

## Session 4 – Motivation, Interesse wecken

### Evaluationsforschung im Feld ‚Schule‘ - Evaluationsdesigns für die Begleitung schulischer Innovationen

Univ.-Prof.<sup>in</sup> Dr.<sup>in</sup> Annette Ostendorf

Institut für Organisation und Lernen (Wirtschaftspädagogik)

Universität Innsbruck

#### 1. Evaluationen im Schulalltag: das Beispiel QIBB

Evaluationen sind mittlerweile fester Bestandteil des schulischen Alltags. Ein Beispiel hierfür ist der berufsbildende Schulbereich, der durch die Qualitätsinitiative Berufsbildung (QIBB, vgl. [www.qibb.at](http://www.qibb.at)), über die seit Jahren die Qualität des beruflichen Bildungswesens gesteuert und verbessert werden soll, mit Evaluationen überzogen ist. Evaluation ist ein sehr wesentlicher Bestandteil des sog. Qualitätsregelkreises. Dieser besteht aus den Elementen Plan-Do-Check-Act, die wie einem Regelkreis aneinander gebunden sind.



Abb. 1: Der Qualitätsregelkreis (<https://www.qibb.at/de/qualitaetsregelkreis.html>)

Bei den Evaluationen wird hier zwischen einem Systemfeedback und dem Individualfeedback unterschieden. Das Systemfeedback zielt auf die Erhebung von Lehr- und Verwaltungsprozessen auf der Organisations- und Systemebene. Hier wurden z.B. im humanberuflichen und höheren land- und forstwirtschaftlichen Schulbereich im Jahr 2010 40.480 Schüler/innen befragt (vgl. bmukk, 2013, S. 21). Allerdings wurden eben weitgehend nur Lernende befragt, wenig jedoch ande-

re Stakeholder. Beim Individualfeedback geht es um die Weiterentwicklung und Reflexion des beruflichen Handelns der Lehrkräfte. Im gerade genannten Schulbereich wurden hier im Schuljahr 2011/12 33.872 Fragebögen versandt. Der Rücklauf war allerdings sehr gering (1283 accounts).

An dieser Stelle werden bereits zwei Besonderheiten deutlich. Diese Art von Evaluationen sind zum einen sehr stark am ‚Messen‘ und weniger am ‚Entwickeln‘ orientiert. Sie sind quantitativ-summativ angelegt.

Zum anderen scheint es auch deutliche Probleme mit dieser Art der Evaluation zu geben. Es lassen sich anscheinend nicht alle Stakeholder einbinden und die geringe Beteiligung der Lehrkräfte, insbesondere aber der mit leitenden Aufgaben betrauten Kollegen/innen – wie Administratoren/innen – erscheint problematisch. Im Bundesqualitätsbericht wird dazu Folgendes berichtet: „Es ergibt sich der Schluss, dass entweder der Bedarf der Betroffenen nach Feedback „gering“ ist, oder dass in den beiden Gruppen (Fachvorstände/innen, Administrator/innen, Anm. A. O.) mehr für die „Akzeptanz“ in Bezug auf die Einholung von Individualfeedback getan werden muss. Möglicherweise sind die angebotenen Instrumente nicht den Bedürfnissen der potentiellen Feedback-Nehmer\_innen angepasst oder Individualfeedback findet in einer anderen Form statt.“ (bmukk, 2013, S. 22)

Obwohl diese auch in QIBB umgesetzte Variante des summativen Evaluationsverständnisses unter Einsatz quantifizierender Instrumente weit verbreitet ist, gibt es auch noch andere Ansätze der Evaluation. Alle Varianten haben ihre Berechtigung, bedienen allerdings unterschiedliche Ansprüche, oder wie Guba & Lincoln (1990: S. 21) dies ausdrücken: „There is no answer to the question, ‚But what is evaluation really?‘ and there is no point in asking it.“

## 2. Alternative Varianten der Evaluation

Im folgenden Beitrag soll eine alternative Variante der Evaluation im Feld ‚Schule‘ vorgestellt werden, die von der Logik des ‚Messens‘ abweicht ohne das ‚Bewerten‘ zu vernachlässigen. Sie ist am ehesten der sog. ‚fourth generation of evaluation‘ zuzuordnen. Guba & Lincoln (1990, S. 22-49) beschreiben insgesamt vier Generationen von Evaluationen:

- I. *Measurement*: Individuumsbezogene Tests und Kompetenzmessung z.B. bei Schüler/innen
- II. *Description*: Stärken- und Schwächenprofile bezüglich bestimmter Ziele werden beschrieben (Programmevaluation)
- III. *Judgement*: Beurteilung steht im Vordergrund; Standards werden überprüft; Evaluatoren/innen als Beurteiler/innen; entscheidungsorientierter Fokus.
- IV. *Responsive constructivist evaluation*: Zwei Merkmale prägen diese Art von Evaluation besonders: a) Fragestellungen und Evaluationziele werden nicht a priori von außen gesetzt, sondern emergieren im Prozess, in den die Stakeholder eingebunden sind. „Responsive evaluation (...) determines parameters and boundaries through an interactive, negotiated process that involves stakeholders and that consumes a considerable portion of the time and resources available. It is for this reason, among others, that the design of a responsive evaluation is said to be emergent.“ (Guba & Lincoln, 1990, S. 38 f.)  
 b) Einbettung in eine konstruktivistische, interpretative methodologische Perspektive und Vorgehensweise. Es geht um die Konfrontation unterschiedlicher Perspektiven und Konstruktionen über bestimmte Evaluationsgegenstände. „As each group copes with the constructions posed by others, their own constructions alter by virtue of becoming better informed and more sophisticated.“ (Guba & Lincoln, 1990, S. 41)

Interessant ist dabei, dass die Einteilung tatsächlich historisch begründet wird. Da das Hauptwerk von Guba & Lincoln aber schon Ende der 1980er Jahre (erste Auflage 1989) geschrieben wurde, muss man darüber nachdenken, ob es nicht schon wieder eine neue Generation gibt. Angesichts von Large scale Messungen wie PISA, TIMMS etc. und der Hinwendung zur testbasierten Kompetenzmessung, Klassenvergleichsarbeiten und Überprüfung von Bildungsstandards kann man feststellen, dass es momentan eine Art von Rolle rückwärts zur ersten Generation gibt. Die bildungspolitische und erziehungswissenschaftliche (Stichwort: Kompetenzmessung) Debatte wird momentan von Evaluationen im Sinne eines ‚Messens‘ beherrscht. „The rise of measurement culture in education has had a profound impact on educational practice, from the highest levels of educational policy at national and supranational level down to the practices of local schools and teacher.“ (Biesta, 2010, S. 12) Der Entwicklungsgedanke, der auch mit Evalu-

ationen verbunden werden kann, wird damit eher marginalisiert. Diesen betonen aber wiederum formative Evaluationen, die eher ein responsives Design haben und der vierten Generation von Evaluationen zugerechnet werden können. Ihre Vorzüge (aber auch ihre Grenzen) sollen deshalb anhand eines Beispiels nochmals genauer verdeutlicht werden.

### 3. Beschreibung eines konkreten Fallbeispiels

Im Auftrag des österreichischen Ministeriums für Unterricht, Kunst und Kultur (bmukk) wurde im Schuljahr 2011/12 an einer Handelsschule der Schulversuch zur Implementierung eines neuen Curriculums an Handelsschulen (im folgenden «Handelsschule NEU» genannt) durch die Universität Innsbruck (Wirtschaftspädagogik) evaluiert.<sup>1</sup> Ähnliche Aufträge ergingen auch an zwei weitere österreichische Universitäten. Die Wahl der Forschungsmethodik war frei wählbar. Ziel war es, Erkenntnisse über Besonderheiten, Stärken und Schwächen des neuen Curriculums in der konkreten schulischen Umsetzung zu gewinnen. Die Evaluation erfolgte im Design einer responsiven Evaluation. Dies bedeutet, dass ähnlich wie im Modell der Aktionsforschung ein Wechsel zwischen Datenerhebung und Rückspiegelung an die Betroffenen stattfand und möglichst viele Stakeholder-Perspektiven einbezogen werden sollten.

Die nachfolgende Abbildung zeigt das methodische Gesamtdesign der Evaluation.

Die gestrichelten Linien verdeutlichen die Rückkoppelungsschleifen. Hier ist anzumerken, dass bei der Rückkoppelung nur die Lehrenden und die Direktion als Stakeholder eingebunden waren. Auch die Auswahl der konkreten Evaluationsaspekte erfolgte in Auseinandersetzung mit diesen beiden Stakeholdern. Dies entspricht nicht ganz dem Modell einer vierten Generation von Evaluation. Auch Elemente der zweiten Generation (Description) finden sich im Ansatz dieser Evaluation wieder, ging es doch um Stärken und Schwächen eines ganzen Programms. Durch die Dominanz der qualitativen Forschungsinstrumente

---

<sup>1</sup> Eine ausführliche Darstellung der Ergebnisse findet sich unter Ostendorf, A. et al. (2012).

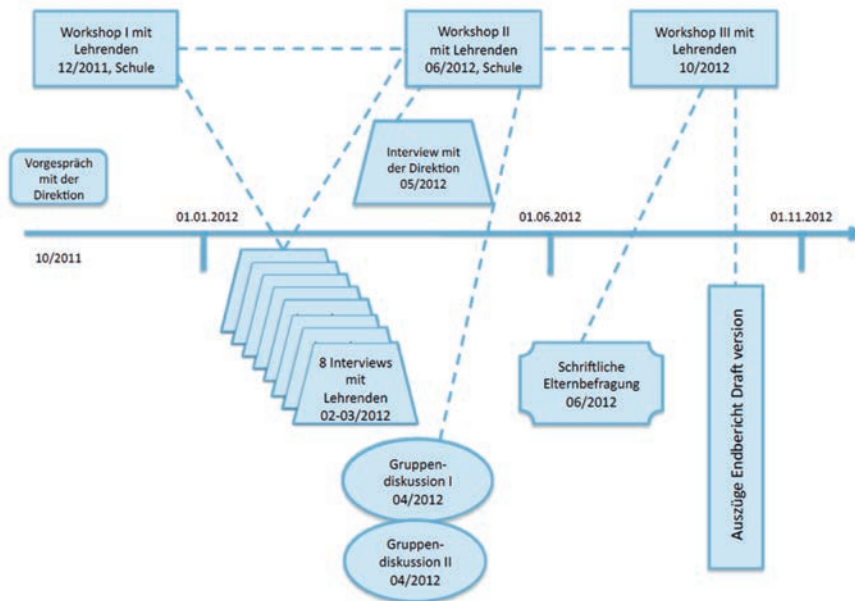


Abb. 2: Gesamtdesign

(qualitative teilstrukturierte Interviews, moderierte Gruppendiskussionen) und die Workshops wurde jedoch dem interpretativen Zugang deutlich Raum gegeben und somit insgesamt einer konstruktivistischen methodologischen Position gefolgt.

#### 4. Nutzen-/Nutzungsprofile von Evaluationen

Es stellt sich nun die Frage, welcher Nutzen durch diese Art der Evaluation gestiftet wird. Prinzipiell lassen sich verschiedene Nutzenarten bestimmen. Saunders (vgl. 2012, S. 425) fasst die verschiedenen Möglichkeiten in folgender Typologie zusammen.

**1. Instrumenteller Nutzen:**

EntscheidungsträgerInnen verändern das Evaluationsobjekt aufgrund der Ergebnisse.

**2. Konzeptioneller Nutzen:**

Die Evaluationsergebnisse helfen den AuftraggeberInnen, das Evaluationsobjekt in einer anderen Weise zu betrachten.

**3. Aufklärung:**

Evaluationen erweitern das Wissen zu einem Feld. Die Ergebnisse sind auch für nicht direkt beteiligte Personen relevant.

**4. Prozessnutzen:**

Durch das Engagement der Beteiligten im Evaluationsprozess kommt es zu Lernprozessen auf individueller und kollektiver Ebene; ein Denken in den Kategorien der Evaluation wird gefördert.

**5. Rechtfertigungsnutzen:**

Bestätigung von Überzeugungen der AuftraggeberInnen, Legitimierung von Entscheidungen.

Nicht zu vergessen ist bei all diesen Nutzenkategorien, dass auch ein negativer Nutzen von Evaluationen bestehen kann, z.B. Datenfriedhöfe, Frustrationen aufgrund eines zu geringen Einflusses auf Entscheidungen, Scheinbegründungen, überbordende Bürokratie.

Im vorhin kurz beschriebenen Beispiel wurden alle fünf Nutzenarten erkennbar, auch wenn teils nicht direkt von der Evaluation auf die Auswirkungen geschlossen werden kann und auch andere Einflussfaktoren bestimmend waren.

Der *instrumentelle* Nutzen zeigt sich z.B. tendenziell darin, dass z.B. zum neu zu integrierenden Betriebspraktikum von der Schulverwaltung Bemühungen zur Strukturierung der Begleitung aufgesetzt und über Stundenverteilungen nachgedacht wurde. Das konkrete Curriculum wurde jedoch in seiner Grundstruktur nicht verändert, was aber auch zu Beginn der Evaluation schon deutlich gemacht wurde. Die Aushandlungsprozesse sind hierfür zu komplex.

Eine alternative Betrachtungsweise für Auftraggebende und somit ein *konzeptioneller* Nutzen zeigt sich darin, dass durch die Diskussion der Evaluationsergebnisse in einer Zusammenschau mit Ergebnissen anderer Evaluationsstandorte

(z.B. auf dem Wirtschaftspädagogik-Kongress 2013 in Graz) unterschiedliche Interpretationen der Lehrpläne und verschiedene Umsetzungsstrategien sichtbar wurden.

Durch Präsentation und Diskussion der Evaluationsergebnisse z.B. bei DirektorInnentagungen wurde auch ein erweiterter Personenkreis aufmerksam und somit auch ein *Aufklärungsnutzen* gestiftet.

Der *Prozessnutzen* ist schwierig zu bestimmen. Allerdings kann aus den Erfahrungen mit dieser Evaluation berichtet werden, dass gerade die hohe Motivation und das Interesse der beteiligten LehrerInnen und SchülerInnen darauf hindeuten, dass sie sich in den Evaluationsprozess einbinden wollten und darin auch einen Sinn für sich selbst sahen. Gerade bei den Lehrpersonen erzeugte die Evaluation auch so etwas wie ein ‚Wahrgenommenwerden‘ durch die EntscheidungsträgerInnen. Sie sahen die Evaluation auch als Möglichkeit der Partizipation und Reflexion. Die konkretere Ausprägung müsste aber genauer empirisch erhoben werden. Die Aussagen können sich momentan nur auf den Eindruck der EvaluatorInnen stützen.

Das Motiv des *Rechtfertigungsnutzens* wurde bereits in der Beauftragung deutlich. Es ging auch um rückwirkende Legitimation von Entscheidungen des Ministeriums.

Welche Art des Nutzens im Vordergrund steht, hängt stark vom jeweiligen Design ab. Prinzipiell zielen aber sowohl summative als auch formative Evaluationsdesigns auf alle fünf Nutzenbereiche. Der Prozessnutzen ist evtl. in responsiven Designs stärker im Fokus, da hier andere Formen der Kommunikation und konzeptionellen Beteiligung der Stakeholder möglich sind.

## 5. Der Nutzen von Evaluationen für die fachdidaktische Forschung

Wenn WissenschaftlerInnen mit der Durchführung von Evaluationen beauftragt werden, steht neben den oben erwähnten Nutzenprofilen noch etwas anderes im Raum. Zu fragen ist nach dem fachdidaktisch-*wissenschaftlichen* Nutzen. Wissenschaftlich abgesicherte Evaluation und Begleitung erfordern einen erhöhten Ressourcenaufwand für WissenschaftlerInnen, der möglicherweise in Konkurrenz

zum eigenen Aufgabenprofil steht. Insofern werden WissenschaftlerInnen, die Forschung als eines ihrer Kerngeschäfte betrachten, auch danach fragen, welchen wissenschaftlichen Output man in und mit Evaluationen erzeugen könnte.

Diese Fragen sind je nach Evaluationsdesign unterschiedlich zu beantworten. Im Folgenden wird eine Antwort vor allem im Hinblick auf responsive Modelle der Evaluation versucht.

Selbstverständlich gibt es eine Vielzahl an nationalen und internationalen Zeitschriften, die Evaluationsstudien und methodologische oder methodische Beiträge der Evaluationsforschung aufnehmen (z.B. *Evaluation*, *Evaluation Review*, *Zeitschrift für Evaluation*, *Educational Evaluation and Policy Analysis*). Wenn Evaluationsergebnisse nur über Abschlussberichte oder andere praxisfokussierte Verbreitungsstrategien vermittelt werden, ist die Sichtbarkeit und Würdigung im wissenschaftlichen Umfeld nur in geringem Umfang möglich. Für eine wissenschaftliche Verwertung sind jedoch andere Standards einzuhalten und insofern im Regelfall erhebliche Überarbeitungen notwendig.

Es gibt aber für WissenschaftlerInnen, gerade auch in der Fachdidaktik, die dem Untersuchungsfeld ‚Schule‘ besonders zugewandt ist, auch noch andere Motive für die Übernahme wissenschaftlicher Evaluationen abseits von Publikationslisten und Impact-Faktoren:

- (1) Fachbezogene und methodische wissenschaftliche Expertise wird ins Feld Schule getragen (und zu bildungspolitischen EntscheidungsträgerInnen). Gleichzeitig eröffnet die Praxis Perspektiven, die wiederum zur Weiterentwicklung von Theorien genutzt werden können. Es handelt sich somit um einen Prozess der wechselseitigen Inspiration. Durch die offeneren und kommunikativeren Strukturen responsiver Evaluationen, in denen verschiedene Stakeholder beitragen, wird dieser Aspekt besonders gefördert.
- (2) Durch das Wissen über innovative Veränderungsoptionen werden nachvollziehbare, fundierte Entscheidungen in spezifischen Kontexten eher möglich. Damit wird Wissenschaft ihrer gesellschaftlichen Verantwortung gerecht.
- (3) Es entstehen gerade in responsiven Designs intensive Vernetzungen mit dem Praxisfeld, die für die Weiterentwicklung der Forschung (z.B. andere



wissenschaftliche Untersuchungen, die eine gute Vertrauensbasis erfordern), aber auch für die LehrerInnenbildung genutzt werden können.

Diese Aspekte begründen nachvollziehbar die Notwendigkeit, in fachdidaktische Forschung auch Evaluationsstudien einzubeziehen. Die Fachdidaktiken sind jedoch gut beraten, wenn sie in Evaluationen – gleich welchen Typs – ihre wissenschaftliche Unabhängigkeit betonen und nicht auf dem Reflexionsniveau der Umsetzung stehenbleiben. So nennen auch Wottawa & Thierau (vgl. 2003, S. 40f.) folgende Zuschreibungen an wissenschaftliche Evaluationen, die profilbildend für diese Art von Evaluationen sind: Neutralität, fächerübergreifende Kompetenzzuschreibung, Explikation von Begründungen, Aufwertung (statusbezogen).

Erst eine theoretisch-konzeptionelle Reflexion der Evaluationsprozesse und -ergebnisse, die auch eine kritische Distanz zum Feld impliziert, steigert den Nutzen einer von WissenschaftlerInnen durchgeführten Evaluation sowohl für die Fachdidaktik als auch die Schulpraxis.

## Literatur

- Biesta, G. J. J. (2010). *Good Education in an Age of Measurement. Ethics, Politics, Democracy.* Boulder, London: Paradigm Publishers.
- bmu (2013). *Bundesqualitätsbericht der humanberuflichen und höheren land- und forstwirtschaftlichen Schulen für den Berichtszeitraum 2010/2011/3023 und den Planungszeitraum 2013/2014/2015*, Wien, Juni 2013.
- Guba, E. G. & Lincoln, Y. S. (1990). *Fourth Generation Evaluation*, third edition, Newbury Park: Cal., SAGE.
- Ostendorf, A. et al. (2012). *Evaluation und Begleitung des Schulversuchs Praxis HAS – Endbericht*. Universität Innsbruck.
- Saunders, M. (2012). The use and usability of evaluation outputs. A social approach. *Evaluation*, 18 (4), pp. 421-436.
- Wottawa, H. & Thierau, H. (2003). *Lehrbuch Evaluation*. 3. korr. Aufl., Bern: Huber.

