

1 Einleitung

Probleme kann man niemals mit derselben
Denkweise lösen,
durch die sie entstanden sind.
(Albert Einstein)

Die Evaluation von Lehrveranstaltungen durch Studierende stellt ein mittlerweile in nahezu allen universitären Kursen, Seminaren und Vorlesungen verwendetes Mittel der Qualitätsoptimierung und -sicherung dar. Diese Methode wird in den USA bereits seit den 1920er Jahren (vgl. Marsh 1987:257) eingesetzt und untersucht und hat sich inzwischen auch im deutschsprachigen Raum als wahrscheinlich *die* Methode zur Verbesserung der Lehre etabliert. Hier fand sie ihren Ursprung als Folge der 68er-Bewegung und den damit verbundenen Reformen an den Hochschulen und »ergießt« sich seit Anfang der 1990er Jahre geradezu über die deutschsprachigen Universitäten, wie Rindermann (2009:32f) dies treffend formuliert.

Ihr Stellenwert im universitären Qualitätsmanagement wird auch durch ihre Verankerung in den jeweiligen Universitätsgesetzen untermauert, welche regelmäßige Evaluationen an deutschsprachigen Universitäten vor gut 20 Jahren zum fixen Bestandteil sämtlicher qualitätsverbessernder Bestrebungen machte (für Österreich siehe Kohler, 2009; für Deutschland siehe Schmidt, 2009 und für die Schweiz siehe Rhy, 2009). Zudem werden die mit Evaluationen verbundenen Ergebnisse vielfach nicht nur innerhalb der Universitäten, sondern auch von Geldgebern, PolitikerInnen und in den Medien diskutiert (siehe z.B. Der Standard 9. Mai 2007). Darüber hinaus hat mittlerweile sowohl die Anzahl der einzelnen Evaluationsmodelle als auch die Literatur hierzu ein Ausmaß erreicht, das – wie auch Mittag/Mutz/Daniel (2012:14) feststellen – »im Rahmen eines qualitativen Literaturreviews nicht mehr zu bewältigen ist«. Evaluation stellt demnach keine Modeerscheinung dar, sondern wird als wesentlicher Bestandteil der Qualitätsbestrebungen an Hochschulen gesehen.

Trotz dieser langen Historie, der Fülle an Publikationen zu diesem Thema und trotz ihres vermeintlichen Stellenwertes in der Hochschulpolitik wird die Lehrveranstaltungsevaluation dennoch nach wie vor von einigen kritisch betrachtet und ihre Effizienz im Hinblick auf die Optimierung der Lehre oftmals in Frage gestellt. Dies hat unterschiedliche Gründe und ist häufig auch darauf zurück-

zuföhren, dass viele Evaluationen mitunter relativ konzeptlos wirken bzw. sind und mehrheitlich nicht von EvaluationsexpertInnen konzipiert oder ohne klar ersichtliche theoretische Basis erstellt wurden, wie Spiel (vgl. 2001:7) feststellt.

Während die einen (siehe z.B. Alphei 2006:7) in Bezug auf Lehrveranstaltungsevaluation von einem der wichtigsten Verfahren zur Unterrichtsoptimierung sprechen, ist es für andere (siehe z.B. Liessmann 2005:15ff) höchst umstritten oder sogar eine Krankheit, »Evaluitis« (siehe z.B. Simon 2000, Frey 2007), die die Lehrveranstaltungen an Universitäten befällt und an welcher Studierende und Lehrende gleichsam leiden, ohne wirklich nachhaltige Verbesserungen zu bemerken. Und in der Tat ist gerade jener Bereich der Evaluationsforschung rar, der sich konkret mit den durch Evaluation ausgelösten vermeintlichen Verbesserungen oder Verschlechterungen der Lehrqualität befasst, wie Rindermann (vgl. 2009:227) feststellt. vgl. Spiel/Gössler (vgl. 2001:13) weisen zudem darauf hin, dass die Konzentration der Evaluation im universitären Kontext primär auf Analyse und Bewertung liegt und nicht auf konkreten Änderungsmaßnahmen. Außerdem wird das »Wie« meist breit diskutiert, während das »Wozu« fast immer ausgeklammert bleibt (vgl. *ibid.*).

Welchen Standpunkt man in diesem mitunter subjektiv und oftmals emotional geföhrt *Krieg der (Evaluations-)Welten* auch vertreten mag, es sind reichlich Argumente *für* und auch *gegen* das Durchföhren von Lehrveranstaltungsevaluationen vorhanden, die für die jeweiligen Evaluationsprojekte sorgfältig abgewogen werden müssen. Eine detaillierte Abhandlung hierzu würde den Rahmen dieses Buches sprengen, es sei u.a. jedoch auf Rindermann (2009), Kromrey (1996), McKeachie (1997), Simon (2000) und Frey (2007) verwiesen.

Während für mich die potentiellen Vorteile einer Evaluation in vielen Bereichen der universitären Qualitätsoptimierung, ganz besonders auch im Bereich des Fremdsprachenunterrichts, klar auf der Hand liegen, so teile ich gleichzeitig auch den kritischen Zugang zur Lehrveranstaltungsevaluation als *der* Methode zur Qualitätsoptimierung der Lehre, denn das Durchföhren einer Evaluation stellt per se noch lange keine hinreichende Bedingung für Qualität oder deren Optimierung dar, wie auch Schöch (vgl. 2005:152) feststellt. Mindestens genauso wichtig wie die Evaluation selbst ist das, was darauf folgt. Nur wenn die mit Hilfe der Evaluation gewonnen Informationen genutzt und umgesetzt werden, lohnt sich der damit zwangsläufig verbundene Mehraufwand.

Zwei Dinge müssen an dieser Stelle mit aller Deutlichkeit kommuniziert werden: Erstens, jede Evaluation ist mit einem gewissen (Mehr-)Aufwand verbunden und braucht daher Ressourcen. Zweitens, eine Evaluation ohne durchdachtes Follow-Up bewirkt keine Qualitätsoptimierung. Nur wenn man sich dieser Tatsachen bewusst ist und den mit einer Evaluation bzw. dem daran an-

schließenden Follow-up verbundenen Mehraufwand in Kauf nimmt, kann eine Qualitätsverbesserung erfolgen.

Evaluation wird im vorliegenden Ansatz – wenn richtig durchgeführt – als zentrale, vielschichtige und wirkungsvolle Methode zur Unterrichtsoptimierung gesehen, jedoch ist auch zu betonen, dass sie kein Allheilmittel darstellt und auch nicht alle Probleme und Defizite, die sie aufzeigt, im Rahmen des Hochschulkontextes beseitigt werden können. Vielfach fehlen hier die nötigen (monetären) Mittel und (zeitlichen) Ressourcen. Zudem muss darauf hingewiesen werden, dass nicht jede aktuell eingesetzte Evaluation jene Wirkung erzielt, die damit intendiert wird.

Evaluert man die an vielen Universitäten eingesetzten Evaluationsmodelle, kann man nicht selten feststellen, dass zahlreiche dieser Werkzeuge und Methoden in Hinblick auf deren Optimierungspotential oftmals fragwürdig sind und mitunter angezweifelt werden muss, ob damit tatsächlich eine wirksame Unterrichtsverbesserung möglich ist. Zentrale Probleme, die bei einer Metaevaluation¹ im Vorfeld detektiert wurden, sollen in Folge kurz diskutiert werden, bevor im anschließenden Abschnitt der in diesem Buch konzipierte Lösungsansatz vorgestellt wird.

1.1 Problemstellung

Die Motivation für diese Arbeit entstand primär aus einer Unzufriedenheit heraus, die ich im Laufe meiner mittlerweile über zehnjährigen Lehrtätigkeit an unterschiedlichen universitären Fremdsprachenzentren und Instituten im In- und Ausland mit den jeweils eingesetzten Lehrveranstaltungsevaluationen verspürte. Diese sind in der Regel mit einem hohen Aufwand verbunden und bringen oftmals aber nur einen kleinen Ertrag. Sie werden z.B. vielfach nach wie vor mit Hilfe von Papierfragebögen durchgeführt, in welchen die LernerInnen den Lehrpersonen am Ende des Semesters ein Feedback zum Kurs geben. Dadurch soll zur Qualitätsverbesserung der betreffenden Lehrveranstaltung beigetragen werden, was mit den jeweiligen Modellen meiner Erfahrung nach und nach Ansicht vieler meiner KollegInnen in den meisten Fällen jedoch nur eingeschränkt möglich ist.

Eine vollständige Auflistung sämtlicher Aspekte, die mit vielen dieser und anderer aktuell eingesetzter Evaluationsmodelle und Methoden einhergehen, wäre verhältnismäßig lang und zu umfangreich, um ihr an dieser Stelle die

1 Evaluation einer Evaluation

nötige Aufmerksamkeit zu schenken. Dies hängt zum einen damit zusammen, dass die Zahl an Evaluationsmodellen und Fragebögen, wie bereits angeführt, mittlerweile nahezu unüberschaubar ist, und zum anderen damit, dass auch die Kriterien, nach welchen man die einzelnen Modelle untersuchen und bewerten könnte, vielfältig und umfangreich sind. Daher wird hier von einer detaillierten Analyse Abstand genommen und es sollen nur einige zentrale Punkte aufgezeigt und kurz diskutiert werden, die das effektive Optimieren des universitären Fremdsprachenunterrichts mit Hilfe der untersuchten Modelle erschweren.

Um die Analyse auf einen für den Rahmen dieses Buches angemessenen Umfang zu beschränken, konzentrierte ich mich auf die Lehrveranstaltungsevaluationen jener fünf österreichischen Fremdsprachenzentren, die zum Untersuchungszeitpunkt (2012) dem *Verband universitärer Sprachenzentren und -institutionen (VUS)* angehören.

Obwohl es sich hierbei um eine vergleichsweise kleine Stichprobe handelt, kann dennoch festgestellt werden, dass sie sehr repräsentativ ist und die hier angeführten zentralen Problematiken auf die Mehrheit der Sprachenzentren im europäischen Hochschulraum – und wahrscheinlich auch darüber hinaus – in der einen oder anderen Weise zutreffen, da viele Sprachenzentren ähnliche Verfahren zur Qualitätsoptimierung einsetzen.

Eine Kollegin am *treffpunkt sprachen*, die in einem, der im Rahmen meiner Dissertation durchgeführten qualitativen Interviews befragt wurde, bringt einige der zentralen Themen aktueller Lehrveranstaltungsevaluationen im Fremdsprachenunterricht wie folgt auf den Punkt:

[...] weil ich mich sowieso seit einiger Zeit schon über diese ständigen Evaluierungen ärgere, nicht weil sie schlecht sind, sind sie nicht, aber es interessiert mich auch nicht zum hundertsten Mal zu hören, dass ich so ein nice teacher bin und so. Irgendwann finde ich, reicht es. *treffpunkt sprachen* muss natürlich für Qualität sorgen, soll mich von mir aus ein Jahr beobachten und dann finde ich, ist es aus und ich habe einfach keine Lust mehr. Ich mag mich nicht mehr evaluieren lassen.

[LP002, 224–230]

Die Inhalte dieser Aussage sind zweifelsfrei auch als einige der Hauptgründe für die von Simon (2000) und Frey (2007) angeführte »Evaluitis« zu sehen, an der viele KollegInnen, vor allem, wenn sie schon mehrere Jahre unterrichten, leiden. Die meisten Lehrenden sind nicht mit einer Evaluation ihres Unterrichts per se unzufrieden und negieren auch nicht deren Wichtigkeit im Hinblick auf ihr Potential, essentielle Informationen zur Qualitätsoptimierung zu generieren. Sie erkennen in der Regel auch die allgemeine Notwendigkeit von Evaluationen in Hinblick auf die Präsentation der betreffenden Institute und Zentren nach

außen an, stellen jedoch, wie die Kollegin aus dem Interview, vielfach fest, dass die *Art und Weise*, mit der diese Evaluationen durchgeführt werden, also der Evaluationsansatz, und die *Informationen*, die sie zum Teil fördern, insuffizient sind, was oftmals zur Folge hat, dass Lehrende mit der Zeit eine Aversion gegen das Evaluieren entwickeln. In der Tat ist diese Abneigung auch bei vielen Studierenden zu bemerken, denn es darf nicht vergessen werden, dass diese in der Regel *viele* Evaluationen pro Semester – in jedem der von ihnen besuchten Kurse zumindest eine – durchführen müssen. Hinzu kommen weitere Befragungen, die beispielsweise einen bestimmten Lehrgang oder die Universität und deren Einrichtungen etc. betreffen. Man kann also nicht behaupten, es würde zu wenig evaluiert.

Auf der Suche nach den Gründen, warum es in Hinblick auf Lehrveranstaltungsevaluationen überhaupt zu den oben genannten und anderen Problemen kommt, können unterschiedliche Ursachen ausfindig gemacht werden. Im Wesentlichen hängen sie jedoch damit zusammen, dass bei vielen Evaluationen oftmals gewisse Standards, wie sie etwa im *Handbuch der Evaluationsstandards* (siehe unten bzw. Sanders 2006) angeführt werden, keine oder eine zu geringe Berücksichtigung finden. Auch die *Standards und Leitlinien für die Qualitätssicherung im europäischen Hochschulraum (ESG)*, die im folgenden Kapitel vorgestellt werden und für die Qualitätsoptimierung bzw. -sicherung durch Evaluation zentral sind, werden in der Regel nicht in das Evaluationsprozedere integriert.

Bevor ich nun einige der zentralen Stellen aufzeige, an welchen die einzelnen Evaluationsmodelle Raum für Optimierung haben, möchte ich einen kurzen Exkurs zu den *Evaluationsstandards* machen, damit die darauffolgenden Ausführungen für den Leser / die Leserin leichter nachvollziehbar sind.

1.2 Evaluationsstandards

Damit sowohl die Anforderungen aus der Praxis als auch wissenschaftliche Standards und Richtlinien beim Durchführen von Evaluationen berücksichtigt bzw. die Qualität von Evaluationen optimiert und gesichert werden können, entwickelte das *Joint Committee on Standards for Educational Evaluation*² Richtlinien, die beim Planen, Durchführen und Bewerten von Evaluationen als Orientierung dienen.

2 Ähnliche Standards wurden auch von der Deutschen Gesellschaft für Evaluation kopiert (siehe DeGEval 2008).

Diese Qualitätsmaßstäbe sind im *Handbuch der Evaluationsstandards* festgehalten und richten sich sowohl an EvaluatorInnen als auch an AuftraggeberInnen. Sie sind nicht als starre mechanische Regeln, sondern als Leitprinzipien zu verstehen und enthalten wichtige Hinweise und Warnungen zur Vermeidung von Fehlern, die bei der Konzeption bzw. der Durchführung von Evaluationen auftreten können. Zudem bezeichnen sie, welche Vorgehensweisen allgemein als akzeptabel bzw. inakzeptabel gehalten werden und zeigen die zurzeit beste Praxis auf (vgl. Sanders 2006:35). Die aktuelle Version sieht folgende Standards für Evaluationen vor:

1.2.1 Inhalte der Evaluationsstandards

Nützlichkeitsstandards	Nützlichkeitsstandards legen fest, ob eine Evaluation den tatsächlichen Informationsbedürfnissen der jeweiligen AdressatInnen gerecht wird. Dazu werden die AdressatInnen und deren Evaluationsbedürfnisse genau ermittelt und die Evaluation demgemäß durchgeführt. Zudem müssen die gewonnenen Informationen bedeutsam sein und rechtzeitig ermittelt werden.
Durchführbarkeitsstandards	Durchführbarkeitsstandards betonen die Praxisbezogenheit von Evaluationen. Diese werden in der Regel nicht in einem Labor, sondern in einem natürlichen Umfeld durchgeführt und verbrauchen daher Ressourcen. Folglich müssen sie so konzipiert sein, dass sie nur so viel Material, Personal, Zeit etc. in Anspruch nehmen, wie erforderlich ist.
Korrektheitsstandards	Die Korrektheitsstandards fordern, dass Evaluationen rechtlich und ethisch korrekt durchgeführt und die Rechte (Privatsphäre, Zugänglichkeit zu Informationen, Schutz der Persönlichkeit etc.) der an Evaluationen beteiligten Personen gewahrt und respektiert werden.
Genauigkeitsstandards	Die Genauigkeitsstandards legen fest, ob eine Evaluation angemessene Informationen hervorbringt. Es sollen die relevanten und als wichtig erachteten Daten erhoben und beurteilt werden. Die Daten sollen technisch angemessen sein und die Urteile müssen in einem logischen Zusammenhang mit den Daten stehen. Damit wird die Güte und Verwendbarkeit der Informationen bestimmt.

Tab. 1: Inhalte der Evaluationsstandards in Anlehnung an Sanders (vgl. 2006:31f)

Natürlich muss an dieser Stelle darauf hingewiesen werden, dass diese Standards ursprünglich in den USA entwickelt wurden und eine Eins-zu-eins-Übertragung nicht für alle Länder und alle Situationen möglich und sinnvoll ist. Es ist ratsam, zuerst den Evaluationskontext zu analysieren und dann eventuell nötige Adaptionen durchzuführen. Insgesamt sind diese potentiellen Anpassungen jedoch,

wie auch Beywl/Widmer (2006:261) festhalten, »eher geringfügiger Art«, weswegen sich die Evaluationsstandards als Basis für die meisten Evaluationen eignen. Dies trifft vor allem auf die Grundstruktur mit den vier genannten Hauptgruppen (Nützlichkeits-, Durchführbarkeits-, Korrektheits- und Genauigkeitsstandards) zu.

Im Bereich des universitären Fremdsprachenunterrichts können diese Standards meiner Erfahrung nach weitgehend unverändert übernommen werden, wengleich der oben angesprochene Kontext durch die Ausführungen in den Theoriekapiteln (Kapitel 3, 4, 5) zuerst genau analysiert und bestimmt werden sollte. Dies ist nötig, um festzustellen, welche der einzelnen Standards für die Evaluation von Fremdsprachenkursen, wie sie hier vorgestellt wird, in welcher Form und in welchem Umfang sinnvoll und plausibel sind. Neben diesen bereits erwähnten sind auch noch weitere Standards wie z. B. die genannten ESG sowie andere Qualitäts- und Bildungsstandards zu beachten.

1.2.2 Evaluationsstandards im Detail

Nützlichkeitsstandards	<ol style="list-style-type: none"> 1. Ermittlung der Beteiligten und Betroffenen 2. Glaubwürdigkeit der EvaluatorInnen 3. Umfang und Auswahl der Informationen 4. Feststellung von Werten 5. Klarheit des Berichts 6. Rechtzeitigkeit der Verbreitung des Berichts 7. Wirkung der Evaluation
Durchführbarkeitsstandards	<ol style="list-style-type: none"> 1. Praktische Verfahren 2. Politische Tragfähigkeit 3. Kostenwirksamkeit
Korrekttheitsstandards	<ol style="list-style-type: none"> 1. Unterstützung der Dienstleistungsorientierung 2. Formale Vereinbarungen 3. Schutz individueller Menschenrechte 4. Human gestaltete Interaktion 5. Vollständige und faire Einschätzung 6. Offenlegung der Ergebnisse 7. Deklaration von Interessenskonflikten 8. Finanzielle Verantwortlichkeit
Genauigkeitsstandards	<ol style="list-style-type: none"> 1. Programmdokumentation 2. Kontextanalyse 3. Beschreibung von Zielen und Vorgehensweisen 4. Verlässliche Informationsquellen 5. Valide Informationen 6. Reliable Informationen 7. Systematische Informationsüberprüfung

	8. Analyse quantitativer Informationen 9. Analyse qualitativer Informationen 10. Begründete Schlussfolgerungen 11. Unparteiische Berichterstattung 12. Meta-Evaluation
--	--

Tab. 2: Evaluationsstandards im Detail (vgl. Sanders 2006)

Um die im folgenden Abschnitt diskutierten Probleme bei den untersuchten Lehrveranstaltungsevaluationen beim Entwurf des in diesem Band vorgestellten Modells weitgehend zu vermeiden, werden sämtliche der hier genannten Standards bei dessen Konzeption berücksichtigt.

1.3 Optimierungspotential bei Evaluationsmodellen

Um nun wieder auf jene Aspekte zurückzukommen, die bei vielen Lehrveranstaltungsevaluationen optimiert werden können, werden, ausgehend von den *Evaluationsstandards*, in Folge einige zentrale Punkte angeführt, die vor allem mit den *Nützlichkeitsstandards*, *Durchführbarkeitsstandards* und *Genauigkeitsstandards* in Konflikt stehen. Die *Korrektheitsstandards* wurden in den analysierten Lehrveranstaltungsevaluationen meiner Ansicht nach nicht verletzt, da alle, soweit von außen nachvollziehbar, zum Zeitpunkt der Analyse rechtlich und ethisch korrekt durchgeführt wurden.

Als ein Hauptkriterium für das erfolgreiche Durchführen von Evaluationen wird im *Handbuch der Evaluationsstandards* unter den *Nützlichkeitsstandards* festgehalten, dass Evaluationen »informativ, zeitgerecht und wirksam« sein sollen (vgl. Sanders 2006:31). Dies ist eine zentrale Forderung, denn mit jeglicher Evaluation stehen gewisse Ziele in Verbindung. Werden Nützlichkeitsstandards nicht erfüllt, sind die Ergebnisse der Evaluation für bestimmte Absichten, wie etwa die Qualitätsoptimierung des Fremdsprachenunterrichts, kaum relevant und der Aufwand steht in keiner vertretbaren Relation zum Informationsgehalt.

Wenngleich der Informationsgehalt, den die untersuchten Evaluationsfragebögen generieren, den Lehrenden bei der Optimierung der Lehre in der einen oder anderen Form durchaus behilflich sein könnte, so ist bei den analysierten Modellen der Zeitpunkt sehr ungünstig, zu welchem die Lehrenden an die jeweiligen Informationen gelangen. Ausnahmslos alle analysierten Evaluationen wurden zum Untersuchungszeitpunkt in Form einer summativen Evaluation am Ende des Kurses / Semesters durchgeführt.

Summative Evaluationen werden im Gegensatz zu *formativen* Evaluationen nach der Durchführung eines Programms (z.B. Fremdsprachenkurs) oder einzelner Maßnahmen eingesetzt und sollen zusammenfassende Aussagen über deren Wirksamkeit tätigen (vgl. Gollwitzer/Jäger 2009:16). Das bedeutet, sie sollen in Erfahrung bringen, wie wirksam bestimmte Maßnahmen/Programme waren. Das Ziel *formativer* Evaluationen hingegen besteht darin, Maßnahmen zu optimieren bzw. Rahmenbedingungen zu schaffen, die die Wirksamkeit einer Maßnahme eher wahrscheinlich machen (vgl. *ibid.*). Bei formativen Evaluationen werden somit in regelmäßigen Intervallen Zwischenergebnisse erstellt, die das Ziel verfolgen, laufende Interventionen zu modifizieren oder zu verbessern (vgl. Bortz/Döring 2006:110).

Im konkreten Fall der Lehrveranstaltungsevaluationen bedeutet dies, dass die betreffenden Lehrpersonen mit den eingesetzten Methoden jene Informationen, die ihnen beim Optimieren des Unterrichts behilflich sein könnten, zu einem Zeitpunkt erhalten, zu welchem es ihnen nicht mehr möglich ist, auf diese Informationen einzugehen und eventuell notwendige Veränderungen durchzuführen, die potentielle Verbesserungen für den jeweiligen Kurs nach sich ziehen könnten, ein Suboptimum, welches u. a. auch von Stieger/Burger (vgl. 2010:163) in einer ähnlichen Analyse kritisiert wird. Wenngleich nicht negiert werden kann, dass einige Informationen durchaus auch für kommende Semester bzw. für die Optimierung der Lehre einzelner Lehrpersonen generell von Relevanz sein können, so ist die Unterrichtsoptimierung für die betreffende Gruppe, die an der Evaluation teilnahm, nicht mehr möglich und es stellt sich die Frage, ob durch diese Herangehensweise nicht das eigentliche Ziel verfehlt wird, welches man mit der Evaluation initial intendierte.

Ein weiterer Grund, warum Evaluationen nicht ihr volles Wirkungspotential ausschöpfen können, wenn sie ausschließlich am Ende des Kurses/Semesters durchgeführt werden, ist, dass die Lehrenden auch nicht mehr direkt zum Handeln angehalten werden, da sie ja im darauffolgenden Semester oftmals mit einer völlig neuen Gruppe konfrontiert sind und eventuell auch bei gleichen Unterrichtsmethoden unterschiedliche Evaluationsergebnisse erwarten können. Dadurch ist fraglich, wie sinnvoll die Lehrenden diese Art der Evaluation einstufen und wie ernst sie die Ergebnisse nehmen bzw. wie motiviert sie sind, Dinge zu verändern. Dies trifft natürlich umgekehrt auch auf die Studierenden zu. Mit welcher Ernsthaftigkeit werden sie die Fragebögen ausfüllen, wenn sich für sie keine unmittelbaren Verbesserungen mehr abzeichnen? Vielfach resultiert diese Methode daher rein in einer Bewertung, die wenige bis keine Konsequenzen in Form einer Optimierung nach sich zieht.

Ein dritter Grund, warum sich die Wirkung der analysierten Evaluationen in einigen Bereichen in Grenzen hält, ist vielfach der Aufbau der Fragebögen selbst, der zwar für eine statistische Auswertung hervorragend geeignet ist, jedoch genau jene Informationen nicht erhebt, die für die Lehrenden in Hinblick auf eine weitreichende Verbesserung der Lehre interessant und notwendig wären. Die in den Bögen gestellten Fragen analysieren vielfach einen von den Studierenden *empfundenen Zustand*, regen die LernerInnen jedoch nicht dazu an, sich an der Verbesserung aktiv zu beteiligen bzw. konkrete Informationen Preis zu geben oder hilfreiche Vorschläge zu unterbreiten, die zu einer Optimierung führen würden, wie dies auch in den ESG explizit vorgeschlagen wird.

Ein Beispiel stellt etwa folgende Frage aus einem der untersuchten Fragebögen dar: *Hat der/die Lehrende die Lern- und Prüfungsziele klar dargelegt?* Abgesehen von der Tatsache, dass das Explizieren der Lern- und Prüfungsziele, wenn überhaupt dezidiert, dann zu Beginn des Semesters erfolgt, kann bezweifelt werden, dass sich alle LernerInnen mehrere Monate danach noch im Detail daran erinnern können, *was* genau die Lehrperson *wie* mitteilte. Des Weiteren waren u.U. in der Lehrveranstaltungseinheit, in welcher die Lern- und Prüfungsziele expliziert wurden, nicht alle LernerInnen anwesend und werden nun dennoch zu einer Antwort *gezwungen*, da alle Fragen zu beantworten sind, damit die Evaluation gültig ist. Darüber hinaus gibt diese geschlossene Frage lediglich Auskunft darüber, wie einzelne LernerInnen diesen Sachverhalt in der jeweiligen Gruppe *empfanden*, wie die Analyse der Fragebögen zeigt. Während es für manche auf der Likert-Skala *nicht klar* war, war es für andere LernerInnen *sehr klar*. Sie offeriert der Lehrperson jedoch keine Anhaltspunkte darüber, wo genau die Probleme lagen, wenn diese Frage etwa mit *nicht klar* beantwortet wurde. Auch bietet der Fragebogen den Studierenden keine Möglichkeit, Vorschläge zu unterbreiten, die für sie eine Verbesserung der Situation bewirken würden. Das bedeutet, die Lehrperson weiß zwar Bescheid darüber, *dass* für einige LernerInnen in diesem Bereich ein Problem war bzw. dass einige dies rückwirkend so empfanden, sie kann jedoch aufgrund fehlender Information, *wo* genau das Problem lag und *wie* dies für die LernerInnen gelöst werden könnte, verhältnismäßig wenig verändern. Diese Tatsache bewahrheitet sich bei sämtlichen Fragen in allen untersuchten Fragebögen.

Man könnte dies nun mit der Sektion der *offenen Fragen* am Ende des Fragebogens rechtfertigen, wo die LernerInnen angeben können, was ihnen am Kurs gut/weniger gut gefallen hat. Jedoch zeigte eine Analyse der Lehrveranstaltungsevaluationen, die von mir in den letzten sechs Jahren in den eigenen Kursen durchgeführt wurden, dass diese Sektion von den Studierenden kaum bis nicht genutzt wird, um davor getätigte Angaben zu verdeutlichen oder um

Verbesserungsvorschläge zu unterbreiten. Es bleibt zu bezweifeln, dass sich dieser Sachverhalt in anderen Kursen bzw. an anderen Instituten/Zentren in dieser Hinsicht markant unterscheidet.

Darüber hinaus ist auch das Stellen der betreffenden Frage nach der Darlegung der Prüfungsziele an sich problematisch. Einerseits, weil diese Art der Fragestellung generell sehr differenzierte Ergebnisse fördert, die in der Praxis schwer nutzbar gemacht werden können, ein Phänomen, auf welches bereits Kromrey (vgl. 1996) hinweist, und andererseits, weil das Verstehen der Erklärung der Lehr- und Prüfungsziele zu sehr auf die Lehrperson zentriert wird. Dies bedeutet im konkreten Kontext bei ersterem, dass einige LernerInnen die betreffende Frage positiv und andere negativ beantworten werden, wie dies oben bereits angeführt wurde, und diese Inkonsistenz nicht nur *zwischen* verschiedenen Lehrveranstaltungen zu bemerken ist, sondern ebenso *innerhalb jeder* (oder fast jeder) Veranstaltung, die eine bestimmte TeilnehmerInnenzahl von ca. 20 überschreitet, wie Kromrey (vgl. *ibid.*) expliziert.

Dieser Umstand wurde auch in meinen eigenen Lehrveranstaltungen mehrfach beobachtet, wenn von mir exakt der gleiche Unterrichtsstoff anhand identer Lehrmethoden in Parallelgruppen vermittelt, aber von diesen unterschiedlich aufgenommen bzw. evaluiert wurde. Das bestätigt die Annahme, dass es folglich nicht nur an der Art und Weise liegen kann, wie die jeweilige Lehrperson Inhalte vermittelt, sondern auch daran, wie die LernerInnen diese aufnehmen und verarbeiten. Dies wiederum verdeutlicht die unterschiedliche Auffassung der Studierenden über *gutes* Lehrverhalten und man sieht sich als Lehrperson mit der Tatsache konfrontiert, dass, wenn man verstärkt auf Vorschläge einer Interessensgruppe eingeht, bei dieser *Pluspunkte* sammelt, während man u. U. bei anderen mit derselben Methode für Unmut sorgt – eine Feststellung, auf die auch eine Lehrende³ in den im Rahmen dieser Arbeit durchgeführten qualitativen Interviews hinwies:

Weil da ist ein Mensch, schreibt das und will über Bolivien reden, aber das ist wirklich ... so extrem individuell. Und das ist eben wo manchmal die Evaluierungen vom Ende vom Kurs ... ist ein bisschen dieses ... gleicher Punkt, komplett auseinander. [...] Also komplett die Meinung. Ich hätte lieber mehr davon, ich hätte lieber mehr davon. Und der andere sagt, ich hätte lieber weniger davon. Und es ist schwer irgendwie ... [LP007: 230–236]

3 Die betreffende Lehrperson ist keine Muttersprachlerin in Deutsch, das Gesagte wurde unverändert transkribiert.