

# Lehrevaluation online oder papierbasiert? Ein empirischer Vergleich zwischen traditionellem Fragebogen und inhaltsgleicher Online-Erhebung

Anke Simon, Yvonne Zajontz, Vanessa Reit

---

Ungeachtet der hohen Bedeutung studentischer Lehrevaluation – einerseits für das interne Qualitätsmanagement einer Hochschule und andererseits für die Außenwirkung – scheinen wenige Erkenntnisse zu potentiellen Effekten der eingesetzten Erhebungsmethoden vorzuliegen. Der vorliegende Beitrag untersucht die Auswirkungen von zwei inhaltsgleichen Erhebungsmethoden (Online- versus papierbasierte Befragung) auf das studentische Evaluationsverhalten. Die Sekundärdatenanalyse basiert auf den Daten von insgesamt 70 Studienkursen und 1027 studentischen Evaluationsbögen der Fakultät Wirtschaft an der Dualen Hochschule Baden-Württemberg Stuttgart. Identifiziert wurden signifikante Unterschiede bei der Responsequote sowie den Bewertungsergebnissen zwischen dem papierbasiertem und dem Online-Verfahren. Die Pilotstudie analysiert nach Kenntnis der Autoren erstmalig einen vollständigen Jahrgang über alle Studiengänge einer Fakultät in Deutschland hinweg im Hinblick auf Methodeneffekte studentischer Lehrevaluation.

---

## 1 Hintergrund der Studie

Während sich in den USA universitäre Lehrevaluationen Mitte der 1950er Jahre etablierten (*Spiel/Gössler 2001, S.11*), entwickelten sich ähnliche Qualitätsinstrumente in Deutschland erst zu Beginn der 1990er Jahre, oftmals als Reaktion auf die unterschiedlichen Krisensymptome in der Hochschulausbildung (*Rindermann 2003, S.233*). Studentische Proteste und Verwaltungsreformen der Bundesländer forderten eine Verbesserung in der Qualität der Hochschullehre. Lehrevaluationen wurden als geeignetes Instrument angesehen, um einerseits die Effektivität und Effizienz der in den Hochschulbereich fließenden Mittel zu kontrollieren, und um andererseits geeignete Maßnahmen zur Verbesserung der Lehrqualität zu entwickeln und umzusetzen. Gleichzeitig sollten anhand von Lehrevaluationen neue Erkenntnisse gewonnen werden, was erfolgreiches Lehren und Studieren an Hochschulen ausmacht (*Kromrey 2003, S.237*). Zudem entfalten in jüngster Zeit die Ergebnisse studentischer Lehrevaluation in Form von veröffentlichten Hochschulrankings eine nicht unbeachtliche Außenwirkung, zum Beispiel im jährlichen Studienführer der Zeitschrift *DIE ZEIT* (o.V. 2011). Da mit dem Begriff „Evaluation“ – ob im Alltagsgebrauch oder im wissenschaft-

lichen Kontext – ein breites Bedeutungsspektrum verbunden ist (*Kromrey 2001, S.21*), lassen sich in vielen Bundesländern die unterschiedlichsten Evaluationsmaßnahmen an Hochschulen zur Sicherung von Mindeststandards und/oder zur Verbesserung der Ausbildungsqualität finden. Mitunter ist schwer zu differenzieren, ob mit Lehrevaluationsmaßnahmen lediglich hochschulpolitischen Regelungen Folge geleistet wird oder ob ein kontinuierlicher Verbesserungsprozess im Sinne des umfassenden Qualitätsmanagements (im englischen Sprachraum bekannt unter Total Quality Management) etabliert wird. Wie auch immer – die einzelnen Betrachtungsgegenstände, ob Lehre, Studienbedingungen oder Hochschulinfrastruktur, erweisen sich je nach angestrebter Intention der Befragung als höchst unterschiedlich.

Wird der Fokus auf das Messverfahren bzw. die Erhebungsmethodik gelenkt, ist eine ebenso große Vielfalt festzustellen. Während vor wenigen Jahren papiergebundene Befragungen noch den Standard bei der Implementierung von Lehrevaluationen definierten, hat sich durch die moderne Informationstechnologie der methodische Verfahrensschwerpunkt verlagert. Onlinebefragungstools spielen eine immer größer werdende Rolle bei der Umsetzung von Lehrevaluationen. Sie lösen papierbasierte Umfragen zunehmend ab (*Lalla/Ferrari 2011, S. 348*). Trotz dieser Entwicklung gibt es immer noch zahlreiche Hochschulen oder Lehrstühle (wie auch die Duale Hochschule Baden-Württemberg), die papierbasierte Befragungen präferieren und an dieser Methode weiterhin festhalten (*Dresel/Tinsner 2008, S.183*).

Obwohl Lehrevaluationsverfahren in der Regel von wissenschaftlichen Institutionen durchgeführt werden, fehlt es der empirischen Methodologie (erstaunlicherweise) oftmals an Rezeption (*Rindermann 2003, S.234*). Erhebungsmethoden und -verfahren werden wenig reflektiert und tendenziell pragmatisch eingesetzt. Die Aussagekraft und Gültigkeit der Umfrageergebnisse gelten damit als wenig abgesichert und werden nicht selten in Frage gestellt (hier exemplarisch *Endrueweit 1992, Kromrey 1999, Kromrey 2003, Maurer/Jandura 2009, Meinefeld 2009, Schnell/Kopp 2001*). An der Dualen Hochschule Baden-Württemberg (kurz DHBW) werden im Rahmen des Qualitätsmanagements einmal im Jahr studentische Feedbacks sowohl mit papierbasierten als auch mit Online-Befragungen erhoben. Die Wahl der Befragungsmethode obliegt dabei den Studiengangsleitern. Die Erhebung studentischen Feedbacks an der DHBW ist einerseits Bestandteil des regelhaft implementierten Qualitätsmanagements (*DHBW 2011, S. 42ff*). Auf der anderen Seite spielt die Erhebung der studentischen Perspektive eine essentielle Rolle im Rahmen der Evaluationsforschung neuer Studienangebote. Hier ist insbesondere das Forschungsprojekt OPEN – OPen Education in Nursing – zu nennen, in welchem die Einführung eines berufsbegleitenden Studienprogramms in der Pflege evaluiert wird (gefördert durch das Bundesministerium für Bildung und Forschung und die Europäische Union, *Simon 2012*).

Vor diesem Hintergrund stellt sich die Frage, ob die Ergebnisse papier- bzw. online-basierter Befragungsmethodik einschlägigen Gütekriterien gleichermaßen gerecht werden. Die Autoren des vorliegenden Beitrags greifen diese Fragestellung auf, indem die Auswirkungen unterschiedlicher Erhebungsmethoden auf das Antwortverhalten bei studentischen Lehrevaluationen näher untersucht werden. Dabei soll der vorliegende Artikel weder die Wirksamkeit studentischer Lehrevaluation auf die Qualitätsverbesserung noch den angemessenen Umfang und Inhalt bestehender Fragebögen beleuchten (*Rindermann 2003, S. 238*). Vielmehr steht die Aufdeckung potentieller Methodeneffekte auf die Evaluationsergebnisse von Online- versus papierbasierten Verfahren im Mittelpunkt dieser Untersuchung.

Daraus abgeleitet sollen die beiden folgenden exploratorischen Forschungsfragen beantwortet werden:

1. Bestehen Unterschiede in der Responsequote in Abhängigkeit von der Erhebungsmethodik?
2. Bestehen Unterschiede in den Evaluationsergebnissen in Abhängigkeit von der Erhebungsmethodik?

## 2 Stand der Literatur<sup>1</sup>

Erste Ansätze zu Lehrevaluationen waren bereits 1990 im deutschen Sprachraum zu beobachten. Dabei wird unter Evaluation im Allgemeinen „[...] *die Erhebung, Aufbereitung und Auswertung bewertender (also evaluativer) Aussagen* [verstanden]“ (*Kromrey 2003, S.234*). Im Hinblick auf die Evaluation der Qualität des Hochschulstudiums lassen sich zwei Perspektiven unterscheiden. Während sich die Lehrveranstaltungs-evaluation ausschließlich auf die Beurteilung von Vorlesungen und Seminaren bezieht, ist die Lehrevaluation veranstaltungs- und dozentenübergreifend auf die Hochschul-lehre insgesamt ausgerichtet, sie umfasst in der Regel das Curriculum, Rahmenbedingungen des Studiums und die Infrastruktur (*Rindermann 2003, S. 235*).

Mit dem Start des World Wide Web haben Markt- und Sozialforscher schnell erkannt, dass das Internet technisch optimale Voraussetzungen bietet, um Befragungen zu generieren und online durchzuführen. Online- (internetbasierte) Befragungen haben in

---

<sup>1</sup>Die Ausführungen in diesem Kapitel basieren auf einer systematischen qualitativen Literaturrecherche im Zeitraum 2001-dato. In die Schlagwortsuche (Methodeneffekt, Evaluation, Lehrevaluation, Befragung, Qualitätsmanagement, papierbasierte Umfrage, Online-Umfrage in Deutsch und Englisch) wurden folgende einschlägige Fachzeitschriften einbezogen: *Research in Higher Education*; *Zeitschrift für Evaluation*; *Assessment in Higher Education*; *Beiträge zur Hochschulforschung*; *Das Hochschulwesen*; *Higher Education*; *Journal of Higher Education*; *American Journal of Evaluation*; *Research Evaluation*; *Practical Assessment*; *Research & Evaluation*; *South Asian Journal of Evaluation* sowie *International Journal of Evaluation and Research in Education*. Hinzu kam eine Handsuche inklusive einer Recherche über Google Scholar. Die Schlagwortsuche ergab insgesamt 103 Treffer. Nach Sichtung der Abstracts wurden 37 Veröffentlichungen näher untersucht, wobei 21 Artikel als relevant für diesen Beitrag eingestuft wurden.

den vergangenen Jahren zunehmend an Bedeutung gewonnen (*Taddicken 2009a*, S. 91). Unter Online-Umfragen werden im Allgemeinen Befragungen verstanden, bei denen der Fragebogen entweder per E-Mail verschickt wird oder der Proband die Möglichkeit hat, den Fragebogen, der auf einem Webserver abgelegt ist, online auszufüllen (*Pötschke 2009*, S. 75f.). Während E-Mail-Befragungen zunehmend an Bedeutung verlieren, gewinnen Umfragen über ein webbasiertes Befragungstool an Dominanz. Lag der Anteil der Online-Erhebungen an allen (kommerziellen) Umfragen im Jahr 1998 bei knapp einem Prozent, wurden im Jahr 2010 nach Aussagen des Arbeitskreis Deutscher Markt- und Sozialforscher 38 Prozent aller Befragungen internetbasiert durchgeführt (*ADM 2011*, S.12). In den USA war schon vor einigen Jahren jede dritte Umfrage onlinebasiert (*Maurer/Jandura 2009*, S. 61). Neue Kommunikationstechnologien führen zu einer Abnahme der klassischen Befragungsformen und forcieren bzw. fördern nicht nur in der Markt- und Sozialforschung, sondern in weitaus mehr Bereichen eine „computerized administration“ (*Richmann et al. 1999*, S. 754).

Seit einigen Jahren zeigt sich die vermehrte Anwendung von Online-Befragungen im tertiären Bildungssektor nicht zuletzt an der Zunahme theoretischer Kapitel zur Online-Forschung in der wissenschaftlichen Literatur (*Zerback et al. 2009*, S. 15). Trotz dieser Entwicklungen werden an Hochschulen weiterhin gedruckte Fragebögen zur studentischen Beurteilung von Lehrveranstaltungen verwendet und zum Teil bevorzugt (*Dresel/Tinsner 2008*, S.183).<sup>2</sup> Als zentrale Vorteile von webbasierten Evaluationen erweisen sich die einfache Handhabung in der Umsetzung von Online-Befragungen im Netz sowie die geringen Kosten für die Datenerhebung und die Schnelligkeit der Datenerfassung (*Pötschke 2009*, S. 77). In Gegensatz dazu kann der Erhebungszeitraum bei papierbasierten Verfahren bis zu zwei Monate in Anspruch nehmen (*Lalla/Ferrari 2011*, S.349). Dazu kommt, dass Fehler in der Datenübertragung bei Online-Erhebungen aufgrund der direkten Datenerfassung entfallen. Die Daten sind unmittelbar nach Beantwortung des Fragebogens verfügbar. Ein weiterer Vorteil gegenüber den klassischen papierbasierten und Telefon-Befragungen sind die hilfreichen Fragebogenwerkzeuge (z. B. Funktion der intelligenten Filterführung, Randomisierung von Fragen und Antwortmöglichkeiten) sowie die Möglichkeit der Einarbeitung von multimedialen Elementen (unter anderem Bild- und Videodateien) (*ebd:78*). Für den Probanden von Vorteil erweist sich außerdem die geringe soziale Präsenz, die wahrgenommene Anonymität sowie die Verminderung potentieller Bewertungsängste (z. B. bei hoch sensiblen Befragungsthemen). Interviewer können die Befragungssituation durch ihre physische und verbale Präsenz nicht beeinflussen (*Taddicken 2009b*, S. 91). Der Effekt von „sozial erwünschten Antworten“ ist bei Online-Erhebungen im Vergleich zu papierbasierten Befragungen geringer, wie vermehrt in wissenschaftlichen Studien

<sup>2</sup> Nach Telefonauskunft vom Mai 2012 des Unternehmens „Electric Paper“, welches Softwarelösungen für Online- und Paper&Pencil-Befragungen für tertiäre Bildungsinstitutionen anbietet, werden ca. 70 bis 80 Prozent aller Befragungen papierbasiert durchgeführt.

untersucht wurde. In der Literatur wird dies als positiver Methodeneffekt bezeichnet (*Taddicken 2009a, S. 92*).

Trotz der Potentiale dieser Primärerhebungsmethode besitzt die Online-Forschung auch Grenzen. Infrage gestellt werden bei Online-Befragungen vor allem die Repräsentativität der Stichprobe sowie die damit zusammenhängende Datenvalidität. Die Beteiligungsquote an Online-Befragungen wird häufig als Gütemaß der erhobenen Daten herangezogen. Dabei wird der Ausfall an Probanden durch Verweigerer als „Unit-Nonresponse“ bezeichnet (*Pötschke 2009, S. 80*). Während Online-Befragungen seit einigen Jahren stetig zunehmen, ist eine gegenläufige Tendenz bei der Teilnehmerquote zu konstatieren (*Shih/Fan 2007, S.59*). In der Studie von *Layne et al. (1999)* fiel die Beteiligungsquote von papierbasierten Lehrveranstaltungsevaluationen (61 % bzw. 92 %) signifikant höher aus als bei der Onlineversion (48 % bzw. 60 %). Zu ähnlichen Ergebnissen kamen *Dresel/Tinsner (2008, S. 193f.)* (27 % vs. 82 %), *Mertler (2003)* (11 % vs. 21 %), *Eberhardinger (2003)* (66 % vs. 94 %). *Leung/Kember (2005)* erzielten gleich hohe Rücklaufquoten, wobei Studierende der Fakultät Technik die Beantwortung über das Internet präferierten. Nur *Simons/Pötschke (2006, S. 239)* erzielten in ihrer Studie bei Online-Befragungen höhere Rücklaufquoten (64 % vs. 44 %). *Adams/Umbach* untersuchten zwar keine papierbasierte Befragung, fanden in ihrer Studie zur Online-Veranstaltungsevaluation jedoch, dass diverse Variablen, z. B. Haupt- und Nebenfach oder Benotung des Fachs, die Responserate der Studierenden beeinflussen (*Adams/Umbach 2012, S.579*). Dieselben Autoren sprechen von einer zunehmenden „Befragungsmüdigkeit“, da Internetuser immer häufiger mit Online-Befragungen im Netz konfrontiert werden (*ebd: S. 579*). *Mertler (2003)* untersuchte in seiner Studie die Ursachen für die Nichtbeantwortung des Fragebogens. Fehlende Zeit zur Beantwortung der Fragen gefolgt von technischen Problemen bei der Öffnung des Fragebogens und die Länge des Fragebogens waren die am häufigsten genannten Gründe. Der Autor empfiehlt, die Beteiligungsquote durch die Schaffung von Anreizen zu steigern. Unter der Annahme, dass die Verweigerer sich nicht systematisch von den Umfrageteilnehmern unterscheiden, wären Verzerrungseffekte nicht zu befürchten. Die Fallzahl sollte jedoch nicht allzu klein werden (*ebd:80*).

Bei systematischen Ausfällen (nicht zufällige Ausfälle) darf die Unit-Nonresponse jedoch nicht vernachlässigt werden (*Adams/Umbach 2012, S. 577*), da statistische Parameter und das Beurteilungsergebnis beeinflusst werden. Insgesamt wirkt sich die Nonresponse-Problematik unter allen Fehlertypen am stärksten auf die Datenvalidität aus (*Lalla/Ferrari 2011, S. 350*). So könnte die selektive Teilnahme dazu führen, dass sich nur Studierende an der Umfrage beteiligen, die besonders positive oder negative Beurteilungen abgeben (*Meinefeld 2009, S. 7*). Eine systematische Über- oder Unterschätzung der eigentlichen Bewertungen wäre die Folge. Die systematische

Verzerrung von Umfragedaten wird in der Literatur auch als negativer Methodeneffekt bezeichnet (*Taddicken 2009a, S. 92*).

Eine weitere in diesem Zusammenhang in der Literatur bekannte Problematik stellt die „Item-Nonresponse“ dar. Dabei handelt es sich um das Fehlen von Antworten, obwohl der Rest des Fragebogens ausgefüllt wurde. Da das Fehlen von Antworten nicht Teil der Forschungsfragen dieser Untersuchung ist, soll dieser Verzerrungsfehler an dieser Stelle nur der Vollständigkeit halber kurz aufgeführt werden. Gründe für die Nichtbeantwortung von Fragen durch Probanden können vielfältig sein und von als zu heikel empfundenen Fragen bis zum Übersehen von einzelnen Aspekten reichen. Die Item-Nonresponse könnte auch an der Wahl der Methode liegen. So stellt *Eberhardinger (2003, S. 8)* in ihrer Studie fest, dass der Ausfüllgrad bei papierbasierten Umfragen fast 100 Prozent entspricht, wobei er bei Online-Umfragen zwischen 18 bis 91 Prozent variiert. Eine weitere Ursache für die Nichtbeantwortung einzelner Fragen könnte auf die Frageform zurückzuführen sein. *Dresel/Tinsner (2008, S. 185)* stellten die Hypothese auf, dass die Verwendung von offenen Fragen seltenere und weniger umfangreiche Antworten begünstigt. Diese Hypothese konnte jedoch nicht verifiziert werden. Auch die Studie von *Simons/Pötschke (2006)* widerlegt diese These. Sie kommt zu dem Schluss, dass die Beantwortung offener Fragen unabhängig von der Befragungsart ist.

Eine weitere Unterart von Messfehlern in Zusammenhang mit Online-Erhebungen ist in der Literatur unter dem Begriff der „sozialen Entkontextualisierung“ bekannt (*Taddicken 2009a, S.102*). Das hohe Anonymitätsempfinden im Internet verbunden mit der als gering wahrgenommenen Präsenz anderer Personen führt hier zur Begünstigung eines sozial eher nicht akzeptablen Antwortverhaltens (z. B. extreme und unhöfliche Kommunikation), da soziale Normen und Verhaltensregeln aufgrund fehlender Sanktionsmechanismen eine geringe Rolle spielen (*Taddicken 2009a, S.101*).

In der Diskussion zu Methodeneffekten auf das Beurteilungsverhalten (web-/papierbasiert) schlagen *Dresel/Tinsner (2008, S. 188f.)* eine weitere Differenzierung vor. Sie unterscheiden zwischen Netto- und Brutto-Methodeneffekt. Ersterer entsteht, wenn intra-individuell unterschiedliches Antwortverhalten bei verschiedenen Erhebungsmethoden auftritt. Beispielsweise antwortet ein Studierender bei Papierumfragen aufgrund der potentiellen Erkennbarkeit seiner Handschrift anders als in Onlinebefragungen. Dem gegenüber steht der Brutto-Methodeneffekt. Er drückt die Gesamtheit der Unterschiede zwischen den Erhebungsmethoden aus, die aus dem Netto-Methodeneffekt sowie aus systematischen Stichprobenausfällen entstehen. Bisher gibt es kaum Studien, die beide Effektarten berücksichtigen.

Im Gegensatz zu den zahlreichen Publikationen, welche Methodeneffekte auf die Rücklaufquote untersuchen, ist die Forschungslage zum Beurteilungsverhalten von Probanden limitiert. Bisherige Studien zeigen höchst unterschiedliche Ergebnisse. Während einige Autoren schlechtere Beurteilungen bei Online-Erhebungen aufgrund von Methodeneffekten aufzeigen (*hier exemplarisch Meinefeld 2009, Taddicken 2009b, Maurer/Jandura 2009*) stellen andere im Gegensatz dazu bessere Ergebnisse bei der Online-Variante fest (z. B. *Carini et al. 2003, Dresel/Tinsner, 2008*). *Layne et al.* können keine Unterschiede im Beurteilungsverhalten aufdecken (*Layne et al. 1999, S. 227*). Zu ähnlichen Resultaten kommen *Gamliel/Davidovitz (2005, S. 585)*, *Leung/Kemper (2005)*. Unterschiedlich sind auch die Untersuchungsgegenstände der einzelnen Studien. Beispielsweise untersuchen *Dresel/Tinsner* Lehrveranstaltungen (sieben ausgewählte Vorlesungen über vier Fächer). Auch die Studie von *Layne et al.* bezieht sich auf die Vorlesungsqualität (74 Lehrveranstaltungen in 5 Fächern). Wohingegen *Carini et al.* Methodeneffekte auf die veranstaltungs- und dozentenübergreifende Lehrevaluation, d. h. die Hochschulqualität als Ganzes, analysieren.

Der vorliegende Artikel soll einen weiteren Beitrag für den empirischen Methodenvergleich liefern und anhand einer Sekundäranalyse von studentischen Lehrevaluationen prüfen, inwieweit sich die Kommunikationsform der Primärerhebung auf die Responsequote und das Beurteilungsverhalten der Studierenden auswirkt.

### **3 Untersuchungsmethode**

#### **3.1 Datenbasis**

An der Fakultät Wirtschaft der Dualen Hochschule Baden-Württemberg am Standort Stuttgart (kurz DHBW Stuttgart) werden im Rahmen des Qualitätsmanagements einmal im Jahr studentische Feedbacks erhoben. Diese Evaluation beinhaltet das Feedback über jeweils zwei Semester (ein Studienjahr). Dabei werden verschiedene Aspekte der Lehre über einen standardisierten schriftlichen Fragebogen entweder über eine ausgedruckte Papierversion (im Englischen *paper & pencil questionnaire*, im folgenden kurz P&P) oder online über einen passwortgeschützten Zugang im Studienportal abgefragt. Der Fragebogen besteht aus zwei Teilen. Teil 1 deckt die Qualitätsmerkmale der Hochschule als Ganzes ab, im Sinne der Lehrevaluation. Im zweiten Teil des Fragebogens, der in dieser Untersuchung keine Rolle spielt, werden einzelne Lehrveranstaltungen bzw. Vorlesungen einer Bewertung unterzogen. Die Studiengangsleiter der DHBW wählen frei, ob der ihnen zugeordnete Kurs online oder papierbasiert befragt wird. Die Erfassung der Daten erfolgt anonym. Es werden bisher keine soziodemographischen Parameter (wie Geschlecht oder Alter) der Studierenden abgefragt. Die Daten werden mit Hilfe der Software Evasys (Electric Paper 5.0) an zentraler Stelle erfasst und im Rahmen des hochschulinternen Qualitätsmanagements pro Studiengang ausgewertet.

Als Basis für die vorliegende Studie wurden alle studentischen Feedbackbögen des Studienjahrgangs 2008 der Fakultät Wirtschaft aufbereitet und für die statistischen Analysen verwendet. Inkludiert sind die Daten folgender Studiengänge: Bank, Finanzdienstleistungen, Gesundheitswirtschaft, Handel, Handwerk, Immobilienwirtschaft, Industrie, Industrie/Dienstleistungsmanagement, International Business, Medien und Kommunikation, Non-Profit-Organisationen, Steuern und Prüfungswesen, Service und Logistik, Sportmanagement und Versicherung.<sup>3</sup> Das Studium an der DHBW weist eine besondere Struktur auf. Jeder Studiengang verfolgt vom ersten bis zum dritten Studienjahr eine fachliche Richtung, so dass es keine großen Grundlagenvorlesungen (z. B. Einführung in die BWL) gibt, sondern die Studenten in kleinen Gruppen, sogenannte Studienkurse (20 bis 30 Studierende), die allgemeinen als auch die spezifischen Inhalte ihres Studiengangs bzw. ihrer Studienrichtung vermittelt bekommen. Es besteht eine Anwesenheitspflicht bei allen Lehrveranstaltungen. Bezugsobjekt dieser Studie ist die Qualität des Studiums insgesamt, im Sinne der Lehrevaluation. Die Evaluation einzelner Vorlesungen bzw. Seminare steht nicht im Fokus dieser Untersuchung, da hier unterschiedliche Fragebatterien im Einsatz sind. Die Datenbasis beinhaltet die gesamte Studienzeit von drei Studienjahren (2008 bis 2011). Es handelt sich um eine exploratorische Studie mit Pilotcharakter. Aus Datenschutzgründen und um die Anonymität der einzelnen Studiengangsleiter zu gewährleisten, wird auf eine Analyse der Methodeneffekte nach Studienkursen verzichtet. Ein Vergleich der Studiengänge im Sinne eines Benchmarking ist nicht intendiert.

### 3.2 Fragebogen

Der Fragebogen zur Lehrevaluation ist standardisiert und unabhängig von der Erhebungsmethode (online oder P&P) inhaltsgleich. Er setzt sich aus insgesamt fünf Befragungskategorien zusammen, die mit 24 weiteren, untergeordneten Items unterlegt werden. Die Konstrukte bestehen aus Bewertungen zum Studium, Rahmenbedingungen des Studiums und der Praxisphasen<sup>4</sup>, ergänzt um die Selbsteinschätzung des eigenen Engagements der Studierenden. Darüber hinaus gibt es Angaben zur Gesamtbewertung des Studiums. Die einzelnen Items werden von den Studierenden auf einer 6er-Skala von 1: „sehr positiv“ bis 6: „sehr negativ“ beurteilt.

<sup>3</sup> Die studentischen Feedbackdaten des Studiengangs Wirtschaftsinformatik wurden nicht in die Analyse einbezogen, da diese nicht vollständig über drei Studienjahre vorlagen.

<sup>4</sup> Das Duale Studium an der DHBW besteht aus sich abwechselnden Theoriephasen mit Lehrveranstaltungen an der Hochschule und Praxisphasen, in denen die Studierenden bei kooperierenden Partnerunternehmen und -organisationen eingesetzt werden (weitere Details siehe unter [www.dhbw-stuttgart.de](http://www.dhbw-stuttgart.de)).



### 3.3 Analysen

Die explorative Sekundärdatenanalyse basiert auf den Primärdaten des vorliegenden studentischen Feedbacks. Ein experimentelles Studiensetting mit vorab vorgenommener Randomisierung (Interventions- und Kontrollgruppe) liegt nicht vor. Die Analysen erfolgten mit Hilfe der Statistik Software SPSS (IBM) in der Version 20. Neben den Darstellungen zur deskriptiven Statistik wurden Mittelwertunterschiede zwischen den Sub-Gruppen (online versus P&P) mit dem t-Test ermittelt (Signifikanzniveau  $p < 0.05$ ). Die Gleichheit der Varianzen wurde über einen Levene-Test und die Normalverteilung der Variablen mit Hilfe eines Kolmogorow-Smirnow Tests überprüft. Obwohl bei größeren Stichproben die Normalverteilungsannahme vernachlässigt werden kann (Bortz 2010, S. 126), diente zur Absicherung der Ergebnisse zudem der Mann Whitney-U-Test als nichtparametrisches Verfahren. Für Variablen mit nominalem Skalenniveau wurde eine Kreuztabelle erstellt.

## 4 Ergebnisse der Erhebung

### 4.1 Stichprobe

Insgesamt lagen von 70 Studienkursen 1027 auswertbare Evaluationsbögen vor (vgl. Tabelle 1). Davon wurde in 45 Studienkursen das Online-Verfahren (642 Studierende) genutzt, 25 Kurse verwendeten papierbasierte Fragebögen (385 Studierende).

**Tabelle 1:** Stichprobe

n	Studienkurse	Studentische Evaluationen
Online	45 (64 %)	642 (63 %)
P&P	25 (36 %)	385 (37 %)
Gesamt	70	1027

### 4.2 Unterschiede zwischen der Online- und der P&P-Erhebung

Wie Tabelle 2 zeigt, liegen deutliche Unterschiede in der Responsequote vor. Die P&P-Erhebung erzielte einen Rücklauf von durchschnittlich 80,42 Prozent, wohingegen bei der Online-Methode nur 57,52 Prozent der Studierenden an der Befragung teilnahmen. Der t-Test zur Überprüfung der Mittelwertunterschiede ergab ein hochsignifikantes Ergebnis mit einem t-Wert von 4,848 auf einem Niveau von  $p \leq .001$ , 2-seitig (bestätigt durch den Mann Whitney-U Test, ebenfalls auf einem Signifikanzniveau  $p \leq .001$ ).

**Tabelle 2:** Unterschiede in der Responsequote

	N	M	SD	Min	Max
Online	45	57,52	16,56	24,00	90,00
P&P	25	80,42	22,65	21,74	100,00

M Mittelwert in %, SD Standardabweichung in %, Min Minimum in %, Max Maximum in %

Wie bereits dargestellt, sind die Fragebögen zur Lehrevaluation inhaltsgleich – unabhängig von den beiden Erhebungsmethoden. Die Bewertungsmöglichkeiten durch die Studierenden erstrecken sich dabei über verschiedene Items in insgesamt fünf Bewertungskategorien:

- Studieninhalte,
- Dozentinnen und Dozenten,
- Organisation und Betreuung des Studiengangs,
- praktische Ausbildung,
- Infrastruktur.

Ergänzend kommt eine Frage zur Selbsteinschätzung des eigenen (d. h. des studentischen) Engagements hinzu (z. B. eigene Vorbereitung, Einstellung, Motivation). Darüber hinaus wird die Gesamteinschätzung zum Studium abgefragt sowie die Wiederwahlabsicht („Würden Sie sich wieder für den Studiengang entscheiden?“). Der Mittelwertvergleich mit Hilfe des t-Tests (abgesichert durch den Mann Whitney-U Test) zeigt auch hier Unterschiede zwischen den Erhebungsmethoden (Tabelle 3). Bis auf die Items zur *Selbsteinschätzung des eigenen Engagements*, der *EDV* sowie der *Bewertung der praktischen Ausbildung* (inklusive der Einzelvariablen *Zielvereinbarung/Ausbildungsplanung*, *Anwendung der Erkenntnisse der Theoriephasen*, *Betreuung*, *Organisation der Praxisphasen*, *Feedback über Ausbildungsfortschritt*) erweisen sich alle Bewertungsunterschiede als signifikant. Insgesamt geben die Nutzer des papierbasierten Fragebogens über nahezu alle Items bessere Bewertungen ab als die Teilnehmer des Online-Verfahrens (Ausnahmen bilden lediglich die Aspekte *Vermittlung berufsrelevanter Qualifikationen* sowie *Förderung von persönlichen und sozialen Kompetenzen*).

Besonders hoch fallen die Unterschiede in den Mittelwerten bei den Variablen *Bewertung der Organisation und Betreuung des Studiengangs* (Online 2,91; P&P 2,00), *Betreuung und Beratung der Studierenden* (Online 2,90; P&P 2,06) sowie *Semesterplanung/-organisation durch StudiengangsleiterIn* (Online 2,78; P&P 1,95) aus. Hier liegen die Bewertungsdifferenzen zum Teil bei fast einer Note – gefolgt von mittleren Bewertungsunterschieden zwischen 0,6 und 0,3 bei den Items *Abstimmung zwischen den Lehrveranstaltungen* (Online 3,59; P&P 3,00), *Bewertung der Dozentinnen und Dozenten* (Online 2,74; P&P 2,31), *methodisch-didaktische Bewertung* (Online 3,20; P&P 2,82), *Infrastruktur* (Online 2,99; P&P 2,59) sowie *Vorlesungs- und Gruppenräume* (Online 3,20; P&P 2,84). Kleinere signifikante Differenzen weisen die Variablen *Bewertung der Studieninhalte*, *Lernziele erreicht*, *Nutzen*, *Relevanz*, *Praxisbezug*, *Fachliche Bewertung* und *Bibliothek* auf.

Bei der Gesamteinschätzung des Studiums über alle Kategorien bewerten die Studierenden des P&P-Fragebogens um 0,4 Punkte besser als die Online-Teilnehmer

(Tabelle 4). 81 Prozent der Nutzer des P&P-Fragebogens geben an, den Studiengang wieder zu wählen, wohingegen nur 67 Prozent der Teilnehmer am Online-Verfahren eine positive Wiederwahlabsicht zum Ausdruck bringen (Tabelle 5).

**Tabelle 3:** Unterschiede zwischen den Bewertungsvariablen

Item	Online M (SD) n=624–641	P&P M (SD) n=350–375	MDiff	t
<b>Bewertung der Studieninhalte</b>	<b>2,58 (0,83)</b>	<b>2,35 (0,73)</b>	<b>0,23</b>	<b>4,64***</b>
Lernziele erreicht	2,53 (0,94)	2,40 (0,76)	0,13	2,38**
Nutzen, Relevanz, Praxisbezug	2,77 (1,07)	2,59 (0,93)	0,18	2,87**
Angemessenes Niveau	2,51 (1,00)	2,34 (0,85)	0,17	2,90**
Abstimmung zwischen den Lehrveranstaltungen	3,59 (1,22)	3,00 (1,18)	0,59	7,49***
<b>Bewertung der Dozentinnen und Dozenten</b>	<b>2,74 (0,80)</b>	<b>2,31 (0,65)</b>	<b>0,43</b>	<b>9,24***</b>
Fachliche Bewertung	2,28 (0,83)	2,01 (0,65)	0,27	5,67***
Methodisch-didaktische Bewertung	3,20 (0,94)	2,82 (0,88)	0,38	6,27***
<b>Bewertung der Organisation und Betreuung des Studiengangs</b>	<b>2,91 (1,29)</b>	<b>2,00 (1,12)</b>	<b>0,91</b>	<b>11,76***</b>
Semesterplanung/-organisation durch Studiengangsleiter/in	2,79 (1,40)	1,95 (1,12)	0,84	10,45***
Betreuung und Beratung der Studierenden	2,90 (1,43)	2,06 (1,16)	0,84	10,23***
<b>Bewertung der praktischen Ausbildung</b>	<b>2,27 (1,05)</b>	<b>2,37 (1,04)</b>	<b>0,10</b>	<b>1,38</b>
Zielvereinbarung, Ausbildungsplanung	2,71 (1,44)	2,63 (1,24)	0,08	0,87
Vermittlung berufsrelevanter Qualifikationen (fachspezifisch)	2,14 (1,03)	2,32 (1,02)	0,18	2,65**
Anwendung der Erkenntnisse der Theoriephasen	3,00 (1,16)	3,04 (1,16)	0,04	0,59
Förderung von persönlichen und sozialen Kompetenzen	2,12 (1,10)	2,35 (1,16)	0,23	3,09**
Betreuung (fachlich und persönlich)	2,34 (1,36)	2,31 (1,24)	0,03	0,36
Organisation der Praxisphasen	2,68 (1,50)	2,62 (1,26)	0,06	0,70
Feedback über Ausbildungsfortschritt (Lernerfolg)	2,77 (1,42)	2,66 (1,35)	0,11	1,20
<b>Bewertung der Infrastruktur</b>	<b>2,99 (0,97)</b>	<b>2,59 (0,96)</b>	<b>0,40</b>	<b>6,30***</b>
Bibliothek	2,76 (1,20)	2,56 (1,11)	0,20	2,65**
EDV	2,91 (1,11)	2,78 (1,30)	0,13	1,51
Vorlesungs- und Gruppenräume	3,20 (1,44)	2,84 (1,27)	0,36	4,12***
Selbsteinschätzung des eigenen Engagements	2,36 (1,00)	2,32 (0,80)	0,04	0,71

M = Mittelwert, SD = Standardabweichung, MDiff = Mittelwertdifferenz

\*\*\* p<0,001 (2-seitig) \*\* p<0,01 (2-seitig) \* p<0,05 (2-seitig)

**Tabelle 4:** Unterschiede in der Gesamtbewertung

Item	Online M (SD) n=595	P&P M (SD) n=353	MDiff	t
Gesamtbewertung	2,66 (0,82)	2,26 (0,81)	0,40	7,21***

M = Mittelwert, SD = Standardabweichung, MDiff = Mittelwertdifferenz  
 \*\*\* p < 0,001 (2-seitig) \*\* p < 0,01 (2-seitig) \* p < 0,05 (2-seitig)

**Tabelle 5:** Kreuztabelle zur Wiederwahlabsicht

Würden Sie sich wieder für den Studiengang entscheiden?	Online	P&P
ja	413 (67 %)	280 (81 %)
nein	206 (33 %)	64 (19 %)
gesamt	619	344

Chi-Quadrat nach Pearson 23,6; df 1; p < 0,001 (2-seitig)

## 5 Diskussion der Ergebnisse und Limitation

### 5.1 Diskussion

Wie bereits anfangs erwähnt, setzen sich Online-Befragungen zunehmend nicht nur im kommerziellen Bereich, sondern auch in der wissenschaftlichen bzw. akademischen Anwendung durch (Zerback *et al.* 2009, S. 15). Diese Tendenz ist ebenso an der DHBW Stuttgart zu beobachten. So führten in der vorliegenden Untersuchung deutlich mehr Studienkurse die Lehrevaluation mit Hilfe der Online-Erhebung durch. Als zentrale Vorteile gelten die einfache Handhabung in der Umsetzung über ein Online-Portal, die geringen Kosten bzw. internen Aufwendungen für die Datenerhebung, sowie der Wegfall von Übertragungsfehlern verbunden mit einer schnelleren Datenerfassung (Pötschke 2009, S. 77).

Da die Responsequote bei Online-Befragungen häufig als Gütemaß der erhobenen Daten gilt, wurde im ersten Schritt der vorliegenden Studie dieser Kennwert analysiert. Generell ist zu beobachten, dass zwar der Einsatz von Online-Befragungen seit einigen Jahren stetig steigt, die Anzahl von Teilnehmern an Online-Befragungen dagegen tendenziell abnimmt (Shih/Fan 2007, S. 59). So ist es nicht weiter überraschend, dass die Responsequote der Online-Erhebung mit 58 Prozent in der vorliegenden Studie deutlich geringer ausfällt als die Beteiligung von 80 Prozent beim P&P-Fragebogen (Forschungsfrage 1). Geht man nun davon aus, dass sich die Nicht-Teilnehmer (Unit-Nonresponse, Pötschke 2009, S. 80) in ihrem Antwortverhalten nicht systematisch von den Umfrageteilnehmern unterscheiden, wären keine gravierenden Verzerrungen zu befürchten. Die dargestellten Analyseergebnisse zeigen jedoch eindeutige

Bewertungsunterschiede (Forschungsfrage 2). Bei der Mehrzahl der Fragebogenitems nahmen die Nutzer des papierbasierten Fragebogens eine signifikant bessere Bewertung als die Online-Teilnehmer vor. Insbesondere die Studieninhalte, die Dozenten, die Studienorganisation und -betreuung sowie die Infrastruktur wiesen deutliche Mittelwertunterschiede zugunsten des P&P-Verfahrens aus. Lediglich bei den Teilaspekten der praktischen Ausbildung, der EDV-Ausstattung und bei der Selbsteinschätzung des eigenen Engagements waren keine signifikanten Unterschiede oder ein leicht gegenläufiger Befund festzustellen. Auch die Gesamteinschätzung des Studiums und die Wiederwahlabsicht fielen beim Papierfragebogen signifikant positiver aus als beim Online-Verfahren.

Die Analyseergebnisse der vorliegenden Studie lassen sich nach bisherigem Stand der Literatur in mehrfacher Hinsicht interpretieren: Die niedrigere Responserate in der Online-Befragung deutet auf Verzerrungen aufgrund selektiver Teilnahme der Probanden hin. Zudem kann durch die unterschiedliche Erhebungssituation in der Online-Befragung, trotz identischer Frageformulierung, ein anderes Antwortverhalten verursacht werden. So scheint die Gefahr der Ablenkung durch andere Informationsquellen im Internet besonders groß (*Baur/Florian, 2009, S.124*). In Bezug auf die Aufmerksamkeit und Ernsthaftigkeit des Probanden in der Beschäftigung mit der Befragung werden Internetinhalte eher flüchtig wahrgenommen, Schnelligkeit und „Wegzappen“ zum nächsten Objekt sind gewöhntes Verhalten. Die Verbindlichkeit und Zuverlässigkeit der Online-Antworten kann entsprechend geringer ausgeprägt sein (*Meinefeld 2009*). Im Vergleich zum „gedruckten Wort“ des Papierfragebogens, der in der Regel an der DHBW Stuttgart persönlich übergeben wird, kann mit der Befragung über das Online-Portal eine größere Distanz und ein weniger aktives Beteiligungsverhalten resultieren. In der Regel sind dann vor allem die Probanden interessiert und zur Teilnahme motiviert, die entweder etwas besonders Positives oder explizite Kritik mitteilen wollen (nach *Meinefeld* die „Fans“ und „Revanchisten“ (*2009, S. 7*)). Die Mehrheit der „halbwegs Zufriedenen“ ist demgegenüber mangels Motivation zur Teilnahme unterrepräsentiert (*Kromrey 2003, S. 17*). Weiterführend kann die Anonymität des Internets, verbunden mit der als gering wahrgenommenen Präsenz anderer Personen, ein sozial eher nicht akzeptables Antwortverhalten begünstigen (z. B. extreme und unhöfliche Kommunikation), da soziale Normen und Verhaltensregeln aufgrund fehlender Sanktionsmechanismen an Bedeutung verlieren. Dieses Phänomen der „sozialen Entkontextualisierung“ wird in der bereits erwähnten Studie von *Taddicken* zu Methodeneffekten von Online-Befragungen im Vergleich zu telefonischen und schriftlichen Erhebungen bestätigt (*Taddicken 2009a, S. 101*).

Schlussendlich sei auf den Einflussfaktor der sozialen Erwünschtheit hingewiesen (*Taddicken 2009a, S. 91*). Der Erklärungszusammenhang hierfür begründet sich darin, dass sozial erwünschte Antworten in Befragungssituationen mit geringer sozialer

Präsenz, hoher Anonymität und verminderten potentiellen Bewertungsängsten weitaus geringer vorkommen. Wenngleich der Effekt sozial erwünschter Antworten vor allem bei Interviewerhebungen zu beobachten ist (*Bortz/Döring 2006, S.247, Pötschke 2010, S.48*), kann dieser Methodeneffekt auch hier eine Rolle spielen. Dies gilt insbesondere dann, wenn der Studiengangsleiter den Papierfragebogen nicht über das Studiensekretariat verteilen lässt, sondern diesen z. B. im Rahmen der letzten Lehrveranstaltung persönlich aushändigt.

## 5.2 Limitation und Ansatzpunkte für die zukünftige Forschung

Die vorliegende Studie weist Pilotcharakter auf. Erstmals wurde nach Kenntnis der Autoren in Deutschland ein vollständiger Jahrgang studentischer Umfragen über alle Studiengänge einer Fakultät hinweg im Hinblick auf Methodeneffekte bei der Lehrevaluation ausgewertet. Die dargestellten Ergebnisse zeigen deutliche Auswirkungen der jeweiligen Erhebungsmethode sowohl auf die Responsequote als auch auf die einzelnen Bewertungsaspekte des Studiums.

Die dargestellten Analyseergebnisse zur Responsequote stehen im Einklang mit vorherigen Untersuchungen (*Layne et al. 1999, Dresel/Tinsner 2008, Mertler 2003, Eberhardinger (2003)*). Die gefundenen Differenzen im Beurteilungsverhalten unterscheiden sich von den Ergebnissen einiger Studien (*Carini et al. 2003, Dresel/Tinsner, 2008, Layne 1999, Gamliel/Davidovitz 2005, Leung/Kemper 2005*). Dies mag mit Unterschieden in Bezug auf das Untersuchungsobjekt (Lehrevaluation oder Lehrveranstaltungsevaluation), die Stichprobe, die Art des Fragebogens (Befragungsdimensionen und Items) sowie der Bildungsinstitution zusammen hängen.

So weist diese Studie einige Besonderheiten und Limitationen auf: Die Ergebnisse basieren auf der Datenbasis eines ausgewählten Zeitraums, an einer Hochschule (single centre) in einem Fachgebiet (Fakultät Wirtschaft). Aufgrund des dualen Studienkonzepts ist die Vergleichbarkeit mit anderen Hochschulen nur bedingt gegeben.

An der DHBW werden bisher keine soziodemographischen, studienbezogenen oder persönlichen Variablen erhoben. Daher war es nicht möglich, entsprechende Effekte zu berechnen oder eine Standardisierung durchzuführen. Auch der Aspekt der Item-Nonresponse konnte nicht ausreichend untersucht werden. Ebenso wurde, wie bereits erwähnt, das Studienfach aus Datenschutzgründen bzw. zur Sicherstellung der Anonymität des Studiengangsleiters nicht ausgewertet. Basis der hier gezeigten Studie ist eine Sekundärdatenanalyse mit den damit einhergehenden Limitationen. Es handelt sich nicht um eine experimentelle Interventionsstudie, die üblicher Weise randomisiert und kontrolliert durchgeführt wird. Ein Experimentaldesign liegt bisher

nicht vor, könnte aber einen nächsten Forschungsschritt darstellen. Kausalanalysen und Moderatoreffekte sollten dann Inhalt der weitergehenden Forschung sein.

Auch wenn der Fokus dieser Analyse auf potentiellen Methodeneffekten lag, sei ein weiterer Hinweis in Bezug auf die deskriptiv dargestellten Mittelwerte am Rande gestattet. Da das Intensivstudium an der DHBW nicht nur mit einem hohen Anforderungsniveau verbunden ist, sondern die Vorlesungen in aller Regel Pflichtveranstaltungen sind, können die Bewertungen der Lehrqualität schwerlich mit anderen Hochschulen verglichen werden.<sup>5</sup>

Insgesamt lässt sich festhalten, dass von einem Methodenmix aus Online-Erhebung und P&P-Verfahren abzuraten ist. Insbesondere sollten Vergleiche zwischen Studienkursen, die ggf. im Rahmen der internen Qualitätssicherung vorgenommen werden, oder auch die Entwicklung von Studienkursen im Zeitablauf über mehrere Jahre generell nur auf Basis eines eindeutig festgelegten Verfahrens durchgeführt werden. Gleiches versteht sich für Vergleiche zwischen Hochschulen, z.B. im Rahmen von Hochschulrankings. Die hohe Responsequote von 80 Prozent spricht tendenziell für das P&P-Verfahren. Die niedrigere Online-Beteiligung von 58 Prozent, die damit verbundenen potentiellen Schwächen in der Repräsentativität und die diskutierten Effekte auf das Antwortverhalten weisen auf einen weiteren Forschungsbedarf hin. Die hier in der Sekundärdatenanalyse gezeigten Unterschiede im Beurteilungsverhalten könnten durch eine geeignete Primärerhebung reflektiert und erweitert werden. Empfehlenswert erscheint hier z.B. eine experimentelle Interventionsstudie unter Einbeziehung soziodemographischer (z.B. Alter und Geschlecht), studienbezogener (z.B. Department und Fach) und psychosozialer (z.B. Vorkenntnisse und Herkunft) Merkmale der Studierenden. Ebenso könnte die Analyse von ausgewählten Lehrveranstaltungen an der DHBW in Bezug auf potentielle Methodeneffekte interessant sein, was zunächst die veranstaltungsübergreifende Vereinheitlichung der entsprechenden Erhebungsinstrumente an der DHBW erfordern würde. Darüber hinaus wären vergleichbare Analysen an anderen Hochschulen im Sinne eines multizentrischen Forschungsdesigns wünschenswert und könnten die Erkenntnisse auf diesem Gebiet entscheidend bereichern. Die Verfasser dieses Beitrages würden entsprechende Forschungsk Kooperationen sehr begrüßen.

Die zum Teil deutlich ausgeprägten signifikanten Bewertungsunterschiede in dieser Studie geben Hinweise, dass es trotz der nicht zu negierenden Vorteile der Online-Befragung nicht genügt, wenn die Zielgruppe der Befragung eindeutig adressierbar und mit Zugangsdaten ausgestattet ist sowie nahezu lückenlos über einen Internetzugang verfügt (ebenso *Meinefeld 2009, S. 21*). Es müssen auch Maßnahmen getrof-

---

<sup>5</sup> Pflichtveranstaltungen erscheinen nach Kromrey in der studentischen Bewertung systematisch negativer, aufgrund der fehlenden Wahlmöglichkeit zwischen Alternativangeboten (*Kromrey 2003, S. 240*).

fen werden, um eine möglichst umfängliche Teilnahme mit aktivem Beteiligungsverhalten zu sichern.

### 5.3 Fazit

Der vorliegende Artikel zeigt anhand einer Sekundärdatenanalyse eindeutige Methodeneffekte bei der studentischen Lehrevaluation auf. Identifiziert wurden signifikante Unterschiede in der Responsequote (Forschungsfrage 1) sowie im Bewertungsverhalten (Forschungsfrage 2) zwischen dem P&P- und dem Online-Verfahren. Bei der P&P-Befragung nahmen durchschnittlich 22 Prozent mehr Probanden teil als bei der Online-Erhebung. Die Bewertung des Studiums durch die P&P-Teilnehmer fiel bei der Mehrzahl der Kriterien besser aus, insbesondere bei den Fragebogen-Items zu Studieninhalten, Dozenten, Studienorganisation und -betreuung sowie zur Infrastruktur. Ebenso wurden bei der Gesamtbewertung des Studiums und der Wiederwahlabsicht positivere Ergebnisse bei den P&P-Teilnehmern erzielt.

Weiterführende Forschungen in Form von Interventionsstudien und multivariaten Analysen sowie zur Übertragbarkeit der Erkenntnisse auf andere Fachgebiete bzw. andere Hochschulen und darauf aufbauend nähere Untersuchungen zur Frage, welches Erhebungsverfahren zur Lehrevaluation methodisch besser geeignet ist, wären wünschenswert.

### Literatur

*Adams Meridith; Umbach, Paul (2012): Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Saliency, Fatigue, and Academic Environments. In: Research in Higher Education, 5, 2012, S.576–591*

*Arbeitskreis deutscher Mark- und Sozialforschungsinstitute e. V. (ADM) (2011): Jahresbericht 2010. Frankfurt (ADM)*

*Baur, Nina; Florian, Michael J. (2009): Stichprobenprobleme bei Online-Umfragen. In: Jakob, Nikolaus u. a. (Hrsg.): Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung. Wiesbaden, S.109–128*

*Bortz, Jürgen; Döring, Nicola (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin*

*Bortz, Jürgen; Schuster, Christof (2010): Statistik für Human- und Sozialwissenschaftler. Berlin/Heidelberg*

*Carini, Robert; Hayek, John; Kuh, Georg; Kennedy, John; Ouimet, Judith (2003): College students responses to web and paper surveys: Does Mode Matter?. In: Research in Higher Education 1, 2003, S.1–19*



DHBW (2011): Qualitätsmanagement-Handbuch. Band I: Strukturen, Grundsätze und Verfahren zur Sicherung und Weiterentwicklung der Qualität des Studiums, [www.dhbw.de](http://www.dhbw.de) (Stand 10.12.2012)

*Dresel, Markus; Tinsner, Karen (2008):* Onlineevaluation von Lehrveranstaltungen: Methodeneffekte bei der Onlineevaluation von Lehrveranstaltungen. In: Zeitschrift für Evaluation, 7(2), S. 183–211

*Eberhardinger, Barbara (2003):* Evaluation von Veranstaltungen der betrieblichen Weiterbildung. Web-basierte Erhebung von Teilnehmerfeedback im Vergleich zum Paper-Pencil-Verfahren. Diplomarbeit Universität Konstanz, [kops.uib.uni-konstanz.de](http://kops.uib.uni-konstanz.de) (Stand 10.12.2012)

*Endruweit, Günter (1992):* Programmevaluation als Laienspiel. Bemerkungen über Meinungsforschung, Sozialforschung und Pusch bei Studentenbefragungen. In: Soziologie 2, 1992, S.107–115

*Gamliel, Eyal; Davidovitz, Liema (2005):* Online vs. traditional teaching evaluation: mode can matter. In: Assessment & Evaluation in Higher Education 6, 2005, S.581–592

*Kromrey, Helmut (1999):* Von den Problemen anwendungsorientierter Sozialforschung und den Gefahren methodischer Halbbildung. In: Sozialwissenschaften und Berufspraxis (SUB), 22, 1999, S.58–77

*Kromrey, Helmut (2001):* Evaluation von Lehre und Studium – Anforderung an Methodik und Design. In: Spiel, Christiane (Hrsg.): Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck. Münster, S.21–60

*Kromrey, Helmut (2003):* Qualität und Evaluation im System Hochschule. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Opladen, S.234–258

*Lalla, Michele; Davide Ferrari (2011):* Web-based versus paper-based data collection for the evaluation of teaching activity: empirical evidence from a case study. In: Assessment & Evaluation in Higher Education 3, 2011, S.347–365

*Layne, Benjamin; DeCrisoforo, Joseph; McGinty, Dixie (1999):* Electronic vs. traditional student rating of instruction. In: Research in Higher Education 2, 1999, S.221–232

*Leung, Doris; Kember, David (2005):* Comparability of Data Gathered from evaluation questionnaires and paper and through the internet. In: Research in Higher Education, 5, 2005, S.571–591

*Maurer, Marcus; Jandura Olaf (2009):* Masse statt Klasse? Einige kritische Anmerkungen zu Repräsentativität und Validität von Online-Befragungen. In: Jakob, Nikolaus u. a. (Hrsg.): Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung. Wiesbaden, S.61–74

*Meinefeld, Werner (2009):* [http://www.sozioologie.phil.uni-erlangen.de/system/files/Lehrveranstaltungsevaluation\\_qua\\_Online-Befragung\\_0.pdf](http://www.sozioologie.phil.uni-erlangen.de/system/files/Lehrveranstaltungsevaluation_qua_Online-Befragung_0.pdf) (Zugriff: 03. April 2012)

*O. V. (2011):* Zeit Studienführer 2010/2011. Die Zeit. S.185–187

*Mertler, Craig A. (2003):* Patterns of response and nonresponse from teachers to traditional and web surveys. In: *Practical Assessment, Research & Evaluation*, 8,22, 2003

*Pötschke, Manuela (2009):* Potentiale von Online-Befragungen: Erfahrungen aus der Hochschulforschung. In: Jakob, Nikolaus u.a. (Hrsg.): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden, S.75–89

*Pötschke, Manuela (2010):* Datengewinnung, Datenaufbereitung und explorative Datenanalyse. In: Wolf, Christoph u.a. (Hrsg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden, S.41–64

*Rindermann, Heiner (2003):* Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. In: *Zeitschrift für Evaluation* 2, 2003, S.233–256

*Richman, Wendy; Kiesler, Sara; Weisband, Suzanne; Drasgow, Fritz (1999):* A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. In: *Journal for Applied Psychology* 5, 1999, S.754–775

*Schnell, Rainer; Kopp, Johannes (2001):* Zur Evaluation von Lehrveranstaltungen in den Sozialwissenschaften. In: *Soziologie* 3, 2001, S.32–40

*Shih, Tse-Hua; Fan Xitao (2007):* Response Rates and Mode Preferences in Web-Mail Mixed-Mode Surveys: A Meta-Analysis. In: *International Journal of Internet Science* 2, 2007, 1, S.59–82

*Simon, Anke (2012):* Informationsbroschüre OPEN Kerninformationen, [www.dhbw-stuttgart.de/pflegewissenschaften](http://www.dhbw-stuttgart.de/pflegewissenschaften) (Stand 10.12.2012)

*Simons, Julia; Pötschke Manuela (2006):* Akzeptanz internetgestützter Evaluation an Universitäten. In: *Zeitschrift für Evaluation*, 2,2006, S.227–248

*Spiel, Christiane; Martin Gössler (2001):* Zwischen Selbstzwecke und Qualitätsmanagement – Quo vadis, evaluatione? In: Spiel, Christiane (Hrsg.): *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck*. Münster, S.9–20

*Taddicken, Monika (2009a):* Die Bedeutung von Methodeneffekten der Online-Befragung: Zusammenhänge zwischen computervermittelter Kommunikation und erreichbarer Datengüte. In: Jakob, Nikolaus u.a. (Hrsg.): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden, S.91–143

*Taddicken, Monika (2009b):* Methodeneffekte von Web-Befragungen: Soziale Erwünschtheit vs. Soziale Entkontextualisierung. In: Weichbold, Martin u.a. (Hrsg.): *Umfrageforschung. Herausforderung und Grenzen*. Wiesbaden, S.85–104

*Zerback, Thomas; Schoen Harald; Jakob Nikolaus; Schlereth Stefanie (2009):* Zehn Jahre Sozialforschung in dem Internet – Eine Analyse zur Nutzung von Online-Befragungen in den Sozialwissenschaften. In: Jakob, Nikolaus u.a. (Hrsg.): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden, S.15–31

### Danksagung

Die Autoren bedanken sich an dieser Stelle insbesondere bei Herrn Dr. Budéus (leitender Verantwortlicher für Lehrevaluation an der DHBW Stuttgart) für die Bereitstellung der Daten aus Evasys. Ebenso geht ein herzlicher Dank an die Hochschulleitung der DHBW Stuttgart, Herrn Rektor Prof. Dr. Weber sowie Herrn Prorektor und Dekan Prof. Dr. Müllerschön, für die ideelle Unterstützung dieses Forschungsprojekts.

### **Anschrift der Verfasserinnen:**

Professor Dr. Anke Simon  
Duale Hochschule Baden-Württemberg Stuttgart  
Leiterin des Studiengangs BWL – Gesundheitswirtschaft und  
des Studiengangs Angewandte Gesundheitswissenschaften  
Fakultät Wirtschaft  
Paulinenstr. 50  
70178 Stuttgart  
E-Mail: [simon@dhbw-stuttgart.de](mailto:simon@dhbw-stuttgart.de)

Dr. Yvonne Zajontz  
Duale Hochschule Baden-Württemberg Stuttgart  
Leiterin des Zentrums für Empirische Forschung  
Fakultät Wirtschaft  
Paulinenstr. 50  
70178 Stuttgart  
E-Mail: [zajontz@dhbw-stuttgart.de](mailto:zajontz@dhbw-stuttgart.de)

Vanessa Reit (M.A.)  
Duale Hochschule Baden-Württemberg Stuttgart  
Leiterin des Zentrums für Empirische Forschung  
Fakultät Wirtschaft  
Paulinenstr. 50  
70178 Stuttgart  
E-Mail: [reit@dhbw-stuttgart.de](mailto:reit@dhbw-stuttgart.de)