

Inhalt

Vorwort	7
1. Warum eine qualitative Evaluation?	11
2. Qualitative Evaluation in sieben Schritten	15
Schritt 1: Evaluationsgegenstand und Evaluationsziele festlegen	16
Schritt 2: Interviewleitfaden und Kurzfragebogen entwickeln	20
Schritt 3: Interviews durchführen, aufnehmen und transkribieren	24
Schritt 4: Daten erkunden, fallweise darstellen	33
Schritt 5: Das Kategoriensystem erstellen und die Interviews codieren	36
Schritt 6: Kategorienbasiert auswerten und Evaluationsbericht erstellen	43
Schritt 7: Fazit erarbeiten, Ergebnisse rückmelden, Bericht abschließen.....	50
3. Reflexion der qualitativen Vorgehensweise	59
3.1 Die Ergebnisse der standardisierten quantitativen Evaluation.....	59
3.2 Der Mehrwert der qualitativen Evaluation	66
3.3 Potenzielle Anwendungsfelder	73
4. Praktische Details der qualitativen Evaluation	77
4.1 Das Material: Ein transkribiertes Interview	77
4.2 Die Codierung: Ein codiertes Interview.....	83
4.3 Das Ergebnis: Kategorienbasierte Auswertung als Herzstück der qualitativen Evaluation	88
5. Ressourcen zum Thema Evaluation	105
Ausgewählte Literatur	105
Ausgewählte Internet-Seiten.....	107
6. Arbeitshilfen für die praktische Umsetzung	109
6.1 Zeitübersicht	109
6.2 Checkliste „Qualitative Evaluation in sieben Schritten“	111
Literatur	117

Vorwort

Die Idee zu dieser Studie entstand im Oktober 2005 anlässlich der Jahrestagung der Deutschen Gesellschaft für Evaluation (DeGEval). Wolfgang Beywl regte dort gegenüber Thorsten Dresing, einem der Autoren der hier vorliegenden Studie an, man solle doch einmal versuchen, möglichst praxisnah zu beschreiben, wie man unter Einsatz digitaler Transkriptionsverfahren und der Methode der computergestützten qualitativen Datenanalyse in vertretbarer Zeit eine qualitative Evaluation durchführen könne. Man einigte sich darauf, dass eine Schritt-für-Schritt-Beschreibung der Vorgehensweise bei einem maximalen Zeitumfang von 100 Stunden das wäre, was man benötige. So ließen sich vielleicht die bei vielen Kolleginnen und Kollegen vorhandenen Schwellenängste gegenüber qualitativen Methoden in der Evaluation überwinden. Schließlich würden viele nicht zuletzt deshalb vor qualitativen Evaluationen zurückschrecken, weil sie einen großen Zeitaufwand vermuten und die einzelnen Schritte bei qualitativen Verfahren nicht so recht klar seien. Die Vorurteile „ausufernde Interpretationen“ einerseits und „immense Zeitbudgets“ andererseits erscheinen nicht unberechtigt, denn in der Tat sind viele der gängigen qualitativen Vorgehensweisen mit hohem Zeitaufwand verbunden (vgl. Flick 2006: 21) und im Rahmen von Evaluationen mit ihrem normalerweise engen Zeitrahmen schwerlich realisierbar. Interessant wäre es nun, möglichst praxisnah zu demonstrieren, dass man auch mit einem abgekürzten Verfahren¹ gute Resultate erzielen kann.

1 Um Missverständnisse zu vermeiden, wird hier auf den Begriff „Abkürzungsstrategie“ bewusst verzichtet. Dieser entstammt dem Umfeld der Objektiven Hermeneutik und wurde unseres Wissens von Christian Lüders (2006) in den Evaluationsdiskurs eingebracht und u. a. von Uwe Flick aufgegriffen. Unser Methodenverständnis bewegt sich nicht im Rahmen der Objektiven Hermeneutik, sondern im Rahmen des theoretischen Codierens, wie man es etwa in der Grounded Theory oder auch in der qualitativen Inhaltsanalyse wiederfindet.

Die Idee zu einem solchen „Demonstrationsvorhaben“ hat in unserer Arbeitsgruppe an der Philipps-Universität Marburg sofort großes Interesse hervorgerufen und wir nahmen uns vor, diese Projektidee möglichst bald zu realisieren. Im Bereich von Bildung und Erziehung, in dem wir tätig sind, haben wir daraufhin nach einem geeigneten Evaluationsgegenstand gesucht. Dabei sollte es sich möglichst nicht um ein „exotisches“ Feld handeln, sondern um eines, das vielen Kolleginnen und Kollegen vertraut ist und sich deshalb auch gut als Demonstrationsbeispiel eignet. Wir entschieden uns für die Evaluation universitärer Lehre, ein Bereich, in dem schon seit etwa zehn Jahren Evaluationen stattfinden und von dem angenommen werden kann, dass viele Rezipientinnen und Rezipienten unserer Studie in der ein oder anderen Weise eigene Erfahrungen mit Evaluationsverfahren gemacht haben.² Lehrevaluation findet geradezu massenhaft statt: Tausende Lehrveranstaltungen, Dozenten und Studierende sind in jedem Semester davon betroffen. Insofern erschienen uns – zumindest theoretisch – die Bedingungen für die Übertragbarkeit unseres Vorgehens auf andere Orte und Situationen günstig zu sein.

Schon Anfang 2006 haben wir dann die Idee in die Tat umgesetzt. Wir haben uns von vornherein ein zeitliches Limit gesetzt, nämlich 100 Stunden, die insgesamt für die Evaluation „verausgabt“ werden sollten. Die knappen zeitlichen Ressourcen sollten allerdings nicht zu Lasten der Qualität gehen, jeder einzelne Schritt des Projektes sollte sorgfältig dokumentiert werden – auch und gerade hinsichtlich der benötigten Zeit. Als Arbeitstitel hatten wir gleich zu Beginn „Quick and Clean in 100 Stunden“ gewählt, worin zum Ausdruck kam, dass wir den Anspruch hatten, trotz der gebotenen Schnelligkeit methodische Strenge an den Tag zu legen. Für alle Projektmitarbeiter stellte diese Studie auch ein Experiment dar, denn wir wussten am Anfang ja tatsächlich nicht, ob das Vorhaben innerhalb dieses engen Zeitrahmens auch realisierbar wäre.

Hier liegt nun das Ergebnis als Buch vor Ihnen. Wir selbst sind durchaus überrascht, wie viel wir mit dieser kleinen qualitativen Studie herausbekommen haben. Aus unserer Sicht hat sich die Arbeit gelohnt und wir glauben, ruhigen Gewissens auch anderen Kolleginnen und Kollegen durchaus zu qualitativen Evaluationsstudien raten zu können. Unser Text

² Auch ein Blick in die letzten Jahrgänge der Zeitschrift für Evaluation zeigt, dass kaum ein Feld so häufig in Artikeln abgehandelt wird wie das der Hochschulevaluation.

richtet sich primär an Personen, die selbst qualitative Evaluation betreiben wollen und eine detaillierte Anleitung für die praktische Umsetzung suchen. Die mit diesem Bericht verbundene Absicht zielt also in erster Linie auf die Methoden der Evaluation und nicht auf den Gegenstandsbereich „Evaluation von Hochschullehre“³, wengleich unsere Studie auch hierzu interessante Aspekte enthält.

An dieser Studie haben neben den vier Autoren Thorsten Dresing, Udo Kuckartz, Stefan Rädiker und Claus Stefer auch Lena Lehmann, Thomas Ebert und Stefanie Zanetti als studentische Hilfskräfte mitgearbeitet. Für ihr großes Engagement möchten wir uns hiermit herzlich bedanken. Dem eingangs bereits erwähnten Wolfgang Beywl danken wir nicht nur für die Geburtshilfe bei der Idee zu diesem Vorhaben, sondern auch für seine sehr hilfreichen Anregungen und die ausführlichen Rückmeldungen zu den ersten Entwürfen des Textes. Sabine Lauber, Katia Tödt, Heiko Grunenberg, Patrick Holland-Moritz, Udo Kelle und Helmut Kromrey danken wir für ihre kritische und konstruktive Lektüre des Manuskripts.

Die Daten der Evaluation stellen wir Interessierten – selbstverständlich anonymisiert – für Sekundäranalysen oder auch einfach nur zum Nachvollziehen der vorgetragenen Ergebnisse gerne zur Verfügung.

3 In diesem Kontext verweisen wir auf Kromrey (2001).

1. Warum eine qualitative Evaluation?

Ausgangspunkt für diese Studie war, wie oben beschrieben, die Idee, ein Demonstrationsvorhaben auszuführen, das im Zeitrahmen von 100 Stunden einen Mehrwert über die üblichen quantitativen Evaluationen hinaus erwirtschaftet. Mit dem Begriff „Mehrwert“ war zunächst nur eine unbestimmte Erwartung verknüpft. Qualitativen Methoden schreibt man üblicherweise eine größere Offenheit und eine Berücksichtigung der Perspektive der Beteiligten zu. So war es also nur natürlich, auf einen Zugewinn gegenüber der standardisierten Evaluation zu hoffen. Wie genau dieser allerdings aussehen würde, darüber waren zu Beginn des Projektes nur Mutmaßungen möglich, die Ergebnisse der Studie sollten es möglich machen, hierzu Präziseres zu äußern. Als Evaluationsobjekt haben wir die vierstündige Lehrveranstaltung „Einführung in die sozialwissenschaftliche Statistik“ ausgewählt, eine Pflichtveranstaltung im Marburger Diplomstudiengang Pädagogik mit annähernd 200 Teilnehmerinnen und Teilnehmern.

Seit vielen Jahren finden an bundesdeutschen Universitäten Evaluationen von Lehrveranstaltungen statt. Nicht anders verhält es sich an der Philipps-Universität Marburg, wo unter Federführung des Instituts für Psychologie ein standardisiertes Verfahren etabliert wurde, bei dem die Studierenden am Semesterende einen zweiseitigen Fragebogen ausfüllen, in dem zu vorgegebenen Statements auf einer vier- oder fünfstufigen Skala Zustimmung oder Ablehnung geäußert werden kann. Die Statements lauten etwa „Der/Die Dozent/in gestaltet die Vorlesung interessant“ oder „Der/Die Dozent/in fördert mein Interesse am Themengebiet“. Die Auswertung der Fragebögen geschieht in einer ebenfalls standardisierten Art und Weise: Als Feedback erhält der Dozent einen mehrseitigen Ausdruck, der die Verteilung und die Mittelwerte der einzelnen Items enthält. Gleichzeitig wird er mit Vergleichsdaten versorgt, kann seine Mittelwerte mit denen des gesamten Fachbereichs oder gar der gesamten Uni vergleichen. Die Antworten auf die offenen Fragen erhält der Dozent als Rohdaten, das heißt in Form von abgeschnittenen Fragebogenschnipseln, so dass er ggf. diese Antwort-

ten nochmal durchlesen kann, sofern er sich denn für diese „qualitativen Daten“ interessiert. Schon diese Rücklieferung in Form nicht bearbeiteter, bloß ausgeschnittener Papierabschnitte zeigt, dass die Bearbeitung qualitativer Daten erheblich aufwändiger ist als diejenige von standardisierten Daten, jedenfalls so aufwändig, dass sie nicht als Service erstellt wird, sondern vom Dozenten selbst vorgenommen werden müsste. Er wird dies normalerweise nur cursorisch tun, denn der Aufwand ist enorm: Die nicht immer gut leserlichen Antworten müssen zunächst transkribiert, ein Kategoriensystem entwickelt und die Antworten codiert werden, d. h. den einzelnen Kategorien zugeordnet werden.

Die quantitativen Daten inklusive der Analyse werden hingegen mehr oder weniger vollautomatisch bearbeitet: Die Fragebögen werden eingescannt, automatisch in eine Datei transferiert, die Mittelwerte von einem Statistikprogramm erstellt und die Vergleiche zu den Resultaten von anderen Lehrveranstaltungen des Fachbereichs und der Gesamtuniversität automatisch generiert. Es lassen sich sogar automatisch entsprechende statistische Tests, z. B. T-Tests oder Varianzanalysen berechnen und signifikante Abweichungen im Text etwa durch Fettdruck, Sternchen o. Ä. hervorheben. Am Ende bleiben allerdings viele Fragen offen, etwa ob denn ein Seminar mit besserem Mittelwert das bessere Seminar ist. Ob man eine Vorlesung „Recht der internationalen Organisationen“ eigentlich mit „Einführung in die Filmanalyse“ vergleichen kann oder was es eigentlich zu bedeuten hat, wenn einer Vorlesung, die dem exakt gleichen Veranstaltungsplan folgt, in zwei aufeinander folgenden Jahren ein signifikant unterschiedlicher Strukturierungsgrad bescheinigt wird.

Das Unterfangen, neben der standardisierten quantitativen Evaluation auch einmal qualitative Methoden einzusetzen, ist vor allem als Versuch zu verstehen, solche Ungereimtheiten zu klären und die offenen Fragen zu beantworten. Was kann eine andere Vorgehensweise für neue, bessere Perspektiven hervorbringen? Wir erhofften uns, über die bloßen Antwortverteilungen und Mittelwerte hinaus ein *besseres Verständnis für die Antworten in ihrem jeweiligen Kontext*. Das bedeutet, den Studierenden nicht in der Rolle des Preisrichters beim Eiskunstlauf zu sehen, der seine 5,6 oder 5,8 als Bewertungsnote aus seinem Köfferchen zieht, sondern seine Motivation zu erfragen, etwas über den Lern- und Arbeitsstil, die persönlichen Ambitionen und Aspirationen in Erfahrung zu bringen. Das bedeutet auch, dass es nicht um eine klassische *Bewertung* geht, sondern um die Verbesserung der

evaluierten Prozesse, also in diesem Fall um eine Verbesserung der Lehrveranstaltung und eine Optimierung des Lernerfolgs der Studierenden.

Unser primäres Ziel war also herauszufinden, ob eine qualitative Evaluation auch unter stark restriktiven Bedingungen (100 Stunden) relevante Ergebnisse erbringen kann. Es ging uns dabei jedoch nicht darum, ein neues Evaluationsdesign oder gar (qualitatives) Paradigma zu entwickeln. Unser Hauptaugenmerk lag vielmehr auf der Praxis, für die wir eine Anleitung für den effizienten und effektiven Einsatz qualitativer Methoden erarbeiten wollten. Deshalb erschien es uns auch wenig sinnvoll, unsere methodischen Vorschläge einem bestimmten Evaluationsmodell zuzuordnen, da qualitative Methoden unabhängig von einem Modell Anwendung finden.⁴

Wenn man, Flick (2006: 18) folgend, qualitative Evaluationsforschung als Anwendung qualitativer Methoden in einem Evaluationsprojekt versteht, dann versteht sich bei Kenntnis der Vielfalt qualitativer Methoden von selbst, dass qualitative Evaluationsforschung diese immense Vielfalt notwendigerweise „geerbt“ hat. Für Autoren eines Buchs über qualitative Evaluationsforschung ergibt sich daraus die Verpflichtung offenzulegen, wo sie sich im Spektrum qualitativer Methoden bewegen. Das Methodenverständnis der Autoren dieses Buches lässt sich als ein pragmatisches kennzeichnen, das sich bspw. in den von Mayring dargestellten Säulen des qualitativen Denkens, den von Flick herausgearbeiteten Charakteristika der qualitativen Forschung und dem pragmatischen Ansatz von Kelle wieder findet. Nach unserem Verständnis ist es die Forschungsfrage, der die zentrale Bedeutung im Forschungsprozess zukommt. Die zur Beantwortung der Forschungsfrage eingesetzten Methoden haben sich als *gegenstandsangemessen* zu legitimieren. Eine prinzipielle Gegenüberstellung von qualitativen und quantitativen Methoden ist deshalb „sachlich“ unangebracht. Kelle hat überzeugend argumentiert, dass die Erklären-Verstehen-Kontroverse dann, wenn man sich eine handlungstheoretische Perspektive zu eigen macht, erheblich an Schärfe verliert, da Handlungsregeln keine universellen Gesetzmäßigkeiten darstellen, sondern zeitlichen und örtlichen Beschränkungen unterliegen. Handlungsbedingungen, Handlungsziele und Handlungsregeln können zwischen verschiedenen Handlungskontexten erheblich variieren und zudem von kompetenten Akteuren in gewissen Grenzen modifiziert

4 Eine Übersicht über die Verwendung qualitativer Methoden in unterschiedlichen Evaluationsmodellen findet sich in Beywl (2006).

werden. Man muss also jeweils damit rechnen, dass unbekannte, nicht antizipierte Wirkungen und Nebenwirkungen auftreten. Die vollständige Erfassung komplexer Wirkungsgefüge mit ausschließlich standardisierten Instrumentarien, d. h. die Erfassung aller wirksamen unabhängigen Variablen, ist nicht möglich. Es bedarf jeweils auch qualitativer Verfahren, um solche im Vorwissen und bei der Konzeption von Evaluationen nicht präsenten Sachverhalte überhaupt entdecken zu können (vgl. Kelle 2006: 120 ff.). Tatsächlich wird ja auch und gerade in der Evaluationsforschung häufig und gerne mit einem Methodenmix („Mixed-Methodology-Forschung“) gearbeitet. Die qualitative Methodik, die in den folgenden Kapiteln beschrieben wird, ist vorrangig an Systematisierung und der Entdeckung von sozialen Regelmäßigkeiten interessiert. Sie ist eher an Techniken des theoretischen Codierens orientiert und weniger an textexegetischen oder sequenzanalytischen Verfahren. Es ist diese, in ihrer – ebenfalls vielfältigen – Tradition auf Max Weber, Paul F. Lazarsfeld, Kurt Lewin, Christel Hopf und viele andere zurückgehende Variante von qualitativen Methoden, deren Ertrag in Evaluationsprojekten durch dieses Demonstrationsvorhaben unter Beweis gestellt werden soll.