

Prof. Dr. Martin Lengwiler

**Evaluation, Qualitätssicherung, Leistungsmessung
in den historischen Wissenschaften**

Zunächst herzlichen Dank für die Einladung, hier einen Input zu formulieren. Ich verdanke diesen Auftritt wahrscheinlich einem Projekt, das ich zusammen mit Christian Simon, John Bendix und Heinrich Hartmann – im Auftrag der CRUS durchgeführt habe – in Kooperation mit Hans-Dieter Daniel (Leiter der Evaluationsstelle der Uni Zürich). Dabei ging es um die Analyse von Qualitätskriterien für die Forschung in den Geisteswissenschaften – unser Basler Teilprojekt hat verschiedene ausländische Modelle zur Forschungsevaluation kritisch untersucht und danach gefragt, über welche spezifischen Verfahren sie verfügen, um mit den Eigenheiten der Geisteswissenschaften umzugehen.

- Projekt ist 2009-2013 durchgeführt worden – Abschlussbericht ist erhältlich.

Mein Input schliesst in vielem an die Ausführungen von Christian Simon an. Im Tenor stehen wir uns sehr nahe – in den Details gibt es auch unterschiedliche Einschätzungen. Ich argumentiere wohl eher aus der Position des Pragmatikers (in Christians Terminologie) und nicht so sehr aus jener fundamentalkritischen Perspektive, die sein Referat am Schluss umrissen hat. Aber mein Input bewegt sich auch auf einer anderen Flughöhe. Ich werde eher aus der Sicht eines Universitäts- und Institutsangehörigen – quasi aus der institutionellen Praxis heraus – argumentieren. Die Vogelschauerspektive des Wissenschaftsrats wird im Folgenden

gleichsam mit einem Blick aus dem universitären Unterholz ergänzt.

Ich möchte den Input in vier Abschnitte und vier Thesen gliedern.

1. These: Qualitätssicherung ist kein neues Thema in den Geisteswissenschaften – die Diskussion reicht auch in unseren Fächern schon bald ein Vierteljahrhundert zurück (in die 1990er Jahre).

(ich benutze den Begriff Qualitätssicherung als einen Sammelbegriff für handlungsorientierte Messverfahren)

Formalisierte Evaluationsverfahren mögen in den Geisteswissenschaften noch relativ jung sein – der Druck, sich durch Leistungen und Produkte zu legitimieren, besteht dagegen schon seit längerem und hat sich in den Geisteswissenschaften spätestens seit den 1990er Jahren breit etabliert.

Die Wissenschaftsforschung hat diese wissenschaftspolitische Epochenwende schon seit längerem diagnostiziert. Noch bis in die 1970er Jahre funktionierten Wissenschaft und Forschung – zumindest in der westlichen Welt – nach einer Art Delegationsprinzip: der Staat investierte in die Grundlagenforschung, überliess es aber der Wissenschaft, wie sie mit diesem Geld genau umging. Der Rechtfertigungsdruck war gering – man ging davon aus, dass die Grundlagenforschung quasi automatisch anwendungsbezogene Erkenntnisse generieren

würde. Oder man unterhielt parallele Forschungseinrichtungen, die eigens der anwendungsorientierten Forschung gewidmet waren (in Deutschland etwa die Fraunhofer-Institute, in der Schweiz die eidgenössischen Forschungsanstalten – heute im ETH-Bereich).

Seit den 1980er Jahren erleben wir international einen Trend hin zu einem Wissenschaftsverständnis, das Wissenschaft zunehmend durch seinen gesellschaftlichen Nutzen legitimiert. Dies ist ein vielschichtiger Prozess, der Elemente der Ökonomisierung, der Bürokratisierung und der Produktivitätssteigerung umfasst. Das Delegationsmodell wurde nicht aufgehoben – wir forschen immer noch mehr oder weniger autonom. Aber wir sind heute gezwungen, unsere Leistungen überprüfen und messen zu lassen. Umgekehrt haben die wissenschaftspolitischen Behörden seit den 1980er Jahren umfangreiche Verfahren zur Sicherung der wissenschaftlichen Integrität und Produktivität entwickelt – nämlich das was wir heute unter dem Begriff der Qualitätssicherung zusammenfassen.

Dieses Denken – dass Forschung durch irgendeine Form von Leistungsausweis legitimiert werden muss – hat auch die Geisteswissenschaften spätestens in den 1990er Jahren erreicht.

- Und zwar nicht nur als Zwang von aussen.
- Wichtig: Qualitätssicherung sollte man als Kommunikationsbeziehung zwischen Förderern und Geförderten verstehen (zwar asymmetrisch, aber doch interaktiv).

Und die Geisteswissenschaften haben dieses Spiel von Beginn weg mitgespielt – nicht weil sie daran glaubten, sondern weil sie sich einen Nutzen versprachen.

Das Historische Seminar der Universität Zürich bietet einen exemplarischen Fall. 1996 wurde dort das 25-Jahr-Jubiläum der Forschungsstelle für Sozial- und Wirtschaftsgeschichte gefeiert (damals unter der Leitung von Hans-Jörg Siegenthaler, Rudolf Braun und Bruno Fritzsche): wie wurde das Jubiläum begangen? Man hat eine Broschüre gemacht, in der fein säuberlich jede Lehrveranstaltung und jede einzelne Liz-Arbeit aller Professoren an der Forschungsstelle aufgelistet wurden. Das Dokument sollte als Leistungsausweis dienen – auf die Zeit hin, in der die drei Professuren wieder besetzt werden sollten. Man wollte sich proaktiv gegenüber der Unileitung, den politischen Behörden und der Öffentlichkeit legitimieren.

Ein sichtbarer externer Druck existierte nach meinem Wissen nicht. 1996 war jedenfalls lange VOR der Einführung formalisierter Evaluationsverfahren an der Uni Zürich. Die Evaluationsstelle wurde an der Uni Zürich erst Ende der 1990er Jahre gegründet.

2. These: Wir sind heute an einem Punkt angekommen, an dem die Instrumente zur Qualitätssicherung hypertroph zu werden drohen und das Ensemble dieser Instrumente so komplex geworden ist, dass eine Übersicht kaum mehr

möglich erscheint, von einer systematischen und in sich logischen Strukturierung der Vielzahl von Verfahren ganz zu schweigen.

In den letzten Jahren haben sich auf verschiedenen Ebenen formalisierte Verfahren zur Qualitätssicherung ausgebreitet und durchgesetzt: die meisten Fachzeitschriften sagen von sich, sie seien peer-reviewed (was immer das bedeutet...); die Universitäten haben Instrumente zur Evaluation von Lehrveranstaltungen und Forschungsleistungen – zudem auch Evaluationen von ganzen Studiengängen, Fächern und Einrichtungen – geschaffen. Einzig die Berufungskommissionen funktionieren in den Geisteswissenschaften noch weitgehend nach altem Muster, ohne Einsatz formalisierter Verfahren (wie etwa bibliometrischer Messungen). Aber auch hier gibt es Ansätze der Formalisierung und Bürokratisierung.

Heute hat dieser Trend eine Eigendynamik entwickelt, die kaum mehr hinterfragt wird. Der Sinn solcher Verfahren wird kaum bestritten – die Diskussion dreht sich kaum mehr ums Ob, sondern vor allem ums Wie – um die technischen Details, um die Frage, wie Qualitätssicherung die geisteswissenschaftliche Wissenschaftskultur angemessen abbilden kann. Die Verfahren sind gleichsam selbstreferentiell geworden. Weil die Qualitätssicherung als allgemein anerkannt gilt, sollten nun auch die Geisteswissenschaften nachziehen. Konkret: Wir haben in Basel gerade eben ein standardisiertes Verfahren zur Lehrevaluation eingeführt. Ein zentrales Argument dafür war, dass unser akademischer Nachwuchs die ausgewerteten

Evaluationen der eigenen Lehrveranstaltungen dringend benötigte
– schliesslich würde bei Berufungsverfahren heute von den
Kandidierenden ein Lehrportfolio erwartet.

Mit anderen Worten: Die Gemengelage an Instrumenten zur
Qualitätssicherung ist heute unüberschaubar geworden.

Dies ist ein aktuelles Beispiel – aus dem Kontext der Uni Basel:

FOLIE

Wir sind permanent und in verschiedenen Rollen solchen
Instrumenten ausgesetzt: beim Publizieren, bei der Einwerbung
von Drittmitteln, bei unseren Bewerbungen, und dann – wenn wir
mal fest angestellt sind – bei unseren Beförderungen; ebenso als
Lehrende, als Uni-Angestellte mit Forschungs- und
Publikationsausweis (durch Füttern von
Forschungsdatenbanken), als Angehörige von Instituten, die als
Gesamteinrichtung evaluiert werden. etc. etc.

Dies führt in der Praxis zu häufigen Redundanzen und
Mehraufwand, der eigentlich unnötig ist.

Auch hier kann man die Basler Erfahrungen anführen.

Unser Rektorat in Basel hat es geschafft, unsere Graduiertenschule
im selben Jahr zweimal unabhängig voneinander zu evaluieren:
einmal im Rahmen einer Departementsevaluation, das andere Mal
spezifisch als Graduiertenschule. Grund war, dass das eine

Vizerektorat nicht wusste, was das andere Vizerektorat grad anstellt. Und wir haben einfach darauf vertraut, dass zwei voneinander unabhängige externe Kommissionen zu einigermaßen denselben Ergebnissen kommen (was dann auch der Fall war).

Der Aufwand an Evaluationen hat heute ein grenzwertiges Ausmass erreicht – und zwar für beide Seiten (Evaluierte wie Evaluierende – irgendjemand muss die Evaluationen ja durchführen und die Gutachten schreiben).

Alleine der interne Aufwand für unsere Institutsevaluation 2013/14 betrug in etwa ein Personenjahr (postdoc, dh. rund 100'000 CHF). Dazu kommen die Kosten für die externe Evaluationsstelle und das Review-Gremium hinzu. Alles in allem hat diese Evaluation an offenen und verdeckten Kosten rund eine Viertelmillion gekostet.

- ich bin sicher, Sie können alle ähnliche Beispiele aufzeigen.

Wichtig scheint mir: Solche Erfahrungen sind symptomatisch für ein Malaise, das im Moment den Verantwortlichen zunehmend bewusst wird.

Der Grenznutzen der Qualitätssicherung wird im Moment zunehmend zum Problem. Die Verfahren sollten dringend verschlankt werden; und die Vielzahl an Instrumenten reduziert und entflochten.

Da und dort lässt sich auch eine Gegenbewegung erkennen. Die evaluationskritische Wende in den Naturwissenschaften – die in Christian Simons Referat angesprochen wurde – kam nicht zuletzt als Reaktion auf den überbordenden Aufwand zustande.

Auch in international anerkannten Evaluationsmodellen – etwa dem Britischen Research Excellence Framework – geht der Trend klar in Richtung Verschlankung der Verfahren, um den Arbeitsaufwand einigermaßen im Rahmen zu halten.

(Hier gab es einen simplen ökonomischen Grund: Das Research Excellence Framework verspricht bei guter Bewertung zusätzliche finanzielle Mittel. Paradoxerweise waren die Bewerberquoten trotz Aussicht auf Zusatzfinanzen lange Zeit rückläufig. Weil die Institute scharf kalkulierten und schnell sahen, dass auch ein moderater Erfolg nicht die enormen Kosten für die Bewerbungsphase decken würde.)

3. These: Die Verfahren der Qualitätssicherung sollten konsequent ihren jeweiligen Zwecken angepasst werden.

Die Frage ist:

Wie lässt sich die Eigendynamik der Qualitätssicherung – diese Aura des Selbstverständlichen – am einfachsten durchbrechen?

Ein sinnvoller Weg scheint mir, dass man konsequent nach dem Zweck der jeweiligen Instrumente fragt und kritisch untersucht, ob diese geeignet sind die entsprechenden Ziele zu erfüllen.

Von dieser kritischen Analyse abhängig könnte man die jeweiligen Verfahren der Qualitätssicherung neu ausrichten.

Diese Überlegung lässt sich gut veranschaulichen am Beispiel der Evaluationen von wissenschaftlichen Einrichtungen (z.B. Uni-Instituten). Wenn die Evaluation auf die Neuallokation finanzieller Ressourcen zielt, werden Sie ein anderes Verfahren haben als wenn es um eine blosse formative Evaluation geht (im Sinne eines unverbindlichen Ratschlags zuhanden des Instituts). Bei ressourcenbezogenen Evaluationen steht mehr auf dem Spiel als bei formativen Evaluationen. Die Verfahren sind im ersten Fall oft aufwändiger, formalisierter, zielen auf stärkere Einbindung der verschiedenen beteiligten Perspektive (um möglichst unparteilich zu sein).

Eine rein formative Evaluation braucht dagegen nicht besonders formalisiert zu sein. Ein paar kluge Expertinnen und Experten, ein vernünftiger Einblick in die Einrichtung, vielleicht auch eine inspirierende Atmosphäre – und schon kann aus der Aussenperspektive eine gewinnbringende Evaluation hervorgehen.

Man könnte das als eine Art Guerilla-Taktik der Qualitätssicherung bezeichnen.

Schliesslich sollte man sich auch der nicht-intendierten und unerwünschten Nebeneffekte bewusst sein. Das Britische Beispiel zeigt etwa, dass ressourcenbezogene Evaluationen – quasi Evaluationen mit harten Sanktionen - zu strategischem Verhalten einladen. Die britischen Universitäten haben schnell gelernt, wie sie das Spiel zu spielen haben, damit sie ihre Erfolgsaussichten optimieren konnten. Dieses Game playing hat schliesslich dazu geführt, dass die Behörden heute die Spielregeln des Research Excellence Framework regelmässig neu definiert, um solchen strategischen Verhaltensweisen zuvor zu kommen.

All dies sollte man bedenken, wenn man aus Sicht der schweizerischen Geschichtswissenschaften sich überlegt, welche Verfahren der Qualitätssicherung zu diesem Fach passen. Wie kann man Qualität überhaupt so definieren, dass man sie auch messen kann? Christian Simon hat dazu bereits Grundlegendes gesagt. Ein Fach wie die Geschichtswissenschaften – das intern recht heterogen ist und sich in verschiedene Teildisziplinen aufteilt – wird es schwer haben, einheitliche quantitative Verfahren zu definieren. Sowieso in einem Land, das mehrere sprachregionale Publikationsformen und Wissenschaftskulturen (je nach Landesteil) kennt. Die Skepsis gegenüber bibliometrischen Verfahren, die in unserem Fach weit verbreitet ist, lässt sich vor diesem Hintergrund gut nachvollziehen. (Man müsste für alle Sprachräume einheitliche Standards für die verschiedenen – sehr heterogenen – Publikationskanäle setzen (zum Beispiel Punkte für verschiedene Zeitschriften je nach ihrem Renommee) und dies dann auf nationaler Ebene abgleichen...).

Qualitativere Zugänge dagegen – wie etwa das Peer review – scheinen für ein Fach wie die Geschichtswissenschaften besser geeignet (sind aber auch aufwändiger).

Keine einfache Ausgangssituation. Aber hoffentlich genug Stoff für die Diskussion.