

Andreas MÜLLER<sup>1</sup> (München)

## Prüfungen als Lernchance – Vorstellung und Überprüfung eines Rahmenmodells

### Zusammenfassung

Universitäre Prüfungen haben im Zuge der Bolognareform eine Bedeutungsaufwertung erfahren, wobei bislang diesbzgl. Lernpotentiale nicht hinreichend thematisiert werden. Der vorliegende Artikel stellt einen erweiterten Prüfungs- und Feedbackzyklus vor, der explizit die Entwicklungsfunktion von der Prüfungsvorbereitung bis zur -nachbereitung im Kontext lernerzentrierter Hochschulveranstaltungen fokussiert. Hierzu wird ein qualitatives Dozentenfeedback ergänzt durch eine quantitative Peerevaluation sowie durch ein Instrument zur Selbstbeobachtung der aktuellen Referenten. Erste Analysen zum Aufwand bei der Implementation, zur Güte des eingesetzten Instruments sowie zur Bewertung durch die Studierenden rechtfertigen den Einsatz im Seminarkontext.

### Schlüsselwörter

Entwicklungsfunktion, Modularisierung, Prüfung, Rückmeldung, Lernen

## Assessments as opportunity to learn – presentation and verification of a conceptual framework

### Abstract

In the course of the Bologna reform assessments at universities have gained increasing recognition. However the underlying learning potential obtained through this new emphasis on assessment has not been addressed sufficiently so far. This paper presents an extended assessment- and feedback-cycle that focuses explicitly on the developmental function of the pre- and postprocessing of assessments in learner-centred courses. Therefore a qualitative feedback of a lecturer is supplemented by a quantitative peer-evaluation and an instrument for self-evaluation of the current speaker. A first analysis regarding the effort of the implementation, the quality of the utilized instruments and the appraisal by the students justifies the implementation in the context of courses at universities.

### Keywords

Development, assessment, feedback, learning

---

<sup>1</sup> e-Mail: [asmueller@edu.lmu.de](mailto:asmueller@edu.lmu.de)

# 1 Bedeutungsaufwertung von Prüfungen im Rahmen des Bologna-Prozesses

Als eine wesentliche Folge des Bologna-Prozesses, lässt sich eine deutliche Veränderung der Prüfungslandschaft an den deutschen Hochschulen konstatieren. Durch die Modularisierung der Studiengänge und die daran gekoppelte Nachweispflicht für belegte Veranstaltungen hat die Prüfungsdichte, vor allem im Vergleich zu traditionellen Magisterstudiengängen, deutlich zugenommen. Lehrende sehen sich im Zuge der Umstellung darüber hinaus vor allem mit zwei Herausforderungen konfrontiert: Zum einen erfolgte im Zuge der Akkreditierungsverfahren der Bachelor-Studiengänge eine Standardisierung und Reglementierung von Prüfungen z.B. hinsichtlich Form oder Umfang.

Zum anderen ergeben sich für die Prüfenden Herausforderungen durch die explizite Kompetenzorientierung der Studiengänge sowie die Kopplung konkreter Kompetenzziele an einzelne Module (BÜLOW-SCHRAMM, 2008). Auch wenn die Definition von Lernzielen eine Grundvoraussetzung für erfolgreiche Lehre und sinnvolles Prüfen darstellt, bringt die Formulierung und Operationalisierung von Kompetenzziele in der Praxis erhebliche Schwierigkeiten mit sich (KURZ, 2010; SIPPEL, 2009; WANNEMACHER, 2009).

Die gewachsene Bedeutung von Prüfungen im Hochschulkontext spiegelt sich in einer steigenden Anzahl von Publikationen wider, die sich explizit mit diesem Themenfeld auseinandersetzen. Vertieft diskutiert werden derzeit vor allem die Umsetzung der Kompetenzorientierung und Prüfungsmethodik (vgl. DANY, SZCZYRBA & WILDT, 2008), Fragen zur Vergleichbarkeit von Prüfungen (KURZ, 2010) sowie zu deren Validität, Reliabilität und Objektivität (DUBS, 2006; STARY, 2006). Mit Blick auf die rein quantitative Zunahme der abzuwickelnden Prüfungen rücken auch die Standardisierung von Testsituationen und automatisierte Prozesse der Erstellung von (Multiple-Choice-)Prüfungen in den Fokus der Aufmerksamkeit (HEES, HERMANN & HUSON, 2008; REEPMEYER, 2008).

Wenngleich diese Aspekte der Prüfungsvorbereitung und -umsetzung bedeutsam erscheinen, um die Qualität universitärer Prüfungen zu gewährleisten, fällt doch auf, dass eine systematische Betrachtung des gesamten Prüfungsprozesses derzeit noch weitgehend aussteht. Auch die funktionale Verortung von universitären Prüfungen mit Bezug auf verschiedene Beteiligte sowie eine Diskussion möglicher Lernpotentiale und deren Realisierung im Rahmen unterschiedlicher Veranstaltungsformen erfolgen bislang erst ansatzweise (MÜLLER & SCHMIDT, 2009).

Der vorliegende Beitrag setzt an diesem Punkt an: Auf der Basis einer kurzen Skizzierung möglicher Funktionen von Prüfungen wird ein prozedural und funktional differenzierendes Rahmenmodell zur Realisierung der Entwicklungsfunktion von Prüfungen in lernerzentrierten Veranstaltungsformen vorgestellt. Dieses wird ergänzt durch praktische Erfahrungen und empirische Befunde zur Implementierung des Modells.

## 2 Funktionen von Prüfungen

Eine grundlegende Systematisierung von Prüfungsfunktionen findet sich bereits bei Flechsig (1976). Er unterscheidet zwischen den in Tabelle 1 aufgelisteten drei Kategorien (vgl. ausführlicher MÜLLER & SCHMIDT, 2009).

Unter der erstgenannten *Herrschafts- und Sozialisierungsfunktion* werden Aspekte subsummiert, die aus systemischer Perspektive primär dem Erhalt und der Reproduktion vorhandener Zustände und Machtkonstellationen dienen. Die zweite Kategorie der *Rekrutierung* umfasst Prozesse der Auswahl oder Selektion. Entweder aus sozialer Perspektive durch Rangreihen (Platzierung) oder mittels einem vorab festgelegten Kriterium bzw. Schwellenwert, der überschritten werden muss (Auslese). In Anlehnung an die funktionale Differenzierung von Evaluationen lässt sich dieser Funktionsbereich auch als summative Bewertung bezeichnen. Summativ bezieht sich hier jedoch weniger auf die zeitliche Anordnung einer Prüfung am Ende einer Veranstaltung oder eines Studiums<sup>2</sup>, sondern beschreibt primär deren Charakter als Instrument der verbindlichen Notengebung und Selektion (vgl. IRONS, 2008).

Die dritte Kategorie umfasst nach FLECHSIG (1976) *didaktische Funktionen*. Prüfungen dienen demnach als zeitliche und inhaltliche Gliederungs- und Orientierungspunkte im Studienverlauf, als Mittel der extrinsischen Lernmotivation (HECKHAUSEN, 1989; NEWSTEAD & HOSKINS, 2001) sowie als Instrument, das aus diagnostischer Perspektive Rückschlüsse auf bis dato realisierte Lern- und Bildungsprozesse sowie mögliche Verbesserungsmaßnahmen zulässt. Zu beachten bleibt, dass die mit dem letztgenannten Aspekt verbundenen Lernchancen im Kontext von Prüfungen nicht nur auf die Prüflinge selbst zu beziehen sind, sondern auch auf die prüfende Person sowie das prüfende System als Gesamtes (vgl. HOCHSCHULREKTORENKONFERENZ, 2006, S. 17).

Analog zur Funktionsdifferenzierung im Evaluationskontext lässt sich die Entwicklungsperspektive im Prüfungskontext mit Blick auf die Studierenden als formative Bewertung beschreiben. IRONS (2008, S. 7) betont unter Bezug auf die universitäre Praxis in modularisierten Studiengängen, dass der Unterschied zu summativen Maßnahmen vor allem darin besteht, ob mit einer Prüfungssituation eine Entwicklungsmöglichkeit und Lernchance für die Studierenden verbunden ist. Realisiert werden kann dieser Förderimpuls über *formative feedback* welches verstanden wird als

„any information, process or activity which affords or accelerates student learning based on comments relating to either formative assessment or summative assessment activities.“

Wie BOUD (2007) in einer vergleichenden Analyse von Prüfungsordnungen im angloamerikanischen Raum feststellt, stehen im universitären Kontext derzeit allerdings Aspekte der Herrschafts-, Sozialisierungs- und Rekrutierungsfunktion von Prüfungen im Vordergrund. Allenfalls bei nachgelagerten Überlegungen

---

<sup>2</sup> Moduleinzelnoten können durchaus als zeitlich-formative Prüfungen im Rahmen des Studienverlaufs angesehen werden.

werden Prüfungen als Informationsquelle und Entwicklungsinstrument für Studierende bedacht. Diese Analyse dürfte auch für deutsche Studiengänge nicht anders ausfallen. Dementsprechend überrascht es nicht, dass der Fokus der Hochschulforschung primär auf der Sicherung von Validität, der Entwicklung reliabler Testverfahren oder der Objektivität von Hochschulprüfungen liegt (DANY, SZCZYRBA & WILDT, 2008; MÜLLER-BENEDICT, 2005). Mit einem solch verengten Blick auf die reine Selektionsfunktion von Prüfungen bleiben jedoch wertvolle Lernpotentiale für alle Beteiligten ungenutzt.

Tabelle 1: Funktionen von Prüfungen (nach FLECHSIG, 1976)

<b>Herrschafts- und Sozialisierungsfunktion (Systemreproduktion)</b>
Initiation
Statusverteilung
Legitimation
<b>Rekrutierungsfunktion (Selektion, Auswahl, Zuordnung, Linearisierung)</b>
Platzierung in Kohorte
Auslese aus Kohorte
Qualifikations- und Kompetenznachweis
<b>Didaktische Funktion (Systemoperationen)</b>
Zeitliche und inhaltliche Gliederung des Studiengangs
Orientierung der Lehrenden und Lernenden über die Studienziele
Extrinsische Lernmotivation
Diagnoseinstrument
Rückmeldung des Lehr-Lernerfolgs an Lehrende und Lernende

Notwendig erscheint hier eine umfassendere Perspektive, die Prüfungen als integralen Bestandteil von Lernprozessen versteht und mögliche didaktische Funktionen aufgreift. Differenziert man nach lehrer- und lernerzentrierten Veranstaltungsformen, können folgende didaktischen Potentiale von Prüfungen diskutiert werden (vgl. FLECHSIG, 1976; ILLER & WICK, 2009): Im Rahmen lehrerzentrierter Veranstaltungsformen, in denen vor allem die Vermittlung deklarativen (Grund-)Wissens im Vordergrund steht, erfolgt durch Prüfungen durchaus eine Gliederung des Studiengangs (in Einzelveranstaltungen bzw. Module) sowie eine Orientierung hinsichtlich inhaltlicher Studienziele (präsentierte Inhalte in der Veranstaltung).

Über motivationale Aspekte hinaus dürfte der Beitrag von Prüfungen in lehrerzentrierten Veranstaltungsformen für den Lernprozess der Studierenden jedoch eher gering ausfallen. Zwar sind im Anschluss an standardisierte Multiple-Choice-Tests eine Vielzahl von Diagnose- und Rückmeldeformen denkbar, derzeit realisiert werden jedoch eher „einfache Feedbackarten“ (NARCISS, 2006, S. 21 ff.) im Sinne von *knowledge of performance* („60% der erreichbaren Punktzahl“), *knowledge of result* („falsch vs. richtig“) oder *knowledge of correct result* (Angabe der korrekten Lösung). Elaboriertere Feedbackformen (z.B. mit Informationen zur

Aufgabenstellung oder Hinweisen auf Lösungsstrategien) sind durchaus denkbar, setzen jedoch eine umfassende Kategorisierung der Testaufgaben, z.B. entsprechend einer Taxonomie von Lernzielen (vgl. ANDERSON & KRATHWOHL, 2001) sowie eine tiefgreifende Analyse der vorliegenden studentischen Antworten voraus und sind daher mit einem beträchtlichen Aufwand verbunden. Darüber hinaus bleibt zu beachten, dass die Hinweise für die Studierenden inhaltlich an den Rahmen der zugrundeliegenden Testsituation gebunden bleiben und somit primär deklaratives Wissen und weniger anwendungsorientierte Kompetenzen zum Inhalt haben können.

Mit Blick auf die in den modularisierten Studiengängen angestrebte Kompetenzorientierung erscheinen demgegenüber vorrangig Seminarveranstaltungen, die den Lerner im Sinne eines *shift from teaching to learning* in das Zentrum der Aufmerksamkeit rücken, geeignet um anwendbares und transferierbares Wissen aufzubauen (SCHMIDT & TIPPELT, 2005; WELBERS, 2005). Ein erweitertes Feedbackmodell für lernerzentrierte Veranstaltungen, das explizit den Lernprozess im Prüfungskontext thematisiert, wird im folgenden Kapitel vorgestellt.

### **3 Modell zur Realisierung der Entwicklungsfunktion von Prüfungen in lernerzentrierten Veranstaltungen**

Erkennt man die Entwicklungsfunktion als eine Wesentliche von Prüfungen an und identifiziert Seminarformen als zentrale Orte zur Umsetzung der Kompetenzorientierung im universitären Kontext, so stellt sich die Frage, wie diese Aspekte im Rahmen lernerzentrierter Veranstaltungen realisiert werden können. Curriculumsunabhängig scheint die Rückmeldung der Prüfungsergebnisse eine, wenn nicht sogar die wesentliche Möglichkeit, Entwicklungs- und Lernprozesse von Studierenden zu initiieren oder zu begleiten (CROSS, 1996; HATTIE, 1987; SIPPEL, 2009).

Das im Folgenden vorgestellte Modell (vgl. Abb. 1) geht zurück auf Überlegungen von MÜLLER & SCHMIDT (2009) sowie einen kürzlich im Neuen Handbuch Hochschullehre erschienen Artikel von MÜLLER (2010a). Das dort präsentierte Instrument zur Realisierung der Entwicklungsfunktion von Prüfungen in lernerzentrierten Veranstaltungen mittels eines erweiterten Feedback-Zyklus wird im Folgenden zusammenfassend vorgestellt, und um einen Baustein zur Selbstevaluation ergänzt.

Das Modell unterscheidet in horizontaler Richtung aus prozeduraler Perspektive zwischen vier Phasen, die sich zeitlich um eine Seminarveranstaltung herum gruppieren. Vertikal kann aus funktionaler Sicht zwischen DozentInnen und SeminarteilnehmerInnen differenziert werden. Erstere agieren im Kontext der neuen BA-Studiengänge verstärkt als Prüfer, die Referate oder Stundengestaltungen zu benoten haben. Da Letztere im Rahmen dieses Modells zwar als Feedbackgeber für die Referenten einbezogen werden, die Beobachtungen der Seminarteilnehmer jedoch *nicht* in die Notengebung einfließen, erscheint der Begriff *Peer-Evaluation* angemessener als der Terminus *peer-assessment*. Dieses bezieht sich primär auf

studentische Bewertungen, die zensurrelevant sind (MACLELLAN, 2001). Zusätzlich implementiert ist eine quantitative Selbsteinschätzung (*self-evaluation*) der Referenten. Durch die Gegenüberstellung der Ergebnisse der Selbst- und Fremdbeurteilung kann die Fähigkeit zur Selbstreflexion gefördert und eine realistische Einordnung des eigenen Kompetenzprofils unterstützt werden (BOUD, 1995; FALCHIKOV, 2007; TAN, 2007, 2008).

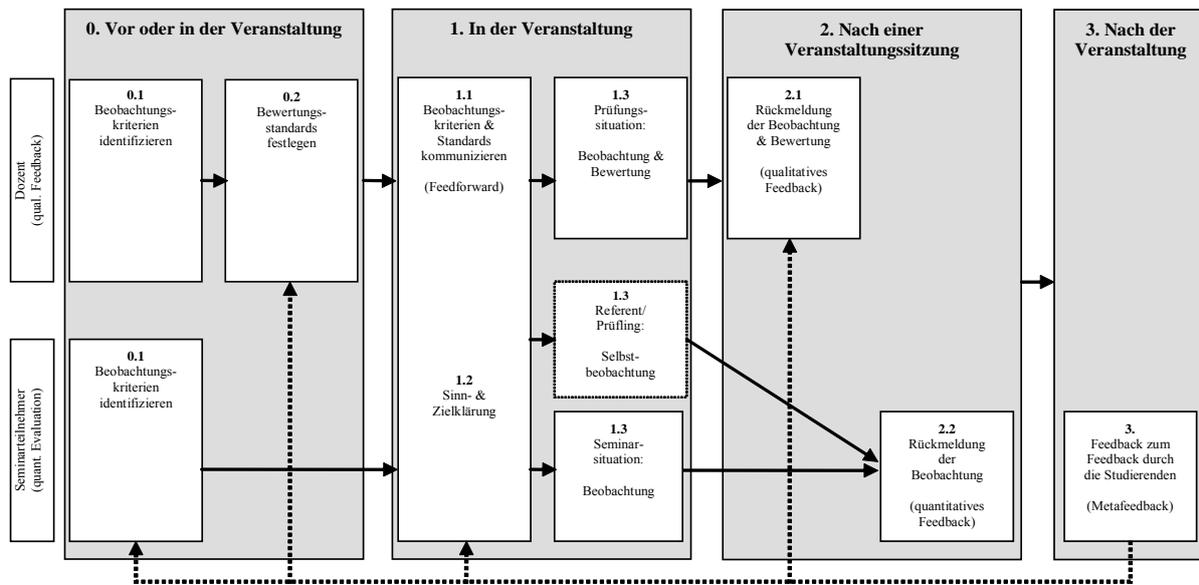


Abb. 1: Erweiterter Prüfungs- und Feedbackzyklus inklusive Selbstevaluation

### 3.1 Vor oder in der Veranstaltung (Schritt 0)

Mit der Umstellung auf Bachelor- und Masterstudiengänge ging die verpflichtende Formulierung von Lern- oder Kompetenzziele für Lehrveranstaltungen und Module einher. Gerade in lernerzentrierten Veranstaltungen, in denen neben inhaltlichen Aspekten auch handlungsnah Kompetenzbereiche vermittelt werden sollen, stellt sich die Frage, wie diese überprüft werden können (SIPPEL, 2009, S. 4 ff.). Hier gilt es zu Beginn eines jeden Lehr-, Prüfungs- und Feedbackprozesses Beobachtungskriterien zu identifizieren, die die Aufmerksamkeit des Dozenten als auch der Seminarteilnehmer systematisieren und „ad-hoc-Verhalten“ (TINNEFELD, 2002, S. 27) reduzieren (Schritt 0.1). Diese Kriterien sollten handlungsnah Operationalisierungen der Kompetenzziele der Veranstaltung darstellen und fließen in einen Beobachtungsbogen ein, der als Grundlage für das qualitative Feedback des Dozenten / der Dozentin dient (vgl. Tabelle 2).

Wenn auch mit vergleichbarer Grundausrichtung bedarf der quantitative Beobachtungsbogen für die Seminarteilnehmer doch in zweifacher Hinsicht einer Modifizierung: Zum einen erscheint insgesamt eine Reduzierung der Beobachtungskriterien notwendig. Die Studierenden sind primär Seminarteilnehmer, für die die Anwesenheit in der Veranstaltung mit einem eigenen Wissens- oder Kompetenzgewinn verbunden sein sollte. Durch die Bearbeitung des Peer-Beobachtungs-

bogens sollte nicht zu viel Aufmerksamkeit gebunden werden, sondern genügend kognitive Ressourcen für die aktive Teilnahme an der Veranstaltung verbleiben.

Bei der Auswahl und Formulierung der Beobachtungskriterien gilt es zudem die individuelle Perspektive der Studierenden als Adressaten der Stundengestaltung hervorzuheben. Wird mit einer Stundengestaltung für die Referenten nicht nur das Ziel verbunden, erarbeitete Inhalte zu präsentieren, sondern auch durch methodische Settings aktivierende teilnehmerzentrierte Lerngelegenheiten zu schaffen, sollten für die Peers dementsprechende Beobachtungskriterien integriert werden (z.B. zu motivierendem Methodeneinsatz, Teilnehmerorientierung, Inhaltsverständnis und Lernerfolg). So stellt die Sicht der Seminarteilnehmer z.B. zum individuellen Verständnis der präsentierten Inhalte eine wesentliche Ergänzung zur Einschätzung des Dozenten hinsichtlich der Verständlichkeit des Referates dar. Aufgegriffen werden hiermit konstruktivistische Überlegungen, die den Lerner selbst im Sinne eines *shift from teaching to learning* in den Fokus der Aufmerksamkeit rücken (MANDL, KRAUSE & RENKL, 2002; WILDT, 2004). Der Bogen zur Selbsteinschätzung der Referenten wird analog zum Peer-Beobachtungsbogen erstellt.

Als Grundlage für den Benotungsprozess des Dozenten gilt es in einem Folgeschritt Bewertungsstandards für die einzelnen Beobachtungskategorien aufzustellen (Schritt 0.2) und somit inhaltliche Zielvorgaben zu formulieren, an denen sich die Studierenden orientieren können (vgl. Tabelle 2).

Tabelle 2: Beispiele für Beobachtungskriterien und Bewertungsstandards

Beobachtungskriterien	Bewertungsstandards
Körpersprache	Unterstützende Gestik
Moderationsfähigkeit	Aktives Eingehen auf Fragen
Inhaltsdarstellung	Wichtige Inhalte identifiziert und korrekt dargestellt
Zielgruppenorientierung	Vorwissen der Teilnehmer berücksichtigt

### 3.2 In der Veranstaltung (Schritt 1)

Mit der Weitergabe der vorab festgelegten Beobachtungskriterien und Bewertungsstandards an die Studierenden (Schritt 1.1) schließt ein wesentlicher Schritt an (HOCHSCHULREKTORENKONFERENZ, 2006; WISSENSCHAFTSRAT, 2008). Aus Sicht der Rückmeldeforschung kann die Präsentation von Standards als Feedforward verstanden werden (BOGART, 1980). Das heißt, die vorgegebenen Standards stellen verhaltenslenkende Ziel- bzw. Orientierungspunkte dar (vgl. Tabelle 1), an denen die Studierenden ihr Agieren im Seminarkontext ausrichten können.

Sicher gestellt werden muss, dass die aufgestellten Ziele von den Studierenden verstanden und inhaltlich entsprechend den Vorgaben des Dozenten interpretiert werden. Wenig erstaunlich liefern Studien Hinweise darauf, dass deutlich unterschiedliche Konzepte von übermittelten Standards bei Dozenten und Studierenden bestehen (HOUNSELL, 1997; NORTON, 1990). Zu beachten bleibt, dass durch die übermittelten Kriterien und Standards für die Studierenden ein Handlungs-

rahmen aufgespannt wird, der ihnen zwar ausreichend eng formulierte Anhaltspunkte liefert, um die Prüfungssituation (hier Referat, Stundengestaltung) zielgerichtet gestalten zu können, ohne sie hierbei zu sehr in ihrer Autonomie und Selbstregulation zu beschneiden (NICOL & MACFARLANE-DICK, 2006). Nur indem innerhalb der definierten Zielvorgaben genügend Freiheitsgrade für die Studierenden verbleiben (z.B. für die Methodenwahl), werden sie als aktiver Konstrukteur von Lernsettings anerkannt und ist langfristig mit einem Kompetenzerwerb im Sinne einer selbstständigen Handlungsfähigkeit zu rechnen (vgl. allgemein BUTLER & WINNE, 1995).

Von besonderer Bedeutung ist das handlungslenkende Vorgeben von Standards zu Beginn der universitären Ausbildung, denn zu diesem Zeitpunkt sind die Studierenden noch nicht mit den Erwartungen und Vorstellungen einer Institution vertraut (BARGEL, MÜBIG-TRAPP & WILLIGE, 2008, S. 313 ff.; POULOS & MAHONY, 2008; YORKE, 2002). Die möglichst frühe Heranführung der Studierenden an Standards sowie Evaluations- und Rückmeldemaßnahmen erscheint auch deswegen relevant, da sich Hinweise darauf finden, dass ein Großteil von ihnen beim Durchlaufen vorgelagerter Bildungsinstitutionen noch nicht vertieft mit Feedbackverfahren in Berührung gekommen ist (BURKE, 2009).

Neben dieser prospektiv handlungssteuernden Funktion von Zielvorgaben oder Standards wird durch Bekanntgabe der Beobachtungs- und Bewertungskriterien auch die Transparenz der Notengebung gefördert – dies dürfte sich positiv auf die Akzeptanz diesbzgl. Prozesse durch die Studierenden auswirken (SIPPEL, 2009, S. 15 ff.).

An die Kommunikation der Beobachtungs- und Bewertungskriterien schließt ein thematisch verallgemeinerter Austausch über Sinn und Ziele von Evaluationsmaßnahmen und Rückmeldungen an (Schritt 1.2). Dies ist bedeutsam, um Bedenken auf Seiten der Studierenden gegenüber Feedbackmaßnahmen abzubauen, aber auch um den Dozenten für auftretende Ängste der Seminarteilnehmer zu sensibilisieren und proaktiv Vorschläge zur Umsetzung der Rückmeldungen einzuholen (VÄRLANDER, 2009; YORKE, 2003).

Es folgt mit einem Referat oder einer Stundengestaltung die Prüfungs- bzw. Seminarsituation selbst (Schritt 1.3), anhand derer das bislang vernachlässigte Spannungsfeld zwischen Beobachtung und Bewertung thematisiert werden soll. Vor allem in der kommunikationswissenschaftlich fundierten und handlungspraktisch orientierten Feedbackforschung wird wiederholt auf die Notwendigkeit verwiesen, bei Rückmeldungen zwischen Beobachtung und Bewertung zu trennen (SCHULZ VON THUN, 2004).

Ersteres kann als wertneutraler Hinweis auf potentielle Entwicklungsfelder verstanden werden, wohingegen letzteres den Zensurcharakter einer Situation betont, wobei subjektive Interpretationen des Feedbackgebers an Gewicht gewinnen. Dies spiegelt sich auch in der Diskussion um die Positionierung von Feedback im Spannungsfeld zwischen Entwicklungs- und Selektions- bzw. Kontrollfunktion wider. In unterschiedlichsten Praxis- und Forschungsfeldern besteht inzwischen weitgehend Einigkeit darüber, dass eine „Realisierung beider Funktionen mit nur einer Methode nicht möglich“ ist (MÜLLER, 2008, S. 6; vgl. SCHNEEWIND, 2007; SCHULER, 2004).

Im Zuge der neuen BA-Studiengänge befinden sich Lehrende jedoch zunehmend in einem Rollenkonflikt: Einerseits müssen sie studentische Leistung per Zensur beurteilen, andererseits sollen und wollen sie fördernd die Entwicklung der Studierenden begleiten (BRENDEL, EGGENSPERGER & GLATHE, 2006; IRONS, 2008, S. 6 ff.). Mit Blick auf die Studierenden erscheint diese ambigüe Situation problematisch, da ein Referat vor allem unter dem Primat der Prüfung und Notenvergabe wahrgenommen wird. Prüfungen weisen somit eine selbstwertbezogene Bedeutung auf, was sich in Mechanismen der Selbst-Verteidigung bzgl. der Rückmeldung äußern kann, wie z.B. externalen Attributionstendenzen oder Ablehnung des Feedbackgebers (HIGGINS, HARTLEY & SKELTON, 2001; ILGEN & DAVIS, 2000).

Auflösen ließe sich diese problematische Kopplung der formativen, entwicklungsbezogenen und summativen, selektionsorientierten Prüfungsfunktion (IRONS, 2008; KNIGHT, 2001) einerseits durch strukturelle Veränderungen, z.B. durch die Einführung nicht-notenrelevanter Lernstandserhebungen mit rein informativem Charakter (NOUNS & BRAUNS, 2008). Auch aktuelle Überlegungen zur Reduzierung der Anzahl der Prüfungen erscheinen vor diesem Hintergrund gewinnbringend. So könnten Veranstaltungen innerhalb eines Moduls (z.B. Begleitübungen zu einer Vorlesung) von der Prüfungspflicht befreit werden – eine Bewertung könnte hier lediglich im Sinne von „bestanden oder nicht bestanden“ in Abhängigkeit von der regelmäßigen Seminarteilnahme erfolgen. Eine eher situative Lösung wäre außerdem die personale Trennung von „Bewerter“ und „Förderer“ im Prüfungskontext. So vielversprechend die erstgenannten strukturellen Vorschläge erscheinen, so idealistisch klingt der Zweite angesichts der angespannten Personalsituation an vielen deutschen Hochschulen.

### **3.3 Nach einer Veranstaltungssitzung (Schritt 2)**

Im Anschluss an eine Veranstaltungssitzung erfolgt zunächst die qualitative Rückmeldung der Beobachtungen und Bewertungen durch den Dozenten (Schritt 2.1). Aus den mittlerweile zahlreichen Reviews zur Gestaltung und Wirkungsweise von Rückmeldungen (HATTIE & TIMPERLEY, 2007; KLUGER & DENISI, 1996; MORY, 2003; MÜLLER, 2008; SIPPEL, 2009) werden an dieser Stelle nur einige Aspekte vorgestellt, die zu einer Reduzierung der oben skizzierten Funktions-, Ziel- und Rollenkonflikte beitragen und die eine möglichst lernwirksame Übermittlung und Verarbeitung des Feedbacks begünstigen.

Wesentlich ist zunächst eine möglichst zeitnahe Rückmeldung der Beobachtungen und Bewertungen. Nur wenn die Referenten das Feedback mit ihrer gezeigten Performance in Verbindung bringen können, werden Verhaltensänderungen wahrscheinlich. Zur Förderung selbstdiagnostischer und selbstregulativer Fähigkeiten empfiehlt es sich, die Studierenden zu Beginn nach ihrem eigenen Eindruck zu Vorbereitung und Ablauf der Sitzung zu befragen. Die mündliche Rückmeldung selbst sollte neben Beobachtungen und Bewertungen auch konkrete Veränderungsvorschläge beinhalten. Förderlich zur Vermeidung einer emotionalen Involvierung der Studierenden sind sachbezogene Rückmeldungen, die sich konkret auf die vorab identifizierten Bewertungskriterien oder beobachtetes Verhalten, nicht aber

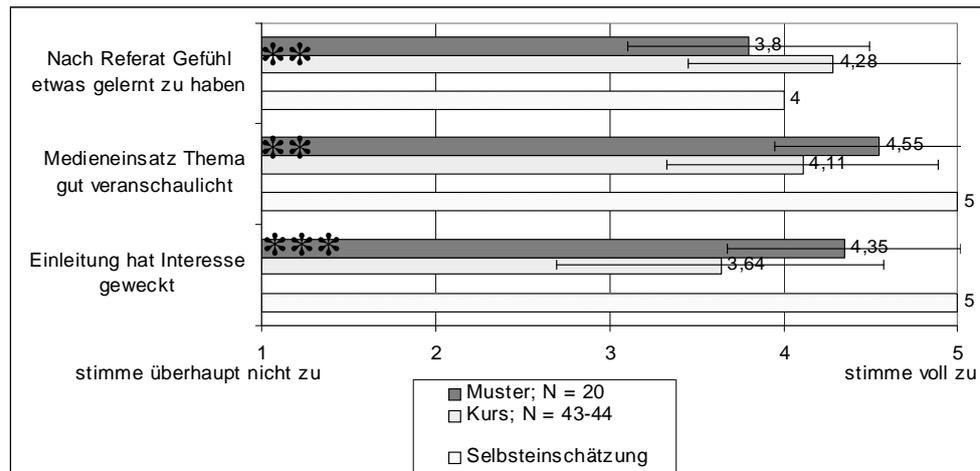
auf die Persönlichkeit des Gegenübers, beziehen (WILLIAM & BLACK, 1998; YORKE, 2003).

Gleichzeitig sollten direkte soziale Vergleiche zwischen den Mitgliedern einer Referatsgruppe vermieden werden (VÄRLANDER, 2009). Um die produktive Rezeption der Rückmeldung zu fördern, wird zudem von verschiedenen Seiten die zeitliche Trennung von Feedback und Note, sprich formativem und summativem Assessment, vorgeschlagen (IRONS, 2008; SIPPEL, 2009). Dies erscheint sinnvoll, da durch die starke Fokussierung auf die Zensur mit der Notenbekanntgabe meist starke Freude oder Enttäuschung verbunden ist und somit wenig kognitive Kapazitäten für die Verarbeitung der qualitativen Rückmeldeinhalte verbleiben.

Vielfach vertreten wird auch die Empfehlung negative Feedbackinhalte in zwei positive Rückmeldungsteile einzubetten (BRINKO, 1993) – wie COE (1998) jedoch betont, geht die Anregung zu diesem Dreischritt über die vorliegenden empirischen Befunde hinaus und bedarf noch einer systematischen Überprüfung. Mit Blick auf die kognitive Verarbeitungskompetenz der Adressaten sollte der Umfang des Feedbacks begrenzt werden – drei unterschiedliche Aspekte scheinen hierbei eine vertretbare Richtgröße darzustellen (NICOL & MACFARLANE-DICK, 2006, S. 209). Zum Abschluss der Rückmeldung sollte den Studierenden Raum für Rückfragen, Anmerkungen sowie einem Feedback an den Dozenten eingeräumt werden.

Die quantitativen Fragebögen der Seminarteilnehmer werden zeitnah ausgewertet und die Ergebnisse separat in schriftlicher Form an die Referenten zurückgemeldet (Schritt 2.2). Da Bezugsnormen wichtig zur Einordnung von Leistungen sind (HECKHAUSEN, 1989), enthält die Rückmeldung neben der Bewertung der aktuellen Stundengestaltung auch einen sozialen Vergleichswert. Dieser gibt in aggregierter Form die Einschätzungen zu den vorangegangenen Seminarsitzungen wieder. Entsprechend den Befunden zur Gestaltung von Rückmeldungen (MÜLLER, 2010b), wird der soziale Vergleichswert jedoch nur dann angegeben, wenn er sich signifikant von der Einschätzung der aktuellen Seminarsitzung unterscheidet (vgl. oberer Abschnitt von Abb. 2). In die schriftliche Rückmeldung integriert ist zudem die Selbsteinschätzung der Referenten (jeweils unterster Balken).

Mit Blick auf die Befunde aus anderen Forschungsbereichen (MÜLLER, 2010b; RHEINBERG, 2001) wäre es ideal, den Studierenden im Rahmen der quantitativen Rückmeldung eine Kombination aus kriterialen, ipsativen und sozialen Bezugsnormen anzubieten. Ersteres dürfte derzeit an der nicht vorliegenden Modellierung von Kompetenzprofilen und -bereichen, wie sie z.B. in Schulvergleichsarbeiten wie PISA eingesetzt werden, scheitern. Ipsative Rückmeldungen zur Entwicklung der Studierenden über die Zeit erscheinen durch die derzeit im Ausbau befindlichen EDV-gestützten Notenverwaltungssysteme durchaus möglich. Sinnvoll werden dementsprechende längsschnittliche Rückmeldungen jedoch nur, wenn die zu Grunde liegenden Prüfungssituationen vergleichbar sind. Auch wenn die Übermittlung sozialer Bezugsnormen durchaus kritisch diskutiert werden kann (vgl. COE, 1998, S. 50), scheinen die möglichen Potentiale (Einordnung der eigenen Leistung; Diskussionsimpulse) gegebene Risiken (z.B. Demotivierung bei negativ ausfallendem Vergleich) im vorliegenden Kontext zu überwiegen.



Signifikanzniveau: \* =  $p < 0,1$ ; \*\* =  $p < 0,05$ ; \*\*\* =  $p < 0,01$

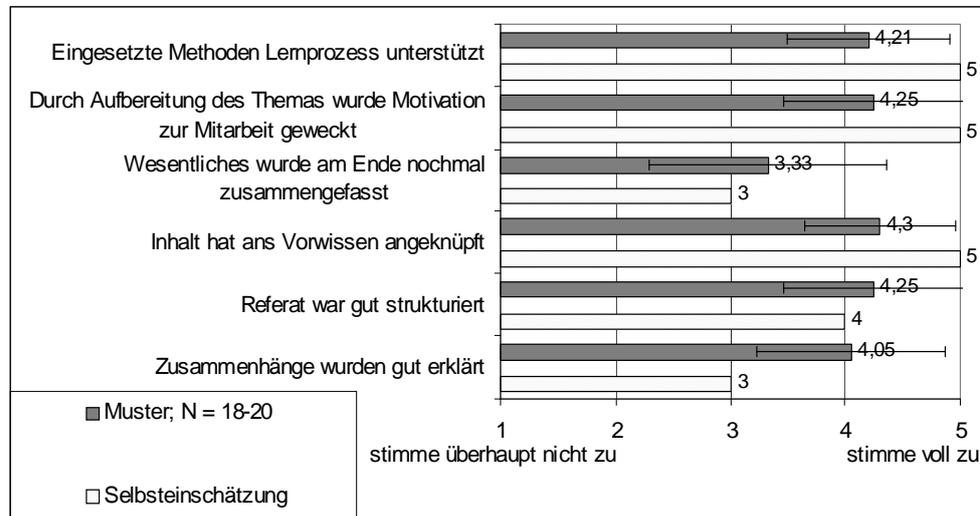


Abb. 2: Beispielerückmeldung zu Aspekten der Referatsgruppe

### 3.4 Nach einer Veranstaltung (Schritt 3)

Abgeschlossen wird das Zyklenmodell durch ein Metafeedback am Ende der Veranstaltung bzw. des Semesters (Schritt 3). Den Studierenden soll hier die Möglichkeit gegeben werden, sich als gleichberechtigte Akteure zu erleben und künftige Rückmeldeprozesse aktiv mit zu gestalten. Mögliche Bezugspunkte für die Rückmeldung zur Rückmeldung ergeben sich entlang des gesamten Prüfungs- und Feedback-Zyklus: Beginnend mit der Auswahl relevanter Beobachtungs- und Bewertungskriterien (Schritte 0 und 0.1) bis zu Aspekten der konkreten mündlichen wie schriftlichen Rückmeldung des Dozenten bzw. der Seminarteilnehmer (Schritt 2.1 und 2.2).

## 4 Praktische Erfahrungen und empirische Befunde zum Einsatz des Feedback-Zyklus

Der in Kapitel 0 vorgestellte erweiterte Prüfungs- und Feedback-Zyklus wird in seiner gesamten Form aktuell im Rahmen mehrerer lernerzentrierter Veranstaltungen am Institut für Pädagogik der Ludwig Maximilians Universität München eingesetzt. Die im Folgenden vorgestellten Auswertungen beziehen sich auf Veranstaltungen aus den Wintersemestern 2009/2010 und 2010/2011.

Zur Implementierung des Modells ergeben sich mit Blick auf die universitäre Praxis zwei Fragestellungen: Zum einen welcher Aufwand mit der Durchführung des mehrstufigen Zyklus verbunden ist. Zum anderen, ob ein Einsatz des Instruments aus empirischer wie adressatenspezifischer Perspektive vertretbar erscheint.

### 4.1 Aufwand bei der Anwendung des Feedback-Zyklus

Bei der Implementierung des Feedback-Zyklus ergeben sich primär zwei Aufwandsarten, die regelmäßig im Zuge einer Veranstaltung auftreten<sup>3</sup>: Einerseits hinsichtlich des Einsatzes im Seminar selbst, andererseits bei der Auswertung und Erstellung der quantitativen Rückmeldungen der Seminarteilnehmer an die Referenten.

Im Rahmen der ersten Seminarsitzungen gilt es *einmalig* die Beobachtungskriterien und Bewertungsstandards zu kommunizieren, den Ablauf der beiden Feedbackformen zu vermitteln und Ziele von Evaluations- und Rückmeldemaßnahmen im Allgemeinen zu thematisieren (Schritte 1.1 und 1.2). Hierfür werden bei eher lehrerzentrierter Vorgehensweise ca. 20 Minuten benötigt. Gegen Ende einer *jeden* Seminarsitzung müssen ca. fünf Minuten eingeplant werden, um den Seminarteilnehmern das Ausfüllen der Evaluationsbögen zur aktuellen Stundengestaltung zu ermöglichen. Für das qualitative Dozentenfeedback direkt im Anschluss an die Sitzung (Schritt 2.1) hat sich bei strukturiertem Vorgehen für zwei bis drei Studierende ein Zeitrahmen von ca. 15 Minuten etabliert.

Die Bearbeitung der quantitativen Rückmeldung der Seminarteilnehmer erfolgt durch eine studentische Tutorin. Für einen Kurs mit ca. 25 Teilnehmern und zwei Referenten können für die Eingabe der Daten ( $N_{Tn} \times \text{ca. } 20 \text{ Items} + N_{\text{Ref\_Selbst}} \times \text{ca. } 15 \text{ Items}$ ), Auswertung der Ergebnisse und Erstellung der schriftlichen Rückmeldungen in etwa 100 Minuten pro Woche angesetzt werden.

### 4.2 Güte der eingesetzten Instrumente

Eine wesentliche Bedingung für belastbare studentische Lehrveranstaltungsevaluationen stellt die Güte der verwendeten Instrumente dar. Die Objektivität des Auswertungsverfahrens ist durch die eindeutige Skalierung der Items und etablierte statistische Verfahren gegeben. Die beiden postulierten Beobachtungsbereiche *Refe-*

---

<sup>3</sup> Der Zeitaufwand für die Erstellung eines Beobachtungs- und Bewertungskatalogs wird hier nicht einbezogen, da dieser schwerpunktmäßig nur einmalig anfällt und über mehrere Veranstaltungen verrechnet werden kann (z.B. ein Bewertungssystem für mehrere parallele Veranstaltungen in einem Modul).

*ratsgruppe* (9 Items) und *Einzelreferent* (3 Items) lassen sich durch eine Faktorenanalyse bei einer insgesamt für dieses Verfahren sehr gut geeigneten Variablenstruktur (Kaiser-Meyer-Olkin-Maß = .87, vgl. BACKHAUS, ERICHSON, PLINKE & WEIBER, 2000) gut reproduzieren. Es ergeben sich zwei Faktoren mit einem Eigenwert größer eins, einem erklärten Anteil an der Gesamtvarianz von 62% und jeweiligen Hauptladungen zwischen .41 und .86. Die gebildeten Skalen *Referatsgruppe* ( $\alpha = .81$ ) und *Einzelreferent* ( $\alpha = .83$ ) weisen eine hohe interne Konsistenz auf.

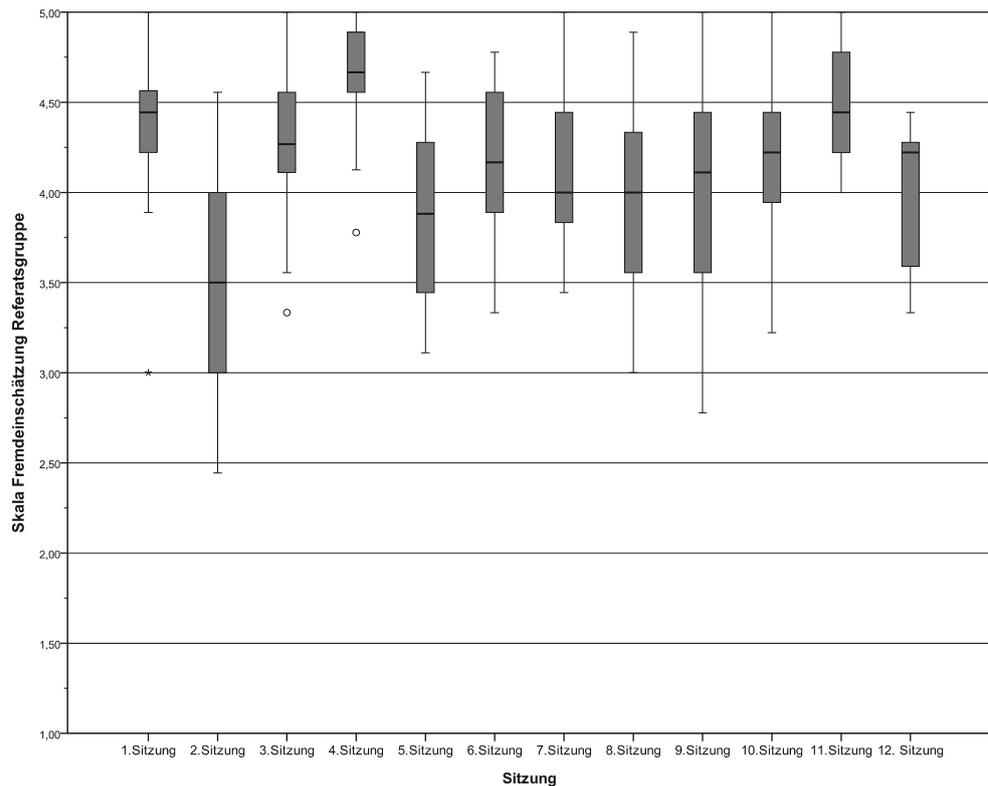


Abb. 3: Peer-Evaluation zu den Referatsgruppen im Semesterverlauf (N = 12-22)

Fragen zur Validität der Bewertung studentischer Seminarleistungen durch Kommilitonen wurden vor allem im angloamerikanischen Raum im Rahmen zensurrelevanter Peer-Assessmentsysteme untersucht (vgl. die Metaanalyse von FALCHIKOV & GOLDFINCH, 2000), mit explizitem Bezug zu mündlichen Präsentationen z.B. von LANGAN et al. (2008)<sup>4</sup>. Hinweise auf die Validität des eingesetzten Instruments ergeben sich durch einen Blick auf das Antwortverhalten der Studierenden über die Zeit hinweg. Differenzierungen im Bewertungsverhalten zwischen einzelnen Sitzungen deuten darauf hin, dass mit dem Beobachtungsbogen die Qualität der einzelnen Veranstaltung und nicht allgemeine Aspekte (z.B. Interesse am Seminar, Dozent etc.) erfasst wurden. Gerade bei kleineren Stichproben eignen sich hierzu Methoden der explorativen Datenanalyse als Ergänzung zu

<sup>4</sup> Die Validität studentischer Lehrveranstaltungsevaluationen werden thematisiert z.B. bei RINDERMANN (1996) oder HOFFMANN (2009).

konventionellen Signifikanztests gut zur Untersuchung von Bewertungsunterschieden (SEDLMEIER, 1996).

Wie eine exemplarische Auswertung für eine Seminarveranstaltung anhand von Boxplots verdeutlicht, fällt die quantitative Peer-Evaluation zwischen den Sitzungen hinsichtlich der Referatsgruppe sehr unterschiedlich aus (vgl. Abb. 3).

Die Mediane als zentrale Werte weichen zwischen den einzelnen Sitzungen teils deutlich voneinander ab, auch die Streuung der Bewertungen ist veranstaltungsspezifisch. Eine einfaktorielle ANOVA mit den 12 Sitzungen als Differenzierungsvariable bestätigt den visuellen Eindruck: Die Bewertung zwischen den Gruppen fällt mit  $F(11,209)=7.10$  ( $p=.000$ ) hochsignifikant unterschiedlich aus – entsprechend dem inkonsistenten Verlauf der Bewertungen über das Semester hinweg lassen sich anhand von post-hoc-Tests (Scheffé, Duncan) keine Tendenzen über die Zeit identifizieren.

Auch die einzelnen Referenten werden zwischen und, in geringerem Ausmaß, auch innerhalb der Seminarsitzungen von den Kommilitonen differenziert eingeschätzt (vgl. Abb. 4). Es finden sich Bewertungsunterschiede sowohl bei den zentralen Werten (Median) als auch hinsichtlich der Einheitlichkeit der Einschätzungen (Streuung). Dieser rein visuelle Eindruck kann mittels einer einfaktoriellen ANOVA statistisch abgesichert werden,  $F(23,417)=4.71$  ( $p=.000$ ). Bewertungstendenzen über den Seminarverlauf lassen sich anhand von post-hoc-Tests (Scheffé, Duncan) auch hier nicht feststellen.

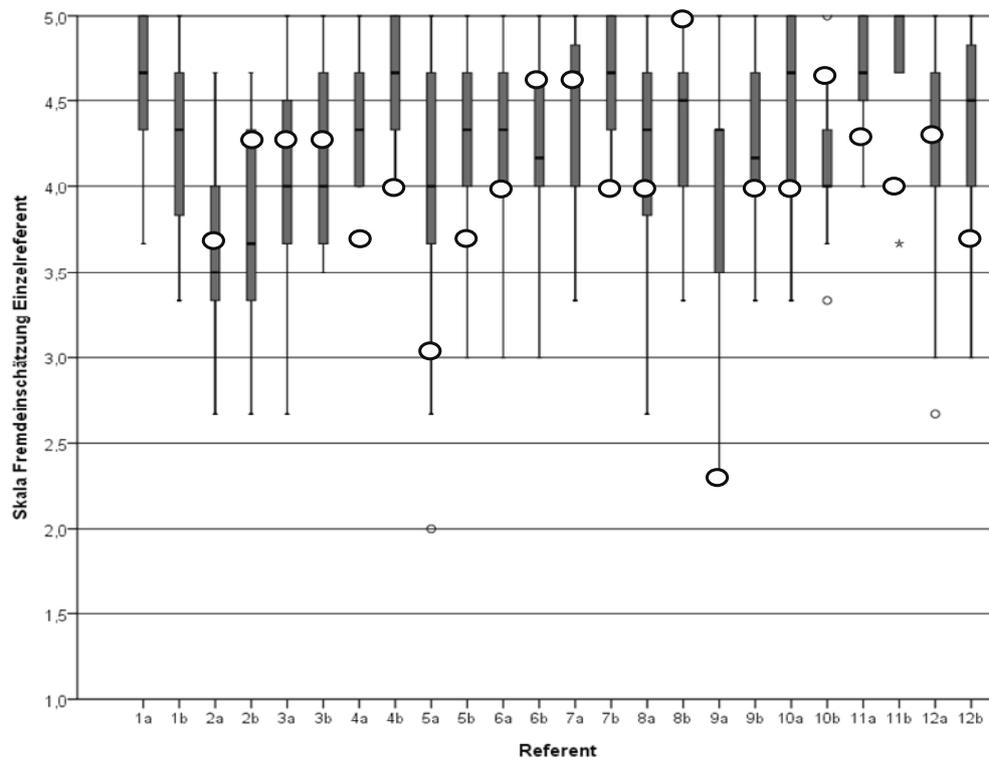


Abb. 4: Peer-Evaluation zu Einzelreferenten (N = 12-22) und Selbstevaluation im Semesterverlauf

Mittels kleiner Kreise zusätzlich visualisiert sind in Abb. 4 die Selbsteinschätzungen der einzelnen Referenten. Es zeigt sich, dass teils deutliche Abweichungen in überschätzender wie unterschätzender Weise zwischen der Selbst- und Fremdbewertung vorliegen (vgl. Referenten 8b und 10b bzw. 5a und 9a).

Weiterführende mehrerebenen- oder varianzanalytische Untersuchungen zur Aufklärung des Einflusses unterschiedlicher Betrachtungsebenen (Referatsgruppe vs. Einzelreferent) oder von Drittvariablen (z.B. Geschlecht, Alter) sind an dieser Stelle nicht möglich. Zum einen weil die quantitativen Peer-Evaluationsbögen nicht individualisiert erfasst wurden, zum anderen weil sich die Datenbasis durch Rollenwechsel (Referenten – Teilnehmer) kontinuierlich verändert.

### 4.3 Einschätzung der Feedbackinstrumente durch die Seminarteilnehmer

Im abschließenden Schritt 3 des vorgestellten Modells werden die SeminarteilnehmerInnen am Ende des Semesters um ihre Einschätzung der verwendeten Rückmeldeinstrumente gebeten. Um das verwendete Feedbackinstrument inhaltlich weiter zu entwickeln, sollten die Studierenden zunächst die Sinnhaftigkeit einzelner Items des quantitativen Beobachtungsbogens einschätzen. Hier ergeben sich über mehrere Kurse<sup>5</sup> hinweg durchweg positive Bewertungen (vgl. ausführlicher MÜLLER, 2010a).

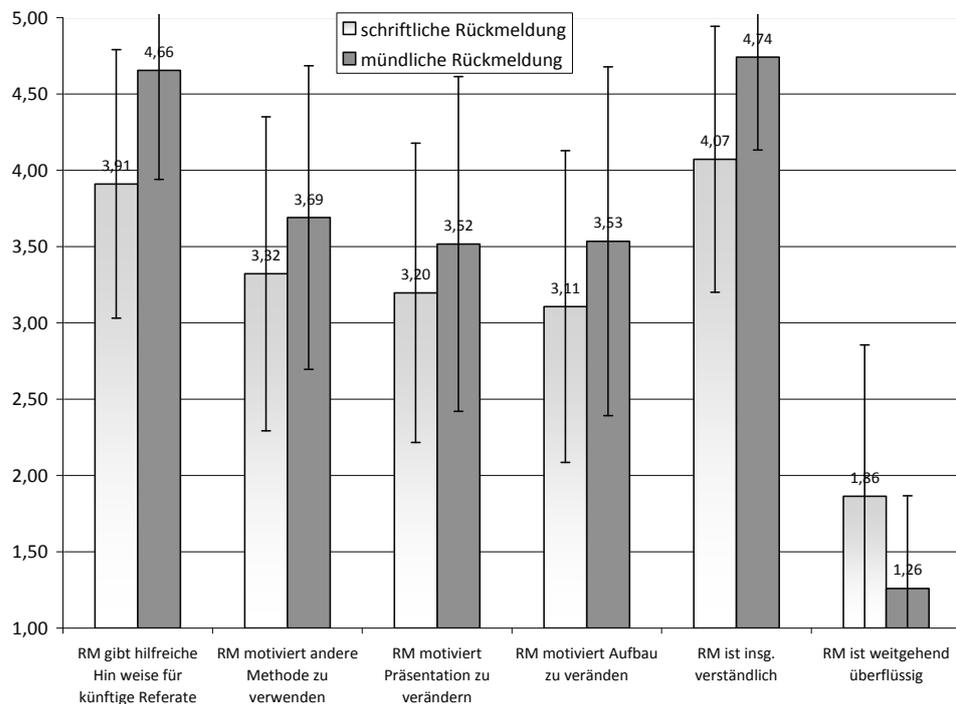


Abb. 5: Bewertung der schriftlichen und mündlichen Rückmeldung (N=56-59)

<sup>5</sup> Die folgenden Auswertungen beziehen sich auf insgesamt drei Kurse aus dem 1., 3., und 5. Semester.

Darüber hinaus wurden die Seminarteilnehmer zu den beiden verwendeten Rückmeldeformen befragt. Sowohl das quantitative Peerfeedback als auch die qualitative Dozentenrückmeldung erfahren insgesamt gute Resonanz (vgl. Abb. 5).

Das mündliche Feedback durch den Dozenten wird in allen sechs Bereichen positiver bewertet – die Unterschiede zur schriftlichen Rückmeldeform fallen durchgehend signifikant bis hochsignifikant aus (t-Test für verbundene Stichproben). Interessant erscheint, dass beide Rückmeldeformate in den Augen der Rezipienten nicht nur hilfreiche Hinweise für künftige Referate beinhalteten, sondern, wenn auch auf niedrigerem Niveau, durchaus auch dazu motivieren, Veränderungen (Methode, Präsentation, Aufbau) ins Auge zu fassen.

Mit Bezug zur quantitativen Peer-Evaluation wurden die Adressaten zusätzlich um ihre Einschätzung des zur Verfügung gestellten sozialen Vergleichs gebeten (vgl. Abb. 6). Entgegen der oft geäußerten Bedenken gegenüber sozialen Vergleichen in Rückmeldungen (vgl. COE, 1998, S. 50), wurde die soziale Bezugsnorm von den Studierenden als sinnvoll bewertet ( $M = 3,95$ ;  $SD = 0,81$ ;  $Min. = 1$ ;  $Max. = 5$ ). Dementsprechend schätzen rund 85% der Studierenden die Bewertung durch ihre Kommilitonen als (eher) unproblematisch ein ( $M = 1,83$ ;  $SD = 1,11$ ;  $Min. = 1$ ;  $Max. = 5$ ). Wenn auch auf etwas niedrigerem Niveau wurde die beim sozialen Vergleich implementierte Signifikanzüberprüfung als nachvollziehbar beurteilt ( $M = 3,69$ ;  $SD = 0,88$ ;  $Min. = 1$ ;  $Max. = 5$ ). Bemerkenswert erscheint, dass dem Einsatz statistischer Elemente (t-test, Konfidenzintervalle), einem Feld demgegenüber sich Studierende der Pädagogik oft sehr reserviert äußern, lediglich von knapp 10% der Befragten mit Zurückhaltung begegnet wird.

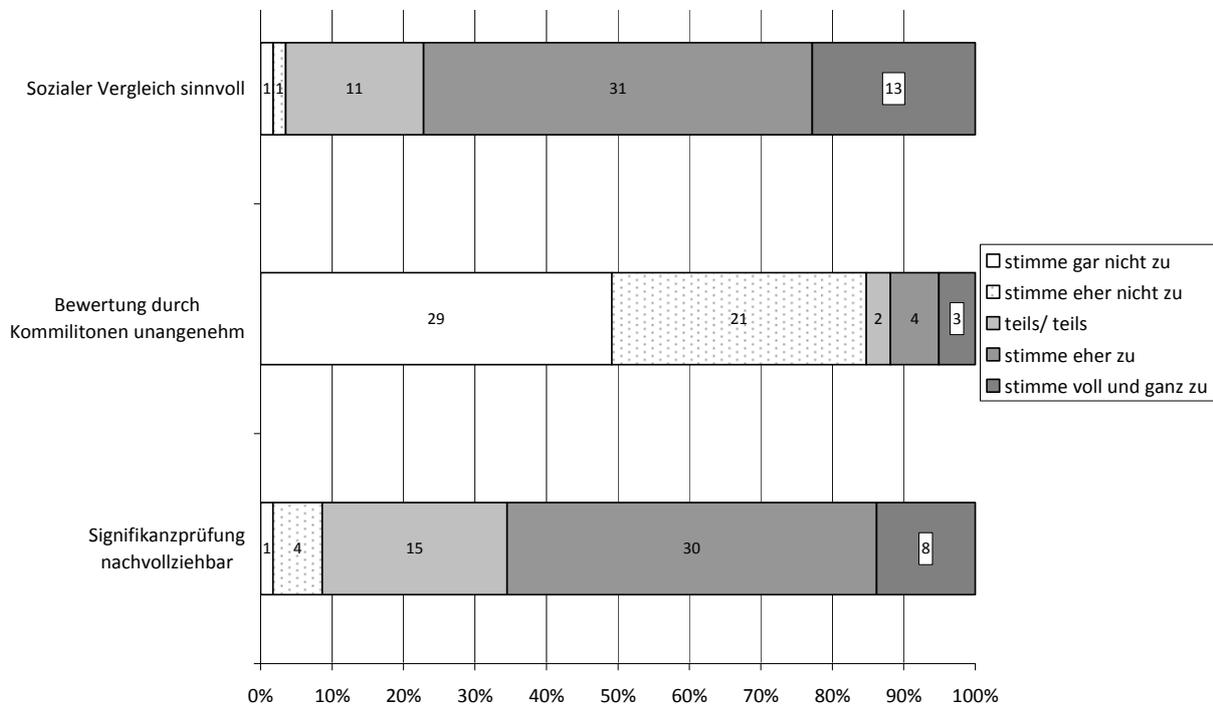


Abb. 6: Einschätzungen des sozialen Vergleichs bei der schriftlichen Rückmeldung (N = 57-59)

Explizit danach gefragt, ob die qualitative Rückmeldung durch den Dozenten ihnen bei der Orientierung an der Universität behilflich war, äußerten sich rund 60% zumindest tendenziell zustimmend, bei nur etwa 10% eher ablehnenden Aussagen ( $M = 3.64$ ;  $SD = 0.95$ ;  $Min. = 1$ ;  $Max. = 5$ ).

## 5 Fazit

Die Modularisierung von Studiengängen im Zuge der Bolognaform geht mit einer Bedeutungssteigerung von Prüfungen im universitären Kontext einher. Im Gegensatz zu Aspekten der Prüfungsvorbereitung und -durchführung, die in den letzten Jahren bereits verstärkt diskutiert wurden, stehen bislang die Lernpotentiale von Prüfungen nur ansatzweise im Fokus der Aufmerksamkeit.

Der vorgestellte erweiterte Feedbackzyklus betrachtet Prüfungen im Sinne des *formative assessment* (IRONS, 2008) primär als Lernchance für Studierende. Unter Bezug auf Facetten der von FLECHSIG (1976) identifizierten didaktischen Funktionen von Prüfungen, ergeben sich für Studierende im Rahmen des Modells von der Prüfungsvorbereitung bis zur -nachbereitung zahlreiche Lerngelegenheiten. Bezieht sich der vorliegende Beitrag mit Blick auf die in den Studienordnungen der Bachelor-Studiengänge verankerte Kompetenzorientierung primär auf lernerzentrierte Veranstaltungen und dort verankerte Prüfungen, lässt sich das vorgestellte Modell in seiner Grundform auch auf andere Veranstaltungsformen übertragen. So sind systematisierte Rückmeldungen von Prüfungsergebnissen auch im Kontext von Multiple-Choice-Prüfungen denkbar.

Die veranstaltungsformunabhängige Bedeutung einer ganzheitlichen Betreuung im Prüfungskontext spiegelt sich auch in Qualitätserwartungen der Studierenden an die Hochschulen wider: Prüfungsvorbereitung und nachfolgendes Feedback werden von den Studierenden mit als wichtigste Beratungs- und Betreuungselemente eingeschätzt (BARGEL, MÜSSIG-TRAPP & WILLIGE, 2008, S. 35 ff.). Mit Blick auf die zunehmende Konkurrenz in der Hochschullandschaft könnte die Implementierung eines auf Betreuung, Förderung und Kompetenzentwicklung ausgerichteten Feedbackinstruments im Prüfungskontext ein gewinnbringender Beitrag zur Profilbildung einer universitären Einrichtung sein. Gerade zu Beginn der universitären Ausbildung kann ein solches System (bei konsequenter Umsetzung!) wertvolle Orientierungshinweise zur Anpassung an institutionelle Standards liefern und somit die Integration an der Hochschule erleichtern.

Die ersten Erfahrungen und empirischen Analysen zur Implementierung des Modells im Rahmen von lernerzentrierten Veranstaltungen rechtfertigen den Einsatz: Aus Ressourcenperspektive ist die Realisierung der einzelnen Prozessschritte nur mit geringem zeitlichen und monetären Aufwand für den einzelnen Dozenten und eine universitäre Einrichtung verbunden. Mit Blick auf forschungstheoretische Aspekte konnte gezeigt werden, dass die Studierenden die Stundengestaltungen ihrer Kommilitonen durchaus differenziert einschätzen – womit Befunde zur Lehrevaluation von Dozenten Bestätigung finden (HOFFMANN, 2009). Dies ist insofern handlungspraktisch bedeutsam, als sich der Aufwand zur Umsetzung eines solchen Instruments nur dann rechtfertigen lässt, wenn die zurückgemeldeten Ergebnisse einen gewissen Differenzierungsgrad aufweisen. Aus Adressatensicht

schließlich, liefern die beiden verwendeten Rückmeldeformen (quantitative Peer-Evaluation; qualitatives Dozenten-Feedback) motivationale wie inhaltliche Impulse für den weiteren Studienverlauf. Auch deuten die teils deutlichen Abweichungen zwischen der Selbst- und Fremdeinschätzung auf das Informationspotential des Instruments für die jeweiligen Referenten hin.

Als wesentlich für die insgesamt positive Einschätzung des verwendeten Instrumentariums wird die Hervorhebung der Lernfunktion im Rahmen des Prüfungs- und Feedbackkontexts angesehen. Entgegen der vor allem im angloamerikanischen Raum vertretenen Ansätze des *peer-assessment* (vgl. MACLELLAN, 2001) werden beim erweiterten Feedbackzyklus die Ergebnisse der *Peer-Evaluation* nicht bei der Notengebung berücksichtigt (vgl. LIU & CARLESS, 2006). Auch wenn an das qualitative Feedback des Dozenten im Kontext der modularisierten Bachelor-Studiengänge häufig eine Zensur gekoppelt ist, so gilt es doch auch hier den formativen Charakter der Prüfung zu betonen. Nur so kann die Wahrscheinlichkeit zur Realisierung der den Prüfungssituationen inhärenten Lernpotentiale maximiert werden.

Als zentrales Kernelement gilt es im Sinne einer Feedbackkultur einen dialogischen Austausch in der konkreten Rückmeldesituation, aber auch bezogen auf den gesamten Prüfungszyklus, anzustreben (VÄRLANDER, 2009; YORKE, 2003). Nur so können trotz aller formalen Vorgaben der modularisierten Studiengänge und gesteigerter Prüfungsanforderungen produktive Lern- und Entwicklungsgelegenheiten an Hochschulen geschaffen werden – für Studierende, Dozenten und Institutionen.

Mit Blick auf Entwicklungsprozesse von Studierenden bleibt zu reflektieren, wie einzelne Prüfungen im Studienverlauf miteinander verknüpft und somit Rückmeldungen zu einer Prüfung möglichst fließend zur Vorbereitung einer nachfolgenden verwendet werden können (MÜLLER, 2010a). Notwendig hierfür erscheint ein über viele Veranstaltungen hinweg synchronisiertes Prüfungskonzept (z.B. identische Beobachtungskriterien und Bewertungsstandards), womit nicht nur die Vergleichbarkeit von Studienleistungen erhöht, sondern auch ein wesentlicher Beitrag zur Professionalisierung der Lehrer geleistet wird (WISSENSCHAFTSRAT, 2008).

## 6 Literaturverzeichnis

- Anderson, L. & Krathwohl, D.** (Hrsg.) (2001). A taxonomy for learning, teaching and assessing. New York: Longman.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R.** (2000). Multivariate Analysemethoden. Berlin: Springer.
- Bargel, T., Müßig-Trapp, P. & Willige, J.** (2008). Studienqualitätsmonitor 2007 - Studienqualität und Studiengebühren. Hannover: HIS: Forum Hochschulen 1/2008.
- Bogart, D.** (1980). Feedback, feedforward, and feedwithin: strategic information in systems. In: Behavioral Science, 25(4), S. 237-249.
- Boud, D.** (1995). Enhancing learning through self-assessment. London: Kogan Page.
- Boud, D.** (2007). Reframing assessment as if learning were important. In: Boud, D. & Falchikov, N. (Hg.), Rethinking Assessment in Higher Education. Learning for the longer term. London, New York: Routledge, S. 15 - 25.
- Brendel, S., Eggensperger, P. & Glathe, A.** (2006). Das Kompetenzprofil von HochschullehrerInnen: Eine Analyse des Bedarfs aus Sicht von Lehrenden und Veranstaltenden. In: Zeitschrift für Hochschulentwicklung, 1(2), S. 55-84.
- Brinko, K.** (1993). The Practice of Giving Feedback to Improve Teaching: What Is Effective? In: The Journal of Higher Education, 64(5), S. 574-593.
- Bülow-Schramm, M.** (2008). Hochschuldidaktische Prüfungskritik revisited under Bologna-Bedingungen. In: Dany, S., Szczyrba, B. & Wildt, J. (Hg.), Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen. Bielefeld: Bertelsmann Verlag, S. 27 - 44.
- Burke, D.** (2009). Strategies for using feedback students bring to higher education. In: Assessment & Evaluation in Higher Education, 34(1), S. 41-50.
- Butler, D. & Winne, P.** (1995). Feedback and self-regulated learning. A theoretical synthesis. In: Review of Educational Research, 65, S. 245-281.
- Coe, R.** (1998). Can Feedback improve teaching? A review of the social science literature with a view to identifying the conditions under which giving feedback to teachers will result in improved performance. In: Research Papers in Education, 13(1), S. 43-66.
- Cross, K.** (1996). Improving teaching and learning through classroom assessment and classroom research. In: Gibbs, G. (Hg.), Improving student learning: using research to improve student learning. Oxford: Oxford Centre for Staff Development, S. 3-10.
- Dany, S., Szczyrba, B. & Wildt, J.** (Hrsg.). (2008). Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen. Düsseldorf: Ulrich Welbers.
- Dubs, R.** (2006). Besser schriftlich prüfen. Prüfungen valide und zuverlässig durchführen. In: Berendt, B., Voss, H.-P. & Wildt, J. (Hg.), Neues Handbuch Hochschullehre (H 5.1). Berlin: Raabe, S. 1-26.
- Falchikov, N.** (2007). The place of peers in learning and assessment. In: David Boud & Falchikov, N. (Hg.), Rethinking Assessment in Higher Education. Learning for the longer term. London, New York: Routledge, S. 128 - 143.

- Falchikov, N. & Goldfinch, J.** (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. In: Review of Educational Research, 70(3), S. 287 - 323.
- Flechsich, K.-H.** (1976). Prüfungen und Evaluation. In: Blickpunkt Hochschuldidaktik, 40, S. 303-336.
- Hattie, J.** (1987). Identifying the salient facets of a model of student learning: a synthesis of meta-analysis. In: International Journal of Educational Research, 11, S. 187-212.
- Hattie, J. & Timperley, H.** (2007). The power of feedback. In: Review of educational research, 77(1), S. 81-112.
- Heckhausen, H.** (1989). Motivation und Handeln. Berlin: Springer Verlag.
- Hees, F., Hermanns, A. & Huson, A.** (2008). Prüfungserstellung mit Total Quality Management (TQM). In: Dany, S., Szczyrba, B. & Wildt, J. (Hg.), Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen. Bielefeld: Bertelsmann Verlag, S. 129 - 141.
- Higgins, R., Hartley, P. & Skelton, A.** (2001). Getting the message across. the problem of communicating assessment feedback. In: Teaching in higher education, 6(2), S. 269-274.
- Hochschulrektorenkonferenz** (2006). Standards und Leitlinien für die Qualitätssicherung im Europäischen Hochschulraum. Abgerufen von HRK [http://www.hrk.de/de/download/dateien/Beitr9-2006-Standards\\_Leitlinien\\_QS.pdf](http://www.hrk.de/de/download/dateien/Beitr9-2006-Standards_Leitlinien_QS.pdf) am 09.11.2010.
- Hoffmann, A.** (2009). Studentische Lehrveranstaltungsevaluation. Ein Zerrbild der Qualität akademischer Lehre? . In: Empirische Pädagogik, 23(1), S. 19-49.
- Hounsell, D.** (1997). Contrasting conceptions of essay-writing. In: Marton, F., Hounsell, D. & Entwistle, N. (Hg.), The experience of learning. Edinburgh: Scottish Academic Press, S. 106-125.
- Ilggen, D. & Davis, C.** (2000). Bearing bad news: reactions to negative performance feedback. In: Journal of Applied Psychology, 48(3), S. 550-565.
- Iller, C. & Wick, A.** (2009). Prüfungen als Evaluation der Kompetenzentwicklung im Studium. In: Das Hochschulwesen, 6, S. 195-201.
- Irons, A.** (2008). Enhancing Learning through formative assessment and feedback. London: Routledge.
- Kluger, A. & DeNisi, A.** (1996). The effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. In: Psychological Bulletin, 119(2), S. 254-284.
- Knight, P.** (2001). Formative and summative, criterion & norm-referenced assessment. Abgerufen von Queen's University Belfast <http://www.qub.ac.uk/directorates/AcademicStudentAffairs/CentreforEducationalDevelopment/Resources/Assessment/ResourcesRelevanttoUniversityAssessmentPolicy/MethodsofAssessmentFormativeandSummativeAssessment/> am 29.04.2009.
- Kurz, G.** (2010). Qualitätssicherung in der Lehre durch angemessene Bewertung von Studienleistungen. In: Berendt, B., Voss, H.-P., Wildt, J. & Tremp, P. (Hg.), Neues Handbuch Hochschullehre. Berlin: Raabe, S. 1-31.

- Langan, M., Shukerb, D., Cullena, R., Penneyc, D., Richard, P. & Wheeler, P.** (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. In: *Assessment and Evaluation in Higher Education*, 33(2), S. 179-190.
- Liu, N.-F. & Carless, D.** (2006). Peer feedback: the learning element of peer assessment. In: *Teaching in Higher Education*, 11(3).
- MacLellan, E.** (2001). Assessment for Learning: the differing perceptions of tutors and students. In: *Assessment & Evaluation in Higher Education*, 26(4), S. 307 - 318.
- Mandl, H., Krause, U. & Renkl, A.** (2002). Lernkompetenz für die Wissensgesellschaft. In: Dobart, A. (Hg.), *Lernen in der Wissensgesellschaft. Beiträge des OECD/CERI-Regionalseminars für deutschsprachige Länder in Esslingen vom 8.-12. Oktober 2001*. Innsbruck: StudienVerlag, S. 239-266.
- Mory, E.** (2003). Feedback Research Revisited. In: Jonassen, D. (Hg.), *Handbook of research for educational communications and technology*. Mahwah: Lawrence Erlbaum, S. 745-783.
- Müller-Benedict, V.** (2005). Sind Universitätsprüfungen objektiv? Der langfristige historische Zusammenhang zwischen Erfolg in akademischen Prüfungen und Überfüllung der akademischen Berufe. In: *Soziologie*, 34(2), S. 191 - 208.
- Müller, A.** (2008). *Feedback: Wirkmechanismen und Einsatz in der Personalentwicklung*. Saarbrücken: VDM Verlag.
- Müller, A.** (2010a). Prüfungen als Lernchance-Rückmeldungen in Hochschulseminaren. In: *Neues Handbuch Hochschullehre*. H 3.3. Berlin: Raabe Verlag, S. 1-26.
- Müller, A.** (2010b). Rückmeldungen nach Vergleichsarbeiten im Kontext des schulischen Qualitätsmanagements. Drei explorative Studien zu Gestaltung und Rezeption im Anschluss an KOALA-S. Berlin: Mensch und Buch Verlag.
- Müller, A. & Schmidt, B.** (2009). Prüfungen als Lernchance: Sinn, Ziele und Formen von Hochschulprüfungen. In: *Zeitschrift für Hochschulentwicklung*, 4(1), S. 23-45.
- Narciss, S.** (2006). *Informatives tutorielles Feedback*. Münster: Waxmann.
- Newstead, S. E. & Hoskins, S.** (2001). Encouraging Student Motivation. In: Fry, H., Ketteridge, S. & Marshall, S. (Hg.), *A Handbook for Teaching & Learning in Higher Education*. London: Kogan Page Limited, S. 70-82.
- Nicol, D. & Macfarlane-Dick, D.** (2006). Formative assessment and selfregulated learning: a model and seven principles of good feedback practice. In: *Studies in Higher Education*, 31(2), S. 199-218.
- Norton, L.** (1990). Essay writing: what really counts? In: *Higher Education*, 20(4), S. 411-442.
- Nouns, Z. & Brauns, K.** (2008). Das Prinzip des Progress-Testing – Detaillierte Leistungsdarstellung und Lehrevaluation auf Basis der kontinuierlichen Wissensentwicklung der Studierenden. In: Dany, S., Szczyrba, B. & Wildt, J. (Hg.), *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen*. Bielefeld: Bertelsmann Verlag, S. 114-128.
- Poulos, A. & Mahony, M.** (2008). Effectiveness of feedback: the students' perspective. In: *Assessment & Evaluation in Higher Education*, 33(2), S. 143-154.

- Reepmeyer, J.-A.** (2008). Was kosten Prüfungen? Vereinfachung durch den Einsatz computergestützter Prüfungssysteme. In: Brocke, J. V. & Becker, J. (Hg.), Einfachheit in Wirtschaftsinformatik und Controlling. München: Verlag Franz Vahlen S. 521-535.
- Rheinberg, F.** (2001). Bezugsnormen und schulische Leistungsbeurteilung. In: Weinert, F. (Hg.), Leistungsmessungen in Schulen. Weinheim: Beltz, S. 59-70.
- Rindermann, H.** (1996). Zur Qualität studentischer Lehrveranstaltungsevaluierungen: Eine Antwort auf Kritik an der Lehrevaluation. In: Zeitschrift für Pädagogische Psychologie, 10, S. 129-145.
- Schmidt, B. & Tippelt, R.** (2005). Besser lehren – Neues von der Hochschuldidaktik. In: Teichler, U. & Tippelt, R. (Hg.), Hochschullandschaft im Wandel. Beiheft der Zeitschrift für Pädagogik, S. 103-114.
- Schneewind, J.** (2007). Wie Lehrkräfte mit Ergebnismeldungen aus Schulleistungsstudien umgehen. Abgerufen von FU Berlin <http://www.diss.fu-berlin.de/2007/252/#contents> am 03.01.2011.
- Schuler, H.** (2004). Leistungsbeurteilung - Gegenstand, Funktionen und Formen. In: Schuler, H. (Hg.), Beurteilung und Förderung beruflicher Leistung. Göttingen: Hogrefe Verlag, S. 1-23.
- Schulz von Thun, F.** (2004). Miteinander reden. 1. Störungen und Klärungen. Allgemeine Psychologie der Kommunikation. Hamburg: Rowohlt Verlag.
- Sedlmeier, P.** (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. In: Methods of Psychological Research Online, 1(4), S. 41-63.
- Sippel, S.** (2009). Zur Relevanz von Assessment-Feedback in der Hochschullehre. In: Zeitschrift für Hochschulentwicklung, 4(1), S. 1-22.
- Stary, J.** (2006). "Doch nicht durch Worte allein..." Die mündliche Prüfung. In: Berendt, B., Voss, H.-P. & Wildt, J. (Hg.), Neues Handbuch Hochschulforschung (H 2.1). Berlin: Raabe, S. 1-17.
- Tan, K.** (2007). Conceptions of self-assessment. What is needed for long-term learning? In: David Boud & Falchikov, N. (Hg.), Rethinking Assessment in Higher Education. Learning for the longer term. London, New York: Routledge, S. 114 - 127.
- Tan, K.** (2008). Qualitatively different ways of experiencing student self-assessment. In: Higher Education Research & Development, 27(1), S. 15-29.
- Tinnefeld, T.** (2002). Prüfungsdidaktik. Zur Fundierung einer neuen wissenschaftlichen Disziplin - am Beispiel der modernen Fremdsprachen. Aachen: Shaker Verlag.
- Värlander, S.** (2009). The role of students' emotions in formal feedback situations. In: Teaching in Higher Education, 13(2), S. 145-156.
- Wannemacher, K.** (2009). Studienbegleitende Modulprüfungen – Möglichkeiten und Grenzen der Assessmentpraxis in Bachelor- und Masterstudiengängen. In: Zeitschrift für Hochschulentwicklung, 4(1), S. 72-89.
- Welbers, U.** (Hrsg.). (2005). The shift from teaching to learning. Konstruktionsbedingungen eines Ideals. Bielefeld: Bertelsmann.
- Wildt, J.** (2004). 'The Shift from Teaching to Learning' – Thesen zum Wandel der Lernkultur in modularisierten Studiengängen. In: Ehlert, H. & Welbers, U. (Hg.),

Qualitätssicherung und Studienreform. Strategie- und Programmentwicklung für Fachbereiche und Hochschulen im Rahmen von Zielvereinbarungen am Beispiel der Heinrich-Heine-Universität Düsseldorf. Düsseldorf: Gruppello-Verlag, S. 168-178.

**Wiliam, D. & Black, P.** (1998). Assessment and classroom learning. In: Assessment in Education: Principles, Policy & Practice, 5(1), S. 7-74.

**Wissenschaftsrat** (2008). Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Abgerufen von <http://www.wissenschaftsrat.de/download/archiv/8639-08.pdf> am 04.01.2011.

**Yorke, M.** (2002). Academic failure: a retrospective view from non-completing students. In: Peelo, M. & Wareham, T. (Hg.), Failing students in higher education. Maidenhead: SRHE/ Open University Press, S. 29-44.

**Yorke, M.** (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. In: Higher Education, 45, S. 477-501.

## Autor



Dr. Andreas MÜLLER || Wissenschaftlicher Mitarbeiter || Lehrstuhl für Allgemeine Pädagogik und Sozialisationsforschung an der Ludwig-Maximilians-Universität München || Forschungsschwerpunkte: Einsatzmöglichkeiten und Wirkungsweise von Rückmeldesystemen und Feedback; Schulqualitätsforschung mit Fokus auf Schulvergleichsarbeiten; Hochschuldidaktik

<http://www.psy.lmu.de/ape/Mitarbeiter/Andreas-Mueller.html>

[asmueller@edu.lmu.de](mailto:asmueller@edu.lmu.de)