

Barbara Gasteiger-Klicpera

Evaluation und Qualitätsentwicklung im Kontext frühpädagogischer Interventionsprogramme: Chancen und kritische Aspekte



Die Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF) ist ein Projekt des Bundesministeriums für Bildung und Forschung und der Robert Bosch Stiftung in Zusammenarbeit mit dem Deutschen Jugendinstitut e. V. Die drei Partner setzen sich dafür ein, im frühpädagogischen Weiterbildungssystem in Deutschland mehr Transparenz herzustellen, die Qualität der Angebote zu sichern und anschlussfähige Bildungswege zu fördern.

© 2011 Deutsches Jugendinstitut e. V.
Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)
Nockherstraße 2, 81541 München
Telefon: +49 (0)89 62306-173
E-Mail: info@weiterbildungsinitiative.de

Herausgeber: Deutsches Jugendinstitut e. V. (DJI)
Koordination: Vera Deppe und Uta Hofele
Lektorat: Jürgen Barthelmes
Gestaltung, Satz: Brandung, Leipzig
Titelfoto: picsfive © Fotolia.com
Druck: Henrich Druck + Medien GmbH, Frankfurt a. M.

www.weiterbildungsinitiative.de

ISBN 978-3-86379-032-5

Barbara Gasteiger-Klicpera

Evaluation und Qualitätsentwicklung im Kontext frühpädagogischer Interventionsprogramme: Chancen und kritische Aspekte

Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)

Vorwort

Ziel und Nutzen der Evaluationen von Sprachförderprogrammen waren in der Expertengruppe Sprachförderung der Weiterbildungsinitiative wiederholt Thema von Diskussionen. Die Expertengruppe stellte fest, dass der häufig verwendete Hinweis auf eine Evaluation suggeriert, die Wirksamkeit des Programmes sei nachgewiesen.

Vor diesem Hintergrund hatte die Expertengruppe Sprachförderung eine Expertise vergeben, die den Begriff der Evaluation klärt, unterschiedliche Evaluationskonzepte einführend darlegt und die Potenziale und Grenzen der jeweiligen Ansätze und Strategien diskutiert. Barbara Gasteiger-Klicpera zeigt in ihrer Expertise am Beispiel der Evaluation des Sprachförderprogramms „Sag’ mal was“, wie die Evaluation eines frühpädagogischen Interventionsprogramms konzipiert und umgesetzt werden kann. Sie beschreibt die Schwierigkeit der praktischen Verwertung von Evaluationsergebnissen und plädiert für eine Verknüpfung von Programmevaluation und Qualitätsentwicklung in der Kindertageseinrichtung.

Die im Auftrag der *Weiterbildungsinitiative Frühpädagogische Fachkräfte* (WiFF) erstellte Expertise möchte zur Klärung von Begriffen und damit verbundenen Erwartungen beitragen. Sie richtet sich vorrangig an Weiterbildnerinnen und Weiterbildner sowie an interessierte Fachkräfte und ist bewusst knapp und leicht verständlich gehalten. Die Verantwortung für die fachliche Aufbereitung der Inhalte liegt bei der Autorin.

München, im Januar 2012



Angelika Diller
Projektleitung WiFF



Bernhard Kalicki
Wissenschaftliche Leitung WiFF

Inhalt

1	Einleitung	6
2	Was ist unter Evaluation zu verstehen?	8
3	Ziele von Evaluation	9
4	Leitprinzipien von Evaluation	10
5	Formen von Evaluation	11
5.1	Formative und summative Evaluation – Was ist das Ziel?	11
5.2	Externe und interne Evaluation – Was nützt wem?	13
6	Voraussetzungen einer Evaluation	14
7	Kritische Fragen zu Evaluation	14
7.1	Fragen der Planung	14
7.2	Fragen der Interpretation und Bewertung von Ergebnissen	16
7.3	Fragen der Umsetzung von Ergebnissen	16
8	Evaluation und Qualitätsentwicklung – Zur nachhaltigen Wirkung von Evaluationen	17
9	Zusammenfassung	19
10	Literatur	20

1 Einleitung

Aufgrund der kritischen Ergebnisse aus Schulleistungsstudien sowie der Suche nach Ursachen und Möglichkeiten, die Chancengleichheit der Kinder zu erhöhen, lag der Versuch nahe, durch eine Förderung der sprachlichen Kompetenzen von Vorschulkindern deren schulische Leistungen zu verbessern.

Der Großteil der schulischen Fächer, insbesondere das Lesen und Schreiben, sind im Wesentlichen an Sprache gekoppelt. Daher wurden vor etwa zehn Jahren in verschiedenen Bundesländern Sprachförderprogramme initiiert. Ein Schwerpunkt dieser Programme lag im Vorschulalter. Diese Altersgruppe eignet sich besonders gut für eine Intervention, kann doch angenommen werden, dass die Kinder zu diesem Zeitpunkt Sprache besonders rasch und effizient erlernen.

In diesem Zusammenhang wurde auch die Notwendigkeit gesehen, die Sprachförderprogramme zu evaluieren, um feststellen zu können, ob die dafür aufgewendeten Mittel sinnvoll und adäquat eingesetzt worden waren. Daher wurde kaum eines dieser Programme ohne Evaluation durchgeführt. Allerdings dominierten meist *interne* Formen der Evaluation, weil für eine externe Evaluation weder die dafür aufzubringenden Mittel noch die Zeit zur Verfügung standen und dies auch nicht als Notwendigkeit angesehen wurde. Das Bewusstsein für die Erfordernis einer regelmäßigen Reflexion und darüber hinaus für die Überprüfung und Bewertung der geleisteten Ergebnisse durch externe Experten hat sich erst langsam durchgesetzt und ist auch nicht widerspruchsfrei geblieben.

Das Programm „sag’ mal was – Sprachförderung für Vorschulkinder“

Für Deutschland erstmalig wurden im Rahmen des Programms *sag’ mal was – Sprachförderung für Vorschulkinder (Baden-Württemberg)* zwei größere Evaluationsstudien durchgeführt. Das Programm *sag’ mal was* wollte Kindern mit Migrationshintergrund oder mit Förderbedarf in deutscher Sprache eine intensive Sprachförderung in kleinen Gruppen ermöglichen. Die Sprachförderung wurde durch eine Erzieherin

durchgeführt, die meist mehrmals pro Woche mit den Kindern arbeitete und im Bereich der Sprachförderung eine besondere Qualifikation nachweisen musste. Inhaltlich und methodisch wurde den Erzieherinnen freigestellt, wie sie vorgehen wollten. Das Programm wurde von der *Baden-Württemberg Stiftung* ins Leben gerufen und als ein „lernendes Projekt“ definiert. Daher wurde von Beginn an eine Gruppe von Experten mit einbezogen, die bei der Konzeption und Durchführung des Programms beratend unterstützen sollten. Ebenso war von Anfang an geplant, dieses Programm zu evaluieren. Diese wissenschaftliche Begleitung übernahmen nach einem Ausschreibungsverfahren zwei Forschergruppen aus zwei wissenschaftlichen Hochschulen. Eine Studie wurde als „Wissenschaftliche Begleitung des Programms *sag’ mal was – Sprachförderung für Vorschulkinder* an der Pädagogischen Hochschule Weingarten von Barbara Gasteiger-Klicpera, Werner Knapp und Diemut Kucharz durchgeführt (Gasteiger-Klicpera u.a. 2010), die zweite Untersuchung als *Evaluationsstudie zur Sprachförderung von Vorschulkindern (EVAS)* von Jeanette Roos und Hermann Schöler an der Pädagogischen Hochschule Heidelberg (Roos u.a. 2010).

Die Evaluation beider Forschungsgruppen unterschied sich jedoch deutlich in ihrer Zielsetzung und Stichprobe sowie im methodischen Vorgehen. In beiden Untersuchungen ging es darum, die Entwicklung der Kinder im Längsschnitt über einen längeren Zeitraum zu beobachten. Zudem sollte erfasst werden, welche Fortschritte die Förderkinder im Vergleich zu Kindern machten, die nicht gefördert wurden und ebenfalls in den Kindertagesstätten betreut wurden.

Die Forschungsgruppe der Hochschule von Weingarten untersuchte die Kinder anhand einer repräsentativen Stichprobe, die, auf unterschiedliche geografische Gebiete im Land verteilt, die städtische sowie die ländliche Bevölkerungsstruktur des Landes angemessen darstellte.

Die Forschungsgruppe in Heidelberg wiederum konzentrierte sich auf die beiden Städte Heidelberg und Mannheim und prüfte dort den Erfolg der Sprachförderung von drei speziellen Sprachförderkonzeptionen (Gasteiger-Klicpera u.a. 2010; Roos u.a. 2010).

Beiden Untersuchungen gemeinsam war die Auswahl eines Teils der untersuchten Variablen, nämlich phonologische Fähigkeiten und Syntax sowie kognitive Fähigkeiten. Allerdings unterschieden sich die

beiden Forschungsgruppen in den angewandten Untersuchungsinstrumenten, d.h. den Kindern wurden unterschiedliche Tests vorgegeben, die teilweise ähnliche Fähigkeiten prüften.

Die Weingartner Gruppe verfolgte ein breiteres methodisches Vorgehen. Neben den vorgegebenen Tests wurden auch die Erzieherinnen in Form von Interviews befragt sowie die Eltern der Kinder mit Hilfe eines Fragebogens. Außerdem wurden Videoaufnahmen der Sprachförderung vorgenommen. Insgesamt wurden ein vielfältiges Spektrum an Untersuchungsmethoden angewandt (Tests, Interviews, Video) sowie unterschiedliche Perspektiven einbezogen. Die Forschergruppe erhielt durch die Befragung von Kindern, Eltern, Erzieherinnen und Leiterinnen der Kindertagesstätten ein umfassendes und buntes Bild von den Möglichkeiten und Schwierigkeiten der Sprachförderung.

Das *methodische Design* war in beiden Untersuchungen ähnlich angelegt. Um Veränderungen in der Entwicklung von Kindern messen zu können, ist es notwendig, die Kinder über einen längeren Zeitraum zu beobachten. Indem man Längsschnittstudien durchführt, kann man verfolgen, wie sich Kinder über die Zeit hin entwickeln.

Bei einem Längsschnittdesign spielt auch die Frage nach der Kontrollgruppe und Vergleichsgruppe eine wichtige Rolle, denn es geht um die Klärung, ob und inwieweit eine Intervention erfolgreich ist oder nicht. Dafür muss man Gruppen von Kindern vergleichen, die ähnliche Voraussetzungen haben, d.h. ähnliche soziale Bedingungen und ähnliche kognitive Fähigkeiten. Die Leistungsentwicklung der Kinder kann nach einiger Zeit (beispielsweise nach einem Jahr) miteinander verglichen werden. Auf diese Weise kann dann der spezifische Effekt einer Intervention von den allgemeinen Entwicklungseffekten getrennt werden sowie der Einfluss des Kindergartens, in dem ja auch sprachliche Förderung stattfindet, vom Einfluss einer spezifischen Sprachförderung, die zusätzlich in kleinen Gruppen angeboten und durchgeführt wird.

Ergebnisse

Ogleich das Untersuchungsdesign der beiden Evaluationen unterschiedlich war, ähnelten sich die Ergebnisse in wesentlichen Bereichen:

- Insgesamt machten alle Kinder in ihrer sprachlichen Entwicklung große Fortschritte.

- Kinder jedoch, die Sprachförderbedarf hatten, also jene, deren Leistungen sich am unteren Ende des Leistungsspektrums befanden, zeigten aufgrund der Sprachförderung keinen deutlichen Vorsprung gegenüber anderen schwachen Kindern, die keine besondere Förderung erhalten hatten. Demnach konnte die spezifische Sprachförderung keinen Vorzug gegenüber der alltäglich im Kindergarten stattfindenden pädagogischen Arbeit nachweisen.

Diese Erkenntnis konnte durch eine Reihe an weiteren Analysen bestätigt werden.

In der Folge war es schwierig, festzustellen, worin denn nun genau die Ursachen dieses Problems lagen und wie die Sprachförderung verbessert werden könnte – was als die wesentlich wichtigere Frage erschien. Für die Beantwortung dieser Frage wären jedoch weiterführende Analysen notwendig gewesen. Insbesondere könnten hier die Videoanalysen, die zusätzlich angefertigt wurden, wichtige Aufschlüsse bringen.

Aufgrund der Ergebnisse dieser Evaluationen erscheint es nicht verwunderlich, dass eine große Skepsis darüber herrscht, ob durch Evaluationen tatsächlich gewinnbringende Erkenntnisse erlangt werden können. Die Gefahr besteht darin, dass aufgrund von kritischen Ergebnissen allzu rasch negative Schlussfolgerungen gezogen werden.

Die beiden aufgeführten Evaluationsdesigns zeigen exemplarisch, mit welcher unterschiedlichen Strategien Evaluationsforscher an sehr ähnliche Fragen herangehen können. Allerdings ist das methodische Design häufig für Laien nicht transparent. Gerade wenn es um die Frage des Nachweises geht, ob ein Programm die erwarteten Effekte bringt, bestehen begründete Zweifel daran, ob die Ergebnisse von Evaluationen auch in einer Weise dargestellt werden, dass das methodische Design mit seinen besonderen Stärken und Schwächen auch für Laien verständlich ist. Allerdings würde es zur wissenschaftlichen Integrität und Ehrenhaftigkeit gehören, die Begrenzung der eigenen Forschung ebenso deutlich zu machen, sowie die Frage, für welche Gruppe denn nun die Ergebnisse zutreffen und wer von diesem Programm profitiert und wer nicht.

Fragestellungen

Da in verschiedenen pädagogischen Bereichen derzeit über Evaluation heftig diskutiert wird, erscheint es

notwendig, eine Standortbestimmung durchzuführen und folgende Frage zu stellen:

- Welche Bedeutung kommt der Evaluation für die pädagogische Praxis zu?
- Welche Aspekte sind bei Evaluationen zu beachten?
- Wie können diese im Sinne von Qualitätssicherung hilfreich für die Praxis eingesetzt werden?

Zum Aufbau der Expertise

Diesen Fragen versucht die vorliegende Expertise nachzugehen.

Im *Kapitel 2* werden in einer *Begriffsklärung von Evaluation* die damit in Zusammenhang stehenden *Begriffe* erläutert und anhand von Beispielen verdeutlicht. Evaluation wird häufig alltagssprachlich in einem Verständnis gebraucht, das wissenschaftliche Kriterien nicht unbedingt einschließt. Es wäre jedoch notwendig, diese Wissenschaftlichkeit bei Evaluationen einzufordern und den Gebrauch des Begriffes entsprechend einzuschränken. Daher wird erörtert, was unter Evaluation zu verstehen ist und welche Qualitätsmaßstäbe sowie Standards an Evaluationen angelegt werden sollten.

Neben den unterschiedlichen *Zielen*, die im Rahmen von Evaluation verfolgt werden und zu Beginn zu klären sind (*Kapitel 3*), gibt es auch eine Reihe an *Leitprinzipien*, die bei einer Evaluation beachtet werden müssen, damit die Betroffenen geschützt bleiben und keinen unnötigen Mühen, ungerechtfertigter Kritik oder zusätzlicher Arbeit ausgesetzt sind (*Kapitel 4*).

Bei diesen Überlegungen ist auch die Frage wichtig, welche Art von Evaluation für eine bestimmte Fragestellung angemessen ist. Daher soll der Unterschied zwischen *formativer* und *summativer Evaluation* genauso diskutiert werden wie die beiden, sich ergänzenden Formen einer *internen* und *externen Evaluation* (*Kapitel 5*).

Um eine Evaluation durchzuführen, sind bestimmte *Voraussetzungen* zu berücksichtigen und zu klären, die gegeben sein müssen, damit diese in sinnvoller Weise stattfinden kann (*Kapitel 6*).

Während des gesamten Prozesses der *Planung*, aber auch bei der Interpretation und der *Bewertung von Ergebnissen* bei Evaluationen gibt es kritische Aspekte, die es in besonderer Weise zu beachten gilt (*Kapitel 7*).

Die Frage, wie die Ergebnisse von Evaluationen *für die Praxis* gewinnbringend eingebracht oder eingearbeitet werden können, wird im *Kapitel 8* thematisiert.

2 Was ist unter Evaluation zu verstehen?

Bezüglich der Definition von Evaluation gibt es divergierende Ansichten.

Die *Deutsche Gesellschaft für Evaluation* (DeGEval) (2008) definiert Evaluation als die „systematische Untersuchung des Nutzens oder Wertes eines Gegenstands“, wobei in einem Nachsatz angemerkt wird, dass der Begriff „Gegenstand“ sehr weit gefasst wird – so werden darunter unter anderem Programme, Maßnahmen, Leistungen, Organisationen oder Forschung verstanden.

In diesem weiteren Sinn der Bewertung oder Beurteilung von Programmen oder Maßnahmen entspricht die Definition weitgehend dem Begriff der Evaluation, der sich inzwischen alltagssprachlich durchgesetzt hat. Allerdings hat die umgangssprachliche Verwendung auch zu einer Verwässerung des Begriffs geführt, da eine allgemeine Bewertung keinen Anspruch an Wissenschaftlichkeit beinhaltet. Oft wird von Evaluation gesprochen, ohne dass der Anspruch auf Wissenschaftlichkeit eingelöst wird.

Aus diesem Grund wird von manchen Autoren empfohlen, zwischen *Evaluation* und *Evaluationsforschung* zu differenzieren. Dies soll den alltagssprachlichen Gebrauch des Begriffs von einer Verwendung in der Forschung, die eigene Maßstäbe anwenden muss, abgrenzen. Während Evaluation im Sinne von Bewertung sich allgemein auf einen Prozess bezieht, der pädagogischem Handeln inhärent sein und jede Form von Intervention kontinuierlich begleiten sollte, ist die zweite Form eine wissenschaftlich begründete, systematische Art von Evaluation, die auf bestimmte methodische Werkzeuge empirischer Sozialforschung zurückgreift.

3 Ziele von Evaluation

Bevor ein Evaluationsprozess beginnen kann, sollten die Ziele festgelegt und die Erwartungen geklärt werden, die Auftraggeber an die Evaluation knüpfen. Aufgrund einer fundierten Zielanalyse können dann das Forschungsdesign entwickelt und die passenden Untersuchungsmethoden festgelegt werden. Die möglichen Ziele von Evaluation lassen sich (nach Wottawa/Thierau 1998) in drei größere Themenbereiche einteilen:

- *Ergebnisevaluation*: die Beurteilung, ob das Ergebnis einer Maßnahme oder eines Programms erreicht wurde.
- *Optimierung und Verbesserung* von Maßnahmen oder Programmen.
- *Verdeckte Ziele*: d.h. jene Ziele, die nicht offen gelegt werden und hinter den offen kommunizierten Zielen stecken, beispielsweise die Unterstützung bei einer Entscheidungsfindung oder die Unterstützung beim Durchsetzen eigener Interessen.

Bei der allgemeinen Bewertung einer Maßnahme geht es um die Frage, ob eine Maßnahme insgesamt die in sie gesetzten Erwartungen erfüllt. Damit in Verbindung steht auch meist eine Kontrolle der Zielerreichung: nämlich die Beantwortung der Frage, ob mit Hilfe eines Programms oder einer Maßnahme das angestrebte Ziel erreicht wurde. Im Wesentlichen handelt es sich dabei um eine *summative* Evaluation.

Ein völlig anderes Anliegen und Vorgehen ist erforderlich, wenn es sich bei den Zielen der Evaluation um die Erarbeitung von Grundlagen und um Anstöße zur Verbesserung von Maßnahmen oder um die Hilfestellung bei Entscheidungen handelt. In diesem Fall muss man davon ausgehen, dass die Ergebnisse wesentlich differenzierter sein müssen und in ihnen nicht nur eine Feststellung des Status quo enthalten sein muss, sondern auch eine detaillierte Analyse einzelner Aspekte, an denen es Anknüpfungspunkte für Veränderungen gibt, sowie Ideen zur Verbesserung. Daher kann diese Art von Evaluation meist als eine *formative* Evaluation konzipiert werden. Hier fänden sich erste Ideen zur Umwandlung des Evaluationsgedankens sowie zur Erweiterung des Konzeptes in Richtung von Qualitätsentwicklung.

Außerdem sind zusätzlich Ziele zu bedenken, die vom Auftraggeber nicht offen kommuniziert werden, und die offen mitgeteilte Ziele untergraben oder aushöhlen können – beispielsweise die Durchsetzung von eigenen Interessen, für die man sich von außen Unterstützung holt, um den eigenen Einfluss zu verstärken, sowie die Delegation von Verantwortung, beispielsweise bei Entscheidungen oder im Rahmen von Veränderungsprozessen. In diesen Fällen wird die Evaluation für andere Ziele instrumentalisiert, die ihr nicht unbedingt inhärent sind und die auch nicht im Vorhinein festgelegt oder offen kommuniziert sein müssen.

Manchmal stellt sich für den Auftraggeber einer Evaluation erst im Nachhinein heraus, dass die eigentlich geplanten Ziele der Evaluation mit denen, die nötig gewesen wären oder denen, die implizit auch vorhanden waren, nicht übereinstimmen. In einem solchen Fall wäre das Evaluationsverfahren zu unterbrechen und nach einer offenen Diskussion der unterschiedlichen Gesichtspunkte hätte eine erneute Klärung der Ziele und eine Veränderung des Fokus der Evaluation zu erfolgen. Sollte dies nicht möglich sein, wäre die Evaluation abzubrechen. Anderenfalls besteht die Gefahr, dass weder die ursprüngliche Intention erreicht wird noch eine zufriedenstellende Erfassung der Aufgabe möglich ist, sodass weder der Auftraggeber noch der Auftragnehmer mit den Ergebnissen zufrieden sind.

4 Leitprinzipien von Evaluation

Evaluationsforschung erfolgt häufig im Kontext praktisch notwendiger Entscheidungen. Ihre Aufgabe besteht dann darin, Grundlagen für Entscheidungsprozesse bereitzustellen, die für die Praxis weitreichende Konsequenzen haben. Daher muss sichergestellt werden, dass die Evaluationsforschung nach anerkanntem Stand der Wissenschaft vorgenommen wird.

In den USA wurden bereits vor 20 Jahren *Leitprinzipien* formuliert, die bei der Durchführung einer Evaluation beachtet werden sollten. Die *American Evaluation Association* (AEA) (2004) schlägt für Evaluatoren bestimmte Prinzipien oder Maßstäbe vor, die deren Arbeit leiten sollen. Diese beziehen sich in erster Linie auf die Anwendung angemessener *Standards methodischer und inhaltlicher Kompetenz* sowie auf die Berücksichtigung bestimmter *ethischer Grundsätze*. Folgende fünf Aspekte wurden von der AEA besonders ausgeführt:

Systematische Befragung und Untersuchung

Die Systematische Befragung und Untersuchung orientieren sich an den höchsten Standards. Eine Evaluation muss methodisch dem aktuellen Stand der Wissenschaft (*state of the art*) entsprechen. Zudem sollten die Evaluatoren die Grenzen und Stärken der Evaluation transparent machen sowie die Ansätze, Methoden und Grenzen präzise und detailliert erklären, sodass andere ihre Arbeit beurteilen können.

Kompetenz

Das gesamte Evaluationsteam muss über *die* Fähigkeiten, Fertigkeiten und Erfahrung verfügen, die der Aufgabe angemessen sind. Dazu gehört auch die interkulturelle Kompetenz und die Berücksichtigung der eigenen Grenzen, d.h. Evaluatoren überschreiten diese nicht. Außerdem achten die Evaluatoren darauf, die eigenen Fähigkeiten weiterzuentwickeln und zu erweitern, damit ihre Evaluation höchsten Ansprüchen genügt.

Integrität und Ehrenhaftigkeit

Diese beiden Aspekte beziehen sich sowohl auf die Person als auch auf den gesamten Evaluationsprozess, d.h.: Evaluatoren sollten in Verhandlungen mit dem Auftraggeber ehrenvoll handeln, aber auch eigene Interessen transparent machen und jede eigene Verwicklung oder persönliche Beziehung offenlegen, die mit dem Auftrag zusammenhängt. Zudem sollten sie jede finanzielle Unterstützung, die für die Evaluation gegeben wird, klarlegen.

Respekt Menschen gegenüber

Evaluatoren verfügen über einen grundlegenden Respekt vor der Würde und dem Selbstwert ihrer Auftraggeber sowie der Klienten und aller Personen, die an der Evaluation beteiligt sind.

Dies betrifft die Vertraulichkeit von Informationen, aber auch die Kommunikation von Untersuchungsergebnissen, bei der jeder Schaden vermieden werden soll. Die Teilnehmenden, die sich während der Evaluation engagieren, sollen danach auch einen Nutzen davon haben.

Verantwortung für die Allgemeinheit und das Gemeinwohl

Evaluatoren sind bereit, Verantwortung für die Allgemeinheit und das Gemeinwohl zu übernehmen und zu tragen, wobei ihnen die Unterschiedlichkeit von Menschen, Interessen und Werten deutlich bewusst ist. Sie sind für eine solche Aufgabe vorbereitet. Dies schließt die Berücksichtigung von Nachhaltigkeit und die Beachtung von längerfristigen Perspektiven mit ein. Ebenso wichtig ist ein Blick für die Gesellschaft als Ganzes, was die vorrangige Beachtung von Allgemeininteressen gegenüber Partikularinteressen einzelner Gruppen bedeutet. In Übertragung dieser Richtlinien auf den deutschen Sprachraum hat die *Deutsche Gesellschaft für Evaluation* (DeGEval 2008) Standards vorgeschlagen, die zum Teil ähnliche Schwerpunkte setzen wie die Standards der AEA, diese jedoch anders gewichten. Insbesondere von Bedeutung sind dabei folgende Kriterien:

- Nützlichkeit
- Durchführbarkeit
- Fairness
- Genauigkeit von Evaluationen.

Die *Nützlichkeit* hat vor allem diejenigen im Auge, die die Ergebnisse nutzen wollen, denn deren Bedürfnisse und Ziele sollten vor allem berücksichtigt werden.

Die *Durchführbarkeit* bezieht sich vorwiegend darauf, dass der vorgesehene Zeit- und Kostenrahmen berücksichtigt wird, sowie auf die Effizienz des Vorgehens.

Die *Fairness* sollte einschließen, dass alle rechtlichen Grundlagen berücksichtigt werden sowie der Schutz individueller Rechte und die Unparteilichkeit in der Durchführung und Berichterstattung gewährleistet sind.

Die *Genauigkeit* schließlich fokussiert die Berücksichtigung von Validität und Reliabilität bei der Datenerhebung sowie insgesamt ein adäquates und korrektes methodisches Vorgehen.

Verglichen mit den amerikanischen Leitlinien fällt auf, dass in den deutschen Standards die Formalisierung von Prozessen einen höheren Stellenwert einnimmt. In den amerikanischen Leitlinien dagegen werden die Berücksichtigung ethischer Prinzipien, die Frage der Ehrenhaftigkeit und persönlichen Integrität der Evaluatoren sowie die Nachhaltigkeit und die Berücksichtigung des Allgemeinwohls stark thematisiert.

Während sich die deutschen Standards auf die einzelnen Schritte der Evaluation konzentrieren, und dabei vor allem ein korrektes methodisches Vorgehen erwartet wird, sprechen die amerikanischen Standards ein höheres Reflexionsniveau an, indem ein größerer Wert auf die Transparenz gelegt wird sowie auf die Berücksichtigung allgemeiner ethischer Prinzipien wie Nachhaltigkeit und Gemeinwohl.

5 Formen von Evaluation

5.1 Formative und summative Evaluation – Was ist das Ziel?

Zwei grundsätzlich unterschiedliche Ansätze werden mit *formativer* und *summativer* Evaluation bezeichnet, die beide unterschiedliche Zielsetzungen verfolgen. Während mithilfe einer *formativen* Evaluation vor allem ein Programm verbessert, angepasst oder implementiert werden soll, besteht das Ziel einer *summativen* Evaluation darin, ein Programm abschließend zu bewerten und zu beurteilen, um beispielsweise seine Wirksamkeit zu erfassen.

Bei einer *formativen* Evaluation ist es daher notwendig, dass die Evaluatoren von vornherein in den gesamten Prozess der Konzeption und Planung einer Maßnahme einbezogen werden, damit sie die Möglichkeit haben, einzugreifen und von Beginn an mitbestimmen können, welche Aspekte notwendig oder wichtig sind.

Bei einer *summativen* Evaluation hingegen wird lediglich das Ergebnis eines Prozesses oder eines Programms bewertet. Deshalb genügt es, wenn die Evaluatoren erst später oder am Ende des Prozesses hinzugezogen werden, und sich weder an der Konzeption noch an der Weiterentwicklung in einzelnen Schritten beteiligen.

Ein Beispiel für eine *summative* Evaluation stellt die (in der Einleitung angesprochene) Evaluation des Programms *sag' mal was – Sprachförderung für Vorschulkinder in Baden-Württemberg* dar. *Summative* Evaluation als paradigmatischer Fall bedeutet: Die Evaluation wurde erst begonnen, als das Programm schon lief; somit waren der Gestaltungsspielraum und die Möglichkeit eines Eingreifens vonseiten der Evaluatoren kaum vorhanden, und es ging hier insbesondere darum, die Effekte des Programms zu messen und zu beurteilen.

Eine *formative* Evaluation hingegen würde anders aussehen. Diese Form wird oft von Schulen initiiert. Vor kurzer Zeit trat beispielsweise eine Schule an uns heran, die ein Projekt zur Peer-Mediation durchführen möchte. Wir wurden eingeladen, im Rahmen eines

Projekttag eine Fortbildung für die Lehrerinnen und Lehrer zu halten. Wir stellten das Programm vor und diskutierten mit den Lehrkräften auch Fragen der Umsetzung und der Implementierung. Künftig werden regelmäßig Treffen stattfinden, in denen wichtige Anliegen und Probleme gemeinsam besprochen werden. Zusätzlich werden die Effekte des Projektes im Rahmen von Masterarbeiten der Studierenden mit Hilfe von Fragebögen und Einschätzskalen gemessen.

Neben der kontinuierlichen Beratung und Begleitung ist also auch eine standardisierte Erhebung der Effekte vorgesehen, was im Rahmen einer *formativen* Evaluation durchaus sinnvoll ist. Damit kann auch genau eingeschätzt werden, ob und welche Erfolge das Programm aufweist. Wichtig ist dabei die enge Zusammenarbeit zwischen dem Evaluationsteam und den Beteiligten an der Schule, die dazu führt, dass das Programm verändert und an die konkreten Gegebenheiten angepasst werden kann.

Im pädagogischen Bereich ist diese Trennung allerdings nicht immer so scharf zu ziehen, was unterschiedliche Gründe hat:

Zum einen ist die Bewertung oder Beurteilung einer Maßnahme häufig mit der Erwartung verknüpft, dass Verbesserungsvorschläge gemacht werden sollten, die zu einer Optimierung führen, d.h. im pädagogischen Bereich geht es kaum um eine einfache Beurteilung von Effekten, sondern meist um eine Weiterentwicklung und eine Verbesserung.

Zum anderen ist Evaluation im pädagogischen Bereich immer an Handlungen von Menschen geknüpft, auf die jede Bewertung wieder zurückwirkt, d.h. jede Bewertung ist gleichzeitig eine Intervention.

Allein das Wissen darum, dass ein Programm evaluiert wird, verändert die Handlungen der Beteiligten und damit das Ergebnis der Evaluation. Dies kann zum *Vorteil* des Programms gereichen, etwa wenn sich die Beteiligten besonders anstrengen. Dies kann aber auch ein *Nachteil* sein, wenn sich die Beteiligten gegen die Evaluation sträuben und diese von vornherein als überflüssig und sinnlos betrachten – wie es manchmal in Schulen der Fall ist.

Ebenfalls ein *Nachteil* kann es sein, wenn eine gute Leistung als Vorwand dient, schwierige Rahmenbedingungen nicht zu verändern, weil die Ergebnisse sowieso schon so gut ausfallen. So könnte beispielsweise die Bereitschaft von Erzieherinnen und Erziehern, ihre

Freizeit für Fortbildungen zu nutzen, dazu führen, dass die Träger keine Notwendigkeit sehen, die pädagogischen Fachkräfte regelmäßig für Fortbildungen freizustellen, da sie sich sowieso schon qualifizieren.

Es wäre ein Trugschluss zu denken, man könne von außen etwas beurteilen, ohne die Menschen mit im Blick zu haben, die ihr Wissen und Können in einen Prozess einbringen oder engagiert an einer Maßnahme arbeiten. Bereits das Wissen darum, dass ein Programm oder eine Maßnahme evaluiert wird, führt zu bestimmten Annahmen und Veränderungen im Denken und Handeln der Menschen. So hat die Vorgabe bestimmter Aufgabentypen bei den größeren Schulleistungsstudien etwa dazu geführt, dass diese Art von Aufgaben von den Lehrkräften im Unterricht stärker berücksichtigt wurde – ein Phänomen, das unter dem Begriff „teaching to the test“ bekannt geworden ist (Klieme/Tippelt 2008).

Für den Elementarbereich würde das bedeuten, dass die Erzieherinnen und Erzieher jene Aufgaben mit den Kindern verstärkt üben, von denen sie wissen, dass sie in einer Evaluation überprüft werden.

Neben der Veränderung des Prozesses durch die Evaluation ist auch eine *Veränderung der Ergebnisinterpretation* durch die Beteiligten möglich. Selbst wenn eine *summative* Evaluation eines pädagogischen Ansatzes zu dem Schluss käme, dieser Ansatz wäre nicht hilfreich oder nicht effektiv, so würde dies noch nicht bedeuten, dass dieser Ansatz aufgegeben würde.

Es wäre ein Irrtum, anzunehmen, man könne in einem pädagogischen Bereich eine Evaluation vornehmen, ohne dass dies bei den Beteiligten bestimmte Reaktionen auslösen würde. Im Allgemeinen wird daher versucht, das methodische Vorgehen so zu planen, dass diese Effekte für die Evaluation genutzt werden, indem beispielsweise die Ziele transparent gemacht und die Folgen erläutert werden, die aus der Evaluation resultieren könnten. Ebenso werden methodische Vorsichtsmaßnahmen getroffen, etwa indem möglichst verschiedene Sichtweisen berücksichtigt oder unterschiedliche Erhebungsmethoden verwendet werden.

Eine weitere Möglichkeit, diese Reaktionen in sinnvoller Weise zu nutzen, liegt in der Verbindung zwischen *externer* und *interner* Evaluation.

5.2 Externe und interne Evaluation – Was nützt wem?

Eine Evaluation durch Personen, die der Organisation oder dem Team angehören, wird als *interne* Evaluation bezeichnet, eine Beurteilung oder Bewertung durch außenstehende Personen als *externe* Evaluation.

Für die Durchführung einer Evaluation bedeutet dies einen grundsätzlichen Unterschied, der eine Reihe von Implikationen und Folgerungen sowohl für die Zielsetzung als auch für den Prozess und schließlich für die Umsetzung der Ergebnisse enthält.

Bei einer *internen* Evaluation besteht das Problem darin, dass es für Menschen, die selbst in alle Vorgänge und Interaktionen involviert sind, schwierig erscheint, eine distanzierte Position einzunehmen, die Objektivität erlaubt und sowohl eine Bewertung als auch Beurteilung ermöglicht.

Bei einer *externen* Evaluation besteht dagegen vor allem die Schwierigkeit der Akzeptanz. In Zusammenhang damit steht die Bereitschaft der Evaluierten, sich in den Prozess einzubringen, Vorschläge für Veränderungen aufzugreifen und diese umzusetzen.

Bei *externen* Evaluationen muss immer damit gerechnet werden, dass die Beteiligten sie als eine Beurteilung von außen erleben und daher weder bereit sind, sich für die Evaluation zu engagieren, noch sich mit den Ergebnissen auseinandersetzen wollen. Daher kann der Lerngewinn nicht allzu groß sein. Allerdings gehören die beiden Formen der Evaluation als Elemente in einem pädagogischen Prozess unabdingbar zusammen – eine *interne* Evaluation ist ohne eine *externe* Evaluation kaum denkbar und umgekehrt. Eine *interne* Evaluation erfordert immer auch eine Begleitung von außen, um den Beteiligten eine erweiterte Sicht sowie andere Erfahrungen und Gewichtungen zu ermöglichen.

Externe Evaluationen sind in jedem Fall notwendig, um eine unbefangene Außensicht zu gewährleisten, um Prozesse in Gang zu bringen, und manchmal vielleicht auch, um schwierige Entscheidungen voranzubringen. Eine *externe* Evaluation ist in einem pädagogischen Kontext jedoch ohne eine *interne* Evaluation fast unmöglich und oft sinnlos, insbesondere wenn es um Veränderungen geht, was meist der Fall ist.

Ohne eine engagierte Beteiligung der Betroffenen ist ein Veränderungsprozess nicht möglich. *Interne* Evaluationen sollten daher einer *externen* Evaluation

vorausgehen, und mit dieser Hand in Hand gehen, um die Beteiligten auf den Umgang mit den Rückmeldeprozessen vorzubereiten. Somit wird gewährleistet, dass die Ergebnisse der *externen* Evaluation umgesetzt und genutzt werden, damit der Erkenntnisgewinn, der daraus gezogen werden kann, möglichst groß ist und die Beteiligten langfristig davon profitieren (Kempfert/Rolf 2005).

Idealerweise wird in pädagogischen Institutionen ein kontinuierlicher Feedbackprozess in Gang gesetzt, der eine ständige Überwachung und Verbesserung der eigenen Tätigkeit ermöglicht und von Zeit zu Zeit durch externe Begleitung ergänzt und unterstrichen wird, sodass *interne* und *externe* Evaluation Hand in Hand gehen und sich als Prozess gegenseitig ergänzen und befruchten.

6 Voraussetzungen einer Evaluation

Heinrich Wottawa und Heike Thierau (1998) definieren in ihrer Darstellung wesentlicher methodischer und inhaltlicher Fragen zum Thema Evaluation jene *Voraussetzungen*, die nötig sind, damit Evaluationsforschung überhaupt sinnvoll sein kann.

Evaluation muss zunächst grundsätzlich *akzeptabel* erscheinen. Für diese Akzeptanz von Evaluationsforschung, die in der Praxis eine nicht zu unterschätzende Rolle spielt, gibt es die drei folgenden *psychologischen Voraussetzungen*:

- Die *Akzeptanz der Veränderbarkeit* und *Veränderungsbedürftigkeit* von Maßnahmen, Interventionen und Programmen, d.h. zu erkennen und zu akzeptieren, dass diese ständig reflektiert und verbessert werden können und sollen, um stets die Möglichkeit einer Innovation oder Weiterentwicklung zu geben.
- Die *Offenheit* für eine empirische, erfahrungsgestützte Grundlage für Entscheidungen, eventuell auch die Fähigkeit und der Mut, übliche Strategien in Frage zu stellen und die Bereitschaft, sich auf die Unsicherheit von Veränderungen einzulassen, die stets auch die Courage zum Scheitern beinhaltet.
- Die *Fähigkeit, Fakten akzeptieren zu können*, gerade wenn man ein Programm einsetzt und versucht, damit einen als unzumutbar angesehenen Zustand zu verändern. So erscheint es beispielsweise schwierig einzusehen, dass ein Programm zur Sprachförderung keinen Erfolg zeigt, wenn man sehr viel Geld in dieses Programm investiert hat, um Kindern mit Migrationshintergrund einen besseren Start in die Schule zu ermöglichen. Deshalb wird in solchen Fällen häufig darauf zurückgegriffen, den Evaluatoren ein unangemessenes oder unzureichendes methodisches Vorgehen vorzuwerfen und die Generalisierbarkeit der Ergebnisse anzuzweifeln.

7 Kritische Fragen zu Evaluation

Im Verlauf einer Evaluation ist eine Reihe an kritischen Punkten zu bewältigen. Diese betreffen zum einen Fragen der Planung und des Forschungsdesigns, zum anderen Fragen der Interpretation und der Bewertung von Ergebnissen, und schließlich sind noch Fragen der Umsetzung und der Implementation zu beachten.

Da im Rahmen der vorliegenden Darstellung nur einige kritische Punkte reflektiert werden können, sei für eine umfassendere Darstellung auf die entsprechende Fachliteratur verwiesen (vgl. Hager u.a. 2000).

7.1 Fragen der Planung

Untersuchungsdesign

Die Frage des *Untersuchungsdesigns* hängt von den Zielen der Evaluation ab. Ein wesentlicher Aspekt ist die Frage, ob es möglich ist, die Untersuchungsziele im Rahmen einer *Querschnittuntersuchung* zu beantworten, oder ob eine *Untersuchung im Längsschnitt* erfolgen muss.

Geht es um die Analyse von Entwicklungsverläufen, wie bei den meisten pädagogisch-psychologischen Programmen, bei denen die Frage beantwortet werden soll, ob sich Kinder durch eine Förderung beispielsweise rascher entwickeln als andere, müssen diese Kinder in jedem Fall über einen längeren Zeitraum beobachtet werden.

Allerdings sind *Längsschnittuntersuchungen* sehr viel aufwendiger und teurer als *Querschnittuntersuchungen*; sie erlauben jedoch eine Analyse der Effekte von Programmen und geben Antworten auf Fragen nach Wirkzusammenhängen.

Um ein angemessenes Design für die Frage nach den Effekten von Programmen zu entwickeln, wäre eigentlich die Bildung von Parallelgruppen nötig, was in einem experimentellen Design möglich ist, in der Praxis der Evaluationsforschung jedoch nur in seltenen Fällen realisiert werden kann. Geht es insbesondere um die Effekte von Programmen oder Interventionen, sollte zumindest ein Vergleich zwischen

einer bestimmten Handhabung mit Personen und einer Gruppe von Personen ohne diese Behandlung ermöglicht werden.

Dies ist insbesondere bei Kindern wichtig, da sich Kinder *von sich aus* weiterentwickeln, allein durch den Kontakt mit anderen Gleichaltrigen, Kindergartenkindern, Mitschülern oder Erwachsenen. Dieses Lernen ist unabhängig davon, ob eine Intervention stattfindet oder nicht.

Die kritische Frage ist daher nicht, ob Kinder durch eine Intervention etwas dazugelernt haben, sondern ob sie mehr gelernt haben, als sie ohne diese Intervention ohnehin schon gelernt hätten. Um diese Frage beantworten zu können, gibt es eigentlich nur die Möglichkeit, eine *Kontrollgruppe* zu bilden. Dies bedeutet, dass man zwei Gruppen von Kindern über denselben Zeitraum untersucht. Allerdings wird es kaum möglich sein, dass sich diese Kinder genau entsprechen.

Eine sogenannte *Parallelisierung* von Kindern wird nur selten möglich sein, wäre aber nötig, um wirklich fundierte Aussagen zu treffen. Daher sollten die Kinder zumindest einen vergleichbaren sozialen Status aufweisen und in ihrer kognitiven Leistungsfähigkeit ähnlich sein. Dies wären im Allgemeinen die wichtigsten Voraussetzungen, um davon ausgehen zu können, dass die beiden Gruppen vergleichbar sind. Doch selbst dann sind bestimmte Fehlerquellen nicht auszuschließen (vgl. dazu Hager u.a. 2000).

Während der Planungsphase ist neben dem Design der geplanten Evaluierung eine Reihe von methodischen Fragen zu lösen: Diese betreffen die Auswahl der Stichprobe und die Entscheidung für die richtigen Untersuchungsinstrumente bzw. die Adaptierung und Entwicklung dieser Instrumente.

Stichprobe

Bei der Auswahl einer angemessenen *Stichprobe* ist neben der Frage der Repräsentativität und der Verteilung der Stichprobe über verschiedene Untersuchungsgruppen vor allem die Größe der Stichprobe zu beachten. Die Gruppe sollte nicht zu klein sein, um bestimmte statistische Verfahren zu ermöglichen, aber aus ökonomischen Gründen auch nicht größer als unbedingt nötig.

Des Weiteren ist die Frage wichtig, welche Personengruppe in die Evaluation einbezogen werden

sollte. Wie kann der Kreis an Personen bestimmt werden, die an der Evaluation beteiligt werden? Geht es beispielsweise um die Evaluation eines Sprachförderprogramms, wer sind dann die entscheidenden Personen: die Kinder, die das Sprachförderprogramm durchlaufen; die Erzieherinnen, die die Förderkurse geben; die Erzieherinnen, die die Kindergartengruppe leiten; die Eltern, die zu Hause die Lernfortschritte des Kindes beobachten?

In aufwendigeren Evaluationsdesigns versucht man meist, Einschätzungen von unterschiedlichen Beurteilern zu bekommen und in die Analyse mit einzubeziehen. Dies verhindert eine einseitige Sichtweise und ermöglicht ein wesentlich umfassenderes und ganzheitlicheres Bild der Situation. Auch hier gilt es, so ökonomisch wie möglich, aber auch so umfassend wie nötig den Kreis der Beteiligten zu definieren.

Messinstrumente

Um ein möglichst umfassendes Bild zu erhalten, wird bei *externen* sowie *internen* Evaluationen zumeist auf unterschiedliche Untersuchungsinstrumente zurückgegriffen. Dies können Tests, Fragebögen, Beobachtungsverfahren und schließlich auch Interviews sein.

Hier geht es vor allem um die Entscheidung der Anwendung adäquater Forschungsmethoden. Da sich in der Zwischenzeit die Auseinandersetzung um die Aussagefähigkeit quantitativer versus qualitativer Methoden weitgehend gelegt hat, werden je nach Fragestellung quantitative oder qualitative Untersuchungsmethoden zum Einsatz kommen, beziehungsweise beide.

Untersuchungsinstrumente

An diese Frage schließt sich die Frage der angemessenen *Untersuchungsinstrumente* an. Zwar stehen unterschiedliche standardisierte Verfahren zur Verfügung, diese werden jedoch häufig nicht in der Form verwendet werden, wie sie zur Verfügung stehen. Demnach wird es notwendig sein, Adaptierungen oder Ergänzungen vorzunehmen, die sich auf die konkrete Fragestellung der Evaluation beziehen. Allerdings stellt sich dann das Problem der Messgenauigkeit, der Reliabilität, der Gültigkeit und der Validität der verschiedenen Verfahren (vgl. dazu Gollwitzer/Jäger 2009).

Information und Motivation der Beteiligten

Eine häufig diskutierte Frage ist, wer von den Beteiligten vorab in welcher Weise informiert werden soll. Bewährt hat sich dabei die Maxime einer höchstmöglichen Transparenz.

Ohne die Mitarbeit aller Beteiligten ist selbst die aufwendigste Forschung zum Scheitern verurteilt. Gerade Evaluatoren sind darauf angewiesen, dass ihnen ganz verschiedene Personen Antwort auf ihre Fragen geben und damit Zeit und Energie in die Evaluation investieren.

Die Frage der Motivation der Personen, die an der Evaluation teilnehmen, ist daher besonders wichtig. Die Beteiligten werden vor allem dann zur Mitarbeit motiviert, wenn die Untersuchungen für sie einen unmittelbaren Nutzen haben, wie es bei den meisten Evaluationen glücklicherweise der Fall ist. Allerdings kann genau dies sie auch davon abhalten. Daher muss ihnen in adäquater Weise das Ziel der Untersuchung vermittelt und die Evaluation begründet werden. Nur wenn alle Beteiligten einsehen, weshalb es diese Untersuchung braucht und welchen Gewinn sie davon haben, werden sie auch bereit sein, sich einzubringen. Demnach ist es sinnvoll, die Ergebnisse immer mit allen Beteiligten zu diskutieren.

7.2 Fragen der Interpretation und Bewertung von Ergebnissen

Sobald die Ergebnisse vorliegen, sind als nächste kritische Etappe die *Interpretation* und die Diskussion der *Bedeutung der Ergebnisse* zu sehen. Eigentlich ist die Interpretation der Ergebnisse eine kooperative Angelegenheit zwischen den Evaluatoren und den an der Evaluation Beteiligten. Meist wissen die Beteiligten wesentlich genauer, welche Bedeutung den Ergebnissen zukommt und welche Ergebnisse in welchen Zusammenhang zu stellen sind. Außerdem sind sie meist dafür verantwortlich, welche Konsequenzen aus den Ergebnissen gezogen und wie diese umgesetzt werden. Eine gemeinsame Diskussion, in der die Ergebnisse präsentiert und besprochen werden, ist daher für alle Beteiligten hilfreich.

Allerdings sind in diesem Zusammenhang auch einige methodische Fragen wichtig und angemessen zu besprechen, beispielsweise die häufig gestellten Fragen:

- „Was sind Effekte?“
- „Wie können Effektstärken interpretiert werden?“ (zur Diskussion von Effektstärken vgl. Bortz/Döring 2006).

Zudem werden häufig Schlussfolgerungen aus Querschnittsuntersuchungen gezogen, d.h. es wird aus korrelativen Zusammenhängen auf Ursachen geschlossen, was nicht zulässig ist. Auf diese Weise können keine kausalen Schlüsse gezogen und keine Effekte von Interventionen festgestellt werden.

7.3 Fragen der Umsetzung von Ergebnissen

Bei jeder Evaluation ist die Frage von Bedeutung, wie man mit den Ergebnissen umgeht. Zunächst haben Ergebnisse unmittelbare Relevanz für die Betroffenen selber sowie für die Auftraggeber. Letzteren wird selbstverständlich zugesprochen, dass sie die Ergebnisse erhalten und darüber verfügen können. Jedoch sollten auch die Betroffenen über die Ergebnisse einer Evaluation informiert werden und damit die Möglichkeit erhalten, sich weiterzuentwickeln. Somit können Rückmeldungen als Anlass genommen werden, eigenes Vorgehen und Handlungsstrategien zu überdenken und zu verbessern. Auch sind bei größeren Evaluationen meist Folgerungen auf breiterer Ebene zu diskutieren, und es ist somit die Frage zu stellen, ob die Ergebnisse auch gesellschaftliche Relevanz besitzen.

Manchmal werden die Ergebnisse so ausfallen, wie man sie sich gerade nicht vorgestellt hätte. Manche Ergebnisse sind beispielsweise politisch nicht erwünscht oder unpassend. Andere Ergebnisse sind für die Beteiligten verletzend oder infrage stellend, beispielsweise wenn ein Interventionsprogramm trotz hohem Engagement der Beteiligten keine Effekte aufweist.

Die kritische Frage ist, wie dann verfahren werden soll. Manche Auftraggeber sehen kritische Ergebnisse als Herausforderung, etwas zu verändern, ein Gespräch zu suchen und die Frage zu stellen, welche Bedeutung nun diese Ergebnisse haben und welche Konsequenzen man daraus ziehen kann. Gerade aus kritischen Ergebnissen kann besonders viel entnommen und besonders viel gelernt werden. Allerdings brauchen Auftraggeber von Evaluationen gerade in

solchen Fällen besondere Unterstützung. Es ist vor allem dem Geschick von Evaluatoren zu verdanken, wenn schwierige Ergebnisse nicht abgelehnt werden, sondern einen positiven Prozess initiieren können.

Wie Lernerfolg nachhaltig gesichert werden kann, wird im folgenden Kapitel diskutiert.

8 Evaluation und Qualitätsentwicklung – Zur nachhaltigen Wirkung von Evaluationen

Bereits bei der Diskussion der Zielsetzungen von Evaluationen wurde deutlich, dass es mannigfaltige Verbindungen zwischen Evaluation und Qualitätsentwicklung gibt. Nur ein kleiner Bereich von Evaluation, nämlich jener der kurzfristigen Bewertung von Maßnahmen, kann unabhängig von Qualitätsentwicklung betrachtet werden. Jedes nachhaltige und eine längere zeitliche Perspektive umfassende Verständnis von Evaluation muss eine enge Verbindung zur Qualitätsentwicklung herstellen und diese Verbindung als notwendig und sinnvoll erachten (Ditton 2009).

Ein Evaluationsschritt kann und sollte meist in einen längeren zeitlichen Prozess eingebunden sein, damit eine Verbindung von Evaluation zur Qualitätsentwicklung möglich ist. Eigentlich sollten die beiden Hand in Hand gehen und voneinander profitieren.

Zur Frage, was denn in pädagogischen Einrichtungen unter Qualität zu verstehen sei, wird eine intensive Diskussion geführt (Fend 2008; Klieme/Tipelt 2008; Harvey/Green 2000). In der Diskussion um Kindertageseinrichtungen wird hierbei häufig auf die Differenzierung zwischen *Strukturqualität*, *Prozessqualität* und *Orientierungsqualität* zurückgegriffen (Tietze u. a. 2007).

Unter *Strukturqualität* werden die strukturellen Bedingungen zusammengefasst, d. h. Gruppengröße, Erzieherinnen-Kind-Schlüssel, Ausbildung und Berufserfahrung der Erzieherinnen und Erzieher, Räume und Ausstattung der Kindertageseinrichtungen.

Unter *Prozessqualität* wird die Qualität der pädagogischen Prozesse verstanden, die Qualität der Interaktionen zwischen Erzieherinnen und Kindern sowie deren Anregungsgrad.

Orientierungsqualität bezieht sich auf die pädagogischen Ziele und Einstellungen, Werte und Vorstellungen des pädagogischen Personals in den Kitas (Tietze u. a. 2007).

Im schulischen Bereich wird die Diskussion, über den Zusammenhang von Schulentwicklung und Eva-

luation schon seit Langem geführt. Ebenso ist dort die Auseinandersetzung mit der Frage der Qualitätsentwicklung bereits im Gang. Wahrscheinlich sind die Voraussetzungen in den Kindertageseinrichtungen ähnlich, und daher wäre es möglich, aus den positiven Erfahrungen im schulischen Bereich zu lernen oder aus den negativen Erfahrungen Konsequenzen zu ziehen. So gehen auch hier einem Prozess der Qualitätsentwicklung viele Ängste und Vorbehalte voran. Erst wenn diese Vorbehalte beseitigt wurden und sich die Erkenntnis durchgesetzt hat, dass eine Weiterentwicklung der gesamten Schule von Interesse für alle Beteiligten ist, kann auch über Instrumente und sinnvolle Prozesse nachgedacht werden.

Evaluation ist hier ein besonders hilfreiches Instrument, um Stärken und Schwächen zu analysieren und Anstöße für eine Weiterentwicklung zu geben. In den meisten Fällen wird man – um den Beteiligten den Sinn der Evaluation zu vermitteln – zunächst mit *internen* Formen von Evaluation beginnen. Allerdings ist das längerfristige Ziel, zu einer *externen* Evaluation überzugehen, da diese in vielerlei Hinsicht eine weitere Perspektive vermitteln kann, über eine bloße Analyse von Stärken und Schwächen hinausgeht und Anstöße für eine Weiterentwicklung bietet. Insbesondere ermöglicht ein Blick von außen einen weniger beschwerten und offeneren Blick auf Strukturen und Prozesse sowie eine klarere und vorurteilsfreiere Sicht auf die verschiedenen pädagogischen Prozesse.

Vielleicht reagieren Erzieherinnen und Erzieher offener und unbefangener als Lehrerinnen und Lehrer, wenn sich ihnen die Möglichkeit bietet, über die eigene Arbeit zu reflektieren und diese zu verbessern. Im Gegensatz zum schulischen Bereich stützt sich Qualitätsentwicklung in Kindertageseinrichtungen bisher stark auf *externe* Evaluationen. Dementsprechend werden verschiedene Konzepte zur *externen* Evaluation in Kindertageseinrichtungen diskutiert. Beispielsweise wurde im Zuge der Entwicklung und Adaptierung der KES-R (Tietze u. a. 2007) das deutsche Kindergarten-Gütesiegel verliehen oder es werden andere Konzepte wie etwa eine Zertifizierung nach DIN/EN ISO 9000 angewendet.

Für die konkrete pädagogische Umsetzung sowie für die sprachliche Entwicklung von Kindern bieten diese Instrumente jedoch kaum und nur unspezi-

fische Anregungen. Zudem führen diese Formen der externen Evaluation mit Recht zu einer kritischen Sicht vonseiten der Erzieherinnen und Erzieher. Die wesentlichen Fragen dabei sind,

- ob sich die tägliche Arbeit mit den Kindern durch eine externe Evaluation verändert,
- ob und wieweit sie davon profitiert und
- welchen Gewinn die Kinder davon haben.

Dies sind für Erzieherinnen und Erzieher die entscheidenden Kriterien. Daher werden Konzepte *externer* Evaluation infrage gestellt und erscheinen nicht sonderlich gewinnbringend. Auf der anderen Seite gibt es bisher nur wenige Überlegungen und wenig Konzepte für die *interne* Qualitätsentwicklung in Kindertageseinrichtungen. Gerade dies jedoch wäre ein Weg, um – ähnlich wie in Schulen – Veränderungen zu initiieren und einen längeren Prozess der Qualitätsentwicklung in die Wege zu leiten, der dann auch zu einer nachhaltigen Veränderung der Arbeit in der Einrichtung führen könnte.

Da im schulischen Bereich eine längere Tradition und Entwicklung aus vorangegangenen Evaluationen vorhanden ist, könnte man auf diesen Erfahrungen aufbauen. Zudem ist – da es sich auch um pädagogische Prozesse und das Lernen von Kindern handelt – eine inhaltliche Nähe gegeben. Wesentlich an den bisherigen Erfahrungen der Qualitätsentwicklung in Schulen ist zum einen, dass für die Beteiligten der unmittelbare Zweck und Nutzen rasch erkennbar sein muss, zum anderen, dass der Prozess vom Vertrauen in die einzelnen Personen getragen ist. Jeder bringt sich mit seinen Fähigkeiten ein und so kommt die Einrichtung ihrem Ziel einen Schritt näher (Kempfert/Rolff 2005).

Die Erfahrungen zeigen, dass zunächst eine *interne* Evaluation erfolgt, und erst dann, wenn die Beteiligten mit dieser Form positive Erfahrungen gesammelt haben, eine *externe* Evaluation. Diese beiden Schritte als miteinander verbunden und gleichzeitig aufeinander aufbauend wahrzunehmen, scheint ein wichtiges Kriterium auch für Kindertageseinrichtungen zu sein. Beide Formen setzen unterschiedliche Ziele voraus, und insbesondere eine externe Evaluation muss auf bestimmten Voraussetzungen aufbauen können, um effektiv zu sein.

Für eine externe Evaluation gelten folgende Voraussetzungen:

- Klare Zielvereinbarung
- Dezidierte Vorstellungen aller Beteiligten: Welchen Gewinn wollen sie aus der Evaluation ziehen?
- Offenheit für die Zusammenarbeit
- Methodische und inhaltliche Expertise der Evaluatoren für das Erreichen der Ergebnisse
- Vorlage von Ideen für die Implementierung von Verbesserungen. Die Ergebnisse sollten nämlich wieder in die Arbeit zurückfließen und dort eine Unterstützung von Veränderungsprozessen bewirken.

9 Zusammenfassung

Die vorliegende Expertise hat es sich zum Ziel gesetzt, einige kritische Fragen im Zusammenhang mit der Evaluation von Sprachförderung zusammenzutragen und verschiedene Argumente gegenüberzustellen.

Evaluationen im Bereich der Sprachförderung sind notwendig, doch dabei sollte eine ganze Reihe an Voraussetzungen und Rahmenbedingungen gegeben sein, damit sie für die beteiligten Personen auch hilfreich erscheint und somit zu einem Prozess der Weiterentwicklung beiträgt.

Anhand von zwei Beispielen einer Evaluation der Sprachförderung in *Baden-Württemberg* wurden unterschiedliche Arten von Evaluation dargestellt und die Probleme der verschiedenen methodischen Zugänge skizziert.

Die Evaluation der Sprachförderung verdeutlicht im Besonderen, dass eine Interpretation von Ergebnissen externer Evaluationen schwierig ist, und dass die Gefahr besteht, Schlussfolgerungen zu ziehen, die für den pädagogischen Prozess nicht förderlich, sondern eher hinderlich sind.

Wesentlich für das Gelingen eines Evaluationsprozesses sind folgende Punkte:

- Genaue Klärung der Ziele und der Voraussetzungen zu Beginn der Evaluation
- Angemessene Berücksichtigung von kritischen Fragen des methodischen Vorgehens und der Interpretation der Ergebnisse während der Evaluation
- Vorsichtige Implementation von Veränderungen infolge der Evaluation.

Um eine Weiterentwicklung durch Evaluationen anzustoßen und zu fördern, ist ein enges Zusammenspiel zwischen Evaluation und Qualitätsentwicklung notwendig. Hier spielt die Verknüpfung von *interner* und *externer* Evaluation eine wichtige Rolle.

Besonders intensiv ist die Frage der Nachhaltigkeit von Evaluationen zu diskutieren, wobei diese nur im Zusammenhang mit der Qualitätsentwicklung realisierbar scheint, da nur auf diese Weise Evaluation einen Beitrag zu einem längerfristigen Entwicklungsprozess leisten kann.

10 Literatur

- American Evaluation Association (2004): Guiding Principles for Evaluators. www.eval.org/Publications/GuidingPrinciples.asp (31.07.2010)
- Bortz, Jürgen/Döring, Nicola (2006): *Forschungsmethoden und Evaluation*. 4. Aufl. Heidelberg
- Deutsche Gesellschaft für Evaluation e.V. (DeGEval) (2008): *Standards für Evaluation*. Mainz
- Ditton, Hartmut (2009): *Evaluation und Qualitätssicherung*. In: Tippelt, Rudolf/Schmidt, Bernhard (Hrsg.): *Handbuch Bildungsforschung*. Wiesbaden
- Fend, Helmut (2008): *Dimensionen von Qualität im Bildungswesen. Von Produktionsindikatoren zu Prozessindikatoren am Beispiel Schule*. In: *Zeitschrift für Pädagogik*, 53. Beiheft: *Qualitätssicherung im Bildungswesen*. S. 190–209
- Gasteiger-Klicpera, Barbara/Knapp, Werner/Kucharz, Diemut/Patzelt, Doreen/Ricart Brede, Julia/Schmidt, Barbara Maria/Vomhof, Beate (2010): *Abschlussbericht der Wissenschaftlichen Begleitung des Programms „Sag’ mal was – Sprachförderung für Vorschulkinder“*. Pädagogische Hochschule Weingarten. www.sagmalwas-bw.de/media/WiBe%201/pdf/PH-Weingarten_Abschlussbericht_2010.pdf (10.10.2010)
- Gollwitzer, Mario/Jäger, Reinhold S. (2009): *Evaluation kompakt*. Weinheim
- Hager, Willi/Patry, Jean-Luc/Brezing, Hermann (Hrsg.) (2000): *Evaluation psychologischer Interventionsmaßnahmen*. Bern
- Harvey, Lee/Green, Diana (2000): *Qualität definieren. Fünf unterschiedliche Ansätze*. In: *Zeitschrift für Pädagogik*, 41. Beiheft: *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule*. S. 17–39
- Holling, Heinz/Gediga, Günther (Hrsg.) (1999): *Evaluationsforschung*. Göttingen
- Kempfert, Guy/Rolff, Hans-Günter (2005): *Qualität und Evaluation: Ein Leitfadens für Pädagogisches Qualitätsmanagement*. 4. Aufl. Weinheim
- Klieme, Eckhardt/Tippelt, Rudolf (2008): *Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz*. In: *Zeitschrift für Pädagogik*, 53. Beiheft: *Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz*. S. 7–13
- Roos, Jeanette/Polotzek, Silvana/Schöler, Hermann (2010): *EVAS Evaluationsstudie zur Sprachförderung von Vorschulkindern. Wissenschaftliche Begleitung der Sprachfördermaßnahmen im Programm „Sag’ mal was – Sprachförderung für Vorschulkinder“*. Pädagogische Hochschule Heidelberg. www.sagmalwas-bw.de/media/WiBe%201/pdf/EVAS_Abschlussbericht_Januar_2010.pdf (10.10.2010).
- Tietze, Wolfgang (2010): *Pädagogische Qualität in Kitas: Schlüsselfaktor Personal*. In: *Journal für LehrerInnenbildung*, 10. Jg., H. 1, S. 54–58
- Tietze, Wolfgang/Schuster, Käthe-Maria/Grenner, Katja/Roßbach, Hans-Günther (2007): *Kindergarten-Skala (KES-R). Feststellung und Unterstützung pädagogischer Qualität in Kindergärten*. Berlin
- Wottawa, Heinrich/Thierau, Heike (1998): *Lehrbuch Evaluation*. Bern: Hans Huber

Zur Autorin







Professorin Dr. Barbara Gasteiger-Klicpera

studierte Psychologie an der Universität Wien, wo sie auch promovierte. Sie habilitierte sich in Psychologie als Charlotte-Bühler-Stipendiatin des Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Sie ist Psychotherapeutin sowie Klinische Psychologin und Gesundheitspsychologin und war Professorin für Pädagogische Psychologie an verschiedenen deutschen Hochschulen. Seit 2010 ist sie Professorin für Integrationspädagogik und Heilpädagogische Psychologie an der Universität Graz und seit 2011 Dekanin der dortigen Umwelt- Regional- und Bildungswissenschaftlichen Fakultät. Ihre Forschungsschwerpunkte liegen bei Lese-Rechtschreibschwierigkeiten, Evaluation von Sprachförderung, Inklusion sowie Prävention von Verhaltens- und Lernschwierigkeiten.

Die Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF) stellt alle Ergebnisse in Form von Print- und Online-Publikationen zur Verfügung.

Alle Publikationen sind erhältlich unter: www.weiterbildungsinitiative.de

WiFF Expertisen	WiFF Studien	WiFF Wegweiser Weiterbildung	WiFF Kooperationen
<p>Wissenschaftliche Analysen und Berichte zu aktuellen Fachdiskussionen, offenen Fragestellungen und verwandten Themen von WiFF</p>	<p>Ergebnisberichte der WiFF-eigenen Forschungen und Erhebungen zur Vermessung der Aus- und Weiterbildungslandschaft in der Frühpädagogik</p>	<p>Exemplarisches Praxismaterial als Orientierungshilfe für die Konzeption und den Vergleich von kompetenzorientierten Weiterbildungsangeboten</p>	<p>Produkte und Ergebnisberichte aus der Zusammenarbeit mit unterschiedlichen Partnern und Initiativen im Feld der Frühpädagogik</p>
Zuletzt erschienen:	Zuletzt erschienen:	Zuletzt erschienen:	Zuletzt erschienen:
			
<p>Band 22: Tina Friederich: Zusammenarbeit mit Eltern – Anforderungen an frühpädagogische Fachkräfte</p>	<p>Band 12: Michael Ledig: Fort- und Weiterbildung von Lehrkräften an Fachschulen für Sozialpädagogik</p>	<p>Band 4: Frühe Bildung – Bedeutung und Aufgaben der pädagogischen Fachkraft</p>	<p>Band 2: Expertengruppe Berufsbegleitende Weiterbildung: Qualität in der Fort- und Weiterbildung von pädagogischen Fachkräften in Kindertageseinrichtungen</p>
<p>Band 21: Angelika Speck-Hamdan: Grundsulpädagogisches Wissen – Impulse für die Elementardidaktik?</p>	<p>Band 11: Vera Deppe: Anforderungen an die Ausbildung von Erzieherinnen und Erziehern</p>	<p>Band 3: Zusammenarbeit mit Eltern</p>	<p>Band 1: Autorengruppe Fachschulwesen: Qualifikationsprofil „Frühpädagogik“ – Fachschule/Fachakademie</p>
<p>Band 20: Franziska Nied/Renate Niesel/ Gabriele Haug-Schnabel/ Monika Wertlein/Joachim Bensel: Kinder in den ersten drei Lebensjahren in altersgemischten Gruppen</p>	<p>Band 10: Katharina Baumeister/Anna Grieser: Berufsbegleitende Fort- und Weiterbildung frühpädagogischer Fachkräfte – Analyse der Programmangebote</p>	<p>Band 2: Kinder in den ersten drei Lebensjahren</p>	
<p>Band 19: Klaus Fröhlich/Gildhoff/Iris Nentwig-Gesemann/Stefanie Pietsch: Kompetenzorientierung in der Qualifizierung frühpädagogischer Fachkräfte</p>	<p>Band 9: Rolf Janssen: Das Profil sozialpädagogischer Fachschulen</p> <p>Band 8: Rolf Janssen: Die Zugangsvoraussetzungen zur sozialpädagogischen Fachschulausbildung von Erzieherinnen und Erziehern</p>	<p>Band 1: Sprachliche Bildung</p>	

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



EUROPÄISCHE UNION

Robert Bosch **Stiftung**



Deutsches
Jugendinstitut

Programme zur sprachlichen Förderung in Kindertageseinrichtungen werden häufig mit dem Zusatz beworben, sie seien „evaluiert“. Damit wird die Assoziation nahegelegt, dass die Wirksamkeit durch ein wissenschaftliches Evaluationsverfahren belegt sei. Tatsächlich wird der Begriff der Evaluation ganz unterschiedlich verwendet und verstanden. Die Autorin zeigt in der Expertise am Beispiel der Evaluation von Sprachförderprogrammen auf, was Evaluation leisten kann und welche unterschiedlichen Formen möglich sind.