Die Evaluation von »Achtung (+) Toleranz« – eine Methodenreflexion

Andreas Hirseland, Oliver Dimbath, Julia von Hayek, Werner Schneider

In diesem Beitrag möchten wir näher auf die im Zusammenhang mit dem Praxisbericht der Evaluation von »Achtung (+) Toleranz« vorgestellte Evaluationsmethode eingehen. Neben einer Beschreibung des Evaluationsprozesses wird es dabei auch um die methodologische Begründung des von uns gewählten Vorgehens, die Verfahren zur Datenerhebung, -auswertung und Ergebnisinterpretation gehen. Abschließend erfolgt eine kurze praxisbezogene Einschätzung der Eignung derartiger Verfahren für die Evaluation von Programmen der Demokratie- und Toleranzerziehung sowie eine Reflexion der erhaltenen Untersuchungsergebnisse als immer auch »methodische Artefakte«.

Im Zentrum des Evaluationsauftrages stand, neben dem Interesse an der Akzeptanz – »wie das Seminarangebot bei den Teilnehmern ankommt« –, die Frage nach den »Wirkungen« von »Achtung (+) Toleranz« und deren Nachhaltigkeit. Der Versuch, eine Antwort auf diese zunächst so klar und eindeutig erscheinende Fragestellung zu finden, erweist sich bei näherer Betrachtung als problematisches Unterfangen:

Eine einfache Kausalität zwischen Maßnahme und »output« oder »outcome« außerhalb experimenteller Untersuchungsdesigns ist kaum nachweisbar (vgl. hierzu Lynen von Berg und Hirseland in diesem Band und die dort angegebene Literatur). Gerade bei Programmen und Maßnahmen der politischen Bildung mit ihren komplexen, eher vage definierten und daher schwer operationalisierbaren Zielsetzungen – wie etwa »Toleranz« – ist nicht per se klar, worin genau die »Wirkung« bestehen solle und welche Wirkungen überhaupt auftreten. Es ist nicht per se klar, wie welche Wirkungszusammenhänge überhaupt feststellbar oder gar »messbar« und da es um Evaluation geht – schließlich bewertbar wären.[1]

1 Zur Unterscheidung von Feststellung und Messung mag folgendes Beispiel dienen: »Wetter« ist im meteorologischen Sinne durchaus anhand der Größen Luftdruck, Luftfeuchtigkeit, Temperatur und Sonnenscheindauer messbar und auf ei-

Derartige Fragestellungen lassen sich weder von den mit Evaluation beauftragten Wissenschaftlern noch von den Auftraggebern allein lösen, sondern bedürfen einer konsensuellen Klärung, d.h. eines Vorgehens, das als »partizipativ« begriffen werden kann (vgl. auch den Beitrag von Bommes et. al. in diesem Band).

Unser Verständnis von Partizipation ist ein »dialogisch-diskursives«, das darauf gerichtet war, in einem beständigen Austausch mit Auftraggebern und Programmentwicklern ein gemeinsames Verständnis der expliziten theoretischen Grundlagen des Programms ebenso wie der diesem unterlegten impliziten Vorannahmen etwa über seine Wirkungsweise zu gewinnen.

Auf diesen Grundlagen erst ließen sich sinnvolle und somit gegenstandsangemessene Kriterien und Forschungsfragestellungen definieren und Transparenz nicht nur des wissenschaftlichen Vorgehens, sondern auch der jeder Methode zugrunde liegenden erkenntnistheoretisch-methodologischen Voraussetzungen schaffen. Dieses Wissen ist Bedingung für die sachgerechte Einordnung und Beurteilung der erhobenen Befunde und ihrer konsensuellen Validierung durch Auftraggeber, Programmentwickler und Forschergruppe (für einen kurzen Abriss der in diesem Prozess zu bearbeitenden Probleme vgl. Wenzel 2003: 18–20).

Beginn der Evaluation: Exploration

Aus den genannten Gründen bestand die erste Stufe der Evaluation darin, zunächst (a) im Rahmen einer qualitativen Vorerhebung Programmmaterialien und andere Dokumente auszuwerten sowie (b) in explorativen Expertengesprächen mit Programmautoren und Seminarleitern eine Rekonstruktion des Programms vorzunehmen.

Bei dieser ersten Stufe ging es darum, das Programm und seine Übungsteile kennen zu lernen, d.h. allgemeines Wissen darüber anzusammeln, was das Programm »Achtung (+) Toleranz« überhaupt »ist«. Der zweite Schritt diente dazu, dieses Wissen zu vertiefen und am Selbstverständnis der Anbieter zu überprüfen, indem die dem Programm zugrunde liegenden Intentionen und Motive[2], Wirkungshypothesen und die daraus resultierenden Er-

ner »Mehr-Weniger-Skala« darstellbar. Subjektiv erfahrene Wirkungen des Wetters hingegen drücken sich in feststellbaren Qualitäten aus (schönes Wetter), die zu messen problematisch ist: War der eine Frühlingstag ein schönerer Frühlingstag als der andere, ist bei ansonsten gleichen klimatischen Bedingungen eines Sommers das Wetter in Italien nicht doch immer irgendwie ein wenig »besser« als am Starnberger See? (vgl. hierzu auch Hirseland in Wenzel 2003: 12–17)

wartungen und Ziele in Bezug auf Effekte einzeln durchgeführter Maßnahmen und Programmbausteine in Erfahrung gebracht, expliziert und zusammen mit Auftraggebern und Programmentwicklern reflektiert wurden.

Auf dieser Grundlage wurden dann (c) gemeinsame Erwartungen an die Evaluation formuliert und ihre Zielsetzung definiert sowie (d) ein Evaluationsdesign entwickelt, das hinsichtlich seiner Gegenstandsangemessenheit und praktisch-organisatorischen Umsetzung an die besonderen Erfordernisse von »Achtung (+) Toleranz« angepasst war.

Ein Ergebnis dieser Auseinandersetzung war die Erarbeitung eines für die Evaluation maßgeblichen Grundverständnisses von Bildungsprozessen in der politischen Bildung, insbesondere im Zusammenhang mit »Achtung (+) Toleranz«. Bildungsprozesse in diesem Bereich funktionieren nicht nach einem behaviouristischen Modell, etwa dergestalt, dass das Seminarprogramm als Input in die (immer unterschiedlichen) Teilnehmer eingespeist wird und dann auf der anderen Seite ein für alle gleiches, beobachtbares Ergebnis als Wirkung herauskäme. In diesem Modell würden die Teilnehmer als »black boxes« begriffen, die durch die Programmteilnahme auf »programmierte« Weise geformt würden. Dies ist weder beabsichtigt, noch wäre es machbar.

Bildungsprogramme wie »Achtung (+) Toleranz« sind im Gegensatz dazu als relativ offene Angebote zu verstehen, die von unterschiedlichen Teilnehmern eben auch unterschiedlich aufgefasst und begriffen, d.h. mit je subjektiver Bedeutung versehen und dementsprechend auf die je eigenen Alltagswirklichkeiten bezogen werden.

Erwartbar ist daher, dass die Ergebnisse einer Teilnahme höchst unterschiedlich ausfallen, denn neues Wissen muss in subjektiven, »eigensinnigen« Aneignungsprozessen immer erst in bereits vorhandenes Wissen »eingebaut« werden, um praktische Bedeutung zu gewinnen. Handeln folgt nicht einfachen Reiz-Reaktion-Schemata, sondern ist stets in typischer Weise mit Intentionalität, mit einem auf andere bezogenen subjektiv gemeinten Sinn versehen.

2 Hier sind im Sinne einer phänomenologischen Soziologie so genannte »Weil-Motive« und »Um-zu-Motive« zu unterscheiden (vgl. Schütze und Luckmann 1994). Weil-Motive beziehen sich auf den Anlass zur Programmentwicklung, Um-zu-Motive auf die erwarteten, erhofften Effekte, die mit dem praktischen Einsatz des Programms verbunden sind.

Tiefenwirkung und Breitenwirkung: multimethodischer Mehrebenenansatz

Für die Evaluation bedeutet dies, dass Wirkungen zunächst dort auftreten, wo sie für den außenstehenden Beobachter unsichtbar sind: in den Veränderungen subjektiv bedeutsamer Wissensbestände. Hier liegt die Tiefenwirkung von durch die Seminarteilnahme hervorgerufenen Bildungsprozessen. Da diese zunächst nur am einzelnen »Fall« festgestellt werden kann, beinhaltete das Evaluationsdesign (a) eine so genannte »Tiefenbohrung«, deren Ziel es war, am Beispiel einer Teilnehmergruppe eines »Achtung (+) Toleranz«-Trainings typische Formen und Muster der Veränderungen von Wissensbeständen und deren alltagspraktische Umsetzung – in veränderten Formen des Deutens und Handelns im Alltag – zu rekonstruieren.

Daneben bestand Interesse daran, einen Überblick über die Akzeptanz und Nachhaltigkeit von »Achtung (+) Toleranz« bei allen bisherigen Teilnehmern zu gewinnen, also gewissermaßen als Ergänzung zur fallbezogenen detaillierten Tiefenbohrung (b) die Fläche oder Breitenwirkung zu erfassen und zu dokumentieren. Hier ging es um die Einbeziehung möglichst großer Fallzahlen, was methodisch mit vertretbarem Aufwand nur durch eine postalische Befragung mittels eines teilstandardisierten Fragebogens möglich war.

Beide Teile zusammengenommen, ergaben das Evaluationsdesign eines multimethodischen Mehrebenenansatzes, mit dessen Hilfe wir eine möglichst umfassende Beschreibung und Bewertung relevanter Prozesse und Wirkungen erreichen wollten (vgl. Abb. 1 auf S. 63).

Teilstandardisierter Fragebogen

Der Einsatz standardisierter Verfahren eignet sich wegen der hohen Reliabilität und Objektivität des Verfahrens vorrangig zu einer Evaluation, die vorab definierte Reaktionen und Effekte mittels statistischer Kennwerte erfasst, und eröffnet darüber hinaus die Möglichkeit statistisch-repräsentativer Aussagen.

Als Nachteil eines derartigen Vorgehens kann sich hingegen erweisen, dass die Befragten eben nur auf das antworten können, was den Evaluatoren vorab als bedeutsam und fragenswert erschienen ist und dies lediglich in der durch die standardisierten Antwortmöglichkeiten vorgegebenen Form. Das ist insbesondere dann von Nachteil, wenn es um die Entdeckung unerwarteter, nicht-intendierter Wirkungen bei den Teilnehmern oder um die Erfassung komplexer Wissensbestände und Handlungskompetenzen gehen

soll. Durch Einbeziehung so genannter offener Fragen, die den Befragten die Möglichkeit selbst gewählter Antworten geben, können die genannten Probleme vermindert, aber letztlich nicht hinreichend gelöst werden, so dass teilstandardisierte Befragungen keine Alternative zu dem erwähnten Einsatz qualitativer Fallstudien darstellen.

Wir haben das Instrument der postalischen Befragung eingesetzt, um in schriftlicher Form alle ehemaligen »Achtung (+) Toleranz«-Teilnehmer zu erfassen und auf diese Weise einen Überblick über allgemeine Einschätzungen des Seminarprogramms zu gewinnen und um anhand offener Fragen auch ausgewählte Felder subjektiver Bedeutungszuweisungen beschreibbar zu machen. Der hierzu entwickelte, sehr umfangreiche und daher eine hohe Motivation und Kooperationsbereitschaft der Respondenten voraussetzende Fragebogen deckte insgesamt sechs Fragebereiche ab:

- Rückblick auf »Achtung (+) Toleranz«
- Erwartungen an das Seminar »Achtung (+) Toleranz«
- Themen und Inhalte von »Achtung (+) Toleranz«
- Die Personen im Seminar
- Bewertungen und Urteile, Feedback
- Angaben zur Person

Die Auswertung der geschlossenen Fragen erfolgte mittels verschiedener statistischer Methoden. Verfahren der deskriptiven Statistik, etwa Häufigkeitsdarstellungen und Mittelwertberechnungen, ermöglichen die Darstellung der Verteilung von Merkmalsausprägungen der erhaltenen Antworten innerhalb des Sample. Auf diese Weise ließen sich etwa Kennzahlen für das Maß der Akzeptanz bestimmen (z.B. »Würden Sie die Veranstaltung weiterempfehlen?«).

In weiteren Untersuchungsschritten wurden diese Angaben auch mit Hilfe von Verfahren der korrelativen Statistik in Bezug gesetzt zu anderen untersuchungsrelevanten Merkmalen des Samples. Auf diese Weise konnten Fragen danach beantwortet werden, inwieweit das Programmkonzept etwa eine geschlechtsspezifische oder bildungsspezifische Ausrichtung hat: Kommt »Achtung (+) Toleranz« bei Frauen oder Männern, bei Teilnehmern mit höherer Bildung »besser« an?

Darüber hinaus haben wir, dort wo aufgrund theoretischer Vorüberlegungen ein »tieferer« Einblick erforderlich war, das Datenmaterial mit Hilfe einer explorativen Faktoranalyse untersucht. Unter inhaltlichen Gesichtspunkten bietet dieses Verfahren die Möglichkeit, latente, gewissermaßen »hinter« den Daten stehende Variablen aufzuspüren. Mit Hilfe dieses Vorgehens ließen sich die statistischen (quantitativen) Daten inhaltlich so interpretieren,

dass unter anderem für die Wirkungsweise von »Achtung (+) Toleranz« bedeutsame mehrdimensionale Typen der Konfliktwahrnehmung gebildet werden konnten (vgl. unseren Beitrag in diesem Band sowie Dimbath et al. 2003).

Dieses Beispiel zeigt, dass sich die Analysetiefe quantifizierender Verfahren in der Evaluation nicht in der bloßen Deskription erschöpft. Im Zusammenspiel von theoretischer Rekonstruktion des Programms und der ihm zugrunde liegenden Wirkungstheorie, deren Operationalisierung für die Verwendung im Fragebogen und dem Einsatz komplexerer statistischer Untersuchungsverfahren lassen sich auch komplexe Wirkungszusammenhänge entdecken.

Die Auswertung der offenen Fragen erfolgte mittels der aus der Grounded Theory (Strauss 1998) entlehnten Technik des Kodierens. Diese zielt darauf ab, die Daten »aufzubrechen« und zu einer gegenstandsbezogenen Theorie zu verdichten. Dazu werden die einzelnen erhaltenen Antworten in Codes »übersetzt« (Kodieren), d.h. ihre semantischen Felder und deren begriffliche Ordnung identifiziert und in sozialwissenschaftlich interpretierbare, begriffliche Kategorien transformiert (Dimensionalisieren und Kategorienbildung).

Kodieren als eine Analysetechnik, die ein Identifizieren, Dimensionalisieren und Kategorisieren der aus den Daten (re-)konstruierten begrifflichen Konzepte und ihrer wechselseitigen Bezüge beinhaltet, führt über den gesamten Analyseprozess hinweg zu einer fortschreitenden interpretativen Verdichtung und Abstraktion des Ausgangsmaterials.[3]

Um beim Kodieren längerer, komplexer Antworten (also solcher, die anstelle von Stichwortnennungen einen argumentativen Zusammenhang präsentierten) den Sinn des ursprünglichen Aussagenzusammenhangs nicht zu verlieren, wurden die betreffenden Textabschnitte zusätzlich einer hermeneutischen Sequenzanalyse unterzogen (vgl. einführend z.B. Reichertz 2000; Wernet 2000), um den Sinn des gesamten Aussagezusammenhangs zu erfassen und in das Kode-/Kategorienschema zu integrieren.

3 Die methodologische Grundlage für die von der Grounded Theory vorgeschlagene Kodiertechnik bildet ein Konzept-Indikator-Modell, nach dem so genannten empirische Indikatoren – d.h. in den Daten enthaltene Beschreibungen, Aussagen, Darstellungen von Verhaltensweisen, Ereignissen etc. – dem Forscher Möglichkeiten für eine Konzeptualisierung, ein Fassen dieser Indikatoren in sie bezeichnende Begriffe (»concepts«) anzeigen.
Indem der Forscher in seinem Analyseprozess eine Vielzahl solcher Indikatoren begrifflich bearbeitet (kodiert), erfolgt ein ständiger Vergleich, eine permanente, an den Daten arbeitende Systematisierung, Gruppierung, Klassifizierung und Dimensionalisierung dieser Indikatoren (zu Kodieren vgl. z.B. Böhm 2000: 475 ff.; Flick 1999: 196 ff.; Strauss und Corbin 1996; Strauss 1998).

Die auf diesem Wege gewonnenen Kategorien umreißen die Konturen der für die Befragten relevanten semantischen Felder und Rahmen, d.h. hier der von den Teilnehmern mit Blick auf das Seminar für relevant (mitteilenswert) erachteten Urteile und Bewertungen, erinnerten Erfahrungen usw.

Problemzentriert-diskursive Interviews

Um die individuellen, thematisch fokussierten Deutungen der Teilnehmer des Evaluationsseminars zu rekonstruieren, wurden mit den ihnen jeweils vor und – zur Feststellung der Nachhaltigkeit von Seminarwirkungen – einige Monate nach der Teilnahme an »Achtung (+) Toleranz« problemzentriert-diskursive Interviews geführt (vgl. Ullrich 1999; Witzel 2000).

Die Befragung sollte zum einen biographische Narrationen (Belegerzählungen) in Bezug auf Erfahrungen und Erlebnisse im Sinn- und Bedeutungskontext von Toleranz erzeugen, um darin die impliziten Wissensbestände und latenten Deutungsmuster der Befragten zu Toleranz, Achtung, Konflikten etc. rekonstruieren zu können. Zum anderen sollten die Interviews auch explizierbares Wissen in offener Form einholen (d.h. bisherige Erfahrungen mit ähnlichen Seminaren, Einstellungen und ihre Begründungen, Meinungsäußerungen etc.).

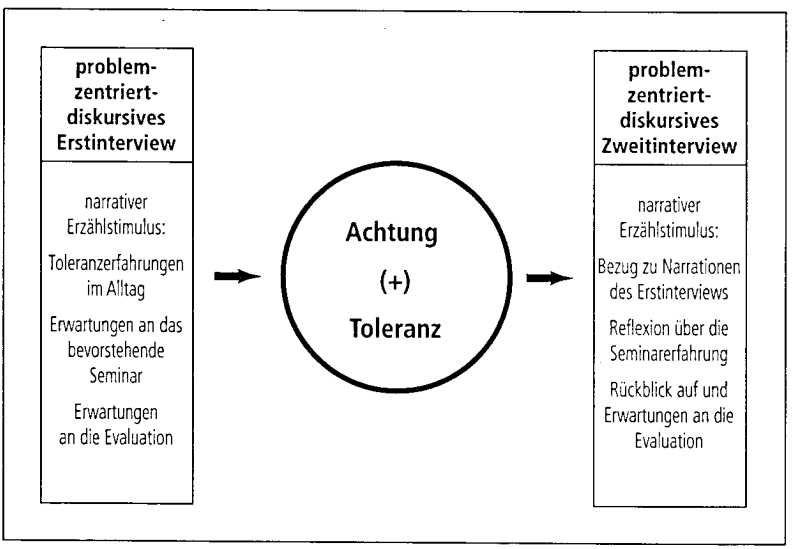
In den Nach-Interviews wurden die im ersten Interview berichteten, teilweise in Erzählungen gepackten konkreten Erfahrungen der einzelnen Seminarteilnehmer auf individuelle Weise in das zweite Interview mit einbezogen, um veränderte Interpretationszuschreibungen und Deutungen dieser Erfahrung aus der Sicht der Einzelnen zu ermitteln. Neben den problemzentrierten und diskursiven Abschnitten des Interviews wurde zum Schluss durch Nachfragen des Interviewers die Evaluation im Gesamtverlauf durch den Befragten reflektiert: Dies fand vor dem Hintergrund des »partizipativen« Interesses an der Perspektive der Teilnehmer statt und lotete nicht nur die subjektiven Einschätzungen hinsichtlich der Arbeit des Evaluationsprojekts, sondern auch mögliche Verbesserungsvorschläge aus.

Die Auswertung der Interviews erfolgte je nach erforderlicher (und damit theoretisch zu begründender) Analysetiefe gemäß unterschiedlicher Auswertungsstrategien. Durch inhaltsanalytische Vorgehensweisen konnten den Interviews wichtige Informationen ohne feinanalytische Textauslegung entnommen werden (z.B. biographische oder seminarbezogene Daten, Thematisierungen). Sequenzanalytische Verfahren der Interpretation wurden eingesetzt, um aus der sequenziellen Ordnung der in den Interviews enthaltenen Narrationen, Berichte und Reflexionen die subjektiven Bedeutungen zu re-

konstruieren, die der Interviewte dem Seminar und seinen Inhalten beimisst. Diese zeigen sich nicht nur auf der manifesten Ebene des inhaltlich Gesagten, sondern latent gerade in der Art und Weise des Erzählens, Darstellens, Begründens oder Bewertens.

Da Sequenzanalysen äußerst (zeit-)aufwändig sind, wurden sie nur an ausgewählten Schlüsselstellen eingesetzt. Die – dem eigenen Erkenntnisinteresse entsprechende und theoretisch begründbare – Auswahl dieser Stellen erfolgte auf der Grundlage einer umfassenden analytischen Bearbeitung jedes Interviews mittels eines an der Methodik der Grounded Theory (Strauss 1998) orientierten Kodierverfahrens (siehe hierzu auch die Auswertungsstrategie der offenen Fragen im Fragebogen).

Abbildung 1: Problemzentriert-diskursive Vorher- und Nachher-Befragung



Videoaufzeichnung und Gruppendiskussion

Der dritte Teil der Mehrebenenevaluation bestand in einem Video-Experiment – ein Weg, der in der qualitativen Evaluationsforschung bislang kaum beschritten wurde. Ziel war der Versuch, Effekte bzw. Wirkungen der Seminarteilnahme nicht nur aus Gesprächen mit Teilnehmern zu rekonstruieren, sondern durch die Beobachtung von Interaktionen. Damit wurde der Erwartung, die Inhalte von »Achtung (+) Toleranz« sollten sich über kognitive Veran-

kerungen und reflektierende Deutungen hinausgehend auch in konkret interaktivem Handeln niederschlagen, methodisch entsprochen.

Die Interaktionsbeobachtung per Video[4] fand zu zwei Zeitpunkten statt – unmittelbar vor dem Seminar (t_1) und einige Monate danach (t_2). Die Seminargruppe erhielt beim ersten Durchgang t_1 den Auftrag, gemeinsam mit Hilfe einer Videokamera (Spot-Cam) einen ca. einminütigen »TV-Spot« zum Thema Toleranz/Intoleranz zu drehen; in t_2 sollte ein Werbespot für »Achtung (+) Toleranz« erstellt werden. Der Zeitrahmen für die Bewältigung der Aufgabe betrug jeweils 30 Minuten.

Zusätzlich wurde die Gruppe aufgefordert, den Gruppenprozess mit Hilfe einer weiteren Videokamera (Doku-Cam) zu dokumentieren. Aus der Art der Dokumentation erwarteten wir, Aufschluss über die Perspektive der Gruppe auf sich selbst und ihren gemeinsamen Arbeitsprozess zu gewinnen.[5]

Ferner wurde in einer Ecke des Raumes eine dritte Kamera (Raum-Cam) installiert, die das ganze Geschehen aus der Vogelperspektive aufzeichnete.[6]

Zum Zeitpunkt t_1 wurden der Gruppe unmittelbar im Anschluss an die Dreharbeiten beide Filme (Spotcam und Dokucam) vorgeführt und von der Gruppe diskutiert. So wurde uns die gemeinsame Reflexion der Gruppe über ihre Arbeit und ihre Erklärung dessen, was in den Filmen zu sehen ist, in einer für weitere Auswertungen protokollierbaren Form zugänglich (Bandaufzeichnung).

4 Gemäß Kai Brauer und Christopher Dehn liegt »der enorme Vorteil einer filmischen Aufzeichnung darin, dass der Ablauf bestimmter Ereignisse konserviert wird und unbegrenzt reproduziert werden kann. ... Es handelt sich hier sozusagen um visualisierte Beobachtungsprotokolle« (Brauer und Dehn 1995: 76). In unserem Fall wurden die Daten noch dazu z.T. in Eigenregie der Teilnehmer (als Selbstbeobachtung) erstellt.

5 Dies lässt sich beispielsweise an folgenden Fragen festmachen: Wie deutet die Gruppe den Dokumentationsauftrag? Wie organisiert sie die Durchführung? Wie gehen die Kameraführer mit den Gruppenbeschlüssen um? Zusammengefasst: Was wird von wem wie mit der DokuCam aufgezeichnet? Insofern dokumentiert das von der Doku-Cam Aufgezeichnete nicht nur den Interaktionsprozess, sondern ist selbst ein Dokument des Interaktionsprozesses, der dieser Aufzeichnung zugrunde liegt.

6 Eine solche »Überwachungskamera«, die, in einer entlegenen Ecke des Raums platziert, während der Gruppenübung völlig aus dem Aufmerksamkeitsbereich der Übungsteilnehmer verschwindet, bedarf einer sensiblen Einführung (vgl. hierzu Renne, Dowrick und Wasek 1983). Sehr schnell kann der Eindruck entstehen, man werde mit dem Dokumentationsauftrag nicht ernst genommen oder solle intransparenten »Psychotests« unterzogen werden. Für eine Teilnehmerin des Projekts war die Einführung durch die Evaluatoren offenbar nicht sensibel genug, was sie im Rahmen des nachbereitenden Gruppengesprächs zu kritischen Äußerungen in Bezug auf die Einrichtung einer Raum-Cam bewog.

In t_2 kehrte sich die Vorgehensweise um. Hier fand das Gespräch nebst Betrachtung der alten Filme (Spot und Doku) in Verbindung mit einer gemeinsamen Rückbesinnung auf die besuchte Seminarveranstaltung vor der Gruppenübung statt. Erst danach erhielt die Gruppe wieder einen Arbeitsauftrag – Erstellen eines Werbespots für »Achtung (+) Toleranz« – und die beiden Kameras (Spot-Cam und Doku-Cam). Aufgrund dieser Umkehrung der Untersuchungsanordnung erwarteten wir, die Gruppe würde durch das vorangehende Gespräch u.a. »Seminarwissen« aktualisieren und dieses in die anschließende Lösung der praktischen Aufgabe einfließen lassen.

Das hier beschriebene Untersuchungsarrangement war auf die Konfrontation unterschiedlicher Perspektiven angelegt: Die Selbstbeschreibung des Arbeitsprozesses mittels der Doku-Cam durch die Gruppe selbst, die »objektive« Beobachtung des gesamten Gruppenprozesses mittels der fest installierten Beobachtungskamera sowie die Reflexion des Handlungsprozesses und des individuellen Handelns im Rahmen der Gruppendiskussionen. Das hätte im Sinne einer »Triangulation« (vgl. Flick 1999) von Beobachtungs- und Untersuchungsperspektiven eine multiperspektivische, sozusagen »360-Grad-Analyse« des Gruppenprozesses ermöglicht.

So ist die Raum-Cam in der Lage, Prozesse zu erfassen, die dem intentionalen Blick der Doku-Cam verborgen bleiben, und damit überhaupt erst deren Selektionsprozesse transparent zu machen. Verdeutlicht wird so der Zusammenhang zwischen dem Handeln »vor der Kamera«, das durch das »Handeln hinter der Kamera« (Doku-Cam) stets nur ausschnitthaft dokumentiert werden kann (vgl. Reichartz 1994). Die Transkripte der Gruppengespräche enthalten zudem zusätzliche Informationen über Gruppen- und Einzelentscheidungen.

Eine systematische Bearbeitung sämtlicher sich hierdurch eröffnender Analyseebenen ist äußerst ressourcenintensiv und war nach Absprache mit den Auftraggebern der Evaluation auch nicht beabsichtigt. Vielmehr handelte es sich bei der Einbeziehung dieses methodischen Konzepts selbst um ein über den eigentlichen Auftrag hinausgehendes, exploratives Experiment.

Evaluatoren wie Auftraggebern war von Anfang an unklar, wie weit dieses im Rahmen der vertraglich vereinbarten Ressourcen ausgewertet werden könnte. Um den Aufwand vertretbar zu halten, mussten wir es bei der Auswertung der Protokolle der Gruppendiskussionen und der Aufzeichnungen durch die Doku-Cam bei einer Ersteindrucksanalyse belassen und haben lediglich die mittels der Raum-Cam erfolgten »neutralen« Aufzeichnungen des Interaktionsprozesses einer genaueren Analyse unterzogen.

Bei der Auswertung des filmischen Materials bietet sich die Arbeit mit Einstellungs- und/oder Sequenzprotokollen an (vgl. hierzu Maier 1995). Da

die Anfertigung eines partiturähnlichen Verlaufsprotokolls, in dem jedes Wort notiert und jede Kamerabewegung bzw. Einstellung erfasst und vor dem Hintergrund ihrer Kontextbedingungen kommentiert wird, enorm aufwändig ist, haben wir für die Raum-Cam auf die kürzeren Sequenzprotokolle zurückgegriffen, mit deren Hilfe die Filmhandlung in sinnvolle Abschnitte gegliedert werden kann.

Entlang der Fragestellung, wie die Gruppe ihren Schaffensprozess strukturiert, wurden auf der Grundlage dieser Sequenzprotokolle Aktionsmemos erstellt, die zu einem späteren Fallvergleich zwischen Doku-Cam und Raum-Cam ebenso wie zwischen den beiden Aufzeichungszeitpunkten herangezogen wurden.

Beim Vergleich der beiden Gruppenübungen wurde der Fokus auf die Entwicklung der Kommunikation im Entscheidungsprozess gerichtet. Anhand einer Kurzbeschreibung der jeweils 30-minütigen Arbeitsphasen und der Konzentration auf die Sequenzprotokolle aus den beiden Raum-Cam-Filmen werden zunächst kommunikative und prozedurale Muster der Aufgabenbewältigung erstellt und anschließend miteinander kontrastiert.

Dieser Vergleich erfolgte vor dem Hintergrund einer Interpretationsfolie, die versuchte, sich auszumalen, welche Veränderungen im Sinne von Achtung (+) Toleranz idealerweise zu erwarten gewesen wäre. Das Ergebnis der Untersuchung ergab sich also durch eine trianguläre Gegenüberstellung der beiden Aufzeichnungszeitpunkte und des vermuteten Idealergebnisses.

Fazit

Den Kern des hier vorgestellten Evaluationskonzepts bildete die partizipative, dialogisch-diskursive Vorgehensweise, in der zusammen mit Auftraggebern und Programmentwicklern sowohl Zielsetzungen als auch Kriterien der Beurteilung von »Achtung (+) Toleranz« entwickelt wurden.

Im Zentrum dieses Prozesses stand die sorgfältige Analyse des Programms und die Rekonstruktion der ihm zugrunde liegenden Leitkonzepte (Welcher Toleranzbegriff wird unterlegt?) sowie der im didaktischen Aufbau enthaltenen Wirkungstheorie (Was soll welche Übung, welcher Trainingsteil erreichen, welches ist sein vermuteter Beitrag zur angestrebten Toleranzbildung?).

Auf der Grundlage dieser gemeinsamen Auseinandersetzung wurde als Rahmen für die spätere Analyse und Einschätzung eine gemeinsame Theorie des Toleranzbildungsprozesses entwickelt. Bereits vorab wurde deutlich, dass die erwartbaren Wirkungen einer Teilnahme nicht vorab eindeutig bestimm-

bar sein würden. Von daher erschien eine bloße Abfrage vorab operationalisierter Wirkungsdimensionen allein als nicht hinreichend. Dem wurde durch die Entwicklung des hier vorgestellten multimethodischen Mehrebenenansatzes Rechnung getragen.

Der Einsatz eines teilstandardisierten Fragebogens erbrachte Einsichten über die allgemeine Akzeptanz des Programms bei den Teilnehmern, über deren Erwartungen und deren Einschätzungen (Beurteilung) der im Seminar ablaufenden Gruppen- und Bildungsprozesse, die Qualität der Vermittlung von Seminarinhalten sowie Hinweise auf besonders gut erinnerte Seminarbestandteile und deren Nachhaltigkeit. Diese Fragen hängen inhaltlich miteinander zusammen: So ist die Akzeptanz der didaktischen Form sicherlich mit dafür ausschlaggebend, ob vermittelte Inhalte auch auf positive Resonanz stoßen und »angenommen« werden; die persönliche Zufriedenheit mit dem Ablauf insgesamt beeinflusst über positive Erinnerungen die Nachhaltigkeit etc. Derartige Erkenntnisse dienen dazu, eine erste Bewertung des Programms zu ermöglichen.

Dennoch erfassen die mit Hilfe derartiger Erhebungs- und Auswertungsverfahren gewonnenen Einsichten eher die »Oberfläche« rückblickend erteilter manifester Meinungsäußerungen und Zustimmungsbekundungen. Ein Einblick in die Wirkungsweise des Programms selbst lässt sich auf diesem Wege nicht gewinnen.

Hier setzte die Fallstudie an. Anhand der Interviewprotokolle konnten Prozesse der durch »Achtung (+) Toleranz« angestrebten Sensibilisierung für Vielfalt und Relationalität von Werturteilen, für das Vorhandensein stillschweigender Annahmen, die der je eigenen Wirklichkeitsdeutung unterlegt sind, und für eine Veränderung des subjektiven Toleranzverständnisses aufgezeigt werden.

Bedeutsam an dieser Form der Rekonstruktion war, dass sie über die Klassifizierung einzelner Veränderungen hinauswies und die Erstellung von Typologien der Wissensaneignung und des Wissenstransfers ermöglichte. Diese machen die Logik des Bildungsprozesses transparent und erlauben eine Einschätzung von Seminarerfolgen auch dort, wo manifeste Verhaltens- oder Einstellungsänderungen nicht oder noch nicht beobachtbar sind.

Wie weit die Teilnahme an einem »Achtung (+) Toleranz« Seminar wirkt, ist demzufolge nicht nur vom Programm und seiner Durchführung selbst, sondern ebenso von den reflexiven und sozialen Kompetenzen der daran Teilnehmenden abhängig. Wie man die erreichten Ergebnisse beurteilt, ist dann nicht nur eine Frage objektivierter Wertungsskalen, sondern bemisst sich an den (aufgrund der Typologien beschreibbar gemachten) Entwicklungsprozessen, die bei den Teilnehmern ausgelöst wurden.

Das Video-Experiment brachte demgegenüber noch eine weitere Dimension zum Vorschein, die mit den beiden genannten Verfahren nicht hinreichend erfasst werden kann, die jedoch gerade für das Feld der Toleranz von ausschlaggebender Bedeutung ist: die Eigendynamik des sozialen Prozesses bzw. sozialer Prozesse. Toleranz ist weniger in Form einer Einstellung bedeutsam, sondern letztlich dort, wo es situativ um das praktische Handeln in sozialen Zusammenhängen geht.

Welche Eigenlogik hier zum Tragen kommt, deuten die auf den ersten Blick paradox anmutenden Ergebnisse der Videobeobachtung an (vgl. unseren Praxisbericht in diesem Band). Soziologisch ist dies nicht weiter verwunderlich. Unserer Auffassung nach hat dieses Ergebnis jedoch weiter reichende methodisch-methodologische Implikationen für die Evaluation von auf soziale Kompetenzsteigerung gerichteten Programmen und Maßnahmen der politischen Bildung.

Es zeigt, dass das Soziale eine eigenständige Wirklichkeitsebene darstellt, die nicht als Summe ihrer einzelnen Teile begriffen werden kann. Von daher sind in diesem Zusammenhang Verfahren, die lediglich individuelle Meinungen, Auffassungen und Beurteilungen erfassen, ungeeignet, das für das Zusammenleben letztlich ausschlaggebende Soziale methodisch adäquat in den Blick zu nehmen – denn: Denken und Handeln sind nicht (immer) ein und dasselbe.

Literatur

Böhm, Andreas (2000): »Theoretisches Codieren. Textanalyse in der Grounded Theory«. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hrsg.): *Qualitative Forschung. Ein Handbuch*. Reinbek. 475–484.

Brauer, Kai, und Christoph Dehn (1995): »Videotechnik. Ein Leitfaden zur Orientierung«. In: Kai Brauer und Christoph Dehn: *Sozialwissenschaftliche Videographie. Erfahrungsberichte, Anwendungsmöglichkeiten und technisches Grundwissen*. Berlin. 57–84.

Dimbath, Oliver, Julia von Hayek, Andreas Hirseland und Werner Schneider (2003): »Toleranz, Konflikte und Wirkungen. Zur multimethodischen Mehrebenen-Evaluation von Toleranzbildungsprogrammen am Beispiel von Achtung (+) Toleranz«. In: Heinz Lynen von Berg und Roland Roth (Hrsg.): *Maßnahmen und Programme gegen Rechtsextremismus*. Opladen 193–227.

Flick, Uwe (1999): *Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften*. Reinbek.

Maier, Birgit (1995): »Zur Methodik der Filmanalyse von ethnographischen Filmen«. In: Edmund Ballhaus und Beate Engelbrecht: *Der ethnographische Film: eine Einführung in Methoden und Praxis.* Berlin. 223–267.

Reichertz, Jo (2000): »Objektive Hermeneutik und hermeneutische Wissenssoziologie«. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hrsg.): *Qualitative Forschung. Ein Handbuch.* Reinbek. 514–523.

Reichertz, Jo (1994): »Selbstgefälliges zum Anziehen. Benetton äußert sich zu Zeichen der Zeit«. In: Norbert Schröer (Hrsg.): *Interpretative Sozialforschung. Auf dem Weg zu einer hermeneutischen Wissenssoziologie.* Opladen. 253–280.

Renne, C. M., Peter W. Dowrick und G. Wasek (1983): »Considerations of the Participant in Video Recording«. In: Peter W. Dowrick und Simon J. Biggs: *Using Video. Psychological and Social Applications.* Chichester. 23–32.

Schütze, Alfred, und Thomas Luckmann (1994): *Strukturen der Lebenswelten.* Band 1. Frankfurt am Main.

Shotter, John (1983): »On Viewing Videotape Records of Oneself and Others. A Hermeneutical Analysis.« In: Peter W. Dowrick und Simon J. Biggs: *Using Video. Psychological and Social Applications.* Chichester. 199– 210.

Strauss, Anselm L. (1998): *Grundlagen qualitativer Sozialforschung. Datenanalyse und Theoriebildung in der empirischen und soziologischen Forschung.* München.

Ullrich, Carsten G. (1999): »Deutungsmusteranalyse und diskursives Interview. Leitfadenkonstruktion«. In: *Interviewführung und Typenbildung.* Mannheimer Zentrum für Europäische Sozialforschung, Arbeitspapier 3.

Wenzel, Florian (2003): »Ist politische Bildung messbar? Zur internen Evaluation und Qualitätssicherung in der außerschulischen Jugendarbeit«. Unveröffentlichte Workshop-Dokumentation. München.

Wernet, Andreas (2000): *Einführung in die Interpretationstechnik der Objektiven Hermeneutik.* Opladen.

Witzel, Andreas (2000): »Das problemzentrierte Interview (26 Absätze)«. In: *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* (Online-Journal) (1) 1 2000, http://qualitative-research.net/fqs.

Ziller, Robert C. (1999): *Photographing the Self. Methods for Observing Personal Orientations.* Newbury Park, Cal.