Sie „Teilweise“ an, wenn nur ein Teil der Inhalte öffentlich besprochen werden. ▪ Kreuzen Sie „Nein“ an, wenn die Lehrenden die Inhalte nach einer ILAP nicht auflösen, oder wenn die Lehrveranstaltung mit einer ILAP endet.

Literaturverzeichnis BEO-LV

- Anderson, J. R. (2007). *Kognitive Psychologie* (6. Aufl.). Heidelberg: Spektrum, Akademischer Verlag.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Ausubel, D. P. (1974). *Psychologie des Unterrichts. Band 1* (Beltz-Studienbuch, Bd. 60). Weinheim [u. a.]: Beltz.
- Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education*, 20, 481-486.
- Biggs, J. B. & Tang, C. (2007). *Teaching for quality learning at university* (3. Aufl.). Maidenhead: Open University Press.
- Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Kompetenzorientierte Lehrevaluation - Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica* 54, 30-42.
- Brezinka, W. (1974). *Grundbegriffe der Erziehungswissenschaft*. München: Rheinhardt.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie*, Serie I, Bd. 3 (S. 177-212). Göttingen [u. a.]: Hogrefe.
- Bruning, R. H., Schraw, G. J., Norby, M. M. & Ronning, R. R. (2004). *Cognitive psychology instruction* (4. Aufl.). Upper Saddle River, NJ: Pearson.
- Butler, A.C. & Roediger, H.L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514-527.
- Chinn, C. A. & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 1-49.
- Deci, E. & Ryan R. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39, 223-238.
- Deslauriers, L., Schelew, E. & Wieman, C. (2011). Improved Learning in a Large-Enrollment Physics Class. *Science*, 332, 862-864.
- Dubs, R. (2009). *Lehrerverhalten. Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht* (Pädagogik, 2. Aufl.). Stuttgart: Steiner.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Hrsg.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (368-395). New York: Springer.
- Gold, A. (1999). Studienabbruch und Studienerfolg. In M. Schröder-Gronostay & H.-D. Daniel (Hrsg.), *Studienerfolg und Studienabbruch* (S. 51-65). Neuwied: Luchterhand.
- Goldstein, E. B. & Vanhorn, D. (2011). *Cognitive psychology* (3. Aufl.). Australia, Belmont, CA: Wadsworth Cengage Learning.
- Good, T. & Brophy J. (1986). School effects. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching. A project of the American Educational Research Association* (3. Aufl.). New York, NY [u. a.]: Macmillan.
- Hattie, J. (2008). *Visible learning. A synthesis of meta-analyses relating to achievement*. London [u. a.]: Routledge.

- Hattie, J. (2012). *Visible learning for teachers. Maximizing impact on learning*. London [u. a.]: Routledge.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (Unterricht verbessern - Schule entwickeln, 4. Aufl.). Seelze-Velber: Klett/Kallmeyer.
- Helmke, A. & Schrader, F.-W. (2010). Hochschuldidaktik. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Heublein, U., Spangenberg, H. & Sommer, D. (2003). *Ursachen des Studienabbruchs. Analyse 2002*. Hannover: HIS.
- HIS-Institut für Hochschulforschung. Universität Konstanz: AG Hochschulforschung (2010). *Studienqualitätsmonitor 2010*. Zugriff am 20.05.2013 unter http://www.his.de/pdf/pub_fh/fh-201204.pdf.
- HIS-Institut für Hochschulforschung. Universität Konstanz: AG Hochschulforschung (2011). *Studienqualitätsmonitor 2011*. Zugriff am 20.05.2013 unter http://www.his.de/pdf/24/sqm2011.Uni_FH.pdf
- Hulleman, C. S. & Harackiewicz, J. M. (2009). Promoting Interest and Performance in High School Science Classes. *Science*, 326, 1410–1412.
- HRK/KMK/BMBF (2005): *Qualifikationsrahmen für Deutsche Hochschulabschlüsse* (Im Zusammenwirken von Hochschulrektorenkonferenz, Kultusministerkonferenz und Bundesministerium für Bildung und Forschung erarbeitet und von der Kultusministerkonferenz am 21.4.2005 beschlossen). Zugriff am 20.05.2013 unter http://www.akkreditierungsrat.de/fileadmin/Seiteninhalte/KMK/Vorgaben/KMK_Qualifikationsrahmen_aktuell.pdf
- Klauer, K. (1984). Intentional and Incidental Learning with Instructional Texts: A Meta-Analysis for 1970-1980. *American Educational Research Journal*, 21, 323–339.
- Klauer, K. J. & Leutner, D. (2007). *Lehren und Lernen. Einführung in die Instruktionspsychologie* (Studentexte, 1. Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Krapp, A. (2010). Interesse. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (Programm PVU, Psychologie-Verlags-Union, 4. Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Ledic, J., Rafajac, B. & Kovac, V. (1999): Assessing the quality of university teaching in Croatia. *Teaching in Higher Education*, 2, 213-235.
- Levie, W. H. & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology*.
- Levin, J. R., Anglin G. J. & Carney R. N. (1987). On empirically validating functions of pictures in prose. In H. A. Houghton & D. M. Willows (Hrsg.), *The psychology of illustration*. New York: Springer.
- Lewalter, D. (1997). *Lernen mit Bildern und Animationen. Studie zum Einfluss von Lernermerkmalen auf die Effektivität von Illustrationen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 2). Münster [u. a.]: Waxmann.
- Mager, R. F. (1994). *Lernziele und Unterricht* (Beltz grüne Reihe, 1977. Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Mähler, C. & Stern, E. (2006). Transfer. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (3. Aufl.). Weinheim [u. a.]: Beltz, PVU.
- Preiss, R.W. & Gayle, B.M. (2006). A Meta-Analysis of the educational benefits of employing advanced organizers. In B. M. Gayle (Hrsg.), *Classroom communication and instructional processes. Advances through meta-analysis*. Mahwah, N.J: Lawrence Erlbaum Associates.

- Prenzel, M. (1996). Bedingungen für selbstbestimmt motiviertes und interessiertes Lernen im Studium. In J. Lompscher & H. Mandl (Hrsg.), *Lehr- und Lernprobleme im Studium. Bedingungen und Veränderungsmöglichkeiten*. Bern [u. a.]: Huber.
- Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51 (2), 237–245.
- Renkl. (2010). Träges Wissen. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (Programm PVU, Psychologie-Verlags-Union, 4. Aufl.). Weinheim [u. a.]: Beltz.
- Rindermann, H. (1996). *Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen* (Psychologie, Bd. 6). Landau: Empirische Pädagogik. e.V.
- Rindermann, H. (2001). *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen ; mit einem Beitrag zur Evaluation computerbasierten Unterrichts* (Psychologie, Bd. 42). Landau: Empirische Pädagogik e.V.
- Roth, G. (2004) Warum sind Lehren und Lernen so schwierig? *Zeitschrift für Pädagogik*, 50, 496-506.
- Sander, P., Stevenson K., King M. & Coates D. (2000) *Studies in Higher Education*, 25, 309-323.
- Schiefele, U. Krapp, A. & Schreyer. I. (1993) Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25, 120–148.
- Schnotz, W. (2010). Visuelles Lernen. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl.). Weinheim [u. a.]: Beltz.
- Seidel, T. & Shavelson R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research* 77, 454–499.
- Seufert, T., Zander S. & Brünken. R. (2007). Das Generieren von Bildern als Verstehenshilfe beim Lernen aus Texten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 39, 33–42.
- Slotte, V. & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, 1–20.
- Spitzer, M. (2006). *Lernen. Gehirnforschung und die Schule des Lebens*. Heidelberg: Spektrum Akad. Verl.
- Springer, L., Stanne. M. E. & Donovan S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 1999, 21–51.
- Strittmatter-Haubold, V. & Ehlail, F. (2012). *Lernen im Aufwind. Methodenreader zur Gestaltung von Lernprozessen* (8., völlig überarbeitete und erweiterte Aufl.). Heidelberg: Inst. für Weiterbildung.
- Symons, C. S. & Johnson, B. T. The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, 121, 371–394.
- Wahl, D. (2006). *Lernumgebungen erfolgreich gestalten. Vom trägen Wissen zum kompetenten Handeln* (2. Aufl.). Bad Heilbrunn: Klinkhardt.
- Weidenmann, B. (1994). Informierende Bilder. In B. Weidenmann (Hrsg.), *Wissenserwerb mit Bildern*. Bern: Verlag Hans Huber.
- Weinert, F. E. (1996). Lerntheorien und Instruktionsmodelle. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Pädagogische Psychologie. Bd 2: Psychologie des Lernens und der Instruktion* (S. 1-8). Göttingen: Hogrefe.

- Wellenreuther, M. (2009). *Forschungsbasierte Schulpädagogik. Anleitungen zur Nutzung empirischer Forschung für die Schulpraxis*. Baltmannsweiler: Schneider Hohengehren.
- Wolf, P., Spiel C. & Pellert A. (2001). Entwicklung eines Fragebogens zur globalen Lehrveranstaltungsevaluation – ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz. In C. Spiel (Hrsg.), *Evaluation universitärer Lehre. Zwischen Qualitätsmanagement und Selbstzweck*. Münster [u. a.]: Waxmann.
- Zumbach, J. Haider, K. & Mandl, H. (2008). Fallbasiertes Lernen: Theoretischer Hintergrund und praktische Anwendung. In H. Mandl & J. Zumbach (Hrsg.), *Pädagogische Psychologie in Theorie und Praxis. Ein fallbasiertes Lehrbuch*. Göttingen [u. a.]: Hogrefe.

Anhang C3. Beobachtungsbogen für Lehrveranstaltungen (BEO-LV)

BEO-LV © Medizinische Fakultät der Universität Heidelberg, Integrative Lehrberatung

Dauer der Lehrveranstaltung →

Welche didaktischen Handlungen treten auf?	1-15	16-30	31-45	46-60	61-75	76-90	90 +	? Anmerkungen
1. Problem/Fall/Herausforderung								<input type="checkbox"/>
2. Visualisierung (Bild, Film, Demonstration...)								<input type="checkbox"/>
3. Struktur/Systematisierung								<input type="checkbox"/>
4. Lernziele								<input type="checkbox"/>
5. Persönlicher Bezug								<input type="checkbox"/>
6. Handlungsrelevanz								<input type="checkbox"/>
7. Wichtige Inhalte								<input type="checkbox"/>
8. Interdisziplinarität								<input type="checkbox"/>
9. Zusammenfassung								<input type="checkbox"/>
Sonstiges 1:								<input type="checkbox"/>
Sonstiges 2:								<input type="checkbox"/>
Kommunikation/Interaktion								
10. Wissensfragen								<input type="checkbox"/>
11. Fragen zu Vorwissen/Vorkenntnissen								<input type="checkbox"/>
12. Fragen zu Interesse an den Inhalten								<input type="checkbox"/>
13. Fragen zum Lernprozess/Verständnis								<input type="checkbox"/>
14. Fragen zu Meinungen/Haltungen								<input type="checkbox"/>
15. Wortmeldungen von Studierenden								<input type="checkbox"/>
16. Studierende präsentieren (Referat/Rollenspiel usw.)								<input type="checkbox"/>
17. Individuelle Lern- und Arbeitsphasen								<input type="checkbox"/>



Methode/Sozialform	Dauer	vollständige Instruktion?	Werden die Ergebnisse eingeholt und zusammengefasst?
(1)		Wie lange <input type="checkbox"/> Wer mit wem <input type="checkbox"/> Was <input type="checkbox"/> Wie <input type="checkbox"/>	Ja <input type="checkbox"/> Teilweise <input type="checkbox"/> Nein <input type="checkbox"/>
(2)		Wie lange <input type="checkbox"/> Wer mit wem <input type="checkbox"/> Was <input type="checkbox"/> Wie <input type="checkbox"/>	Ja <input type="checkbox"/> Teilweise <input type="checkbox"/> Nein <input type="checkbox"/>
(3)		Wie lange <input type="checkbox"/> Wer mit wem <input type="checkbox"/> Was <input type="checkbox"/> Wie <input type="checkbox"/>	Ja <input type="checkbox"/> Teilweise <input type="checkbox"/> Nein <input type="checkbox"/>

Anhang C4. Handanweisung zur Benutzung des BEO-LV

BEO-LV	Beschreibung und Positivbeispiele	Nicht gemeint:
1. Problem/Fall/ Herausforderung	Fall (-Beispiel) oder Herausforderung wird vorgestellt (mit dem Ziel zu veranschaulichen bzw. eine Lösung/Antwort/Stellungnahme zu finden) z.B. Falldarstellung/Schlagzeile/Kontroverse Frage/Zitat Zeitungsartikel/Krankengeschichte/Studie (Aufbau und Ergebnisse) Muss in inhaltlichen Zusammenhang mit dem Inhalt der Lehrveranstaltung stehen.	<ul style="list-style-type: none"> ➤ Stellt Titel oder Struktur des Themas vor (dann evtl. 3) ➤ Erlebnis des Lehrenden (dann 5) ➤ Problem/Fall der Lernenden (dann 6) ➤ Überfachlicher Bezug (dann 8) ➤ Übungsaufgabe für Lernende (dann Methode) ➤ Einfaches Beispiel ohne Lösungsansätze
2. Visualisierung	Mit inhaltlichem Bezug: Bild, Film, Grafiken, etwas vorführen, Demonstration, Landkarte, etwas wird herumgegeben... Auch haptische Reize sind hier möglich („Tasten Sie mal...“)	<ul style="list-style-type: none"> ➤ Strich, Tabelle, Schaubild, Diagramm (dann evtl. 3) ➤ Bild, welches nicht zum Verständnis oder zur Erinnerung der Inhalte beitragen: z.B. Dekoratives Bild (Natur, Stadt)
3. Struktur/ Systematisierung	Stellt Zusammenhang dar/Gliederung/Vergleiche z.B. mit vergleichender Tabelle, Mind-Map, Concept-Map, Advance Organizer, Schemata, Diagramme usw.	<ul style="list-style-type: none"> ➤ Beiläufiger Zusammenhang/Vergleich ➤ Aufzählung/Auflistung unter Überschrift ➤ Zusammenfassung wichtiger Punkte (dann 9) ➤ Sagt eine Regel auf/ Merkhilfe (dann evtl. 7 oder 9)
4. Lernziele	Stellt Lernziele der Lehrveranstaltung ausdrücklich dar (Evtl. auf Folien). Nennt Thema/Inhalt und zu erwerbende Fähigkeit: z.B. Bluthochdruck erklären/definieren/behandeln Auch ungenaue Aussagen gelten: „wissen/kennen/verstehen“	<ul style="list-style-type: none"> ➤ Sagt was wichtig ist (dann 7) ➤ Unklare Aussage wie: „Sie müssen das Buch lesen“ ➤ Stellt zu behandelndes Thema dar (ohne Kompetenz)
5. Persönlicher Bezug	Äußert beruflichen Bezug/Erlebnisse/eigenes Interesse zu dem Thema der Lehrveranstaltung: „Ich habe dazu promoviert“/„Die Frage beschäftigt mich selbst sehr intensiv...“ „Damit habe ich mir damals selbst schwergetan und habe es erst verstanden, als....“	<ul style="list-style-type: none"> ➤ Allgemeine Aussage, die nicht nah an der Person oder nahe an dem Thema sind: „Ich habe das natürlich auch lernen müssen...“ ➤ „Ich habe zwei Kinder“
6. Handlungsrelevanz	Weist auf (potentielle) praktische Handlungsrelevanz für die Teilnehmer hin, sowohl beruflich als auch privat „Das ist wichtig für Ihre spätere Berufstätigkeit“ „Wenn Sie später arbeiten, dann könnten Sie XY machen“ „In Ihrem Praxiseinsatz haben Sie das schon angewendet...“	<ul style="list-style-type: none"> ➤ Ausschließlich persönlicher beruflicher Bezug des Lehrenden (dann 5) ➤ Allgemeinaussagen: z.B. „Das muss man behandeln“ ➤ Allgemeine Falldarstellung (dann 1) ➤ Ausformulierte Lernziele (4)
7. Wichtige Inhalte	Relevanz der Inhalte in Bezug auf Prüfungen/Promotion: „Das sind typische Prüfungsfragen...“ „So können Sie die Prüfungsfrage leicht beantworten...“	<ul style="list-style-type: none"> ➤ Ausformulierte Lernziele (4) ➤ Praktische Handlungsrelevanz (6) ➤ Allgemeine Aussagen zur Wichtigkeit von XY ➤ Verneinung von Wichtigkeit („...müssen Sie nicht wissen“) ➤ Allgemeiner Hinweis zu Prüfungen (Termine, Prozedur...)
8. Interdisziplinarität	Wissenschaftliche Bezüge, die über das eigentliche Thema der Veranstaltung hinausgehen, werden kurz erklärt. Z.B. Bezug zu Nebenthemen, Erklärungsmodelle/Ansätze anderer Disziplinen, Ergebnisse zu ähnlichen Themen anderer Fächer. „Was sagt die Psychologie zu dem Thema Stress?“	<ul style="list-style-type: none"> ➤ Bezug sehr nahe dem Hauptthema ➤ Bezug zu nicht fachlichen Themen ➤ Praktische Handlungsrelevanz (6) ➤ Praktisches Beispiel ➤ Bezug ohne tiefere Erklärungen
9. Zusammenfassung	Deutliche Wiederholung zuvor genannter und zunächst abgeschlossener Inhalte (auch bei Bezug auf vorherige Lehrveranstaltungen der Lehrenden). Auch in Kombination mit neuen Aspekten des Themas. Evtl. als Zusammenfassung oder als Take-Home-Message: „Hier noch einmal die wichtigsten Fakten...“	<ul style="list-style-type: none"> ➤ Erstmalige Aufzählung/Auflistung von Inhalten ➤ Aspekt des Themas wird in Beziehung zu anderem Themen (dann 8) oder zu handlungsrelevantem Aspekt (dann 6) gesetzt ➤ Wichtiges steht ohne Erklärung an der Tafel/auf der Folie

Fortsetzung Anhang C4. Handanweisung zur Benutzung des BEO-LV

BEO-LV	Beschreibung und Positivbeispiele	Nicht gemeint:
10. Wissensfragen	Stellt Fragen, auf die es eine richtige Antwort geben kann/soll: Definitionen, Fakten, Regeln, Erklärungen, Übersetzungen, Zusammenfassen in eigenen Worten, Analysen, Hypothesen aufstellen, Abwägen, Beurteilen... Unabhängig von der Schwere oder Komplexität der Fragen.	<ul style="list-style-type: none"> ➤ Fragen, die von Lehrenden selbst beantwortet werden ➤ Arbeitsauftrag von individuellen Lernphasen ➤ Fragen, die unter 11-14 fallen
11. Fragen zu Vorwissen/Vorkenntnissen	Frägt, wer die Lernenden sind, was die Lernenden schon wissen/erlebt haben/gelernt haben usw. „Wer hat schon in der Neurologie ein Praktikum gemacht?“ „Wer hatte in der Schule Biologie-Leistungskurs?“	<ul style="list-style-type: none"> ➤ Arbeitsauftrag von individuellen Lernphasen ➤ Fragen ohne inhaltlichen Bezug ➤ Reine Wissensfragen ohne persönlichen Bezug
12. Fragen zu Interesse an den Inhalten	Stellt Fragen danach, was die Lernenden besonders viel/wenig interessiert. „Wer möchte sich mit diesem Thema in Zukunft noch intensiver beschäftigen?“ „Wer möchte später bei XY arbeiten?“	<ul style="list-style-type: none"> ➤ Arbeitsauftrag von individuellen Lernphasen
13. Fragen zum Lernprozess/Verständnis	Thematisiert wie Teilnehmer/innen in der Lehrveranstaltung lernen und dort Verständnis aufbauen. Fragt wie Inhalte verstanden werden. Ermuntert dazu, Fragen zu stellen, um den Inhalt zu verstehen. „Haben Sie noch Fragen?“ „Ist Ihnen klar geworden, was ein XY ist?“ „Fällt es Ihnen leicht, diesen Zusammenhang zu verstehen?“ „Ist Ihnen klar, warum dies so ist?“ „Können Sie XY erklären?“	<ul style="list-style-type: none"> ➤ Dozent referiert über typische Missverständnisse/Fragen anderer Personen. ➤ Angebot, Verständnisfragen im Anschluss oder in der Sprechstunde zu klären. ➤ Frage: „Haben Sie noch Fragen?“ als rhetorische Frage (ohne auf eine Antwort zu warten) ➤ Arbeitsauftrag von individuellen Lernphasen
14. Fragen zu Meinungen/Haltungen	Stellt Fragen, auf die es keine eindeutig richtige Antwort gibt: „Wie stehen Sie zur Stammzellentherapie?“ „Welche Meinung haben Sie zu...?“ (persönliche Meinung der Lernenden)	<ul style="list-style-type: none"> ➤ Fragen zu Positionen, die nicht unbedingt die persönliche Meinung widerspiegeln „Nennen Sie drei typische Pro-Argumente für Stammzellentherapie“ ➤ Arbeitsauftrag von individuellen Lernphasen
Für alle Fragen gilt:	Die Frage muss so gestellt sein, dass jemand der Teilnehmer darauf antworten <u>kann</u> (mehr als eine Sekunde warten). Auch Aufforderungen zählen hier als Frage („Bitte nennen Sie die Definition von XY“). Biographische Fragen ohne direkten inhaltlichen Bezug können unter <i>Sonstiges</i> protokolliert werden. Fragen an einzelne Teilnehmer während individueller Lernphasen werden hier nicht	
15. Studierende präsentieren	Alle inhaltsrelevanten Handlungen von Teilnehmern vor der gesamten Gruppe oder vor Kleingruppen: Referate, Rollenspiele, praktische Demonstrationen, Untersuchungen usw.	<ul style="list-style-type: none"> ➤ Alle Teilnehmer sind gleichzeitig aktiv (dann 17) ➤ Wortmeldungen von Teilnehmern ➤ Diskussionen unter den Teilnehmern
16. Wortmeldungen von Studierenden	Alle Wortmeldungen von den Lernenden (Mehrere Sätze hintereinander im Gespräch zählen als insgesamt eine Wortmeldung) Beiträge zu Diskussionen. Auch Antworten auf Fragen der Lehrenden Auch Aussagen wie „Ich weiß nicht“.	<ul style="list-style-type: none"> ➤ Gespräche unter den Teilnehmern ➤ Aussagen, die nichts mit den Inhalten der Lehrveranstaltung zu tun haben: „Können Sie die Folie noch mal zeigen?“
17. Individuelle Lern- und Arbeitsphasen (ILAP)	Arbeitsaufträge <u>an alle</u> Lernenden: Individuelle Lernphasen mit der Folge mentaler Aktivierung. Z.B. Aufgaben in Einzel-, Partner oder Gruppenarbeit.	<ul style="list-style-type: none"> ➤ Lehrende/r stellt Fragen (10-14) an die Gruppe ➤ Teilnehmer schreiben etwas ab ➤ Abstimmungen ohne direkten Lernbezug: z.B. Vorwissen oder zu Interesse ➤ Einzelne Lernende sind aktiviert (dann 15 oder 16)
Instruktion und Synchronisation	Wird der Arbeitsauftrag umfassend/klar instruiert? - Was? (soll getan werden) - Wie? (schriftlich, mündlich...) - Mit wem? (alleine, zu dritt...) - Wie lange? (Zeit) Die Ergebnisse werden komplett eingeholt, wenn alle Ergebnisse der ILAP besprochen werden.	<ul style="list-style-type: none"> ➤ Lehrfilme ➤ Etwas wird ohne Arbeitsauftrag herumgegeben/ausgeteilt (evtl. 2)

Anhang D. Empirische Studie 2, Vorstudie 2: Beschreibung der Videos

Video	Thema	Dauer (in Minuten)	Anzahl didaktischer Handlungen	<i>p</i>		
				Experten	Novizen	Gesamt
1	Burnout	8.23	6	.83	.73	.78
2	Geschichte der Kindheit (Teil1)	9.18	4	.95	.80	.87
3	Basisemotionen	6.21	4	.78	.71	.75
4	Entwicklungspsychologie (Teil 1)	8.23	4	.92	.77	.84
5	Entwicklungspsychologie (Teil 2)	10.13	4	.88	.74	.81
6	Geschichte der Kindheit (Teil 2)	12.02	2	.80	.81	.81
7	Wissenschaftliche Hypothesen	8.07	5	.70	.70	.70
8	Vegan-vegetarische Erziehung	14.33	3	.82	.77	.79
9	Shell-Jugendstudie (Teil 1)	6.51	5	.92	.84	.89
10	Shell-Jugendstudie (Teil 2)	9.41	4	.87	.76	.81
Mittelwert		9.20	4,1	.85.	.76	.80

Anmerkungen. Die p-Werte geben die Schwierigkeit je Video an (Anteil richtiger Antworten). Alle Videos sind abrufbar unter:

<https://sites.google.com/site/beobachtungsbogen/home/testvideos>

Anhang E. Beobachtungsbogen für kollegiale Hospitationen

Selbstbeobachtungsbogen - Dozenten

Lehrende/r:
Beobachter/-in:
Thema:

Wenn Sie eine didaktische Handlung **nicht eingesetzt** haben, dann lassen Sie die entsprechende Antwortkategorie bitte einfach offen!

Welche didaktischen Handlungen habe ich in der Lehrveranstaltung eingesetzt?

Einleitung/Hinführung zum Thema

Wie?

Visualisierung

Wie?

Strukturierung des Themas

Wie?

Nennung der Lernziele

(Wenn einmal „Etwas“, wenn mehrmals „Viel“)

Wie?

Zusammenfassung einzelner Abschnitte

Wie?

Darstellung der Praxisrelevanz der Inhalte

Wie?

Darstellung der Prüfungsrelevanz der Inhalte

Wie?

Abschließende Zusammenfassung

Wie?

Sonstiges:

DOS-Nr.:

Datum:

Veranstaltungsformat:
Anzahl der Teilnehmer:
Dauer:

Welche Kommunikation und Interaktion fand zwischen mir und den Teilnehmern statt?

Etwas Viel

Wissensfragen gestellt

Wie?

Nach Vorwissen/Vorkenntnissen gefragt

Wie?

Nach Interesse an den Inhalten gefragt

Wie?

Fragen zum Lernprozess/Verständnis gestellt

Wie?

Meinungen/Haltungen aufgegriffen

Wie?

Diskussionen entstehen (lassen)

Wie?

Den Studierenden Feedback gegeben

Wie?

Feedback von den Studierenden erhalten

Wie?

Habe ich individuelle Lern- und Arbeitsphasen eingesetzt?

Ja Nein

Habe ich die individuellen Lernphasen klar instruiert?

klare Instruktion unklare Instruktion

Habe ich die Ergebnisse individueller Lernphasen synchronisiert?

ja, sehr ausführlich nein, gar nicht

Fortsetzung Anhang E. Beobachtungsbogen für kollegiale Hospitationen

Zusammenfassung:

Was ist gut gelungen?

Die Inhalte sind vom Umfang her ...

zu umfassend erweiterbar

Was meinen Sie, haben die meisten Studierenden die Lernziele erreicht?

voll erreicht nicht erreicht

Insgesamt sollte die Lehrveranstaltung ...

genau so bleiben stark überarbeitet werden

Anteil Dozentenzentrierter Lehre (Kollektive Lernphasen):

0% 100%
50 %

Was hätte Ihrer Meinung nach anders gemacht werden können?

Sonstige Anmerkungen (Lernziele, Methodeneinsatz, Synchronisation...):

Anmerkung: Bitte beachten Sie, dass ein Abschluss der Praxisberatung nur im Tandem und nach gegenseitiger Hospitation möglich ist. Bringen Sie bitte zur Abschlussreflexion die ausgefüllten Evaluationsbögen (Selbst- und Fremdbeobachtungsbogen) Ihrer Hospitationen mit. Die Beobachtungsbögen werden im Rahmen der wissenschaftlichen Begleitung von DOS ausgewertet. Dabei werden die Daten garantiert unabhängig von Ihrem Namen gespeichert. Behalten Sie am besten auch eine eigene Kopie. Sie können diese Ergebnisse in die schriftliche Abschlussreflexion von DOS einfließen lassen, falls Sie das Gesamtzertifikat Hochschuldidaktik Baden-Württemberg anstreben.

Anhang F1. Items der Studierendenbefragung

Skala	Item im Wortlaut (in Klammern die Abkürzung)
Akzeptanz aktivierender Methoden in Vorlesungen (AM-V) ^a	Ich bevorzuge ausschließlich frontalen Unterricht in Vorlesungen*. (AMV1)
	Ich mag kleine Einzelarbeitsphasen in Vorlesungen. (AMV2)
	Ich mag kleine Diskussionsphasen in Vorlesungen. (AMV3)
	Vorlesungen sind nicht der richtige Ort für aktivierende Lehrmethoden*. (AMV4)
	In einer Vorlesung sollte sich der Dozentenvortrag regelmäßig mit aktivierenden Lehr- und Lernmethoden abwechseln. (AMV5)
	Ich mag es <u>nicht</u> , wenn ich in Vorlesungen kleine Aufgaben mit anderen bearbeiten soll (Kleingruppenarbeit)*. (AMV6)
	Ich mag es <u>nicht</u> , in Vorlesungen den Auftrag zu bekommen, mit den Sitznachbarn zu diskutieren (Partnerarbeit)*. (AMV7)
	Ich mag es <u>nicht</u> , in Vorlesungen Fragen der Lehrenden zu beantworten*. (AMV8)

Anmerkungen. Antwortformat: a = 1 (Stimme nicht zu), 2 (Stimme eher zu), 3 (Stimme teilweise zu), 4 (Stimme eher nicht zu), 5 (Stimme zu), Kann ich so pauschal nicht sagen.

Fortsetzung Anhang F1. Items der Studierendenbefragung

Skala	Item im Wortlaut (in Klammern die Abkürzung)
Aktivierende Methoden in Vorlesungen und Seminaren (AM-VS) ^b	<p>Wenn Sie die Didaktik in den "theoretischen" Lehrveranstaltungen (Vorlesungen und Seminare) mit mindestens 30 Teilnehmern Ihres Studiums für sich selbst bestimmen könnten, welche Verteilung würden Sie wählen? Geben Sie bitte jeweils eine Prozentzahl an (zwischen 0 und 100).</p> <ul style="list-style-type: none"> • Die Lehrenden vermitteln Fachwissen ('Frontalunterricht'). (AMVS1) • Die Lehrenden stellen Fragen und diskutieren mit den Studierenden (oder regen Diskussionen an). (AMVS2) • Sie erarbeiten in Einzelarbeit Ergebnisse (Übungsaufgaben, Fallbearbeitung usw.). (AMVS3) • Sie erarbeiten in Einzelarbeit Ergebnisse (Übungsaufgaben, Fallbearbeitung usw.). (AMVS4) • Sie erarbeiten zu dritt oder viert Ergebnisse (Diskussion, Übungsaufgaben, Fallbearbeitung usw.). (AMVS5)
Aktivierende Methoden in Modulen (AM-M) ^c	<p>Stellen Sie sich bitte eine zweiwöchige Lehrveranstaltungsreihe vor, an deren Ende eine Prüfung steht. Unter welchen Umständen würden Sie einen eher frontalen Unterricht bevorzugen und unter welchen Umständen einen eher aktivierenden Unterricht mit vielen individuellen Lernphasen?</p> <ul style="list-style-type: none"> • Interessantes Fach - Einfache Prüfung. (AMM1) • Uninteressantes Fach - Einfache Prüfung (AMM2) • Interessantes Fach - Schwierige Prüfung (AMM3) • Uninteressantes Fach - Schwierige Prüfung (AMM4)

Anmerkungen. Antwortformat: b = Prozentsummenskala. c = 1 (Durchgehend Frontalunterricht), 2 (Eher Frontalunterricht), 3 (Beides etwa gleich), 4 (Eher aktivierenden Unterricht), 5 (Durchgehend aktivierenden Unterricht), Kann ich so pauschal nicht sagen.

Fortsetzung Anhang F1. Items der Studierendenbefragung

Probleme bei der Umsetzung aktivierender Methoden (AM-P) ^d	<p>Bei aktivierenden Lehrmethoden sollen die Studierenden selbst aktiv werden. Beispielsweise durch Übungsaufgaben oder Diskussionen in einer Kleingruppe. Nun folgen mögliche Hindernisse, welche theoretisch bei der erfolgreichen Umsetzung von aktivierenden Lehrmethoden in Vorlesungen auftreten können.</p> <p>Haben Sie in Ihrem Studium bereits solche Erfahrungen gemacht?</p> <p>Die Kommilitonen waren unmotiviert (in dem Moment). (AMP1)</p> <p>Die Darstellung der Ergebnisse hat zu lange gedauert. (AMP2)</p> <p>Die Aufgabe war uninteressant. (AMP3)</p> <p>Ich selber hatte keine Lust mitzuarbeiten (in dem Moment). (AMP4)</p> <p>Gespräche haben mich abgelenkt. (AMP5)</p> <p>Die Aufgabe war zu einfach. (AMP6)</p> <p>Die Aufgabe war unklar instruiert. (AMP7)</p> <p>Die Aufgabe war zu schwer. (AMP8)</p> <p>Ich hatte keinen geeigneten Partner. (AMP9)</p> <p>Mir war unklar, mit wem ich die Aufgabe bearbeiten soll. (AMP10)</p>
---	---

Anmerkungen. Antwortformat: d = 1 (Nein, habe ich nie erlebt), 2 = (Nur selten erlebt), 3 (Gelegentlich erlebt), 4 (Ja, habe ich oft erlebt), Habe keine Erfahrungen damit.

Fortsetzung Anhang F1. Items der Studierendenbefragung

Skala	Item im Wortlaut (in Klammern die Abkürzung)
	Im Studium geht es mir darum,...
Lernzielorientierung (LO) ^e	... ein tiefes Verständnis für die Inhalte zu erwerben. (LO1) ...zum Nachdenken angeregt zu werden. (LO2) ...so viel wie möglich zu verstehen. (LO3) ...komplizierte Inhalte genau zu verstehen. (LO4)
Annäherungs- Leis- tungszielorientierung (AL) ^e	...zu den Besten des Semesters zu gehören. (AL1) ... sehr gute Noten oder Beurteilungen zu erzielen. (AL2) ...zu zeigen, dass ich die Inhalte beherrsche.(AL3)
Vermeidungs- Leistungszielorientierung (VL) ^e	...dass niemand denkt, ich sei weniger klug als andere. (VL1) ...dass niemand merkt, wenn ich etwas nicht verstehe. (VL2) ...mich nicht zu blamieren (z. B. durch falsche Ergebnisse oder triviale Fragen). (VL3)
Arbeitsvermeidung (AV) ^e	...mit wenig Arbeit durch das Studium zu kommen. (AV1) ...den Arbeitsaufwand stets gering zu halten. (AV2) ...nicht so schwer zu arbeiten. (AV3)
Interesse (INT) ^e	Ich finde mein Studium richtig interessant. (INT1) Ich beschäftige mich gerne mit den Inhalten meines Studi- ums. (INT2) Im Studium lerne ich Inhalte, die mich begeistern. (INT3)
Soziale Eingebundenheit (SE) ^e	In meiner Uni fühle ich mich sozial gut eingebunden. (SE 1) Mit meinen Kommilitonen verstehe ich mich insgesamt sehr gut. (SE 2) Ich habe im Studium viele Leute, mit denen ich mich gut verstehe. (SE 3)

Anmerkungen. Antwortformat: e = 1 (Stimmt nicht), 2 (Stimmt eher nicht), 3 (Stimmt teils), 4 (Stimmt eher), 5 (Stimmt genau), Kann oder will ich nicht beurteilen. * Items wurden rekodiert.

Fortsetzung Anhang F1. Items der Studierendenbefragung

Studienabschnitt	In welchem Studienabschnitt sind Sie jetzt in diesem Sommersemester 2012? Vorklinik/ Propädeutik/ Klinik/ Praktisches Jahr (PJ)/ Sonstiges
Noten im Studium	Bitte denken Sie nun an Ihre bisherigen Benotungen im Studium insgesamt. Diese entsprechen etwa folgendem Notenschnitt: 1.0 – 1.4/ 1.5 – 1.9/ 2.0 – 2.4/2.5 – 2.9/3.0 – 3.4/ 3.4 oder höher/ Weiß nicht/ Keine Angabe
Geschlecht	Weiblich/ Männlich/ Keine Angabe
Wie ist Ihre Meinung zu diesem Fragebogen?	Gut/ Mäßig /Schlecht/ Sonstiges
Einverständnis zur wissenschaftlichen Nutzung der Daten	Ja, ich bin einverstanden Nein, ich bin nicht einverstanden

Anhang F2. Studierendenbefragung: Deskriptive Statistiken der in Hauptstudie eingesetzten Items

Item	<i>M</i>	<i>SD</i>	Median	Min	Max	<i>N</i>
AMV1	2.39	1.16	2	1	5	364
AMV2	2.46	1.37	2	1	5	372
AMV3	3.03	1.23	3	1	5	372
AMV4	2.66	1.32	3	1	5	357
AMV5	2.81	1.2	3	1	5	364
AMV6	2.11	1.28	2	1	5	370
AMV7	2.34	1.39	2	1	5	370
AMV8	3.24	1.39	3	1	5	370
AMVS1	53.37	22.72	50	0	100	374
AMVS2	22.73	15.13	20	0	80	374
AMVS3	8.19	9.40	5	0	60	374
AMVS4	8.75	9.07	5	0	50	374
AMVS5	6.92	9.39	5	0	80	374
AMM1	2.58	1.07	2	1	5	353
AMM2	3.17	1.19	3	1	5	356
AMM3	2.54	1.19	2	1	5	349
AMM4	2.72	1.33	2	1	5	344
AMP1	2.85	0.88	3	1	4	355
AMP2	2.62	0.81	3	1	4	354
AMP3	3.17	0.91	3	1	4	356
AMP4	3.43	0.73	4	1	4	366
AMP5	2.92	0.95	3	1	4	364
AMP6	3.00	0.86	3	1	4	368
AMP7	2.47	0.98	2	1	4	366
AMP8	3.02	0.85	3	1	4	362
AMP9	2.79	0.93	3	1	4	364
AMP10	1.76	0.91	2	1	4	353

Anmerkungen. AM-V = Aktivierende Methoden in Vorlesungen. AM-VS = Aktivierende Methoden in Vorlesungen und Seminaren. AM-M = Aktivierende Methoden in Modulen; AM-P = Probleme bei der Umsetzung aktivierender Methoden.

Fortsetzung Anhang F2. Studierendenbefragung: Deskriptive Statistiken der in der Hauptstudie eingesetzten Items

Item	<i>M</i>	<i>SD</i>	Median	Min	Max	<i>N</i>
LO1	4.58	0.69	5	1	5	374
LO2	4.22	0.84	4	1	5	374
LO3	4.65	0.64	5	1	5	374
LO4	4.29	0.81	4	1	5	373
AL1	2.69	1.22	3	1	5	370
AL2	3.50	1.09	4	1	5	373
AL3	3.67	1.06	4	1	5	374
VL1	2.18	1.23	2	1	5	365
VL2	1.97	0.98	2	1	5	371
VL3	2.70	1.20	3	1	5	373
AV1	2.12	1.00	2	1	5	373
AV2	2.42	1.01	2	1	5	371
AV3	2.25	1.01	2	1	5	371
INT1	4.34	0.83	5	1	5	374
INT2	4.22	0.77	4	2	5	372
INT3	4.13	0.88	4	1	5	372
SE1	3.95	1.09	4	1	5	374
SE2	4.15	0.93	4	1	5	371
SE3	4.08	1.09	4	1	5	373

Anmerkungen. AM-VS = Aktivierende Methoden in Vorlesungen und Seminaren. AM-M = Aktivierende Methoden in Modulen. AM-V = Aktivierende Methoden in Vorlesungen. AM-P = Probleme bei der Umsetzung aktivierender Methoden. LO = Lernzielorientierung. AL = Annäherungs-Leistungszielorientierung. VL = Vermeidungs-Leistungszielorientierung. AV = Arbeitsvermeidung. SE = Soziale Eingebundenheit.

Anhang G1. Trainerbefragung zur Effektivität der Dozentenschulung

Die Lehrenden schätzen vor DOS 1 und ein Jahr später ein, inwiefern sie die folgenden didaktischen Handlungen umsetzen (indirekte Veränderungsmessung). Zusätzlich frage ich die Lehrenden, ob sich die Lehrhandlungen in den letzten 12 Monaten verändert haben (direkte Veränderungsmessung). So erhalten wir ein genaueres Bild von den Veränderungen, welche möglicherweise von der DOS angeregt werden. Jetzt geht es darum, zu prüfen, auf welche didaktischen Handlungen die DOS womöglich einen unmittelbaren Einfluss haben kann und auf welche eher nicht. Bitte überlegen Sie nun, **welche unmittelbare Wirkung die DOS Ihrer Meinung nach insgesamt auf die Teilnehmer hat?**

Didaktische Handlung (Aus Sicht der Lehrenden formuliert)	Negativer Einfluss (-3) bis positiver Einfluss (+3) Einfluss von DOS
Ich versuche, die Studierenden für mein Fach zu begeistern.	
Ich ermutige Studierende dazu, eigene Lösungen zu finden (z. B. in Gruppenarbeit).	
Ich lehre mit einem ausgefeilten Präsentationsstil und sehr guter Rhetorik.	
Ich teile den Studierenden die Lernziele mit.	
Ich fördere die anwendungsorientierte Auseinandersetzung der Studierenden mit den Inhalten.	
Ich fördere die besten Studierenden.	
Ich richte meine Lehre an den Lernzielen aus.	
Ich zeige die praktische Relevanz der Inhalte auf.	
Ich kommuniziere intensiv mit den Studierenden.	
Ich reflektiere mit den Studierenden gemeinsam deren Lernprozess.	
Ich gestalte die Lehre methodisch sehr abwechslungsreich.	
Ich helfe (schwächeren) Studierenden, die Unterstützung benötigen.	
Ich lasse die Studierenden wichtige Inhalte selbst zusammenfassen.	
Ich strukturiere die Inhalte klar.	
Ich unterbinde auftretende Störungen konsequent.	
Ich binde alle Studierenden aktiv ein (Übungsaufgaben, Gruppenarbeit etc.).	
Ich fasse wichtige Inhalte zusammen.	
Ich bereite die Studierenden auf die Prüfung(en) vor.	

-3,-2,-1 = Negativer Einfluss: Die Lehrenden zeigen das Verhalten nach DOS seltener

0 = Kein Einfluss: DOS beeinflusst das Verhalten nicht

+3, +2, +1 = Positiver Einfluss: Die Lehrenden zeigen das Verhalten nach DOS häufiger.

Hohe Werte bedeuten stärkeren negativen bzw. stärkeren positiven Einfluss

Fortsetzung Anhang G1. Trainerbefragung zur Effektivität der Dozentenschulung

Eine weitere Datenquelle stellen die Hospitationen dar. Untersucht wird hier, inwiefern sich die Lehre von Lehrenden ohne DOS von der Lehre von Lehrenden mit DOS (aktuell) bzw. von DOS-Absolventen unterscheidet. Hier sind verschiedene Verhaltensweisen aufgelistet, welche den Beobachtungen zugrunde lagen. Bitte überlegen Sie nun, welche unmittelbare Wirkung DOS Ihrer Meinung nach insgesamt auf die Teilnehmer hat?

Didaktische Handlung (Aus Beobachtersicht formuliert)	Negativer Einfluss (-3) bis positiver Einfluss (+3) Einfluss von DOS
Lehrende zeigen Fallbeispiele oder Herausforderungen auf.	
Lehrende visualisieren die Inhalte.	
Lehrende zeigen die Struktur der Inhalte auf/ Systematisieren die Inhalte.	
Lehrende teilen den Studierenden die Lernziele mit.	
Lehrende äußern den persönlichen Bezug zu den Themen.	
Lehrende zeigen die praktische Handlungsrelevanz der Inhalte auf.	
Lehrende weisen auf die Prüfungsrelevanz von Inhalten hin.	
Lehrende stellen einen (interdisziplinären) Bezug zu anderen Themen her.	
Lehrende fassen einzelne Abschnitte/Wichtiges zusammen.	
Lehrende stellen Wissensfragen.	
Lehrende stellen Fragen zum Vorwissen/zu Vorkenntnissen.	
Lehrende stellen Fragen zum Interesse der Studierenden an den Inhalten.	
Lehrende stellen Fragen zum Lernprozess oder zum Verständnis.	
Lehrende stellen Fragen zu Meinungen oderhaltungen.	
Lehrende setzen individuelle Lern-und Arbeitsphasen in den Lehrveranstaltungen um.	

-3,-2,-1 = Negativer Einfluss: Die Lehrenden zeigen das Verhalten nach DOS seltener

0 = Kein Einfluss: DOS beeinflusst das Verhalten nicht

+3, +2, +1 = Positiver Einfluss: Die Lehrenden zeigen das Verhalten nach DOS häufiger.

Hohe Werte bedeuten stärkeren negativen bzw. stärkeren positiven Einfluss

Anhang G2. Trainerbefragung: Deskriptive Statistiken aller Items

Studierendenzentrierte Lehre	<i>M</i>	<i>SD</i>
Ich ermutige Studierende dazu, eigene Lösungen zu finden	1.64	0.50
Ich fördere die anwendungsorientierte Auseinandersetzung der Studierenden mit den Inhalten.	1.36	0.92
Ich kommuniziere intensiv mit den Studierenden.	1.45	0.52
Ich reflektiere mit den Studierenden gemeinsam deren Lernprozess.	0.64	0.50
Ich gestalte die Lehre methodisch sehr abwechslungsreich.	1.64	0.50
Ich lasse die Studierenden wichtige Inhalte selbst zusammenfassen.	1.09	0.54
Ich binde alle Studierenden aktiv ein	1.64	0.67
Lernziele in der Lehre		
Ich teile den Studierenden die Lernziele mit.	2.09	0.94
Ich richte meine Lehre an den Lernzielen aus.	1.82	0.60
Allgemeine didaktische Handlungen		
Ich versuche, die Studierenden für mein Fach zu begeistern.	0.91	0.94
Ich lehre mit einem ausgefeilten Präsentationsstil und sehr guter Rhetorik.	0.55	0.82
Ich fördere die besten Studierenden.	0.18	0.60
Ich zeige die praktische Relevanz der Inhalte auf.	1.27	0.47
Ich helfe (schwächeren) Studierenden, die Unterstützung benötigen.	0.73	0.79
Ich strukturiere die Inhalte klar.	2.09	0.83
Ich unterbinde auftretende Störungen konsequent.	0.09	0.30
Ich fasse wichtige Inhalte zusammen.	1.73	0.65
Ich bereite die Studierenden auf die Prüfung(en) vor.	0.82	0.87

Anmerkungen. Didaktische Handlungen (aus Sicht der Lehrenden formuliert) auf einer Skala von -3 (Negativer Einfluss) bis +3 (positiver Einfluss).

Fortsetzung Anhang G2. Trainerbefragung: Deskriptive Statistiken aller Items

	<i>M</i>	<i>SD</i>
Lehrende zeigen Fallbeispiele oder Herausforderungen auf.	1.45	0.52
Lehrende visualisieren die Inhalte.	1.09	0.70
Lehrende zeigen die Struktur der Inhalte auf/ Systematisieren die Inhalte.	1.73	0.90
Lehrende teilen den Studierenden die Lernziele mit.	2.27	0.65
Lehrende äußern den persönlichen Bezug zu den Themen.	0.91	0.83
Lehrende zeigen die praktische Handlungsrelevanz der Inhalte auf.	1.27	0.47
Lehrende weisen auf die Prüfungsrelevanz von Inhalten hin.	0.73	0.90
Lehrende stellen einen (interdisziplinären) Bezug zu anderen Themen her.	0.64	0.67
Lehrende fassen einzelne Abschnitte/Wichtiges zusammen.	1.27	0.65
Lehrende stellen Wissensfragen.	1.00	1.55
Lehrende stellen Fragen zum Vorwissen/zu Vorkenntnissen.	1.73	0.65
Lehrende stellen Fragen zum Interesse der Studierenden an den Inhalten.	0.73	0.65
Lehrende stellen Fragen zum Lernprozess oder zum Verständnis.	1.00	0.45
Lehrende stellen Fragen zu Meinungen oder Haltungen.	1.18	0.75
Lehrende setzen individuelle Lern- und Arbeitsphasen in den Lehrveranstaltungen um.	2.18	0.60

Anmerkungen. Didaktische Handlungen (aus Beobachtersicht formuliert) auf einer Skala von -3 (Negativer Einfluss) bis +3 (positiver Einfluss).