

Landwehr, Norbert

Lehrevaluation als Anstoss zur Unterrichtsentwicklung

Beiträge zur Lehrerbildung 23 (2005) 3, S. 321-333



Quellenangabe/ Reference:

Landwehr, Norbert: Lehrevaluation als Anstoss zur Unterrichtsentwicklung - In: Beiträge zur Lehrerbildung 23 (2005) 3, S. 321-333 - URN: urn:nbn:de:0111-pedocs-135776 - DOI: 10.25656/01:13577

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-135776>

<https://doi.org/10.25656/01:13577>

in Kooperation mit / in cooperation with:

Zeitschrift zu Theorie und Praxis der Aus- und
Weiterbildung von Lehrerinnen und Lehrern

**BEITRÄGE ZUR LEHRERINNEN-
UND LEHRERBILDUNG**

Organ der Schweizerischen Gesellschaft für
Lehrerinnen- und Lehrerbildung (SGL)

ISSN 2296-9632

<http://www.bzl-online.ch>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Lehrevaluation als Anstoss zur Unterrichtsentwicklung

Norbert Landwehr

An vielen Hochschulen und Fachhochschulen bildet die Lehrevaluation einen festen Bestandteil des internen Qualitätsmanagements. Im folgenden Artikel wird diese Thematik aufgegriffen: Nach einer Klärung zur Funktion und zu den inhaltlichen Schwerpunkten von Lehrevaluationen wird eine kurze Bewertung von fragebogengestützten Evaluationsverfahren vorgenommen – mit speziellem Blick auf die im Hochschulunterricht sehr verbreiteten quantitativen Evaluationsmethoden. Im Zentrum des Artikels steht dann aber eine eher unbekanntere Evaluationsform, welche quantitative und qualitative Evaluationselemente miteinander zu verbinden sucht: die «Ratingkonferenz». Zu diesem Evaluationsverfahren werden die einzelnen Verlaufsschritte dargestellt und Hinweise zur praktischen Umsetzung gegeben. Zudem werden Chancen und Schwierigkeiten beschrieben, die mit dem Einsatz dieses Verfahrens der Lehrevaluation verbunden sind.

1. Einleitung

Die Evaluation der Lehre gehört heute zu den selbstverständlichen Ansprüchen an eine professionelle Gestaltung von Lehr-Lernveranstaltungen an einer Fachhochschule: Sie gilt zu Recht als eine der wichtigsten Grundlagen für die Qualitätsentwicklung der eigenen Unterrichtspraxis. Mag sein, dass dieser Anspruch für viele Dozierende schon immer eine Selbstverständlichkeit war. Was sich aber in den letzten Jahren unter der Vorgabe eines systematischen Qualitätsmanagements deutlich geändert hat, sind die Ansprüche an das Instrumentarium, das zu diesem Zweck eingesetzt wird. Waren es früher in erster Linie die regelmässigen Überprüfungen des Lernerfolges durch Lernkontrollen sowie die intuitive, möglichst sensible Wahrnehmung der Reaktionen, welche die Studierenden – oft durch entsprechende Rückfragen und informelle Gesprächsanlässe – gestützt auf den erlebten Unterricht zum Ausdruck gebracht haben, ist gerade das Ungenügen dieser beiden traditionellen Evaluationsformen der Hauptgrund für die Forderung nach einer neuen Ausrichtung der unterrichtsbezogenen Evaluationspraxis:

Lernkontrollen und Prüfungen – vor allem wenn sie klassen- oder kursintern gestaltet werden – haben bezüglich ihrer Aussagekraft für die Prozessqualität der Lehre eine sehr hohe Interpretationsbeliebigkeit: Prüfungsergebnisse lassen bekanntlich offen, wie das gute oder schlechte Abschneiden der Studierenden zustande kommt. Es bleibt in der Beliebigkeit der einzelnen Lehrperson, die Gründe für den jeweiligen Lernerfolg bei sich bzw. bei der eigenen Unterrichtsqualität oder aber beim Lernenden zu attribuieren. (Der Qualifizierungs- und Selektionsauftrag, der mit der gängigen Prüfungspraxis

verbunden ist, führt verständlicherweise zur Tendenz, die Lerner-Attribuierung in den Vordergrund zu stellen!)

Die intuitive Beobachtung/Erfassung der Studierendenreaktionen – die zweit genannte Evaluationsform – ist ebenfalls eine wenig zuverlässige Evaluationsgrundlage: Die Wahrnehmungspsychologie hat nachgewiesen, dass verschiedene Formen von Verzerrungen die Zuverlässigkeit der eigenen Wahrnehmung stark einschränken und dass die subjektive Absicht, sich nach bestem Wissen und Gewissen auf eine möglichst unverfälschte (objektive) Wahrnehmung einzulassen, keinesfalls ausreicht, um die Quellen der möglichen Wahrnehmungsverfälschungen auszuschliessen und um die Qualitätsbeurteilung der eigenen Praxis auf einer objektiven Grundlage durchzuführen. Wir benötigen daher Instrumente, welche den meist unbewusst wirksamen Verzerrungsquellen und -mechanismen entgegenwirken. Wenn wir heute im Rahmen eines systematischen und kritisch reflektierten Qualitätsmanagements von «Lehrevaluation» sprechen, findet dies im Bewusstsein statt, dass die Qualitätsbeurteilung von Lehre eine sehr anspruchsvolle Angelegenheit ist, bei der es darum geht, die Qualitätsbeurteilung der Lehre auf eine objektivere Basis zu stellen.

2. Klärungen

Ich möchte zunächst auf zwei Ebenen eine begriffliche Differenzierung vornehmen, die mir als notwendig erscheint, um in die aktuelle Diskussion um den Sinn und den Nutzen der Lehrevaluation mehr Klarheit zu bringen.

2.1 Zur Funktion der Lehrevaluation

Zunächst ist es hilfreich, zwei Grundfunktionen der Evaluation zu unterscheiden: die Funktion der individuellen, selbstverantwortlichen Qualitätsverbesserung (Entwicklungsfunktion) und die Funktion des Qualitätsnachweises gegenüber einer institutionell übergeordneten Instanz (Rechenschaftsfunktion). Mit Blick auf diese beiden Grundfunktionen kann es im Rahmen des Qualitätsmanagement einer Schule/Hochschule aus zwei unterschiedlichen Gründen angezeigt sein, eine regelmässige Evaluation der Lehre einzuführen:

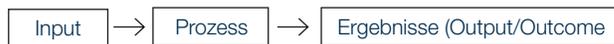
- *Entwicklungsfunktion*: Die *Lehrpersonen* selber wünschen differenzierte Informationen zur (eigenen) Unterrichtsqualität, weil sie Einsicht gewinnen möchten in mögliche Diskrepanzen zwischen eigenen Qualitätsansprüchen und der von den Studierenden wahrgenommenen/erlebten Unterrichtswirklichkeit. Sie möchten mehr erfahren über praxisrelevante Gelingensbedingungen des eigenen Unterrichts, um mit Hilfe dieser Informationsgrundlage den eigenen Unterricht schrittweise zu verbessern.
- *Rechenschaftsfunktion*: Die *Schulleitung/Ausbildungsleitung* wünscht einen Einblick in die Qualität der einzelnen Lehr-Lernangebote, die in ihrem Verantwortungsbereich liegen – beispielsweise im Sinne der Frage: «Erfüllt die betreffende Lehrveranstaltung die geltenden Qualitätsstandards?» Eine Evaluation mit dieser

Ausrichtung dient der Schul- und Ausbildungsleitung dazu, die Personenqualifikation mit der notwendigen datengestützten Fundierung vornehmen zu können; andererseits wird damit die Möglichkeit eröffnet, bei sich abzeichnenden Qualitätsdefiziten rechtzeitig geeignete Massnahmen anzuordnen bzw. mit den betreffenden Personen zu vereinbaren.

Auch wenn sich diese beiden Grundfunktionen nicht zu widersprechen brauchen, hat es sich doch als hilfreich erwiesen, jeweils Klarheit über die Ausrichtung der Evaluation und der eingesetzten Evaluationsinstrumente zu gewinnen: Dies insbesondere deshalb, weil eine rechenschaftsorientierte Evaluation bei den betroffenen Lehrpersonen Ängste und Befürchtungen wecken kann, die für eine offene und lernbereite Auseinandersetzung mit den eigenen Stärken und Schwächen, d. h. für den eigenen unterrichtsbezogenen Lern- und Entwicklungsprozess, hinderlich sein können.

2.2 Zum inhaltlichen Schwerpunkt der Lehrevaluation (Evaluationsfokus)

Eine zweite Klärung betrifft die inhaltlichen Schwerpunkte (Fokus) der Evaluation. In diesem Zusammenhang kann es hilfreich sein, ein einfaches, aus der Analyse von Produktions- und Transformationsprozessen gewonnenes Praxismodell zu verwenden, das die folgenden drei Grundkomponenten unterscheidet:



Gemäss diesem Modell kann der Fokus einer Evaluation grundsätzlich drei unterschiedliche Ausrichtungen haben:

- Die Evaluation kann die *Inputvariablen* betreffen: d. h. die von der Institution vorgegebenen Rahmenbedingungen, innerhalb derer der Unterricht jeweils stattfindet.
- Die Evaluation kann die Variablen der *Prozessgestaltung* in den Vordergrund rücken: z. B. die Strukturierung des Lehrstoffes, den Aufbau der Unterrichtsarrangements, die Gestaltung der Kommunikation mit den Studierenden usw.
- Die Evaluation kann auf die erreichten *Ergebnisse*, d. h. auf die kurz- und langfristigen Auswirkungen des Lehr-Lernprozesses wie Lernerfolg, Sozialisierungseffekte, Zufriedenheit, Auswirkung auf das künftige Lerninteresse usw., ausgerichtet sein.

Von einer umfassenden Praxisbeurteilung wird erwartet, dass alle drei Komponenten in die Evaluation einbezogen werden. Dies würde allerdings voraussetzen, dass für alle drei Komponenten die notwendigen Bewertungsgrundlagen (Standards, Kriterien) in der notwendigen Operationalisierung vorliegen müssen. Vor einer Evaluation sollte also jeweils geklärt werden, wie der Evaluationsfokus definiert ist. Diese Klärung ist nicht zuletzt deshalb wichtig, weil je nach gewähltem Evaluationsschwerpunkt auch unterschiedliche Zuständigkeitsinstanzen betroffen sein können. So liegt beispielsweise die Qualitätsverantwortung für den Inputbereich in der Regel eher bei der Institution, für den Bereich der Praxisgestaltung dagegen eher bei der einzelnen Lehrperson. Dies bedeutet, dass auch der Adressat der Evaluationsergebnisse entsprechend festgelegt werden muss.

3. Fragebogengestützte Evaluationsverfahren

Bezüglich der methodischen und instrumentellen Gestaltung des Evaluationsprozesses ist es ebenfalls sinnvoll, eine Differenzierung vorzunehmen. Hier lassen sich unterscheiden: schriftliche Fragebogen, Interviews (mündliche Befragungen), Beobachtungen, Dokumentenanalysen, Leistungstests. In unserem Zusammenhang interessieren vor allem die *fragebogengestützten Evaluationsverfahren*, wobei üblicherweise zwischen qualitativen und quantitativen Verfahren unterschieden wird.

3.1 Quantitative und qualitative Evaluationsverfahren

Die *qualitativen Evaluationsverfahren* arbeiten mit Instrumenten, welche einen grossen Spielraum zulassen für die individuell unterschiedlichen Darstellungen / Formulierungen der Wirklichkeitswahrnehmung durch die befragten Personen selber. Mit anderen Worten: Qualitative Verfahren verwenden so genannte «offene Fragen», die zu freien Antworten, Beschreibungen und Eigenschaftsaufzählungen durch die Befragten führen. Die Antworten werden dann in einem nachfolgenden Verarbeitungsschritt bestimmten Kategorien zugeordnet oder auf ihre inhärente «Konstruktionslogik» hin analysiert.

Qualitative (offene) Verfahren werden in der Evaluation von Lehr-Lernveranstaltungen eingesetzt, wenn man herausfinden möchte, welche unterschiedlichen «Wirklichkeitskonstruktionen», welche unterschiedlichen Wahrnehmungs- und Deutungsmuster zum Unterricht innerhalb einer Studierendengruppe überhaupt existieren. Die Frage nach der quantitativen Verteilung der Antworten (d.h. die Häufigkeit einer Aussage) steht dabei klar im Hintergrund; sie dient allenfalls als zusätzliche Gewichtungsinformation (die allerdings mit gewissen Zufälligkeiten behaftet bleibt).

In *quantitativen Evaluationsverfahren* finden die Befragungen mit Fragebogen statt, die mit vorgegebenen Antwortalternativen arbeiten. Mit Blick auf die vordefinierten Antworten wird auch von «geschlossenen» Fragen gesprochen. Die befragten Personen werden dazu aufgefordert, die für sie zutreffenden Antworten auszuwählen oder mit einem Zahlenwert – gemäss der subjektiv eingeschätzten Zustimmung oder Bedeutsamkeit – zu versehen. In diesem Sinne liefern quantitative Verfahren konkrete Zahlen bzw. Messwerte, mit denen man rechnen kann – die Resultate sind also quantifizierbar (Anzahl, Ausmass) und statistisch auswertbar.

Bei quantitativen (geschlossenen) Verfahren werden bestimmte Wahrnehmungs- und Deutungsmuster zu einem bestimmten Sachverhalt durch den Evaluator/die Evaluatorin vorweg definiert, um herauszufinden, wie deren quantitative Verteilung in einer bestimmten Population aussieht. Es wird hier davon ausgegangen, dass die bedeutsamen Varianten an Wahrnehmungen und Deutungen inhaltlich bereits hinreichend bekannt sind: Es soll in Erfahrung gebracht werden, wie die vordefinierten Wahrnehmungs- und Deutungsmuster in einer Kohorte (z.B. innerhalb einer Kursgruppe) quantitativ verteilt sind. Fragebogen mit vorgegebenen Antwort-Items können für die Lehrevaluation

vor allem dann hilfreich sein, wenn das Interesse auf die quantitative Verteilung von bestimmten Wahrnehmungs-, Deutungs- und Urteilkategorien innerhalb der Studierenden gerichtet ist.

3.2 Chancen und Gefahren von quantitativen Lehrevaluationen

Der Einsatz von Fragebogen – insbesondere in der quantitativen/geschlossenen Form – ist gegenwärtig hoch im Kurs; auch im Bereich der Lehrevaluation kommen quantitative Evaluationsinstrumente immer häufiger zur Anwendung. Ich gehe davon aus, dass das praxisbezogene Wissen zur Erstellung von Fragebogen, zur Datenerfassung, zur Auswertung und Interpretation der Daten, zum Datenfeedback an die Studierenden als bekannt vorausgesetzt werden darf. Andernfalls kann hier auf die reichhaltige, z. T. sehr praxisnahe Literatur verwiesen werden (z. B. Burkhard & Eikenbusch, 2000; Konrad, 2005; Landwehr u. a., 2003; Schratz u. a., 2000)¹. An dieser Stelle möchte ich mich darauf beschränken, auf ein paar wichtige Chancen und Gefahren dieses Evaluationsverfahrens hinzuweisen.

(a) Chancen von quantitativen Lehrevaluationen

- Quantitative Befragungen führen zu Ergebnissen, die einerseits eindeutig sind (Zahlenwerte!), und die andererseits als «objektiv» gelten. Nicht zuletzt auf Grund einer grossen (z.T. naiven) Zahlengläubigkeit und eines unkritischen Objektivitätsverständnisses haben quantitative Daten oft eine erstaunlich provokative Wirkung: Sie regen zu intensiven Auseinandersetzungen über die Qualität an – selbst bei Personen, welche die Ergebnisse anzweifeln. Die Qualitätsdiskussion wird dadurch verbindlicher.
- Quantitative Ergebnisse sind hilfreich, um die verschiedenen Auffassungen zu einem Sachverhalt im Meinungsspektrum der betreffenden Gruppe zu gewichten. Minderheiten und Mehrheiten werden erkennbar. In quantitativen Evaluationen erhält die Auffassung von lautstarken Meinungsführern dasselbe Gewicht wie die Meinung der zurückhaltenden Kursteilnehmerinnen und -teilnehmer.
- Die Anonymität der Antworten, wie sie in quantitativen Fragebogen-Erhebungen möglich ist, erleichtert es den Teilnehmerinnen und Teilnehmern, die eigene Meinung offen zum Ausdruck zu bringen (ohne Rücksicht auf soziale Gefälligkeit). Dadurch werden die Ergebnisse weniger verfälscht, d. h. glaubwürdiger.
- Quantitative Befragungen ermöglichen es, die Verteilung der Auffassungen/Bewertungen in verschiedenen Gruppierungen miteinander zu vergleichen bzw. die Abweichungen von einem «Normwert» festzustellen. Dadurch erhält die Qualitätsdiagnose einen zusätzlichen Aussagewert und die Qualitätsdiskussion einen zusätzlichen Anregungsimpuls.

¹ Beispiele für Evaluationsinstrumente aus dem Bereich der Hochschule finden sich unter: www.evaluation.unizh.ch/verfahren/dokumente; Musterfragebogen www.vfh-hessen.de/Spectrum/98-1-2.htm; Fragebogen zur Evaluation der Lehre: www.fbw.hs-bremen.de/changem/frames-evaluation.htm; Formular für den Fragebogen: Ein wissenschaftlich gut untermauertes Instrument ist das «Heidelberger Inventar zur Lehrveranstaltungs-evaluation», publiziert in Rindermann (2001).

- Bei qualitativen Befragungen führt die fehlende Information bezüglich der quantitativen Verteilung der jeweiligen Meinungsäußerung dazu, dass die Aussagegewichtungen nach eigenem Gutdünken oder nach intuitiver Erfassung der Gesamtstimmung vorgenommen werden kann: Aussagen der Studierenden werden beispielsweise als relevant eingestuft, wenn sie dem eigenen Konzept gut entsprechen oder wenn sie von einer Person geäußert wurden, die als «signifikanter Meinungsträger» eingestuft wird. Bei einer quantitativen Befragung wird dagegen eine Gewichtungsinformation mitgeliefert, die für die Dateninterpretation hilfreich sein kann.
- Ergebnisse von Fragebogenuntersuchungen mit ihrer Quantifizierung machen es möglich, den Ist-Zustand über einen längeren Zeitraum zu erfassen und durch den Datenvergleich den (positiven oder negativen) Entwicklungstrend aufzuzeigen.

(b) Gefahren von quantitativen Lehrevaluationen

Trotz der klaren Vorzüge lauern bei der praktischen Umsetzung von quantitativen Evaluationen verschiedene Gefahren, welche die genannten Chancen torpedieren können.

- Unterrichtsevaluationen, die sich auf quantitative Befragungen abstützen, können dazu verführen, die erhobenen Zahlen zu verabsolutieren. Es können inadäquate Schlussfolgerungen gezogen werden, weil – z. B. auf Grund einer naiven Zahlenfaszination – den qualitativen Informationen zuwenig Beachtung geschenkt wird. Zudem besteht die Gefahr, dass die Quantifizierbarkeit zum bestimmenden Faktor dafür wird, welche Wirklichkeitsaspekte in die Evaluation (und in den anschließenden Optimierungsprozess) einbezogen werden.
- Unterrichtsevaluationen nehmen – explizit oder implizit – immer Bezug auf Vorstellungen von gutem Unterricht. Wird diese normative Basis im Vorfeld der Evaluation (vor der Fragebogenerstellung!) nicht genügend geklärt – z. B. über ein Qualitätsleitbild –, so besteht die Gefahr, dass bei vorliegenden Evaluationsergebnissen die Diskussion über Ist-Soll-Diskrepanzen behindert wird (z. B. weil Diskussionen über die Normenlegitimation und über Normenabweichungen miteinander vermischt werden). Gerade bei quantitativen Fragebogen besteht die Gefahr, dass Items aus bestehenden Fragebogen übernommen werden, ohne dass die normativen Implikationen im Vorfeld hinreichend geklärt werden.
- Quantitative Evaluationen führen zu Daten, die von der konkreten Situation losgelöst, d. h. kontextneutral sind und die in diesem Sinne einen geringen konkreten Aussagegehalt haben. Gerade für die Suche nach Optimierungsschritten kann sich das Fehlen der kontextbezogenen Informationen erschwerend auswirken. Es überrascht daher nicht, dass neuere Untersuchungen auf die Bedeutung einer *diskursiven Einbettung von quantitativen Evaluationsergebnissen* hinweisen: Eine quantitative Evaluation scheint vor allem dann wirksam zu sein, wenn im Anschluss an die Erhebung und Auswertung der Daten eine qualitative Dateninterpretation anschliesst, die gemeinsam mit den befragten Personen gestaltet wird. Auf Grund dieser – empirisch gestützten – Feststellung postulieren beispielsweise Webler (1992) und Rindermann (2001) ein «hochschuldidaktisches Diskursmodell», in welchem die Besprechung

der Evaluationsergebnisse mit den Studierenden einen festen Bestandteil des Evaluationsverfahrens bildet. In den folgenden Ausführungen zur Ratingkonferenz spielt gerade diese Überlegung eine entscheidende Rolle.

4. Die Ratingkonferenz: Ein Verfahren zur Integration von quantitativen und qualitativen Evaluationselementen

Die Ratingkonferenz ist ein datengestütztes Evaluationsverfahren, bei welchem die qualitativen und quantitativen Aspekte der Evaluation und deren Vorzüge miteinander verbunden werden. Das Verfahren basiert im Wesentlichen auf zwei sich ergänzenden Prozesselementen:

- der individuellen Einschätzung (dem «Rating») von vorgegebenen Aussagen zum jeweiligen Evaluationsthema / -gegenstand (quantitatives/geschlossenes Element),
- der unmittelbar anschliessenden, gemeinsam mit den Befragten vorgenommenen Auswertung und Interpretation der erhobenen Daten (qualitatives/offenes Element).

Mit Blick auf die Lehrevaluation lässt sich das Vorgehen wie folgt umreissen (siehe Kasten): Die Kursteilnehmerinnen und Kursteilnehmer erhalten einen kurzen Fragebogen zur Qualität einer Lehrveranstaltung/eines Lehrmoduls. Die Items werden mit Hilfe einer vorgegebenen Skala eingeschätzt. Anschliessend übertragen die Teilnehmerinnen und Teilnehmer ihre individuellen Einschätzungen auf ein vorbereitetes Plakat oder auf eine Pinnwand (z. B. mit Markier-Punkten), so dass die Einschätzergebnisse für alle Anwesenden auf einen Blick ersichtlich sind. Diese Ergebnisse werden nun in der Gruppe nach vorgegebenen Fragen interpretiert und diskutiert, wobei der Dozent/die Dozentin als Moderator/Moderatorin dieser «partizipativen Dateninterpretation» fungiert. Ziel dieser diskursiven Phase ist es, einerseits die quantitativen Einschätzungen mit qualitativen Informationen anzureichern und andererseits gemeinsam Massnahmen zu generieren, die geeignet sein können, die Qualität der Lehrveranstaltung/des Lehrmoduls weiter zu entwickeln.

4.1 Hinweise zur Praxis der Ratingkonferenz

(a) Formulierung des Rating-Instruments

Die Aussagen (Items), die den Teilnehmerinnen und Teilnehmern zu Beginn der Ratingkonferenz zur individuellen Einschätzung vorgelegt werden, bilden die Basis des Verfahrens: Hier entstehen die Daten, auf die im Verlauf des diskursiven Auseinandersetzungsprozesses Bezug genommen wird. Ein idealer Umfang liegt bei etwa einem Dutzend Items. Auf keinen Fall sollte das Ratinginstrument mehr als zwanzig Items umfassen, denn zu viele Items bringen die Gefahr einer zu oberflächlichen qualitativen Interpretation mit sich. Die Items sollten so ausgewählt beziehungsweise formuliert sein, dass sie die wichtigsten Aspekte ansprechen, zu denen die Erfahrungen und Meinungen der Schülerinnen und Schüler gefragt sind. Für die Formulierung der Items sind grundsätzlich dieselben Hinweise zu beachten, die für die Gestaltung von Frage-

Zur Durchführung der Ratingkonferenz: Verlaufsmodell

- *Einstieg (Anwärmen)*: Das Thema, zu dem die Rating-Konferenz durchgeführt wird, und das Vorgehen werden kurz erläutert. Hinweise auf die Rolle des Moderators/der Moderatorin und die weitere Verwendung der Ergebnisse nicht vergessen! Evtl. folgt eine grundsätzliche und allgemeine Frage zum Evaluationsthema. (Z. B. «Wo sehen Sie die Stärken und Schwächen des Unterrichts im Fach ...?»)
- *Durchführung der Befragung*: Die Teilnehmerinnen und Teilnehmer erhalten ein Blatt mit den Items, die es mit Hilfe der vorgegebenen Skala zu bewerten gilt. Es wird genügend Zeit eingeräumt, um in Ruhe die individuelle Einschätzung vorzunehmen.
- *Visualisierung der Ergebnisse*: Die Ergebnisse der individuellen Einschätzung werden auf ein vorbereitetes Plakat übertragen (Items und Punkteskala in grosser Schrift). Für die Übertragung der Einschätzungen erhält jeder Teilnehmer/jede Teilnehmerin eine entsprechende Anzahl Markierungspunkte.
 - *Allgemeiner Gedankenaustausch*: Die Teilnehmer und Teilnehmerinnen werden zu einer ersten allgemeinen Interpretation des entstandenen «Bildes» aufgefordert:
 - «Sie haben ihre Klebepunkte gesetzt und sehen nun, wie die Aussagen von der Gruppe beurteilt werden und wo Sie mit ihrem Urteil im Vergleich zu den übrigen Gruppenmitgliedern stehen. Wollen Sie dazu etwas sagen?
 - Haben Sie Punktierungen vorgenommen, zu denen Sie selber das Bedürfnis haben, noch etwas zu kommentieren?
 - Wenn Sie das Ergebnis gesamthaft betrachten: Was fällt ihnen auf? Gibt es Einschätzungen, von denen Sie überrascht sind?» (Die allgemeinen Auffälligkeiten, die genannt werden, werden auf Karten festgehalten.)
- *Fokussierte Interpretation der Daten*: Einzelne auffällige Daten werden herausgegriffen: z. B. extrem hohe oder tiefe Einschätzungen, grosse Streuungen, grosse Ist-Soll-Diskrepanzen (falls eine Ist-Soll-Einschätzung vorgenommen wurde.) Leitfrage: «Welche Gründe sehen Sie für die auffällig hohe (tiefe, streuende,...) Einschätzung?» Die Interpretationsvorschläge der Teilnehmerinnen und Teilnehmer werden auf Flipchart festgehalten.
- *Fragen nach Stärken und Schwächen sowie nach Massnahmenvorschlägen*: Die Teilnehmenden werden aufgefordert, durch die Befragung manifest gewordene Stärken und Schwächen zusammenfassend zu nennen (auflisten) und anschliessend Massnahmen vorzuschlagen, die als geeignet erscheinen, (a) um die vorhandenen Stärken zu fördern und (b) um die Schwächen zu beseitigen.
- *Abschluss*: Die Teilnehmer und Teilnehmerinnen erhalten Gelegenheit, abschliessende Gedanken zum Verlauf der Rating-Konferenz und zu den Ergebnissen zu äussern. («Was ich noch sagen möchte», «Was bei der Weiterentwicklung des Unterrichts unbedingt beachtet werden sollte»)

bogen gelten: Es braucht einfache, gut verständliche, präzise, neutrale (nicht suggestive) und eindimensionale Formulierungen. Für die Einschätzung der Items sollte eine mindestens vierstufige Skala vorgegeben werden – beispielsweise mit der folgenden Abstufung: 1 = trifft nicht zu; 2 = trifft teilweise zu; 3 = trifft überwiegend zu; 4 = trifft vollständig zu.

Beispiele

Dem Dozenten/der Dozentin gelingt es, in diesem Modul ein Gleichgewicht zwischen selbst gesteuertem Lernen und geleiteter Wissensvermittlung herzustellen.

trifft nicht zu	1	2	3	4	trifft vollständig zu
-----------------	---	---	---	---	-----------------------

Wie zufrieden sind Sie mit dem Lernprofil in dieser Veranstaltung?

sehr unzufrieden	1	2	3	4	sehr zufrieden
------------------	---	---	---	---	----------------

(b) Zur Moderation der Dateninterpretation

Nach dem individuellen Ausfüllen des Fragebogens werden die Einschätzungen auf ein vorbereitetes Plakat übertragen. Die dadurch entstandene Übersicht über die Befragungsergebnisse bildet die Grundlage für ein moderiertes Evaluationsgespräch.

Die moderierende Lehrperson sollte in der Moderationstechnik geschult sein: Sie sollte einerseits Visualisierungstechniken beherrschen und andererseits die gemeinsame thematische Auseinandersetzung so strukturieren und leiten, dass möglichst offene und unbeeinflusste Äusserungen der Teilnehmerinnen und Teilnehmer zustande kommen.

Handlungsleitende Grundsätze für diese partizipative/diskursive Phase des Evaluationsprozesses sind:

- *Konsequente Visualisierung* der Äusserungen der Teilnehmerinnen und Teilnehmer auf Flipchart oder mit Hilfe von Karten auf Pinnwand. (Leitfragen für das Gespräch so formulieren, dass eine Visualisierung der Antworten möglich ist!)
- *Keine Beeinflussung*: Die moderierende Person sollte als thematisch neutral wahrgenommen werden. Wenn der Verdacht aufkommt, dass der Moderator/ die Moderatorin gewisse Statements hören bzw. nicht hören will, kann sich dies auf die Gesprächsoffenheit der Teilnehmerinnen und Teilnehmer negativ auswirken.
- *Keine konsensorientierte Diskussionen*: Diskussionen zur Frage, welche Auffassung richtig oder falsch sei, sind zu vermeiden. In erster Linie geht es darum, qualitative Statements einzuholen. Die einzelnen Statements können nebeneinander stehen bleiben, auch wenn sie sich widersprechen. Letztlich kann die Gewichtung und Einordnung von unterschiedlichen Statements bei der Lehrperson bleiben, die für den betreffenden Unterricht verantwortlich ist.
- *Anonymisierung mit Möglichkeit zur persönlichen Stellungnahme*: Die Ratingkonferenz ist grundsätzlich als anonymes Verfahren angelegt: Die Evaluationsergebnisse sollen einzig auf Grund ihres sachlichen Aussagegehalts interpretiert werden und nicht mit Blick auf die Person, welche eine bestimmte Einschätzung vorgenommen hat. Beispielsweise gilt für die moderierende Person die Regel, dass sie nicht nach dem Urheber einer Einschätzung fragen darf. (Also nicht: «Wer hat diesen Punkt hierhin geklebt», sondern: «Welche Überlegung/Erfahrung könnte eine Person dazu bewegt haben, diesen Punkt hierhin zu kleben?») Es muss den Befragten überlassen bleiben, ob sie sich «outen» und die persönlichen Gründe für eine bestimmte positive oder negative Einschätzung bekannt geben möchten.

4.2 Bewertung der Ratingkonferenz als Verfahren zur Lehrevaluation

Die Ratingkonferenz eignet sich ausgezeichnet für ein Unterrichtsfeedback, da die quantitative Datengrundlage in Verbindung mit einer qualitativen Interpretation durch die Betroffenen ein ausserordentlich reichhaltiges Material ergibt, um den Unterricht zu überdenken und um gemeinsam mögliche Optimierungsmassnahmen zu erörtern.

Die folgenden Punkte haben sich in der praktischen Umsetzung als wichtige Stärken des Verfahrens erwiesen:

- *Verbindung von quantitativen und qualitativen Verfahren*: Die Ratingkonferenz versteht sich als Verfahren, welches quantitative und qualitative Evaluationselemente zu integrieren vermag. Damit können die zentralen Vorteile von zwei normalerweise getrennten Evaluationsverfahren miteinander verbunden werden: sich Einblick verschaffen in die quantitative Verteilung von bestimmten Wahrnehmungen und Auffassungen, verbunden mit der Möglichkeit, bei den Betroffenen vertiefende qualitative Informationen in Form von Erläuterungen, Erklärungen, Interpretationen einzuholen.
- *Möglichkeit zur klärenden Nachfrage*: Sowohl der Dozent/die Dozentin als auch die Studierenden haben die Möglichkeit, bei Unklarheiten und Vieldeutigkeiten präzisierend nachzufragen. Dies bedeutet, dass der strenge Massstab an die Eindeutigkeit der Fragen, der bei der Fragebogenkonstruktion zu beachten ist, etwas weniger ins Gewicht fällt und dass interessante Aspekte, die in einem geschlossenen Item keinen Platz haben, mündlich eingeholt werden können.
- *Einbezug der emotionalen Ebene*: Eine Befragung löst in der Regel bei den Befragten jeweils auch Emotionen aus. Nicht zuletzt durch die enge zeitliche Verknüpfung der Fragenbeantwortung mit der Ergebnissichtung und -interpretation kann durch die Ratingkonferenz eine relativ starke emotionale Dynamik entstehen. Diese Emotionalität kann für ein adäquates Verständnis und die angemessene Gewichtung der Evaluationsdaten von grosser Wichtigkeit sein.
- *Hohe Akzeptanz der Ergebnisse bei den Betroffenen*: Die Ratingkonferenz versteht die quantitativen Ergebnisse der Evaluation als das, was sie vermutlich in produktiver, entwicklungsorientierter Betrachtung immer auch sind: als Reflexions- und Diskussionsimpulse, die zu einem vertieften Verständnis der Situation und zum Austausch unterschiedlicher Wahrnehmungen und Erfahrungsweisen anregen können. Nicht zuletzt durch die Unmittelbarkeit von gemeinsamer Datengenerierung und gemeinsamer Dateninterpretation kann eine produktive Gesprächs- und Reflexionssituation entstehen, die zu Evaluationsergebnissen mit einer breiteren Akzeptanz bei den Betroffenen führt.
- *Zeitsparendes Verfahren*: Verglichen mit dem Aufwand, den datengestützte Evaluationen normalerweise mit sich bringen, ist die Ratingkonferenz ein sehr ökonomisches Verfahren: Abgesehen von der Fragebogenentwicklung beträgt der Aufwand für die Durchführung, die Auswertung, die Dateninterpretation sowie die Massnahmengenerierung ca. 30–60 Minuten (was natürlich auch eine Beschränkung der Fragenmenge auf ca. 15–20 Items bedeutet!). Allerdings muss auf der Seite der

Studierenden mit einem grösseren Zeitgefäss gerechnet werden als in den üblichen Fragebogenverfahren, wo ja die Datenrückmeldung und die gemeinsame Dateninterpretation oft in unzulässiger Weise vernachlässigt werden. Nicht zuletzt aus dieser Überlegung heraus empfiehlt es sich, Ratingkonferenzen – trotz der hohen Effizienz – sparsam einzusetzen.

Trotz der aufgeführten Stärken muss beachtet werden, dass auch das Verfahren der Ratingkonferenz gewisse Schwierigkeiten und Stolpersteine mit sich bringt. Zumindest bezüglich der folgenden beiden Punkte hat es sich gezeigt, dass ein vorsichtiger Umgang angezeigt ist:

- *Konfusion von Moderationsanspruch und Betroffenheitsperspektive*: Wie bereits erwähnt, ist es für die Moderation der Dateninterpretation ausserordentlich wichtig, dass eine wertfreie Entgegennahme der Teilnehmerinnen- und Teilnehmerbeiträge erfolgt: Die Teilnehmenden sollen unzensuriert das sagen können, was ihnen angesichts der jeweiligen Datenkonstellation durch den Kopf geht. Nun befindet sich die Dozentin/der Dozent bei einer unterrichtsbezogenen Ratingkonferenz in der Regel in einer Doppelrolle: einerseits als Moderator, andererseits als Betroffener, der sich durch die Rückmeldung direkt angesprochen fühlt. Die von der moderierenden Person geforderte Wertneutralität wird durch die Betroffenheitsperspektive in zweifacher Hinsicht bedroht:
 - Erstens widerspricht die emotionale Betroffenheit dem Anspruch an eine neutrale, sachorientierte Moderation: Die moderierende Person steht vor dem Dilemma, entweder authentisch oder wertneutral zu sein.
 - Zweitens kann von Seiten der Studierenden für bestimmte (vor allem negativ bewertete) Phänomene nach Erklärungen gefragt werden, was dann relativ rasch in einen Rechtfertigungsdialog mit gegenseitiger Schuldzuweisung einmünden kann (was sich erfahrungsgemäss – nicht zuletzt angesichts der faktischen Machtverteilung zwischen Studierenden und Dozierenden – auf den weiteren Verlauf des Evaluationsgesprächs sowie auf künftige Evaluationen nicht sehr förderlich auswirkt!).
- Der adäquate Umgang mit dieser «Rollenkonfusion» erweist sich in der Praxis als ausserordentlich anspruchsvoll. Gerade in kritischen Beurteilungssituationen mit negativen Evaluationsergebnissen kann es günstig sein, wenn die Moderationsrolle von einer externen, nicht betroffenen Person übernommen wird.
- *Unmittelbarkeit der Datenrückmeldungen*: Bei schriftlichen Befragungen liegt in der Regel eine gewisse Zeitspanne zwischen der Datenauswertung (bzw. der Zurkenntnisnahme der Evaluationsergebnisse durch die Dozierenden) und dem Datenfeedback an die Studierenden. Damit hat der Dozent/die Dozentin auch die Möglichkeit, die Evaluationsergebnisse emotional zu verarbeiten und die eigene Reaktion gegenüber den Studierenden vorzubereiten. In der Ratingkonferenz dagegen fallen diese beiden Phasen zeitlich zusammen: Die Zurkenntnisnahme der Ergebnisse und das Datenfeedback an die Studierenden finden hier gleichzeitig statt. Diese Unmittelbarkeit kann von den Dozierenden – vor allem bei negativen Evaluationsergebnis-

sen – evtl. als Überforderung empfunden werden. Der Dozent/die Dozentin sollte sich daher zugestehen, eine differenzierte Kommentierung der Ergebnisse eventuell erst zu einem späteren Zeitpunkt (z. B. im Rahmen der nächsten Lehr-Lernveranstaltung) vorzunehmen.



Insgesamt zeigen die Erfahrungen, dass die Ratingkonferenz eine wertvolle Alternative ist sowohl zu den herkömmlichen quantitativen/geschlossenen Fragebogenevaluationen als auch zu den offenen/qualitativen Veranstaltungsevaluationen. Zwei Charakteristika des Verfahrens dürften vermutlich dafür ausschlaggebend sein: einerseits der Mix von qualitativen und quantitativen Daten, der zu einer aussagekräftigen Bewertungsgrundlage führt, andererseits der konsequente Einbezug der diskursiven Auseinandersetzung zwischen den prozessgestaltenden Partnern. Insbesondere für entwicklungsorientierte Evaluationen entsteht auf diese Weise eine Qualitätsdiagnose, die sich als gute Basis für die Planung und Durchführung von Unterrichtsentwicklungsmassnahmen erweist.

Literatur

- Altrichter, H.** (1998). Reflexion und Evaluation in Schulentwicklungsprozessen. In H. Altrichter, W. Schley, & M. Schratz (Hrsg.), *Handbuch zur Schulentwicklung* (263–335). Innsbruck: Studienverlag.
- Becker, G., Ilsemann, C. & Schratz, M.** (2001). *Qualität entwickeln: evaluieren*. Friedrich Jahresheft XiX.
- Burkhard, C. & Eikenbusch, G.** (2000). *Praxishandbuch Evaluation in der Schule*. Berlin: Cornelsen Scriptor.
- Konrad, K.** (2005). *Mündliche und schriftliche Befragung. Ein Lehrbuch*. Landau: Verlag empirische Pädagogik.
- Landwehr, N., Steiner, P. & Keller, H.** (2003). Schritte zur datengestützten Schulevaluation. Eine Anleitung zur systematischen Datenerhebung mit Fragebogen. Bern: hep-verlag.

Lehrevaluation

Müller, S. (2000). Evaluation von Unterricht. In W. Böttcher & E. Philipp (2000). *Mit Schülern Unterricht und Schule entwickeln*. (S. 52–71). Weinheim: Beltz.

Schratz, M., Iby, M. & Radnitzky, E. (2000). *Qualitätsentwicklung. Verfahren, Methoden, Instrumente*. Weinheim: Beltz.

Rindermann, H. (2001). *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen*. Landau: Verlag Empirische Pädagogik.

Webler, W.-D. (1992). Evaluation der Lehre: Praxiserfahrung und Methodenhinweise. In D. Grün & H. Gattwinkel (Hrsg.), *Evaluation von Lehrveranstaltungen. Überfrachtung eines sinnvollen Instrumentes?* (S. 143–161). Berlin: FU-Dokumentationsreihe.

Autor

Norbert Landwehr, Dr. phil., Leiter Zentrum Schulqualität im Institut Wissen und Vermittlung der Pädagogischen Hochschule Aargau, Küttigerstr. 21, 5000 Aarau, norbert.landwehr@fh-aargau.ch