

## **Entwicklung einer Kurzsкала zur Lehrevaluation**

*Jörg Zumbach, Birgit Spinath, Joachim Schahn, Monika Friedrich und Melanie Kögel*

In Michael Krämer, Siegfried Preiser & Kerstin Brusdeylins, (Hrsg.), *Psychodidaktik und Evaluation Bd. VI* (S. 317-325). Göttingen: V & R unipress.

Mittlerweile gibt es eine Vielzahl an Verfahren, die der Evaluation und somit der Qualitätssicherung von Lehrveranstaltungen dienen. Dabei scheint es eine breite Vielfalt an gestalterischen Ideen zur Entwicklung dieser Instrumente zu geben: Von offenen Befragungen hin zu standardisierten Fragebögen, die teils psychometrisch konstruiert, teils rational generiert oder ohne weitere nachvollziehbare Rationale geschaffen wurden (einen Überblick über verschiedene Instrumente zur Lehrveranstaltungsevaluation und deren psychometrische Qualitäten – so weit vorhanden – geben z.B. Braun & Gusy, 2006).

Entsprechend der Vielfalt an Konstruktionsmöglichkeiten liegen den Verfahren auch unterschiedliche Dimensionen zugrunde, die das Konstrukt der „Qualität von Lehre“ erfassen sollen. Während beispielsweise das Trierer Inventar zur Lehrevaluation die vier Dimensionen „Struktur und Didaktik“, „Anregung und Motivation“, „Interaktion und Kommunikation“ sowie „Anwendungsbezug“ aufweist (TRIL; vgl. Gollwitzer, Kranz & Vogel, 2006; Gollwitzer & Schlotz, 2003) und das Heidelberger Inventar zur Lehrevaluation sieben Dimensionen zugrunde legt (HILVE; vgl. Rindermann & Amelang, 1994a; 1994b), sind es bei Schulz, Greve, Koch, Koops und Wilmers (2006) bereits die folgenden zehn Kategorien: Organisation, Lehrkompetenz, Dozentenverhalten, Klima, Inhalt, Eigenes Studierverhalten, Gesamteinschätzung, Zufriedenheit und Nebenqualifikationen.

Ein noch breiteres Spektrum eröffnet sich bei der zugrunde liegenden Zahl an Items, aus denen sich die Verfahren zusammensetzen. Dabei reicht das Spektrum bei psychometrisch konstruierten Verfahren von zwanzig (Fragebogen zur Evaluation von Vorlesungen, FEVOR; Staufenbiel, 2000) über 42 beim HILVE (Rindermann & Amelang, 1994a) bis hin zu 61 (beim Fragebogen zur Beurteilung von Lehrveranstaltungen von Winteler und Schmock, 1979; zit. n. Braun & Gusy, 2006).

Insbesondere die Konstruktion reliabler Subskalen macht es erforderlich, umfangreichere Itemsammlungen aufzunehmen. Dies ist jedoch problematisch, weil ein großer Umfang sowohl zu langer Dauer als auch zu aufwändiger Auswertung führt. Hier ist mitunter auch mit systematischen Verzerrungen zu rechnen, da die Rücklaufquote mit zunehmender Länge der

Fragebögen abnehmen oder zu einschlägig bekannten Problemen wie Frustration, Tendenz zur Mitte oder Extremwerten führen kann. Auch die Auswertung innerhalb des zeitlichen Rahmens von Lehrveranstaltungen kann die zuletzt genannten Probleme nicht umgehen und nimmt zum Teil benötigte Ressourcen der Unterrichts- bzw. Lernzeit in Anspruch.

Eine weitere Unklarheit bei den bisherigen Verfahren liegt auch in der Frage, was eigentlich das Ziel des Verfahrens ist und in welcher Form die Anwendung stattfindet.

### **Ziele und Durchführung von Lehrveranstaltungsevaluationen**

Die Ziele von Lehrveranstaltungsevaluationen und den dabei zugrunde liegenden Verfahren sind nicht immer ganz eindeutig. So scheinen einige Instrumente zumindest sekundär zur Erfassung des Konstruktes „Lehrqualität“ entwickelt worden zu sein (z.B. HILVE oder TRIL). Problematisch ist dabei, dass die Skalenkonstruktion bei diesen Verfahren kaum theoretisch fundiert ist und entsprechend zugrunde liegende Modelle fehlen. Gollwitzer et al. (2006, S. 91) kommen hier zu einem vorläufigen Fazit: „Ein Grund für die Heterogenität (...) liegt sicherlich auch darin begründet, dass es keine klare und theoretisch fundierte Antwort auf die Frage gibt, aus welchen notwendigen und hinreichenden Kriterien sich das Konstrukt Lehrqualität zusammensetzt“.

Abgesehen von diesem wissenschaftstheoretischen Aspekt für die Entwicklung von Skalen zur Lehrevaluation gibt es noch weitere Aspekte, die das übliche Vorgehen bei der Fragebogenevaluation problematisch erscheinen lassen. So ist ein primäres Ziel von Lehrevaluationen sowohl die Erfassung eines Ist-Standes als auch die Aufdeckung von Möglichkeiten zur Optimierung dieser Situation, also zur Erreichung eines Soll-Zustandes. Dies kann im Sinne des Instruktionsdesigns als formative Evaluation aufgefasst werden (vgl. Schott, 1991). Eine weitere Anwendungsmöglichkeit ist die systematische Variation von instruktionellen Komponenten oder des kompletten instruktionellen Designs ein und derselben Lehrveranstaltung und den sich daraus ergebenden Konsequenzen auf das Erleben der Lernumgebung. Diese summative Evaluation findet aber wohl eher selten statt. Stattdessen kommen hier im Sinne einer summativen Evaluation (vgl. auch hier Schott, 1991) eher Vergleiche zwischen unterschiedlichen Lehrveranstaltungen vor, die sich auf unterschiedliche Inhalte, Dozierende und/oder beidem ergeben. Dabei kommt allerdings das Problem zum Tragen, dass die Vergleichbarkeit dessen, was evaluiert wird, fraglich ist. Zum einen liegt dies an dem bereits aufgeführten Dilemma, dass es bislang noch kein theoretisch wie empirisch fundiertes Modell dessen gibt, was eigentlich Lehrqualität ausmacht (vgl. hierzu auch Rindermann, 2003). Zum anderen liegt es daran, dass hier

unterschiedliche Merkmale wie Eigenschaften des oder der Dozierenden, Anzahl der Studierenden, Raum der Veranstaltung, fachliche Einordnung eines Kurses, Betreuungsschlüssel, Pflichtveranstaltung oder freiwilliger Besuch, Renommee der Veranstaltung und unzählige weitere Störgrößen erheblichen Einfluss auf die Beurteilung einer Lehrveranstaltung nehmen können. Zudem sind aus messtheoretischer Sicht Kohorteneffekte nicht auszuschließen.

Erscheint der summative Evaluationsnutzen also eher zweifelhaft, so stellt sich die Frage, warum die meisten Verfahren zur Lehrveranstaltungsevaluation denn tatsächlich „summativ“ zumeist in einer der letzten oder tatsächlich (nach) der letzten Lehrveranstaltung zum Einsatz kommen (vgl. Braun & Gusy, 2006; einige wenige Verfahren wie der Medical School Learning Environment Survey werden zu Beginn und am Ende eingesetzt; vgl. Marshall, 1978; Zumbach, 2003). Eine formative Evaluation mit Hilfe standardisierter Verfahren bietet zwar ansatzweise Hinweise zur Überarbeitung eines Lehrkonzeptes, allerdings sind die Rahmenbedingungen hierfür eher ungünstig: Standardisierte Verfahren können die Wahrnehmung der Lernenden nicht adäquat abbilden und der Zeitpunkt der Evaluation ist ungünstig gewählt. Werden auch offene Fragen zur Optimierung der Lehre erfasst, besteht zumeist nach Erhebung in der letzten Sitzung innerhalb der Vorlesungszeit keine Möglichkeit zur Rücksprache und Klärung. Dies betrifft auch die standardisierten Fragen, da diese und ihre summarischen Ergebnisse nochmals zum Gegenstand eines Dialoges zwischen Lehrenden und Lernenden werden sollte. Auch systematische Verzerrungen summativer Evaluationsziele sind in Kauf zu nehmen, da Kursabbrecherinnen nicht mehr erreicht werden und auch subjektive Verzerrungen durch Attributionen möglich sind (z.B. „Ich habe den Kurs geschafft, so schlecht kann das Lehrangebot ja nicht gewesen sein“). Sehr oft ist der oder die Dozierende allerdings an organisatorische Rahmenbedingungen gebunden, die diesem Mehrwert der Lehrevaluation entgegen stehen können: Hierzu gehören mitunter das Fehlen offener Fragen, die zentrale Vergabe und Auswertung von Fragebögen oder die fehlenden Möglichkeiten und Mittel zur Durchführung einer systematischen Evaluation, der zu große Umfang und damit verbundene Zeitaufwand bei der Durchführung und Auswertung etablierter Verfahren usw.

Im Folgenden gehen wir auf die Anforderungen ein, die bei der Entwicklung der hier beschriebenen Kurzskala ausschlaggebend waren. Auch hier nehmen organisatorische Rahmenbedingungen einen zentralen Stellenwert ein, wurden aber durch die Erfahrungen beim Einsatz in der Lehrevaluation deutlich reglementiert.

## **Rationale für die Konstruktion einer Kurzsкала zur Lehrevaluation**

In diesem Beitrag nimmt die Entwicklung und Konstruktion einer Kurzsкала zur Lehrevaluation eine zentrale Rolle ein. Hintergrund der Entwicklung einer eigenen Skala trotz der breiten Verfügbarkeit alternativer Instrumente war in erster Linie die mangelhafte Praktikabilität und Ökonomie bisheriger Verfahren. Da die Durchführungsdauer und die Durchführungsmodalität etablierter Verfahren zur Lehrevaluation zum einen zu lang und zum anderen mit zu ausgeprägtem summativen Charakter ist, war hier ein praktikableres Instrument mit alternativer Durchführungsmodalität beabsichtigt, die summative wie formative Evaluation vereint. Die summative Evaluation mit kurzer Durchführungsdauer war deswegen beabsichtigt, weil eine leistungsabhängige Mittelvergabe im universitären Kontext auch die Qualität der Lehre integrieren kann und sollte. Damit dies dann in der breiten Anwendung erfolgen kann, muss die Bereitschaft zur Anwendung eines Instrumentes zur Lehrevaluation sowohl auf Seiten der Lernenden als auch auf Seiten der Lehrenden gegeben sein. Ein wesentliches Mittel, diese Bereitschaft zu erhöhen, ist die Beschränkung der Bearbeitungsdauer. Ein weiteres Mittel ist zudem die Erhöhung der Selbstwirksamkeit: Durch die Möglichkeit, negative wie positive Aspekte der Lehre und Verbesserungsmöglichkeiten anhand offener Fragen zu äußern und dadurch aktiv Einfluss auf das Lehrgeschehen zu nehmen, resultiert ein weiterer Anreizfaktor zur Bearbeitung von Fragebögen. Dozierende erhalten dadurch auch ein breiteres Feedback, welches im Sinne einer formativen Rückmeldung einen deutlichen Mehrwert gegenüber summarischen Punktwerten auf verschiedenen Dimensionen beinhaltet. Um die Möglichkeiten dieses interaktiven Prozesses zwischen Lehrenden und Lernenden zu optimieren, war es auch ein Hauptanliegen, den „traditionellen“ Erhebungszeitraum innerhalb der letzten Woche(n) der Vorlesungszeit zu verwerfen und mitten in der Veranstaltungsphase das Instrument einzusetzen. Dadurch wird zum einen ein repräsentativerer Eindruck von der Lehrveranstaltung erfasst, da weniger Verzerrungen aus der Retrospektive zu erwarten sind. Zum anderen können die Ergebnisse der standardisierten Fragen wie offenen Fragen vor dem Ende der Veranstaltung ermittelt werden und nochmals dialogisch zwischen Lehrenden und Lernenden eingesetzt werden.

Auf Basis dieser Überlegungen wurde die im Folgenden beschriebene Entwicklung einer Kurzsкала zu Lehrevaluation vorgenommen.

## **Methode**

Da wie bei dem Gros der verfügbaren Verfahren zur Lehrevaluation kein solides theoretisches Modell zur Lehrevaluation zugrunde gelegt werden konnte, wurde auch hier eine rationale Skalenkonstruktion durchgeführt. Diese sollte dann empirisch bestätigt und optimiert werden. Basis dieser psychometrischen Fragebogenkonstruktion war eine Sammlung verschiedenster Items zur Beurteilung der Facetten guter Lehre. Hierzu wurde eine Sichtung der Literatur zur Qualität universitärer Lehre vorgenommen, verschiedene Verfahren und deren zugrunde liegende Dimensionen wie Items zur Lehrveranstaltungsevaluation gesichtet und zusätzliche, fehlende Aspekte der Bewertung von Lehre in Form von Items formuliert. In einem zweiten Zyklus wurden so weit wie möglich redundante Formulierungen eliminiert. Als Ergebnis resultierte ein Fragebogen mit insgesamt 38 zentralen Items mit einer fünfstufigen Likertskala (von „Trifft voll zu“ bis „Trifft gar nicht zu“) zur Bewertung der Qualität guter Hochschullehre. In einem ersten empirischen Durchlauf zur Semestermitte anhand einer studentischen Stichprobe aus dem Fachbereich Psychologie der Universität Heidelberg (N=212) wurde eine erste Faktorenanalyse berechnet, die auf den ersten Blick einen g-Faktor der Qualität guter Lehre nahe legt, aber dennoch in vier zugrunde liegende Dimensionen aufgefaltet werden konnte. Auf Basis dieser vierfaktoriellen Lösung wurden nur die Items mit zufrieden stellender Ladung beibehalten und so die Gesamtanzahl an standardisierten Fragen auf 14 eingeschränkt, wobei zusätzlich die numerische Absolutbewertung in Form einer Note integriert wurde. Zudem wurden die offenen Fragen „Was hat Ihnen gut gefallen?“, „Was hat Ihnen nicht gut gefallen?“ und „Welche Verbesserungsvorschläge und weiteren Kommentare haben Sie?“ aufgenommen.

In einem weiteren empirischen Durchlauf, ebenfalls inmitten eines laufenden Semesters, wurde das so resultierende Instrument in verschiedenen Seminaren, Übungen, Praktika und Vorlesungen am Psychologischen Institut der Universität eingesetzt (N=648). Zudem wurden die dabei involvierten Dozierenden um Äußerungen zum entwickelten Fragebogen als Mittel zur Evaluation guter Lehre und, draus folgernd, als Werkzeug zur Mittelvergabe gebeten. Hierzu wurden folgende Fragen vorgegeben:

1. Können Sie sich vorstellen, dass der von uns angebotene Fragebogen dazu verwendet werden kann, finanzielle Mittel auf Basis der Lehrevaluation zu verteilen? Welche Grenzen, Vor- und Nachteile sehen Sie hier in Bezug auf den Fragebogen und generell? Welche Verbesserungsvorschläge haben Sie?

2. Wie sehen Sie die Möglichkeiten - auf Basis der Lehrevaluation - Ihre Lehre zu verbessern? Wie setzen Sie das Instrument selbst ein, um Rückmeldung an Ihre Studierenden zu geben? Welche Schlüsse ziehen Sie selbst aus den Ergebnissen?

### **Ergebnisse**

Der resultierende Fragebogen setzt sich aus den folgenden Dimensionen mit den zugehörigen Items zusammen:

#### *Form und Struktur:*

1. Der Stoff wurde angemessen veranschaulicht (z.B. durch Beispiele, Visualisierungen etc.).
2. Die Lehrveranstaltung war klar strukturiert.
3. Die Lernziele der Lehrveranstaltung waren klar definiert.
4. Es wurden zusätzliche hilfreiche Ressourcen (Handapparat, Literatur, Internet-Anbindung etc.) zur Verfügung gestellt.

#### *Merkmale des/der Dozenten/Dozentin:*

5. Der Dozent / die Dozentin war offen für Kritik.
6. Der Dozent / die Dozentin regte zur kritischen Auseinandersetzung mit den behandelten Themen an.
7. Der Dozent / die Dozentin war im Umgang mit den Studierenden freundlich und aufgeschlossen.
8. Der Dozent / die Dozentin war hilfsbereit (z.B. bei Vorbereitung studentischer Beiträge).

#### *Umfang und Relevanz:*

9. Die Relevanz der angebotenen Lehrinhalte war hoch (z.B. für Prüfungen, Beruf, Disziplin etc.).
10. Der Stoff der Lehrveranstaltung wurde in einem angemessenen Tempo behandelt.
11. Die Stoffmenge in dieser Lehrveranstaltung war zu umfangreich.

#### *Lernerfolg:*

12. Die Inhalte in dieser Lehrveranstaltung waren, wo möglich, aktuell.
13. Meinen Lernzuwachs durch diese Lehrveranstaltung schätze ich hoch ein.
14. Die Lehrveranstaltung förderte mein Interesse an dem Thema.

Eine Überprüfung der internen Konsistenz der Subskalen ergab ein Cronbachs Alpha für *Form und Struktur* von .73, *Merkmale des/der Dozenten/Dozentin* von .71, *Umfang und Relevanz* von .58 und *Lernerfolg* von .73. Mit Ausnahme der vorletzten Dimension *Umfang und Relevanz* ergaben sich somit unter Berücksichtigung der geringen Itemzahl akzeptable Werte. Da die interne Konsistenz bei der Kategorie *Umfang und Relevanz* durch die Eliminierung des Relevanz-Items nur geringfügig auf Alpha von .63 steigen würde, bleibt die faktorenanalytisch und rational ermittelte Struktur beibehalten.

Neben den oben geschilderten offenen Items wurde auch das folgende Item erfasst „Ich gebe der Gestaltung der Lehrveranstaltung durch den Dozenten/die Dozentin insgesamt folgende Note:“, wobei hier Noten von 1 in Schritten von 0,3 (bzw. 0,4) in Anlehnung an universitäre Notenvergaben bis 5 vergeben werden konnten (1; 1,3; 1,7; 2 usw.). Die regressionsanalytische Vorhersage dieser summarischen Bewertung auf Basis der Skalenmittelwerte ergab eine Varianzaufklärung von 64%.

Hinsichtlich der offenen Fragen an die beteiligten Dozierenden ergab sich folgende Befundlage: Eine leistungsabhängige Mittelvergabe auf Basis eines Instrumentes zur Lehrveranstaltungsevaluation wird kategorisch abgelehnt. Die wesentlichen Gründe hierfür liegen in der Frage nach der Beurteilungskompetenz von Studierenden: Können zum Beispiel Studierenden in den ersten beiden Semestern ihres Studiums überhaupt die Qualität der akademischen Lehre adäquat beurteilen? Außerdem ist die Vergleichbarkeit von Inhalten und Form kaum gegeben. So ist die psychologische Methodenlehre beispielsweise eher weniger beliebt; das Seminar mit wenigen Teilnehmenden wird der Massenvorlesung vorgezogen. Zudem wird der Konkurrenzdruck unter den Dozierenden erhöht und eine Senkung von Anspruchsniveaus zur Erhöhung der eigenen Beurteilung wird befürchtet. Personenübergreifend wird hier befürchtet, dass die Bewertungsfunktion der Dozierenden unter diesem Aspekt der Lehrevaluation leiden könnte.

Hinsichtlich des Einsatzes des Fragebogens wird insbesondere die Möglichkeit der Erhebung im laufenden Semester als positiv empfunden und in Kombination mit den Antworten auf die offenen Fragen als ideales Instrument zur Optimierung der eigenen Lehre verstanden.

### **Lessons learned: Zum Stand der Lehrevaluation**

Die hier skizzierten Resultate zur Entwicklung einer Kurzsкала zur Lehrevaluation sind recht heterogen. Zum einen zeigt das hier berichtete Instrument durchaus Ähnlichkeit mit bereits vorhandenen Verfahren, was

Items und Dimensionen betrifft. Dass die meisten dieser Fragebögen sich nicht grundlegend unterscheiden, liegt wohl in der ähnlichen Vorgehensweise bei der Skalenkonstruktion begründet und damit verbunden auch dem, was man sich unter dem Konstrukt guter Hochschullehre vorstellt. Die hier berichteten Skalenwerte zeigen aber, dass man dabei nicht immer auf zeitlich aufwändige Durchführungen zurückgreifen muss, um durchaus auch passable Werte zu erhalten. Selbstverständlich profitieren die psychometrischen Gütekriterien von umfangreicheren Itemsammlung, messen jedoch praktisch das gleiche wie vergleichbare Kurzfassungen (vgl. Schulz et al., 2006). Die hier berichtete Regressionsanalyse, mit der die seitens der Lernenden vergebene Schulnote einer Lehrveranstaltung durch die der Kurzsкала zugrunde liegenden Dimensionen mit respektabler Güte vorhergesagt werden kann, ist ebenfalls ein Indiz dafür, dass die zugrunde liegende Itemstruktur eher einen übergreifenden, allgemeinen Faktor widerspiegelt.

Allerdings bieten die einzelnen Subskalen die Möglichkeit, ein differenziertes Bild über den Stand der Lehrqualität einer Hochschulveranstaltung zu erhalten, um daraus Handlungen zur Optimierung abzuleiten. In diesem Sinne verliert auch der standardisierte Teil des Verfahrens seinen summativen Charakter und wird formativ eingesetzt. Dies ist auch ein Hauptabliegen, was sich aus den bereits gesammelten Erfahrungen mit dem Einsatz der Kurskala zur Lehrevaluation ergeben hat: Die Nutzung des Verfahrens anhand einer sozialen Bezugsnorm zur leistungsabhängigen Mittelvergabe erscheint eher unpraktikabel und unfair. Systematische Fehlerquellen wie eine mangelnde Beurteilungsqualität, wechselnder didaktischer Methodeneinsatz in Lehrveranstaltungen, Rahmenbedingungen wie Teilnehmergröße, Art der Lehrveranstaltung, Philosophie und Renommee des Faches lassen die summarische Vergleichbarkeit unterschiedlicher Lehrveranstaltungen fraglich erscheinen. Auch die Ziele der Dozierenden sind aufgrund fehlender Lehrpläne und daraus abgeleiteten Lehrzielen – auch wenn vorhanden – kaum zwischen verschiedenen Disziplinen vergleichbar. Wie sollte man einen Kurs mit zehn Studierenden, bei dem der Lehrende die Veranstaltung zwar interessant und anregend gestaltet, aber kaum auf die Kernpunkte einer Disziplin eingeht, mit einer Massenvorlesung vergleichen können, bei der systematisch auf die Examensprüfung vorbereitet wird? Auch die hier skizzierte Dozierendenbefragung zeigt, dass dieses mögliche Ziel der Lehrevaluation kritisch zu hinterfragen ist und eine etwaige Mittelvergabe auf Basis solcher Evaluationen kategorisch abgelehnt wird.

Vielmehr ist es sinnvoll, eine individuelle Bezugsnorm anzulegen und mit Hilfe von Lehrevaluation die eigene Lehre zu optimieren. In diesem Sinne bietet die Erhebung in der laufenden Veranstaltungszeit die Gelegenheit, einen Gesamteindruck zu erhalten und daraufhin nochmals die Rücksprache



mit den Lernenden zu suchen. Die Kombination aus standardisierten und offenen Fragen unterstützt dabei die Suche nach guten wie verbesserungswürdigen Aspekten der eigenen Lehre. Wird eine Lehrveranstaltung mehrfach in Folge abgehalten, können die standardisierten Fragebogenteile auch zur summativen Evaluation der Entwicklung eines oder einer Dozierenden herangezogen werden, um so den Nutzen etwaiger Veränderungsmaßnahmen beim Lehrkonzept systematisch zu evaluieren. Den Vergleich mit den Lehrveranstaltungen anderer muss man dabei nicht völlig verwerfen, sondern kann diesen zur Förderung einer besseren Lernkultur heranziehen. Ein Beitrag bei der Dozierendenbefragung macht den Vorschlag: „Bei institutsweitem Einsatz könnten jedes Semester die drei besten Lehrveranstaltungen gekürt werden (keine silbernen Zitronen, sondern best practice).“ Auf diese Weise kann eine positive Anreizstruktur geschaffen werden, welche die Möglichkeit zum Modelllernen bietet und Bewusstsein für die Verantwortlichkeit zur Überprüfung und Optimierung der eigenen Lehre schaffen kann.

Um dies zu unterstützen, ist der Einsatz zeitlich wie finanziell ökonomischer Verfahren indiziert. Während die hier vorgestellte Kurzsкала zur Lehrevaluation bereits vom Zeitumfang und von den Kosten für Kopien und Auswertung gegenüber längeren Verfahren von Vorteil ist (der vierseitige Fragebogen passt verkleinert und beidseitig gedruckt auf ein einziges A4-Blatt), kann eine Online-Lösung noch weitere Vorteile mit sich bringen. So kann die Bearbeitung des Fragebogens zum Teil des Leistungsnachweises gemacht werden, wobei spezielle Zertifikate eines Online-Evaluationswerkzeuges als anonymer Nachweis gegenüber den Dozierenden erbracht werden können. Die Auswertung der Fragebögen kann weitestgehend automatisch erfolgen und so weitere Ressourcen schonen. Auf diese Weise kann die Lehrevaluation ein fester Bestandteil der universitären Unternehmenskultur werden und einen wertvollen Beitrag zur Weiterentwicklung der Lehrqualität führen.

### **Literaturverzeichnis:**

- Braun, E. & Gusy, B. (2006): Perspektiven der Lehrevaluation. In: G. Krampen & H. Zayer (Hg.), Psychologiedidaktik und Evaluation. Göttingen, S. 152-166.
- Gollwitzer, M., Kranz, D., & Vogel, E. (2006): Die Validität studentischer Lehrveranstaltungsevaluationen und ihre Nützlichkeit für die Verbesserung der Hochschullehre: Neuere Befunde zu den Gütekriterien des „Trierer Inventars zur Lehrevaluation“ (TRIL). In: Krampen, G.; Zayer, H. (Hg.): Psychologiedidaktik und Evaluation. Göttingen, S. 90-104.
- Gollwitzer, M. & Schlotz, W. (2003): Das „Trierer Inventar zur Lehrveranstaltungsevaluation“ (TRIL): Entwicklung und erste testtheoretische Erprobungen. In: Krampen, G.; Zayer, H. (Hg.): Psychologiedidaktik und Evaluation IV. Bonn, S. 114-128.
- Marshall, R. E. (1978): Measuring the medical school learning environment. *Journal of Medical Education*. 53: 98-104.
- Rindermann, H. (2003): Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*. 3: 233-256.
- Rindermann, H. & Amlang, M. (1994a): Das Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE): Handanweisung. Heidelberg.
- Rindermann, H. & Amelang, M. (1994b): Entwicklung und Erprobung eines Fragebogens zur studentischen Veranstaltungsevaluation. *Empirische Pädagogik*. 8: 131-151.
- Schott, F. (1991): Instruktionsdesign, Instruktionstheorie und Wissensdesign: Aufgabenstellung, gegenwärtiger Stand und zukünftige Herausforderung. *Unterrichtswissenschaft*. 4: 195-217.
- Schulz, N., Greve, W., Koch, U., Koops, T., & Wilmers, N. (2006): Wie gut erfassen Fragebögen die Qualität der Lehrer? In: Krampen, G.; Zayer, H. (Hg.): Psychologiedidaktik und Evaluation. Göttingen, S. 75-89.
- Staufenbiel, T. (2000): Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende. *Diagnostika*. 46: 169-181.
- Zumbach, J. (2003): Problembasiertes Lernen. Münster.