
Annette Kluge, Kerstin Schüler (Hrsg.)

Qualitätssicherung und -entwicklung
an Hochschulen:
Methoden und Ergebnisse



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

Onlinebefragung in der Lehrveranstaltungsevaluation: Ein faires, verzerrungsfreies und ökonomisches Verfahren?

Karen Tinsner und Markus Dresel

Zusammenfassung: Der Einsatz von Onlinebefragungstools in der Lehrveranstaltungsevaluation verspricht eine ökonomische Erfassung und Verarbeitung einer großen Zahl studentischer Beurteilungen. Allerdings bestehen dabei einige methodische Bedenken, vor allem in Bezug auf niedrige Beteiligungsquoten, mangelnde Repräsentativität sowie geringere Datenqualität. Im vorliegenden Kapitel wird ein Überblick über die Befundlage zu derartigen Methodeneffekten gegeben. Die vorliegenden Ergebnisse zeigen, dass Befürchtungen hinsichtlich geringer Datenqualität unberechtigt sind. Allerdings werden mit der Onlineevaluation sehr viel weniger Studierende erreicht als mit der papierbasierten Befragung, wobei die Stichprobenselbstselektivität systematisch von Merkmalen der Studierenden abhängt. Darüber hinaus existieren Indizien für schwache Methodeneffekte auf die Bewertung der Lehrqualität.

Schlagwörter: Onlinebefragung, Lehrveranstaltungsevaluation, Methodeneffekte

Abstract: The use of online survey tools promises an effective approach to assess and process huge numbers of teacher evaluations. However, there are some methodical objections, chief among them are low response rates, a lack of representativeness and low information quality. In the present chapter, we review the existing literature concerning method effects of online evaluations of teaching quality. The results were not able to confirm the suspicion of low data quality. Admittedly, the online survey was only able to reach substantially less students than the paper based survey. Furthermore, evidence was found that this method of evaluation affects – albeit rather slightly – the student's rating of teaching quality.

Key Words: Online Survey, Evaluation of Teaching Quality, Method Effects

Einführung

Im Zuge der Anforderung an Hochschulen, Qualität öffentlichkeitswirksam nachvollziehbar zu belegen und gleichzeitig zu steigern, hat sich als Qualitätsbewertungsmaßnahme in der Lehre die studentische Beurteilung von Lehrveranstaltungen breit etabliert (Lehrveranstaltungsevaluation). Bisher werden dazu die studentischen Beurteilungen meist papierbasiert mittels gedruckter Fragebögen eingeholt, woraus ein hoher Aufwand bei der Organisation der Befragungen sowie der Erfassung und Analyse der Daten resultiert. Vor dem Hintergrund dieser hohen Ressourcenbelastung werden an Hochschulen im deutschsprachigen Raum zunehmend internetbasierte Lehrveranstaltungsevaluationen durchgeführt (z. B. Gaul, Bomhardt & Schmidt-Mänz, 2004; Göritz, Soucek & Bacher, 2005). Denn für eine auf einem Online-evaluationssystem basierende Lehrveranstaltungsevaluation spricht eine Reihe an Vorteilen hinsichtlich Verfahrensökonomie, Verarbeitungsgeschwindigkeit und Verfahrenobjektivität. Da die studentischen Beurteilungen mittels eines Web-Formulars eingeholt werden, entfallen zeitaufwändige manuelle oder halbautomatische Eingaben der Daten und somit auch Fehlerquellen bei der Datenerfassung. Da zudem die Fragebogeninstruktion ausschließlich schriftlich am Bildschirm dargeboten wird, kann von praktisch vollkommener Durchführungs- und Auswertungsobjektivität ausgegangen werden. Da in vielen Systemen Funktionalitäten implementiert sind, die eine automatische Auswertung der Daten ermöglichen, ist bei der Online-Evaluation eine Ergebnisrückmeldung an die DozentInnen in „Echtzeit“ möglich (vgl. Göritz et al., 2005).

Insbesondere die rasche Datenverarbeitung, die mit einem Onlineevaluationssystem organisatorisch und finanziell realisierbar ist, stellt eine wichtige Bedingung für die meist angestrebten Verbesserungen der Lehrqualität dar. Denn bei einer zeitfernen Ergebnisrückmeldung an die Lehrenden dürften Transferprobleme die Umsetzung der Implikationen der Evaluationsergebnisse für das Lehrhandeln zusätzlich erschweren (vgl. Dresel, Tinsner & Schröck-Pauli, 2006; Dresel, Rindermann & Tinsner, in diesem Band). Die genannten Aspekte sprechen zusammenfassend dafür, dass die Umstellung von papierbasierten auf Onlinebefragungsmethoden mit Objektivitäts- und Ökonomiegewinnen bei der Lehrveranstaltungsevaluation einhergeht.

Trotz der offensichtlich gegebenen Vorteile von online durchgeführten Lehrveranstaltungsevaluationen ist allerdings vor deren flächendeckendem Einsatz zunächst die Frage zu klären, ob und wenn ja, welche Effekte die Erhebungsmethode auf die Befragten, die realisierte Befragungsstichprobe und die gewonnenen Beurteilungen der Lehrqualität hat. Zu den möglichen Methodeneffekten der Onlinebefragung in der Lehrveranstaltungsevaluation zählen insbesondere potenzielle Limitationen des Verfahrens im Hinblick auf die Rücklaufquote, die Selbstselektivität und die Repräsentativität der realisierten Studierendenstichprobe, die Qualität der gewonnenen Daten sowie mögliche Einflüsse der Erhebungsmethode auf die Evaluationsergebnisse, also auf die Bewertung der Lehrqualität.

Während Methodeneffekte von Onlinebefragungstechniken in der Empirischen Sozialforschung bereits relativ breit untersucht wurden, liegen zu möglichen Verzerrungen bei online durchgeführten Lehrveranstaltungsevaluationen bisher nur wenige Erkenntnisse vor. Im Folgenden stellen wir zunächst methodische Besonderheiten der internetgestützten Sozialforschung vor und leiten potenzielle

Probleme der internetgestützten Lehrveranstaltungsevaluation ab. Im Anschluss daran referieren wir die bisher vorliegenden, für die Onlineevaluation der Lehrqualität einschlägigen Studien. Dabei stellen wir etwas ausführlicher eine eigene empirische Arbeit vor, mit der Einschränkungen bisheriger Studien überwunden werden sollten.

Internetbasierte Methoden in der Empirischen Sozialforschung

Ebenso wie an eine internetbasierte Lehrveranstaltungsevaluation sind an den Einsatz des Internet bei der Durchführung von Forschungsstudien vor allem Erwartungen an eine Ökonomisierung der Datenerhebungsphase gerichtet. In diesem Zusammenhang werden die Orts- und Zeitunabhängigkeit internetbasierter Befragungen, die Möglichkeit zur schnellen Massenaussendung von Teilnahmeaufforderungen per E-Mail, die automatisierte Datenerfassung, die schnelle Verfügbarkeit und damit Auswertung der Daten sowie das Entfallen von Porto- und Druckkosten genannt (Überblick bei Bandilla, 1999). Diese Aspekte versprechen somit bei vergleichsweise geringem finanziellen Aufwand innerhalb eines kurzen Zeitraums, Befragungen mit einer hohen Fallzahl zu realisieren – Online-Befragungsmethoden haben deshalb mittlerweile ihren festen Platz in der Empirischen Sozialforschung (Überblick bei Batinic, Reips & Bosnjak, 2002).

Allerdings sind Einsatzmöglichkeiten, Bedeutung sowie Limitationen von internetbasierten Befragungsverfahren in der sozialwissenschaftlichen Forschung nicht unumstritten und werden noch diskutiert. Hierbei wird insbesondere der Frage nachgegangen, mit welcher Datenqualität zu rechnen ist. Studien der letzten Jahre zeigten, dass internetbasierte Befragungen hinsichtlich Reliabilität sowie kriterialer und faktorieller Validität mit klassischen Methoden vergleichbar sind (z. B. Hertel, Naumann, Konradt & Batinic, 2002). Allerdings zeigten sich auch einige – bisher kaum überwundene – Methodenprobleme, welche die Belastbarkeit der gewonnenen Daten reduzieren (je nach verfolgter Fragestellung in unterschiedlichem Ausmaß). Für die Bedingungen der Lehrveranstaltungsevaluation sind die folgenden Aspekte bedeutsam:

- *Niedrige Rücklaufquoten:* Über verschiedene Studien hinweg zeigte sich, dass im Falle der internetbasierten Befragung mit deutlich niedrigeren Teilnahmebereitschaften als bei traditionellen Verfahren (z. B. postalische Befragung, Telefoninterviews) zu rechnen ist (z. B. Tuten, Urban & Bosnjak, 2002).
- *Eingeschränkte Repräsentativität:* Aufgrund von systematischen Stichprobenausfällen in Abhängigkeit bestimmter biographischer, infrastruktureller aber auch psychologischer Merkmale, also aufgrund einer systematischen Selbstselektivität in der Stichprobe ist diese in Bezug auf wichtige Merkmale häufig nicht repräsentativ für die Grundgesamtheit (Bandilla, 1999).
- *Reduzierte Datenqualität:* Einige Studien erbrachten, dass die Befragten bei internetbasierten Umfragen Antworten in den Extremwertbereichen der Antwortskalen vermeiden und dazu tendieren, mehr Antworten im mittleren Skalenbereich zu geben (Taylor, 2000). Zudem liegen Indizien für eine erhöhte Anzahl an fehlenden Antworten vor, deren Ursachen durch die generelle Unkontrollierbarkeit der Untersuchungssituation nicht eruiert werden kann (Bandilla, 1999).

Zu befürchten bleibt, dass diese Methodenprobleme beim Einsatz eines Online-systems zur Lehrveranstaltungsevaluation ebenfalls auftreten. Insbesondere eine unzureichende Repräsentativität und eine Stichprobenselbstselektivität in Abhängigkeit von Beurteilungen der Lehrqualität stellen potenzielle Probleme dar, welche die Anwendbarkeit der Onlineevaluation fraglich erscheinen lassen. Dies gilt umso mehr, als es bei der Lehrveranstaltungsevaluation weniger um die Prüfung von relativen Merkmalszusammenhängen, denn um den Erhalt von absoluten Bewertungen der Lehrqualität auf letztlich individualdiagnostischer Ebene der Lehrpersonen geht.

Befundlage zu Online-Befragungsmethoden bei der Lehrveranstaltungsevaluation

Bisher wurden potenzielle Methodeneffekte der Onlineerfassung studentischer Lehrveranstaltungsbeurteilungen nur selten systematisch untersucht. Dies ist insofern als relevantes Forschungsdefizit zu werten, als die Befunde zur internetbasierten Sozialforschung nicht ohne Weiteres auf die Bedingungen der Lehrveranstaltungsevaluation übertragbar sind, beispielsweise weil es sich dabei um Fremdbeurteilungen (des Handelns einer Lehrperson) und nicht um die Selbstbeurteilung eigener Verhaltensweisen oder Einstellungen handelt. Im Folgenden besprechen wir die derzeit vorliegenden Studien, wobei wir auf die Rücklaufquoten, etwaige Selbstselektionseffekte, die Repräsentativität der gewonnenen Stichproben, die Datenqualität sowie mögliche Verzerrungen der Evaluationsergebnisse eingehen.

Layne, DeCristoforo und McGinty (1999) verglichen online erfasste mit papierbasierten Veranstaltungsbeurteilungen in einer Studie, die sie an einer US-amerikanischen Universität durchführten. Studierende verschiedener Studiengänge wurden dabei innerhalb von Lehrveranstaltungen randomisiert einer Online- oder eine Papierbedingung zugewiesen. Die Rücklaufquote in der Onlinebedingung betrug eine Woche nach Evaluationsbeginn 35%, mit einer Erinnerungsmail konnte sie auf 48% erhöht werden. In der Papierbedingung konnte demgegenüber der größere Rücklauf von 61% erreicht werden. Gegenüber diesen differenziellen, scheinbar auf eine geringere Akzeptanz der Onlinebefragung hindeutenden Rücklaufquoten, ergab eine Zusatzbefragung, dass die Zufriedenheit mit der Onlineerhebung größer war als mit der papierbasierten Erhebung. Die Anonymität wurde allerdings bei der herkömmlichen Befragung besser bewertet. Im Hinblick auf eine systematische Selbstselektivität in der Stichprobe ergab die Studie von Layne et al. (1999), dass Studierende mit besseren Noten häufiger an der Online-Evaluation teilnahmen als Studierende mit schlechteren Noten. Weiterhin nahmen Studierende im zweiten Studienjahr eher an der Online-Evaluation teil als Studierende höherer Semester. Schließlich waren Studierende der Naturwissenschaften eher zur Teilnahme an der Online-Evaluation bereit als Studierende aus anderen Studiengängen. Hinsichtlich möglicher Verzerrungen der Evaluationsergebnisse erbrachten die Ergebnisse von Layne et al. (1999) keine signifikanten Abhängigkeiten der Lehrqualitätsbewertungen von der Befragungsmethode. Ein weiterer Befund war schließlich, dass Studierende der Onlinebedingung mehr Antworten auf offene Fragen gaben als Studierende der Papierbedingung.

Dommeyer, Baum und Hanna (2002) gingen in ihrer ebenfalls an einer US-amerikanischen Universität durchgeführten Studie der Frage nach den individuellen

Gründen der Studierenden für eine Nicht-Teilnahme an der Online- bzw. papierbasierten Evaluation nach. Die Autoren ließen dazu Kurse evaluieren, die ohnehin in zwei Studierendengruppen stattfanden. Eine Gruppe wurde gebeten, einen Onlinefragebogen auszufüllen, die andere Gruppe wurde der Papierbedingung zugewiesen. Nach Abschluss des Kurses wurde allen Studierenden ein Papierfragebogen vorgelegt, mit dem biographische Merkmale, die erwartete Note im Kurs, Gründe für die Nicht-Teilnahme an der Evaluation sowie eine Einschätzung der jeweiligen Erhebungsmethode erfasst wurden. Ähnlich wie in der Studie von Layne et al. (1999) ergab sich für die Onlineevaluation eine wesentlich niedrigere Beteiligungsquote als für die papierbasierte Evaluation (60% vs. 92%). Die Analyse des Teilnahmeverhaltens erbrachte für keine der beiden Erhebungsmethoden eine Stichprobenselbstselektivität in Abhängigkeit des Geschlechts der Studierenden, der erwarteten Note oder der Bewertung der Lehrqualität. Die angegebenen Gründe für die Nicht-Teilnahme an der Kursevaluation waren für die jeweilige Erhebungsmethode spezifisch: Im Zusammenhang der papierbasierten Befragung wurde als Grund vor allem fehlende Anwesenheit am Evaluationstermin genannt. Hinsichtlich der Onlineevaluation wurde von den Nicht-TeilnehmerInnen dagegen am häufigsten angegeben, dass sie den Evaluationsstichtag verpasst oder Probleme mit dem Computer hatten. Zudem hing die niedrigere Rücklaufquote bei der Onlineerhebung damit zusammen, dass die Studierenden deren Anonymität in Frage stellten (vgl. Layne et al., 1999), die Onlinemethode unbequem und zeitaufwändig empfanden oder technische Probleme hatten. Dommeyer et al. (2002) vertreten die Auffassung, dass trotz der Vorteile der Onlineevaluation vor allem das Problem der niedrigen Rücklaufquoten gelöst werden müsse, bevor Lehrveranstaltungsbeurteilungen flächendeckend via Internet erfasst werden können.

Darauf aufbauend fokussierte die Autorengruppe mit einer weiteren Studie an derselben Universität die Wirkung verschiedener Anreize zur Verbesserung des Rücklaufs bei der Onlineevaluation (Dommeyer, Baum, Hanna & Chapman, 2004). Zum Einsatz kam ein ähnliches quasi-experimentelles Design wie in der ersten Studie der Autorengruppe: Jeweils eine Untergruppe des jeweiligen Kurses evaluierte online außerhalb der Veranstaltung, die jeweils andere Untergruppe evaluierte papierbasiert während der Veranstaltungszeit. Zusätzlich wurden die Studierenden, die den Onlinefragebogen beantworteten, einer von vier Anreizbedingungen zugewiesen: (1) Kontrollgruppe (Online-Evaluation ohne zusätzlichen Anreiz), (2) Notenbonus für das Ausfüllen des Onlinefragebogens, (3) Demonstration der Beantwortung des Onlinefragebogens sowie (4) schnelle Rückmeldung der Kursnoten wenn zwei Drittel der TeilnehmerInnen den Kurs online bewertet haben. Die Ergebnisse von Dommeyer et al. (2004) zeigten Unterschiede zwischen den Rücklaufquoten in den einzelnen Bedingungen. Zunächst war der Unterschied zwischen der Onlineevaluation ohne zusätzlichen Anreiz (Kontrollgruppe) und der papierbasierten Evaluation deutlicher als in früheren Untersuchungen (29% vs. 70%). Am effektivsten zur Steigerung des Rücklaufs bei der Onlineevaluation erwies sich der Notenbonus: Nur mit diesem Anreiz konnten vergleichbare Rücklaufquoten wie bei der papierbasierten Erhebung erreicht werden (87%). Die Demonstration des Onlinesystems und die Rückmeldung der Kursnoten bewirkten eine leichte Steigerung der Rücklaufquoten (53% und 51%), waren aber weniger effektiv als der Notenbonus.

Die bisherigen Studien konzentrierten sich vor allem auf die Rücklaufquoten, die bei der Evaluation von Lehrveranstaltungen mit einem Onlinesystem zu erzielen sind. Sie erbrachten übereinstimmend deutlich niedrigere Rücklaufquoten als bei der

papierbasierten Befragung. Eine uneinheitliche Befundlage resultierte im Hinblick auf die zentrale Frage nach einer systematischen Selbstselektivität und damit der Repräsentativität der Befragungsstichprobe. Während Layne et al. (1999) hier einige Merkmale identifizierten, die die Wahrscheinlichkeit zur Teilnahme an der Onlineevaluation beeinflussten und wohl auch mit der Bewertung der Lehrqualität im Zusammenhang stehen (z. B. Leistungen; vgl. Rindermann, 2001), konnte in der Studie von Dommeyer et al. (2002) keine derartige Selbstselektivität nachgewiesen werden. Vor dem Hintergrund dieser uneinheitlichen Befunde sind die in den vorliegenden Studien nicht nachweisbaren Abhängigkeiten der Evaluationsergebnisse von der Erhebungsmethode vorläufig nicht völlig auszuschließen. Dies gilt umso mehr, als die bisherigen Studien auch in methodischer Hinsicht Einschränkungen aufweisen. So wurde in den Studien von Dommeyer et al. (2002, 2004) ein quasi-experimentelles Design realisiert, bei dem unterschiedliche, nicht randomisiert zugewiesene Studierende unterschiedliche Lehrheiten bewerteten (die sich aber auf die selben Inhalte bezogen). Damit kann die Äquivalenz der Lehrqualität nicht kausal auf die Äquivalenz der Erhebungsmethode zurückgeführt werden. Einschränkungen ergeben sich auch aus der fraglichen Übertragbarkeit der Befunde, die an US-amerikanischen Universitäten mit spezifischer Lehr- und Evaluationskultur gewonnen wurden, auf die Bedingungen an deutschsprachigen Hochschulen. Beispielsweise wäre der Anreiz des Notenbonus zur Verbesserung der Rücklaufquote, der an Hochschulen in den USA im Rahmen des dortigen Credit-Point-Systems durchaus praktikabel erscheint, an deutschen Hochschulen derzeit wohl kaum vorstellbar (vgl. Dommeyer et al., 2004). Darüber hinaus bleibt in den berichteten Untersuchungen offen, inwieweit das im Zusammenhang der internetbasierten Sozialforschung angesprochene Methodenproblem der geringeren Datenqualität bei der Onlineevaluation von Lehrveranstaltungen zum Tragen kommt.

In einer eigenen Studie untersuchten wir Rücklaufquoten, Selbstselektionseffekte, Datenqualität sowie mögliche Verzerrungen der Evaluationsergebnisse bei online erfassten und papiergebundenen Beurteilungen von Lehrveranstaltungen an der Universität Ulm (Dresel & Tinsner, 2006). Realisiert wurde ein Messwiederholungsdesign, bei dem sieben Vorlesungen aus unterschiedlichen Fächern und Studienabschnitten wiederholt evaluiert wurden, einmal papiergebunden und einmal mit einem Onlineevaluationssystem. Um Sequenzeffekte und Effekte des Befragungszeitpunktes zu kontrollieren, wurde ein Teil der Vorlesungen zuerst papiergebunden und ein anderer Teil zuerst mit einem Onlinesystem evaluiert. Die Befragungssequenz wurde in allen Analysen kontrolliert. Das realisierte Messwiederholungsdesign ermöglichte es, über die Effekte der Selbstselektivität hinaus Methodeneffekte zu identifizieren, die darauf zurückzuführen sind, dass dieselben Personen Onlinefragebögen in anderer Weise beantworten als Papierfragebögen. Dies könnte beispielsweise dann der Fall sein, wenn das Onlineformat aufgrund seiner spezifischen Darstellungsform zu einer anderen Beurteilung führt als das Papierformat. Eingesetzt wurde ein Fragebogen, der an der Universität Ulm zur studentischen Evaluation von Vorlesungen entwickelt und dessen Reliabilität sowie Validität nachgewiesen wurden (vgl. Dresel et al., 2006). Bei der papierbasierten Erhebung wurden die Fragebögen während der Vorlesungszeit ausgeteilt und von den Studierenden ausgefüllt. In der Versuchsgruppe der Onlineevaluation wurde ebenso während der Vorlesungszeit den Studierenden das Prozedere der Onlineevaluation ausführlich erklärt und demonstriert. Sie erhielten zusätzlich ein Informationsblatt, auf dem die Onlineevaluation erklärt war.

Insgesamt lagen in unserer Untersuchung nach Abschluss der Datenerhebung 864 ausgefüllte Papier- und 326 ausgefüllte Onlinefragebögen vor. Relativiert auf die an den jeweiligen Vorlesungssitzungen anwesenden Studierenden resultierte für die Papierbefragung eine Rücklaufquote von 82% und für die Onlinebefragung eine Rücklaufquote von 29%. Dieser sehr deutliche Unterschied war auch auf Ebene der Lehrveranstaltungen statistisch signifikant. Die Bedenken hinsichtlich der Qualität der Daten konnten mit unserer Studie nicht bestätigt werden. So war die Anzahl fehlender Werte bei der Onlinebefragung nicht statistisch signifikant erhöht. Im Hinblick auf die befürchtete Tendenz zur Skalenmitte erbrachte die Analyse sogar bei vier von 14 Konstrukten der Lehrqualität signifikant größere Varianzen bei der Onlineevaluation. Bei weiteren acht Konstrukten waren größere Varianzen auf deskriptiver Ebene zu beobachten. Bei der Onlineevaluation schöpften die Befragten die Antwortskalen also eher stärker aus als bei der papierbasierten Evaluation. Die Urteile fielen somit bei der Beantwortung des Onlinefragebogens differenzierter aus statt wie befürchtet weniger differenziert. Die Überprüfung, ob die aus den beiden Erhebungsmethoden resultierenden Stichproben im Hinblick auf Merkmale der Befragten vergleichbar waren, zeigte, dass in der Onlinestichprobe Studierende der ersten beiden Fachsemester überrepräsentiert und jene höheren Fachsemesters unterrepräsentiert waren. Weiterhin waren Studierende, die die Lehrveranstaltungen aus Interesse besuchten, in der Onlinestichprobe überrepräsentiert. In Bezug auf andere Besuchsgründe, den individuellen Aufwand für die Veranstaltung, das Geschlecht und das Vorwissen der Studierenden ließen sich keine statistisch signifikanten Unterschiede beobachten. Da die mit der papiergebundenen Befragung erreichte Stichprobe die jeweilige Grundgesamtheit aller VeranstaltungsteilnehmerInnen weit mehr ausschöpft als die mit dem Onlinesystem gewonnene Stichprobe, können die genannten Unterschiede zwischen den beiden Stichproben vorsichtig als Indikatoren für eine unzureichende Repräsentativität vor allem der Onlinestichprobe interpretiert werden. Unter multivariater Perspektive und unter Einbezug der Beurteilungen der Lehrqualität wurde die Teilnahme an der Onlinebefragung einer logistischen Regression unterzogen. Diese Analyse bestätigte zunächst die genannten Unterschiede zwischen den beiden Stichproben: Die Wahrscheinlichkeit für die Teilnahme an der Onlineevaluation war mit höherem Fachsemester reduziert und mit Vorliegen des Besuchsgrunds „Interesse“ erhöht. Daneben ließen sich Einflüsse dahingehend nachweisen, dass Studierende umso wahrscheinlicher an der Onlineevaluation teilnahmen, je besser sie die betreffende Veranstaltung insgesamt beurteilten (Allgemeinbeurteilung), aber auch je unfreundlicher sie den bzw. die DozentIn einschätzten (soziales Klima). Im Hinblick auf mögliche Verzerrungen der Evaluationsergebnisse erbrachte unsere Untersuchung schließlich einen – mit einer Effektstärke von $\eta^2 = .02$ allerdings sehr schwachen – Effekt der Befragungsmethode auf die Beurteilung der Lehrqualität. Dieser Effekt betraf nicht die summarische Bewertung des Lehrerfolgs (Lernzuwachs, Allgemeinbeurteilung), sondern einzelne Prozessmerkmale der Lehre (Rhetorik, Interessantheit der Darstellung, Medieneinsatz, soziales Klima). Diese wurden bei der Onlinebefragung ungünstiger bewertet als bei der papierbasierten Befragung.

Diskussion

Der Einsatz von Onlinebefragungsmethoden bei der Evaluation von Lehrveranstaltungen verspricht eine Ökonomisierung von Datenerfassung und Datenanalyse

und eröffnet damit unter anderem die Möglichkeit, Evaluationsergebnisse zeitnah an Lehrende und Studierende zurückzumelden. Weitere Vorteile bestehen in Bezug auf die Objektivität von Durchführung und Auswertung der Evaluation.

Betrachtet man die derzeit vorliegende Befundlage zur Onlinebefragung in der Lehrveranstaltungsevaluation zusammenfassend, lässt sich zunächst festhalten, dass sich die Bedenken im Hinblick auf geringe Rücklaufquoten bei der Onlineevaluation nicht zurückweisen lassen. Alle Untersuchungen zeigten übereinstimmend, dass Studierende sich an der Onlineevaluation gravierend seltener beteiligen als an der papiergebundenen Evaluation, sofern keine zusätzlichen Anreize dafür gegeben werden. Der Notenbonus, der sich in der Studie von Dommeyer et al. (2004) als effektivster Anreiz für die Teilnahme an der Onlineevaluation erwiesen hat, erscheint für die Bedingungen der Lehrveranstaltungsevaluation an deutschsprachigen Hochschulen jedoch als wenig praktikabel. Von daher erscheint die Identifikation wirkungsvoller und praktikabler Anreize bzw. Maßnahmen zur Verbesserung der Teilnahmebereitschaft nach wie vor als wichtiger Gegenstand zukünftiger Arbeiten. Erste Ansatzpunkte dafür lieferte die Analyse der individuellen Gründe für die Nicht-Teilnahme an der Onlinebefragung, wobei hier auf Seiten des eigentlichen Online-systems insbesondere Verbesserungen der (wahrgenommenen) Anonymität und der Bedienungsfreundlichkeit zu nennen sind (vgl. Dommeyer et al., 2002; Layne et al., 1999).

Eine geringe Beteiligung an der Onlineevaluation wäre grundsätzlich unproblematisch, wenn die damit gewonnenen Stichproben repräsentativ für die jeweiligen Grundgesamtheiten aller VeranstaltungsteilnehmerInnen wären. Allerdings wurden Indikatoren dafür gefunden, dass dies nicht gilt. In der Untersuchung von Layne et al. (1999) und in unserer Untersuchung (Dresel & Tinsner, 2006) zeigte sich eine Stichprobenselbstselektivität dahingehend, dass Studierende niedrigen Fachsemesters, Studierende mit günstiger Motivation (Interesse) und Studierende mit guten Leistungen in Onlinebefragungsstichproben stärker repräsentiert sind als andere Studierende. Unsere Studie lieferte zudem einen Indikator für eine Selbstselektion in Abhängigkeit von der Beurteilung der Lehrqualität. Besonders stark scheint der Selbstselektionseffekt in Bezug auf das Fachsemester zu sein. Dieser kann dahingehend interpretiert werden, dass sich mit zunehmender Studiendauer bei Studierenden eine zunehmende „Evaluationsmüdigkeit“ ausbreitet. Diese resultiert möglicherweise daraus, dass der Sinn und die Effekte der Evaluation auf die Lehrqualität nicht erkannt werden, da Evaluationsergebnisse in der Praxis selten an Studierende zurückgemeldet werden und dadurch initiierte Veränderungen im Lehrhandeln nicht transparent gemacht werden. Bei der papiergebundenen Evaluation kommt eine „Evaluationsmüdigkeit“ nicht oder nur kaum zum Tragen, da hier meist Vorlesungszeit für die komplette Beantwortung zur Verfügung gestellt wird und dabei ein größerer sozialer Zwang zum Ausfüllen der Fragebögen besteht. Aus dieser Interpretation lässt sich auch die Hypothese ableiten, dass mit einer regelmäßigen Rückmeldung der Evaluationsergebnisse an die Studierenden, die möglicherweise um eine veranstaltungsinterne Diskussion der Ergebnisse und etwaiger Verbesserungsvorschläge ergänzt wird (vgl. Webler, 1992), die Rücklaufquoten bei der Onlineevaluation langfristig verbessert werden können. Vielerorts impliziert dies aber eine veränderte Evaluationskultur (vgl. Rindermann, 2001).

Insgesamt uneinheitlich stellt sich die Befundlage im Hinblick auf mögliche Verzerrungen der Evaluationsergebnisse durch den Einsatz von Onlinebefragungs-

methoden dar. Derartige Verzerrungen, die aufgrund der geschilderten Stichproben-selbstselektivität erwartbar sind, waren in US-amerikanischen Studien nicht nachweisbar, fanden sich aber in geringem Ausmaß in der von uns an einer deutschen Universität durchgeführten Studie in Richtung ungünstigerer Beurteilungen. Diese Verzerrungen waren allerdings sehr schwach und auf einige Prozessmerkmale der Lehrqualität begrenzt, so dass im Hinblick auf die resultierenden Evaluationsergebnisse praktisch von der Äquivalenz der beiden Erhebungsmethoden ausgegangen werden kann.

Zurückgewiesen werden können Bedenken hinsichtlich einer geringeren Datenqualität bei der Onlineevaluation. Unsere Studie erbrachte für die Onlinebefragung keine signifikant höhere Anzahl an fehlenden Werten und sogar differenziertere Urteile als für die Papierbefragung.

Anhand der berichteten Befundlage lassen sich für den Einsatz von Onlinebefragungsmethoden bei der Lehrveranstaltungsevaluation spezifische Vor- und Nachteile gegenüber der herkömmlichen papierbasierten Befragung ableiten, die in Tabelle 1 zusammengefasst sind.

Tabelle 1. *Vor- und Nachteile der Onlineevaluation gegenüber der papierbasierten Evaluation*

Vorteile	Nachteile
- Ökonomischere Datenerfassung und Datenanalyse	- Deutlich geringere Beteiligungsquoten
- Auswertungen in „Echtzeit“ möglich	- Selbstselektion vor allem im Hinblick auf Fachsemester und Motivation
- Bessere Durchführungs- und Auswertungsobjektivität	- Daraus resultierende Einschränkungen der Repräsentativität der Befragungsstichproben für die Grundgesamtheiten aller VeranstaltungsteilnehmerInnen
- Keine Anzeichen für geringere Datenqualität	
- Praktisch keine Verzerrungen der Evaluationsergebnisse	
- Mehr Antworten auf offene Fragen, die zudem nicht in Handschrift abgegeben werden (Anonymität)	

Für die abschließende Beantwortung der im Titel angesprochenen Frage, ob die Onlineevaluation von Lehrveranstaltungen ein faires, verzerrungsfreies und ökonomisches Verfahren ist, lassen sich nicht in Bezug auf alle drei Aspekte endgültige Antworten geben. Die Ökonomiegewinne gegenüber der papierbasierten Befragung können sicherlich kaum von der Hand gewiesen werden. Sie sind der zentrale Mehrwert des Einsatzes von Onlinebefragungsmethoden. Fair ist die Onlineevaluation in dem Sinne, als dass alle VeranstaltungsteilnehmerInnen die gleiche Chance haben, ihre Beurteilungen abzugeben. Aufgrund des gegenüber der Papierbefragung längeren Evaluationszeitraums haben sogar Studierende die Chance teilzunehmen, die an dem Termin, an dem das Verfahren vorgestellt wird, nicht anwesend sind. Technische Hürden dürften in der Praxis ebenfalls – zumindest in der Population der Studieren-

den – kein Problem darstellen, so dass auch hieraus keine Einschränkungen der Fairness resultieren dürften. Praktisch verzerrungsfrei scheint die Onlineevaluation von Lehrveranstaltungen in Bezug auf die resultierenden Evaluationsergebnisse zu sein, wobei aber schwache Verzerrungseffekte hinsichtlich einzelner Aspekte der Lehrqualität resultieren können. Größere Verzerrungen treten dagegen im Hinblick auf die Zusammensetzung der Stichprobe auf. Aus der systematischen Stichprobenselbstselektivität hinsichtlich des Fachsemesters ließe sich eine gewisse Unfairness gegenüber LeiterInnen von Lehrveranstaltungen ableiten, die für spätere Studienabschnitte vorgesehen sind. Wie stark die Selbstselektivität und die deutlich reduzierte Rücklaufquote gewichtet werden, ist letztlich nicht allgemein zu entscheiden, sondern hängt vom jeweiligen Evaluationszweck ab. Bei der Entscheidung für oder gegen den Einsatz der Onlineevaluation von Lehrveranstaltungen gilt es deshalb stets, ökonomischen Nutzen und Einschränkungen bei der Stichprobenzusammensetzung gegeneinander abzuwägen.

Literatur

- Bandilla, W. (1999). WWW-Umfragen – Eine alternative Datenerhebungstechnik für die empirische Sozialforschung? In B. Batinic, A. Werner, L. Gräf & W. Bandilla (Hrsg.), *Online Research. Methoden, Anwendungen und Ergebnisse* (S. 9-20). Göttingen: Hogrefe.
- Batinic, B., Reips, U.-D. & Bosnjak, M. (Eds.) (2002). *Online Social Sciences*. Cambridge, MA: Hogrefe & Huber Publishers.
- Dommeyer, C. J., Baum, P. & Hanna, R. W. (2002). College students' attitudes towards two methods of collecting teaching evaluations: in-class vs. online. *Journal of Education of Business*, 78, 11-15.
- Dommeyer, C. J., Baum, P., Hanna, R. W. & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29, 611-623.
- Dresel M. & Tinsner K. (2006). *Methodeneffekte bei der Onlineevaluation von Lehrveranstaltungen*. Manuskript eingereicht.
- Dresel, M., Tinsner, K. & Schröck-Pauli, C. (2006). *Universitätsweite Lehrveranstaltungsevaluation an der Universität Ulm*. Abschlussbericht der Arbeitsgruppe Lehrveranstaltungsevaluation. Ulm: Universität Ulm.
- Gaul, W., Bomhardt, C. & Schmidt-Mänz, N. (2004). Einsatz von computergestützter Lehrveranstaltungsevaluation. *Zeitschrift für Evaluation*, 4, 35-50
- Göritz, A. S., Soucek, R. & Bacher, J. (2005). Webbasierte Lehrevaluation. In J. Wiemeyer (Hrsg.), *Education, Research and New Media. Chances and Challenges for Science* (S. 218-222). Hamburg: Feldhaus.
- Hertel, G., Naumann, S., Konradt, U. & Batinic, B. (2002). Personality Assessment via Internet: Comparing online and paper-and-pencil questionnaires. In B. Batinic, U.-D. Reips & M. Bosnjak (Eds.), *Online Social Sciences* (pp. 115-133). Cambridge, MA: Hogrefe & Huber Publishers.
- Layne, B. H., DeCristoforo, J. R. & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40, 221-232.

- Rindermann, H. (2001). *Lehrevaluation – Einführung und Überblick zur Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen*. Landau: VEP.
- Taylor, H. (2000). Does Internet research work? *International Journal of Market Research*, 42, 51-63.
- Tuten, T. L., Urban, D. J. & Bosnjak, M. (2002): Internet Surveys and Data Quality: A review. In B. Batinic, U.-D. Reips & M. Bosnjak (Eds.), *Online Social Sciences* (pp. 7-26). Cambridge, MA: Hogrefe & Huber Publishers.
- Webler, W.-D. (1992). Evaluation der Lehre: Praxiserfahrungen und Methodenhinweise. In D. Grünh & H. Gattwinkel (Hrsg.), *Evaluation von Lehrveranstaltungen. Überfrachtung eines sinnvollen Instruments* (S. 143-161)? Berlin: FU-Dokumentationsreihe.