

## Langfristige Effekte von regelmäßigem Feedback aus studentischen Lehrveranstaltungsevaluationen

*Jonas W. B. Lang und Martin Kersting*

**Zusammenfassung:** Erkenntnisse über die Wirkung von Feedback aus studentischen Lehrveranstaltungsevaluationen zur Qualitätsverbesserung der Hochschullehre sind angesichts von vermehrt eingeführten Qualitätssicherungsmaßnahmen in der deutschen Hochschullandschaft von hohem Interesse. Eine kurzfristige Wirkung von Feedback aus Lehrveranstaltungsevaluationen ist empirisch gut belegt. Bislang ist aber nur wenig darüber bekannt, ob die kurzfristigen Effekte auch anhalten. Im vorliegenden Beitrag wird das Thema mit Konzepten und Untersuchungen aus anderen Bereichen der Psychologie verknüpft, die die Vermutung nahe legen, dass der kurzfristige positive Effekt nicht langfristig anhält. Neuere empirische Belege stützen diese Hypothese. Abschließend werden Implikationen für die zukünftige Forschung sowie für die Evaluationspraxis der Hochschulen diskutiert.

**Schlagwörter:** Feedback, Studentische Lehrevaluation, Lehrqualität, Langzeiteffekte

**Abstract:** In the last years, German universities have increased their efforts on quality management practices. Therefore, research findings on the effects of student ratings on the improvement of instructional quality are of particular interest for German higher-education organizations. While there are sufficient empirical data on short-term effects of student ratings of instruction, research on the effects of regular feedback from student ratings of instruction over longer periods of time is sparse. Based on theoretical considerations and empirical studies from other psychological areas, the authors suggest that the initially found positive effects of student ratings are non-durable. The authors review new empirical evidence on this hypothesis. Future research directions as well as implications for higher-education organizations are discussed.

**Key Words:** Feedback, Student Ratings, Teaching Effectiveness, Long-term Effects

In vielen Ländern sind studentische Lehrveranstaltungsevaluationen heute eine gängige Praxis in Hochschuleinrichtungen. Seit den 1970er Jahren werden im deutschen Sprachraum Lehrveranstaltungen in nennenswertem Umfang durch Studierende evaluiert (Diehl, 2001). Durch die vorgenommenen Umstrukturierungen der Hochschulen und den zunehmenden Wettbewerb unter den Hochschulen gab es vor allem in den letzten Jahren einen starken Trend zum vermehrten Einsatz von studentischen Lehrveranstaltungsevaluationen. In Deutschland hat dieser Trend auch einen gesetzgeberischen Hintergrund. Eine Evaluation der Lehre unter der Beteiligung der Studierenden wird u. a. im Hochschulrahmengesetz (§ 6 des HRG in der Fassung vom 20.8.1998) und einigen Landeshochschulgesetzen vorgeschrieben (Diehl, 2003).

Für Hochschuleinrichtungen haben studentische Lehrveranstaltungsevaluationen in der Regel drei verschiedene Funktionen (Cohen, 1980; Howell & Symbaluk, 2001). Erstens sollen studentische Lehrveranstaltungsevaluationen Personalentscheidungen über Gehaltserhöhungen, Beförderungen und Berufungen von Hochschullehrenden erleichtern, indem sie präzise Messungen der Lehrqualität erlauben (Carter, 1989).

Zweitens können studentische Lehrveranstaltungsevaluationen den Studierenden bei der Wahl von Kursen und Dozierenden helfen. Die Forschung hat gezeigt, dass die Mehrzahl der Studierenden in studentischen Lehrveranstaltungsevaluationen eine wichtige Informationsquelle für derartige Entscheidungen sieht (Coleman & McKeachie, 1981; Howell & Symbaluk, 2001; Wilhelm, 2004). Dagegen haben Hochschullehrende häufig Vorbehalte gegen eine Veröffentlichung von Lehrveranstaltungsevaluationsergebnissen (Howell & Symbaluk, 2001).

Eine dritte Funktion solcher studentischer Evaluationen wird in der Verbesserung der Lehre gesehen. Der vorliegende Beitrag konzentriert sich auf diese Funktion von studentischen Lehrveranstaltungsevaluationen und beschäftigt sich insbesondere mit der Frage, ob Feedback aus studentischen Lehrveranstaltungsevaluationen geeignet ist, die Hochschullehre nachhaltig zu verbessern. Der Beitrag beginnt mit einem Überblick über die vorhandene Forschung zur Effektivität von Feedback aus Lehrveranstaltungsevaluationen. Vor diesem Hintergrund werden anschließend Implikationen aus dieser Forschung für Hochschullehrende und Verantwortliche in Hochschuleinrichtungen, sowie Perspektiven für zukünftige Forschungsvorhaben diskutiert.

#### Verbessert Feedback aus studentischen Lehrveranstaltungsevaluationen die Hochschullehre kurzfristig?

Seit den 60er Jahren wurden die kurzfristigen Effekte von Feedback aus studentischen Lehrveranstaltungsevaluationen auf die Qualität der Lehre intensiv untersucht. In den beiden verfügbaren Metaanalysen (Cohen, 1980; L'Hommedieu, Menges & Brinko, 1990) werden mäßige Effekte von studentischem Feedback auf die Qualität der Lehre berichtet (Cohen, 1980:  $d = .38$ ; L'Hommedieu et al., 1990:  $d = .34$ ). Diese Ergebnisse decken sich mit den Resultaten von Metaanalysen aus anderen Bereichen der Psychologie. Kluger und DeNisi (1996) untersuchten beispielsweise

die Effekte von verschiedenen Feedback-Interventionen auf die Leistung aus allen Bereichen der Psychologie. Guzzo, Jette und Katzell (1985) analysierten die Effekte von unterschiedlichen psychologisch basierten Interventionsprogrammen auf die Produktivität von Arbeitern. Bei Kluger und DeNisi (1996) ergab sich ein Effekt von  $d = .41$  für Feedback-Interventionen und bei Guzzo et al. von  $d = .35$ , was sich mit den Ergebnissen von Cohen (1980) und L'Hommedieu et al. (1990) deckt.

Das typische Versuchsdesign der in den Metaanalysen von Cohen (1980) und L'Hommedieu et al. (1990) berücksichtigten Studien zur Überprüfung der Effektivität von studentischen Bewertungen ist ein Zweigruppen-Vortest-Nachtest-Design. In den meisten Fällen wurde eine Gruppe von Dozierenden, die Feedback mittels eines Vortests erhalten hatten, mit den Dozierenden verglichen, die kein Feedback erhalten hatten. Typisch für diese Studien ist, dass sie lediglich den Zeitrahmen von einem Semester abdecken, in dem für die Experimentalgruppe nur eine Evaluation (meistens zur Mitte des Semesters) stattfand. Keine der Studien in den Metaanalysen bezog langfristige (mehr als zwei Semester) Lehrveranstaltungsevaluationen ein. In beiden Metaanalysen zeigte sich, dass die Effekte des Feedbacks stärker waren, wenn das Feedback durch eine Beratungsintervention ergänzt wurde (Cohen, 1980:  $d = .20$  für die reine Feedbackbedingung und  $d = .64$  für die Bedingung Feedback mit Beratung; L'Hommedieu et al., 1990:  $d = .18$  für die reine Feedbackbedingung und  $d = .55$  für die Bedingung Feedback mit Beratung).

Zusammenfassend lässt sich feststellen, dass es in der Literatur überzeugende Belege dafür gibt, dass Feedback aus studentischen Lehrveranstaltungsevaluationen zumindest eine initiale Verbesserung in späteren studentischen Evaluationen bewirkt. Dieser Befund hat dazu geführt, dass studentische Lehrveranstaltungsevaluationen in der Forschung und besonders in der Praxis als sinnvolles und wirkungsvolles Instrument zur Verbesserung der Hochschullehre gesehen werden. Verschiedene Autoren haben jedoch darauf hingewiesen, dass sowohl kurzfristige als auch langfristige Effekte berücksichtigt werden sollten, wenn der Einfluss von Feedback auf studentische Lehrveranstaltungsevaluationen bewertet werden soll (Armstrong, 1998; Greenwald & Gillmore, 1998; Stevens & Aleamoni, 1985). Langfristige Effekte sind insbesondere sehr bedeutsam, da die Mehrheit der Hochschullehrenden über viele Jahre hinweg in Hochschulinstitutionen arbeitet und während dieser Zeit regelmäßig Feedback erhält.

Trotz der hohen Bedeutung von langfristigen Effekten konzentriert sich die vorhandene Forschung nahezu ausschließlich auf die kurzfristigen Effekte von Lehrveranstaltungsevaluations-Feedback. Daher werden im nachfolgenden Abschnitt theoretische Überlegungen zu den dauerhaften Effekten von Feedback aus studentischen Lehrveranstaltungsevaluationen sowie die vorhandenen empirischen Untersuchungen zu dieser Frage rekapituliert. Es zeigt sich, dass die empirische Forschung zu den langfristigen Effekten von Feedback aus Lehrveranstaltungsevaluationen nur in sehr begrenztem Umfang Hinweise darauf liefert, wie sich die Ergebnisse aus Lehrveranstaltungsevaluationen nach dem initialen Anstieg nach dem ersten Feedback weiterentwickeln.

## Verbessert Feedback aus studentischen Lehrveranstaltungsevaluationen die Hochschullehre dauerhaft?

### Theoretische Überlegungen

Aus theoretischer Sicht gibt es drei denkbare längerfristige Verläufe der Studierendenbewertungen als Funktion von wiederholtem Feedback aus Lehrveranstaltungsevaluationen. (Darüber hinaus sind andere Veränderungsmuster denkbar, aber schwierig zu interpretieren.)

- Die erste Möglichkeit ist, dass das allgemeine Niveau der studentischen Lehrveranstaltungsevaluationen nach dem zweiten Feedback stagniert und stabil bleibt, obwohl weiterhin Evaluationen stattfinden.
- Die zweite Möglichkeit ist, dass sich die Bewertungen der Studierenden nach dem zweiten Feedback immer weiter verbessern und ein stabiles Niveau, wenn überhaupt, erst nach mehreren zusätzlichen studentischen Lehrveranstaltungsevaluationen erreicht wird.
- Eine dritte Möglichkeit ist, dass es einen qualitativen Rückgang nach der zweiten Evaluation gibt, was gleichbedeutend damit ist, dass sich die Lehre durch studentisches Feedback auf einen längeren Zeitraum hin nicht verbessert.

Die ersten beiden Möglichkeiten sind konsistent mit klassischen Lernkurven, die anzeigen, dass Feedback aus studentischen Lehrveranstaltungsevaluationen die Dozierenden anfangs dazu befähigt, ineffektive Lehrmethoden abzuschaffen. Der dritte Ansatz deckt sich mit dem theoretisch von Wood und Locke (1990) sowie von Kluger und DeNisi (1996) beschriebenen Phänomen des „Working harder-Working smarter“. Die Autoren argumentierten, dass Individuen auf Feedback, das nicht ausschließlich positiv ist, normalerweise zunächst mit höherer Anstrengung reagieren. Das größere Maß an Anstrengung äußert sich durch erhöhten Aufwand, der Fokussierung der Aufmerksamkeit auf die Aufgabe und durch die Aktivierung von aufgabenspezifischen Programmen oder Handlungsskripten, die sich in der Vergangenheit bei der gleichen oder ähnlichen Aufgaben bewährt haben. Wenn härteres Arbeiten nicht zu dem gewünschten Erfolg führt, versucht das Individuum statt nur härter auch geschickter zu arbeiten. Eine Möglichkeit geschickter zu arbeiten ist es, neue aufgabenspezifische Strategien zu suchen und auszuprobieren. Wenn eine neue Strategie fehlschlägt, wird versucht eine weitere neue aufgabenspezifische Strategie zu entwickeln. Ein derartiges Experimentieren mit der Aufgabe führt häufig zu negativen Effekten auf die Leistung in dieser Aufgabe, da die neuen aufgabenspezifischen Pläne möglicherweise weniger adäquat sind als die Strategie, die vorher von dem Individuum eingesetzt wurde (Kluger & DeNisi, 1996; Wood & Locke, 1990). Nachteilige Effekte von wiederholtem Feedback sind insbesondere dann wahrscheinlich, wenn die Aufgabe bereits gut gelernt ist, da Menschen typischerweise gut evaluierte und hoch automatisierte Strategien verwenden, um gut gelernte Aufgaben zu bewältigen (Kluger & DeNisi, 1996).

Im Bezug auf die Auswirkungen von Lehrveranstaltungsevaluationen bedeutet dies, dass die Dozierenden nach dem ersten Feedback motiviert sind, ihre Evaluationen zu verbessern und aus diesem Grund härter arbeiten, mit dem Ergebnis besserer Lehrveranstaltungsevaluationen. Die Dozierenden erwarten jedoch, dass diese Verbesserungen deutlich größer sind und schneller eintreten. Darüber hinaus verbessern sich ihre Kollegen auch, was bedeutet, dass die Verbesserung eines

einzelnen Dozierenden weniger gravierend erscheint. Eventuell ist es sogar so, dass trotz der Anstrengung und der individuellen Verbesserung relativ zu den Kollegen keinerlei Verbesserung erzielt wird. Aus diesem Grund kommen die Dozierenden zu der (impliziten) Überzeugung, dass grundsätzliche Veränderungen für weitere Verbesserungen notwendig sind und experimentieren mit völlig neuen Lehrstrategien („working smarter“). Diese neuen Strategien sind jedoch in der Regel schlecht gelernt und erprobt und wirken sich somit negativ auf die Leistungen aus.

### Empirische Datenbasis

Bislang erlaubt die vorhandene empirische Forschung noch keine abschließende Bewertung darüber, wie sich Lehrveranstaltungsbewertungen von Dozierenden über längere Zeiträume entwickeln. Zu dieser Fragestellung konnten nur zwei empirische Studien ermittelt werden, deren Ergebnisse nachfolgend rekapituliert werden.

Die erste Studie stammt von Stevens und Aleamoni (1985). Die Autoren verglichen in ihrer Studie eine Lehrveranstaltungsevaluation ohne vorheriges Feedback mit Lehrveranstaltungsevaluationen nach Feedback über mehrere Semester. In dieser Studie wurden Lehrveranstaltungsevaluationen an vier Zeitpunkten über einen Gesamtzeitraum von zehn Jahren untersucht ((1) initiale Lehrveranstaltungsevaluation ohne vorheriges Feedback und dann in Abständen von (2) 0.5 Jahren, (3) 4 – 7.5 Jahren und (4) 7.5 – 10 Jahren nach dem ersten Feedback). Zwar ergaben sich Unterschiede zwischen Dozierenden, die Feedback bekamen und Dozierenden, die neben dem Feedback an zusätzlichen Interventionsprogrammen (Interpretation der Ergebnisse oder Interpretation der Ergebnisse und Rücksprache zur zielführenden Verbesserung der Lehre) teilnahmen, aber es gab keine signifikanten qualitativen Veränderungen in den Veranstaltungsevaluationen über die Zeit hinweg. Aus dem Blickwinkel der bereits oben erwähnten Ergebnisse der beiden Metaanalysen erscheint die fehlende signifikante Verbesserung zwischen der initialen Evaluation und der ersten Lehrveranstaltungsevaluation nach Feedback in der Studie von Stevens und Aleamoni (ebd.) untypisch. Die Ergebnisse der beiden Metaanalysen, die beide klare Anstiege nach dem ersten Feedback gefunden haben, sprechen hier dafür, dass es sich um einen singulären Befund handelt. Als Folge davon ist es schwierig, die Ergebnisse der Studie über diese bereits initial fehlende Veränderung hinaus zu interpretieren.

Die zweite Studie (Lang & Kersting, in Druck) erlaubt mehr Rückschlüsse darüber, wie sich Veranstaltungsevaluationen nach mehrmaligem Feedback entwickeln. Die Studie wurde in einer Institution durchgeführt, in der zuvor keine Evaluationen der Lehrqualität stattgefunden hatten. In vier aufeinander folgenden Semestern evaluierten die Studierenden am Ende jedes Semesters die Dozenten, und die Dozenten erhielten ein entsprechendes Feedback. Zum Schluss der Studie waren vollständige Daten über den Zeitraum von vier Semestern von zwölf Dozenten verfügbar. Dies entspricht weitgehend einer Vollerhebung des kleinen Instituts. Für diese zwölf Dozenten lagen insgesamt 3122 Fragebogen vor. Eingesetzt wurde der um einige selbstkonstruierte Items ergänzte Evaluationsbogen von Diehl (Diehl, 2002, 2003; Diehl & Kohr, 1977). Die Fragebogen enthielten jeweils allgemeine und spezifische Items, die zu einer Gesamtskala zusammengefasst wurden. Empfehlungen aktueller Arbeiten aus der Mehrebenenetheorie (Bliese, 2000; Bliese & Jex, 2002;

Klein, Dansereau & Hall, 1994) folgend, wurden die Bewertungen der Studierenden auf der Dozentenebene aggregiert. Statistische Analysen mit der Hilfe von Gruppenübereinstimmungskoeffizienten und Intraklassenkorrelationen ergaben, dass die Aggregation auch statistisch gerechtfertigt war.

Die Ergebnisse der Studie zeigten einen deutlichen Anstieg vom ersten Semester (kein Feedback) auf das zweite Semester. Die Effektstärke des Anstiegs zwischen dem ersten und dem zweiten Semester war mit  $d = 0.65$  deutlich höher als die mittleren Effektstärken in den beiden Metaanalysen zu kurzfristigen Effekten von Feedback aus Lehrveranstaltungsevaluationen. Dieser Unterschied ist jedoch auf die Aggregation auf der Dozentenebene zurückführbar, da die in den Metaanalysen verwendeten Studien vorwiegend Analysen auf der Ebene der studentischen Bewertungen vornahmen. Lehrveranstaltungsevaluationen auf der Studierendenebene weisen in der Regel höhere Standardabweichungen und damit in der Folge auch niedrigere Effektstärken auf, da in die Standardabweichungen auch individuelle Merkmale der Bewertenden eingehen. Auf der Ebene der studentischen Bewertungen entsprach die Effektstärke des Anstiegs zwischen dem ersten und dem zweiten Semester etwa den in den Metaanalysen gefundenen Effektstärken ( $d = .17$ ). Nach dem typischen initialen Anstieg sanken die Veranstaltungsevaluationen von Semester zu Semester ab ( $d = -.30$  vom zweiten zum dritten und  $d = -.29$  vom dritten zum vierten Semester), obwohl die Dozenten weiterhin jedes Semester Feedback erhielten.

Um diesen Verlauf inferenzstatistisch zu analysieren, wurde eine Serie von linearen Mischeffektmodellen (Pinheiro & Bates, 2000) mit polynominalen Trends getestet. Ein einfaches lineares Mischeffektmodell mit einem linearen Veränderungstrend offenbarte keinen signifikanten linearen Trend über die Gesamtzeit. Dagegen zeigte sich in einem quadratischen Modell ein signifikanter quadratischer Trend über die Zeit. Um den Anstieg zwischen dem ersten und dem zweiten Semester und um den kontinuierlichen Abfall ab dem zweiten Semester genauer zu modellieren, wurde außerdem ein stückweises Mischeffektmodell (Raudenbush & Bryk, 2002) getestet. Das Modell offenbarte einen signifikanten linearen Anstieg vom ersten auf das zweite Semester und einen signifikanten Abfall ab dem zweiten Semester. Ein Modellvergleich auf der Basis von informationstheoretischen Kennwerten ergab, dass das stückweise Modell geringfügig besser auf die Daten passte als das quadratische Modell.

Die beschriebenen empirischen Studienergebnisse bieten noch keine breite empirische Datenbasis zur Frage, ob studentische Lehrveranstaltungsevaluationen dauerhaft die Lehre verbessern. Erste wichtige und ernsthafte Hinweise sind aber dennoch aus der zweiten Studie ableitbar. Die erste Studie liefert keine Hinweise darauf, dass sich durch Feedback studentische Lehrveranstaltungsevaluationen überhaupt verbessern und steht damit im Widerspruch zu den Ergebnissen der Metaanalyse zu den kurzfristigen Effekten von Feedback. Eine Interpretation im Hinblick auf die dauerhafte Verbesserung durch Feedback aus Lehrveranstaltungsevaluationen ist daher nur schwer möglich. Die beschriebenen Ergebnisse der zweiten Studie stehen dagegen im Einklang mit den Befunden aus den Metaanalysen zu den kurzfristigen Effekten von Feedback. Darüber hinausgehend liefern diese Ergebnisse deutliche erste Hinweise darauf, dass das von Kluger und DeNisi (1996) sowie von Wood und Locke (1990) beschriebene Phänomen des „Working harder – Working smarter“ für wiederholtes Feedback aus studentischen Lehrveranstaltungsevaluationen zutreffen könnte.

## Forschungsperspektiven

Fünf Fragen erscheinen für die zukünftige Forschung besonders interessant: Erstens sollten weitere Studien durchgeführt werden, in denen die langfristigen Effekte von Feedback aus studentischen Lehrveranstaltungsevaluationen über mehr als zwei Semester untersucht werden. Dabei erscheint es besonders wichtig, dass derartige Studien in unterschiedlichsten universitären Kontexten durchgeführt werden. Zweitens sollten bei diesen Untersuchungen die Vorgehensweisen der Dozierenden (beispielsweise Änderungen der Lehrstrategie) erfasst und dokumentiert werden. Zukünftige Untersuchungen sollten drittens versuchen, zusätzlich zu den Studierendenbewertungen auch andere Kriterien für die langfristige Qualität von Lehrveranstaltungen zu etablieren. Beispiele sind Wissenstests, Studienerfolgskriterien oder Expertenbeobachtungen. Viertens erscheint es sinnvoll, in zukünftiger Forschung näher zu untersuchen, ob Interventionen, die das Feedback mit Beratung oder Training der Dozierenden verbinden, effektiver sind, als reines Feedback oder reine Beratungsinterventionen. Es ist denkbar, dass Feedback in Kombination mit Training oder Beratung über das alleinige Feedback hinaus inkrementell wirksamer ist. Eine fünfte Forschungsfrage schließlich bezieht sich auf die organisationale Perspektive der Hochschulen. Hier wäre es wichtig zu klären, inwiefern die Bedeutung von erfolgreicher Lehre in der Organisationskultur der Hochschule den dauerhaften Erfolg von Feedback-Interventionen beeinflusst. Bislang hat die Lehrveranstaltungsevaluation im deutschsprachigen Raum kaum Konsequenzen auf die Vergabe von Mitteln und Stellen.

## Implikationen für die Hochschulen

Obwohl viele Forscher vorgeschlagen haben, das Feedback von studentischen Veranstaltungsbewertungen durch Beratung zu ergänzen (Marsh & Roche, 1993), wird oft argumentiert, dass studentische Ratings allein ein wertvolles Instrument sind, um die Qualität der Lehre in Hochschulen zu verbessern. So findet sich zum Beispiel in der Satzung zur Qualitätssicherung in Studium und Lehre an der Otto-von-Guericke-Universität Magdeburg (2006) der Satz „Die studentische Lehrveranstaltungsevaluation dient der Verbesserung der Lehr- und Lernprozesse“. Von einer Unterstützung der Lehrenden durch die Hochschule oder andersartigen Maßnahmen außer der regelmäßigen Lehrveranstaltungsevaluation selbst ist in der gesamten Satzung nicht die Rede.

Vor dem Hintergrund der in diesem Beitrag beschriebenen Forschung erscheinen derartige Überlegungen übertrieben optimistisch. Es besteht die Gefahr, dass Hochschuleinrichtungen in der Erwartung der langfristigen Effektivität von Lehrveranstaltungsevaluationen Finanzmittel in deren Durchführung investieren, ohne dass allein durch diese Maßnahme eine dauerhafte Verbesserung der Lehre auftritt. Kontrolliert wird in diesen Fällen nicht die Effektivität der Lehrveranstaltungsevaluation, sondern die fristgemäße Abgabe der Evaluationsbogen.

Ungeachtet der aktuell noch fraglichen langfristigen Wirksamkeit von Feedback aus studentischen Lehrveranstaltungsevaluationen kann es für Hochschulen sinnvoll sein, studentische Lehrveranstaltungsevaluationen auch ohne andere Maßnahmen zur Verbesserung der Lehre einzusetzen. Wie eingangs erwähnt können Veranstaltungs-

evaluationen zur Messung der didaktischen Leistung von Hochschullehrenden eingesetzt werden. Die Hochschulen können den Studenten dann mit diesen Leistungsinformationen die Auswahl von Kursen erleichtern und selbst sinnvolle Personalentscheidungen treffen.

Insofern ist die vermehrte Nutzung von studentischen Lehrveranstaltungsevaluationen in den letzten Jahren zu begrüßen. Möglicherweise führt der vermehrte Einsatz von Lehrveranstaltungsevaluationen langfristig dazu, dass neben der häufig allein zur Bewertung von Dozierenden und Professoren/Professorinnen herangezogenen Forschungsleistung auch ihre Lehrleistungen verstärkt gewürdigt werden.

#### Literatur

- Armstrong, S. J. (1998). Are student ratings of instruction useful? *American Psychologist*, 53, 1223–1224.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Bliese, P. D. & Jex, S. M. (2002). Incorporating a multilevel perspective into occupational stress research: Theoretical, methodological, and practical implications. *Journal of Occupational Health Psychology*, 7, 265–276.
- Carter, R. E. (1989). Comparison of criteria for academic promotion of medical-school and university-based psychologists. *Professional Psychology: Research and Practice*, 20, 400–403.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis. *Research in Higher Education*, 13, 321–341.
- Coleman, J. & McKeachie, W. J. (1981). Effects of instructor/course evaluations on student course selection. *Journal of Educational Psychology*, 73, 224–226.
- Diehl, J. M. (2001). Studentische Lehrevaluation in den Sozialwissenschaften: Fragebögen, Normen, Probleme. In E. Keiner (Hrsg.), *Evaluation (in) der Erziehungswissenschaft*. Weinheim: Deutscher Studien Verlag.
- Diehl, J. M. (2003). Normierung zweier Fragebögen zur studentischen Beurteilung von Vorlesungen und Seminaren. *Psychologie in Erziehung und Unterricht*, 50, 27–42.
- Diehl, J.M. & Kohr, H.-U. (1977). Entwicklung eines Fragebogens zur Beurteilung von Hochschulveranstaltungen im Fach Psychologie [Development of a psychology course evaluation questionnaire]. *Psychologie in Erziehung und Unterricht*, 24, 61–75.
- Greenwald, A. G. & Gillmore, G. M. (1998). How useful are student ratings? Reactions to comments on the current issues section. *American Psychologist*, 53, 1228–1229.
- Guzzo, R. A., Jette, R. D. & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology*, 38, 275–291.
- Howell, A. J. & Symbaluk, D. G. (2001). Published student ratings of instruction: Revealing and reconciling the views of students and faculty. *Journal of Educational Psychology*, 93, 790–796.
- Klein, K. J., Dansereau, F. & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, 19, 195–229.
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Lang, J. W. B. & Kersting, M. (in Druck). Regular Feedback from Student Ratings of Instruction: Do College Teachers Improve their Ratings in the Long Run? *Instructional Science*, .
- L'Hommedieu, R., Menges, R. J. & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology*, 82, 232–241.
- Marsh, H. W. & Roche, L. A. (1993). The use of student evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30, 217–251.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Stevens, J. J. & Aleamoni, L. M. (1985). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science*, 13, 285–304.
- Wilhelm, W. B. (2004). The relative influence of published teaching evaluations and other instructor attributes on course choice. *Journal of Marketing Education*, 26, 17–30.
- Wood, R. E. & Locke, E. A. (1990). Goal setting and strategy effects on complex tasks. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 12, p. 73–109). Greenwich, CT: JAI Press.

---

*Annette Kluge, Kerstin Schüler (Hrsg.)*

---

Qualitätssicherung und -entwicklung  
an Hochschulen:  
Methoden und Ergebnisse

---

(2007)



PABST SCIENCE PUBLISHERS  
Lengerich, Berlin, Bremen, Miami,  
Riga, Viernheim, Wien, Zagreb