

**Systematic Evaluation Analysis (SEA)
als Instrument der Wirkungsevaluation**

Reinhard Stockmann

Stockmann, Reinhard

**Systematic Evaluation Analysis (SEA) als Instrument
der Wirkungsevaluation**

Saarbrücken, Centrum für Evaluation, 2019.

(CEval-Arbeitspapiere; 22)

NICHT IM BUCHHANDEL ERHÄTLICH

SCHUTZGEBÜHR: 5 €

**Bezug: Centrum für Evaluation (CEval)
Universität des Saarlandes
Postfach 15 11 50
D-66041 Saarbrücken
info@ceval.de**



**oder kostenfrei zum Download:
<http://www.ceval.de>**

Inhalt

1.	Evaluation und Wirkungsforschung	1
2.	Kausalproblematik in der Wirkungsforschung.....	3
3.	Metaevaluation, Systematic Review und Systematic Evaluation Analysis (SEA)	6
4.	Literaturverzeichnis	15

1. Evaluation und Wirkungsforschung

Bevor ich zu meinem eigentlichen Thema komme, lassen Sie mich zuerst etwas zu dem Tagungsthema: "Evaluation und Wirkungsforschung in der Berufsbildung" sagen. Zunächst einmal ist festzuhalten, dass es sich bei der Evaluation um das breitere Wissenschaftskonzept handelt. Während die Wirkungsforschung eine klassische Disziplin ist, die seit alters her die Zusammenhänge von Ursache und Wirkung untersucht, ist die Evaluation ein Kind der Moderne. (vgl. STOCKMANN 2007; STOCKMANN/MEYER 2014; STOCKMANN/MEYER 2017) Diese zeichnet sich dadurch aus, dass traditionell und religiös bestimmte Glaubensvorstellungen von der Ordnung der Welt durch das Vertrauen auf Rationalität und Fortschritt ersetzt werden (LERNER 1968, S. 387). Damit verbunden ist die Überzeugung, dass die Moderne gestaltet und gesteuert werden kann. Sollen soziale Veränderungsprozesse eingeleitet werden, um gesellschaftlich akzeptierte Ziele zu erreichen, z.B. Chancengleichheit, soziale Gerechtigkeit, nachhaltige Entwicklung etc., sind Informationen (Daten, Fakten, Evidences) notwendig, um solche Prozesse steuern zu können. Es gilt zu erkennen, welche Faktoren welche Wirkungen hervorrufen, damit Interventionen unter gegebenen Rahmenbedingungen gezielt ausgewählt und implementiert werden können.

Hier kommen nun die Evaluation und Wirkungsforschung ins Spiel. Beide sind geeignet – neben anderen Instrumenten – Evidences zu produzieren, mit denen nicht nur die Effektivität – also die Zielerreichung – sondern auch die Wirksamkeit von Policies, Strategien, Programmen, Projekten und einzelnen Maßnahmen bestimmt und auf ihre kausalen Ursachenfaktoren zurückgeführt werden können. Darüber hinaus können mit Hilfe von Evaluationen auch Kriterien wie Relevanz, Effizienz, Kohärenz, Signifikanz, Nachhaltigkeit etc. untersucht und bewertet werden.

Evaluation ist anders als die Wirkungsforschung nicht nur retrospektiv aufgestellt, sondern wird in Form von Ex-Ante-Evaluationen auch planerisch tätig. Mit Ex-Ante-Evaluationen werden Informationen und Bewertungen für die Planungsphase, für die aktive und prozessorientierte Gestaltung von Strategien und Prozessen geliefert. Es handelt sich um „Analysis for policy“ oder „Science for action“. In der Implementierungsphase kann diese formative Prozessperspektive beibehalten und zusätzlich um eine summative Komponente erweitert werden, indem die Durchführungsprozesse und ihre Resultate beobachtet werden (on-going-Evaluation). Zum dritten – und da trifft sich die Evaluationsforschung wieder mit der Wirkungsforschung – kann Evaluation dazu dienen, summativ, bilanzierend und ergebnisorientiert Wirkungen zu beobachten und ihren Ursachenfaktoren zuzuschreiben. In diesem Sinne betreibt Evaluation dann "Analysis of policy" oder "Science for knowledge". Werden solche

Evaluationen Ex-post durchgeführt, lässt sich auch die Nachhaltigkeit von Interventionen messen und bewerten.

Schaubild 1: Gemeinsamkeiten und Unterschiede zwischen Evaluation und Wirkungsforschung

	Evaluation	Wirkungsforschung
Herkunft	neue Disziplin	alte Disziplin
Spektrum	Planung, Implementation, Wirkung formativ + summativ	Wirkung summativ
Erkenntnisinteresse	Science for action + Science for knowledge	Science for knowledge
Ziel	Nutzen für Entscheider stiften, Bewertungen und Empfehlungen geben	Kausalität offenlegen
Methodik	empirisch wissenschaftliche Forschungslogik + Handlungsforschung	Empirisch wissenschaftliche Forschungslogik

Evaluation beinhaltet demnach immer eine Analyse- und eine Bewertungskomponente. Für die Analyse bedient sie sich – wie die Wirkungsforschung – aller Instrumente der empirischen Sozialforschung und ist auch deren wissenschaftlichen Qualitätsstandards wie Objektivität, Reliabilität und Validität verpflichtet. Doch über diese wissenschaftliche Wirkungsanalyse hinaus ist jede Evaluation mit einer Bewertung verbunden. Jede gängige Evaluationsdefinition enthält deshalb im Kern diesen Aspekt. So z.B. in der vielzitierten Definition von Donna Mertens (1998, S. 219): "Evaluation is the systematic investigation of the merit or worth of an object (program)" zumeist zu dem Zweck: "of reducing uncertainty in decision making".

Für die Bewertung des Wertes oder Nutzens eines Evaluationsobjektes werden Kriterien benötigt. Diese können, im Unterschied zu Normenreihen wie ISO, ganz verschieden sein. Einige, wie die Relevanz, Effektivität, Effizienz, Wirksamkeit, Nachhaltigkeit von Projekten, Programmen etc. habe ich schon genannt. Wichtig ist, dass die Bewertungskriterien offengelegt und klar definiert werden, so dass sie intersubjektiv nachvollzogen werden können. Wichtig ist weiterhin, dass Evaluation Bewertungen (z.B. anhand solcher Kriterien) vornimmt, aber keine Werturteile fällt. Ansonsten wäre jede Bewertung subjektiv und von dem Wertesystem des Bewertenden abhängig.

Zusammenfassend kann festgehalten werden, dass Evaluation im Vergleich zur Wirkungsforschung das breitere Wissenschaftskonzept ist, das nicht nur summativ Wirkungen erfasst und ihren Kausalfaktoren zuschreibt, sondern sich auch auf andere Programmphasen oder Aspekte beziehen kann. Zudem kommt Wirkungsforschung auch ohne die Bewertung des Nutzens der untersuchten Interventionen aus. Sie kann sich allein auf die Wirkungsbeobachtung und Kausalanalyse beschränken. Anders als die Evaluation kann Wirkungsforschung auch als Grundlagenwissenschaft betrieben werden, also ohne das Ziel, in einem Programm oder Projekt einen konkreten Nutzen zu stiften, z.B. um wissenschaftlich gesicherte Erkenntnisse für Entscheidungsprozesse zu liefern.

Schaut man sich an, wie die beiden Konzepte in der Berufsbildungsforschung verwendet werden, dann fällt auf, dass Wirkungsforschung häufig dem Kontext der empirisch analytischen Grundlagenforschung und Evaluation eher dem Konzept der zumeist qualitativ ausgerichteten Handlungsforschung zugeschrieben wird (SEVERING/WEIß 2012, S. 6; EULER 2018, S. 42-43,45). Diese Zuschreibung wird jedoch der Evaluationsforschung nicht gerecht, die zwar auch das Konzept der Handlungsforschung nutzt, sich aber grundsätzlich den Prinzipien der empirisch wissenschaftlichen Sozialforschung verpflichtet sieht, wie man z.B. den Standards für Evaluation der DeGEval entnehmen kann.

2. Kausalproblematik in der Wirkungsforschung

Wenn es um Wirkungsforschung oder Wirkungsevaluation geht, muss zunächst einmal geklärt werden, was denn die Wirkung einer Intervention darstellt. Häufig werden Wirkungen mit Outputs verwechselt. Outputs sind die Leistungen, die als Folge von Programmaktivitäten erbracht werden, also zum Beispiel die Zahl Kranker, die in einer Gesundheitsstation behandelt werden oder die Zahl durchgeführter Umweltberatungen einer Handwerkskammer.

Was uns aber in der Wirkungsevaluation oder Wirkungsforschung interessiert, sind die daraus entstandenen Wirkungen, also zum Beispiel die Verbesserung des Gesundheitszustands der behandelten Kranken oder das veränderte Umweltverhalten der Beratenen. Dabei steht die Wirkungsevaluation vor zwei Herausforderungen: Zum einen der Erfassung möglichst aller messbaren Wirkungen, d.h. sowohl der intendierten als auch nicht intendierten, sowie zweitens der Identifizierung der Ursachenfaktoren. Bei der Kausalanalyse geht es vor allem darum, alternative Erklärungen auszuschließen. So könnte ein krisenhafter Zustand von alleine wieder verschwinden. In der Medizin nennt man das "spontane Remission". Diese wäre zum Beispiel dann gegeben, wenn nicht die Behandlung des Arztes, sondern die Selbstheilungskräfte des Körpers zu einer Verbesserung des Gesundheitszustands geführt haben.

Darüber hinaus können auch exogene Ereignisse für beobachtete Veränderungen verantwortlich sein. So könnten zum Beispiel nicht die Umweltberatungen das beobachtete veränderte Verhalten ausgelöst haben, sondern gesetzliche Vorgaben oder allgemeine strukturelle Trends.

Fassen wir zusammen: Ziel der Wirkungsevaluationen ist es einerseits, möglichst viele, im Idealfall alle intendierten und nicht-intendierten Wirkungen zu erfassen. Hierfür ist der Einsatz von möglichst vielen unterschiedlichen Daten-Erhebungsmethoden, wie zum Beispiel standardisierte Interviews, leitfadengestützte Interviews, Focus-Gruppen, Beobachtungen, Sekundärdaten- oder Textanalysen etc. notwendig.

Andererseits geht es bei der Wirkungsevaluation darum, konkurrierende Erklärungen auszuschließen, um Ursache-Wirkungszusammenhänge optimal zu belegen. Für diesen Zweck bietet sich eine Reihe von Untersuchungsdesigns an, auf die ich hier nicht alle eingehen kann. Wegen ihrer hohen internen Validität werden gerne experimentelle Verfahren bevorzugt, am besten als Randomized Controlled Trials, bei denen man sich sicher sein kann, dass die beobachtete Wirkung auf eine bestimmte Intervention zurückgeführt werden kann. Allerdings ist die externe Validität von Experimenten nur schwach ausgeprägt, da sich alles darum dreht, in einer künstlichen, laborartigen Situation alle Rahmenbedingungen konstant zu halten, so dass sich Versuchs- und Kontroll-(Vergleichs-)Gruppe nur in einem Punkt unterscheiden: der Intervention. In der realen Welt lassen sich jedoch solche Rahmenbedingungen häufig nicht herstellen, so dass die externe Validität, also die Sicherheit, dass der Kausalzusammenhang, der zwischen zwei Variablen festgestellt wird, auch in anderen Rahmenbedingungen gültig ist.

Grundsätzlich geht es bei allen Kausaluntersuchungen immer darum, alternative Erklärungen auszuschließen und die Frage zu beantworten, was wäre gewesen, wenn es keine Intervention gegeben hätte:

"The goal of a good impact evaluation is to define a logically sound counterfactual that can test for and eliminate alternative explanations." (World Bank Institute 2008).

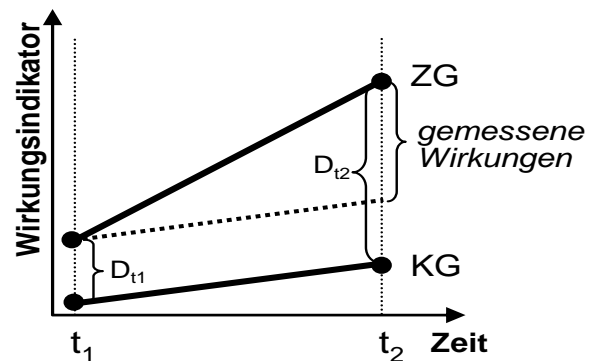
Eine solche kontrafaktische Situation ist gegeben, wenn es gelingt, eine Gruppe (Kontroll- oder Vergleichsgruppe) zu bilden, die so gut wie irgend möglich der Versuchsgruppe ähnelt, das heißt am besten mit dieser in allen wichtigen Charakteristika übereinstimmt. Wenn wir dann feststellen, dass sich Veränderungen nur bei der Versuchsgruppe einstellen, nicht aber bei der Kontrollgruppe, dann kann man mit großer Sicherheit sagen, dass der Unterschied auf die Intervention zurückzuführen ist. Um sicherzustellen, dass Kontroll- und Versuchsgruppe sich nicht unterscheiden, wird die Technik der Randomisierung angewendet. Ist diese

nicht anwendbar, können Kontrollgruppen auch anders gebildet werden, dann spricht man von Vergleichsgruppen und einem quasi-experimentellen Design. Ungeachtet dessen wird ein so genannter Double-Difference-Ansatz eingesetzt, bei dem aktuell erhobene Daten (t_2) mit Baseline- und Monitoringdaten (t_1) bei der Versuchs- und Kontroll- bzw. Vergleichsgruppe miteinander verglichen werden, um Veränderungen feststellen zu können.

Schaubild 2: Experimenteller Ansatz

Double-Difference Ansatz (t_1 und t_2)

Vergleich aktuell erhobener Daten (t_2) mit Baseline- & Monitoringdaten (t_1) bei der Ziel- und Vergleichsgruppe zur Bestimmung des tatsächlichen Ausmaßes der Wirkungszuschreibung



ZG: Zielgruppe, KG: Kontrollgruppe,
D: Differenz ZG – KG, t: Zeitpunkt

Zeitpunkt (t_1) z.B. Baseline
(t_2) z.B. Survey Daten

Nun gibt es in der Realität leider häufig das Problem, dass kein randomisiertes Experiment oder auch kein Quasi-Experiment mit Vergleichsgruppen eingesetzt werden kann.

Zum Beispiel

- wenn eine Intervention sich auf ein ganzes Land bezieht, weil von einem Gesetz, einer Regelung oder einer Versicherungsleistung alle betroffen sind,
- oder wenn aus ethischen Gründen keine Personen zu Versuchszwecken ausgeschlossen werden können, (zum Beispiel im Kontext humanitärer Hilfe oder Medikamentenversorgung),
- oder wenn sich keine Gruppen finden lassen, die nicht in irgendeiner Form gefördert wurden (zum Beispiel in Ländern mit einer sehr hohen Geburtenrate),
- oder wenn Baseline-Daten nicht erhoben wurden und sich auch nachträglich nicht mehr valide retrospektiv erheben lassen.

In solchen, nicht seltenen Fällen, müssen andere Untersuchungsdesigns für die Kausalanalyse herangezogen werden, zum Beispiel ein Single-Difference-Ansatz wie die

Querschnittsanalyse, bei dem nur ein Vergleich von Versuchs- und Kontrollgruppe zu einem Zeitpunkt (t_2) stattfindet, oder ein Pretest-Posttest Design, bei dem nur ein Vorher-Nachher-Vergleich bei der Versuchsgruppe stattfindet, oder ein um weitere Mess-Zeitpunkte erweitertes Paneldesign, Zeitreihenanalyse oder eben ein Verfahren, das ich Ihnen heute vorstellen möchte, die Systematic Evaluation Analysis (SEA).

3. Metaevaluation, Systematic Review und Systematic Evaluation Analysis (SEA)

Zuerst ist eine kleine Begriffsexplikation vonnöten, da die folgenden Begriffe in einer verwirrenden, teils sich widersprechenden Vielfalt verwendet werden. Dennoch haben Metaevaluation, Metaanalyse, Literature Review, Systematic Review, Research Synthesis ein Ziel gemeinsam, sie wollen Wissen systematisch kumulieren.

Nachdem der Begriff Metaevaluation zunächst sowohl für die Kumulation inhaltlicher Befunde aus Evaluationsstudien als auch für die Bewertung ihrer methodischen Qualität verwendet wurde, hat sich mittlerweile eine terminologische Differenzierung durchgesetzt. (STOCKMANN/SILVESTRINI 2013) Demnach wird der Begriff Metaevaluation nur noch für die Evaluation von Evaluationen¹ (SCRIVEN 1991, S. 228ff.; WIDMER 1996, S. 4, CASPARI 2009, S. 2f.) als eine Form der Qualitätsprüfung von Evaluation verwendet. Diese erfolgt häufig entlang der Standards für Evaluation.

Der Begriff Metaanalyse bezeichnet hingegen eine Vorgehensweise, die systematisch die Ergebnisse verschiedener Einzelstudien aus einem Forschungsbereich integriert und subsumiert, also rein inhaltlich ausgerichtet ist. Diesem Ansatz ähnelt das Systematic Review, das das Ziel verfolgt, "to locate, appraise and synthesize the best available evidence relating to a specific research question in order to provide informative and evidence-based answers" (BOLAND/CHERRY/DICKSON 2017, S. 2), Systematic Reviews "typically involve a detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic" (UMAN 2011, S. 57).

¹ Schon 1940 hat Orata seinen Aufsatz mit „Evaluating Evaluation“ (ORATA, S. 641) für diesen Sachverhalt beschrieben.

Mitunter werden Systematic Reviews als Untertyp des Literature Review bezeichnet (COOPER 2017, S. 1ff.), die bereits publiziertes Material auswerten „not based primarily on new facts and findings but on publications containing such primary information, whereby the latter is digested, sifted, classified, simplified, and synthesized" (MANTEN 1973, S. 75).

Von diesen Begriffen wird wiederum der Begriff Research Synthesis mit dem Verweis eines anderen Focus und Ziels abgegrenzt: „Research syntheses attempt to integrate empirical research for the purpose of creating generalizations. (...). Also, research synthesis almost always pay attention to relevant theories, critically analyze the research they cover, try to resolve conflicts in the literature, and attempt to identify central issues for future research." (COOPER/HEDGES 2009, S. 6)

Hier wird eine Analysemethode angewendet, die sowohl Elemente der Metaevaluation als auch der Metaanalyse bzw. der damit verwandten Reviewkonzepte umfasst und deshalb als Systematic Evaluation Analysis (SEA) bezeichnet werden soll. Mit der Metaevaluation hat SEA gemeinsam, dass es sich um die Analyse von Evaluationen handelt, deren methodische Qualität geprüft wird, um dann inhaltliche Ergebnisse nach ausgewählten Kriterien oder Hypothesen zu synthetisieren, zu strukturieren, zu klassifizieren und zu subsumieren, um letztlich kumulierte, verdichtete Befunde zu generieren, die Generalisierungen zulassen. Eine Evaluation der methodischen Qualität der dafür herangezogenen Studien ist notwendig, da eine inhaltliche Kumulation von Ergebnissen nur auf der Basis methodisch einwandfrei durchgeführter Studien Sinn macht, die eine hohe Reliabilität und Validität aufweisen.

Bei der hier vorgenommenen Analyse handelt es sich nicht um ein Literature-Review, sondern um die Synthetisierung einer Reihe von selbst durchgeführten Evaluationen zur deutschen Berufsbildungszusammenarbeit (vgl. Tabelle 1).

Tabelle 1: Empirische Grundlage: Analyse bilateraler Berufsbildungsk Kooperationen

Jahr	Autoren / Autorinnen	Analysierte bilaterale Berufsbildungsmaßnahmen
1992	Stockmann	9 in Lateinamerika
1996	Stockmann	15 in Lateinamerika
1997	Stockmann, Leicht	1 in Afrika
1998	Stockmann, Kohlmann	6 in Afrika 6 in Lateinamerika 5 in Asien (davon eine in China)
2000	Stockmann, Meyer, Krapp, Köhne	8 in Asien (alle in China)
2013	Stockmann, Silvestrini	6 in Asien (davon eine in China) 3 in Afrika 2 in Europa 1 in Lateinamerika
2017	Stockmann, Meyer	21 in Asien (alle in China)

Ziel ist es, die Erfolgsfaktoren zu identifizieren, die für die Effektivität, Wirksamkeit und Nachhaltigkeit von Berufsbildungsprojekten verantwortlich sind. Während es in den ausgewählten Studien selbst vor allem darum ging, die interne Validität der Aussagen zu stärken, und dementsprechend – wenn möglich – quasi-experimentelle Untersuchungsansätze verwendet wurden, geht es jetzt darum, ein Verfahren anzuwenden, dass eine hohe externe Validität aufweist, um zu generalisierbaren Aussagen zu kommen. Hierfür wird so vorgegangen, dass systematisch untersucht wird, welche Faktoren für den 'Erfolg' eines Projekts oder Programmes zwingend erforderlich sind. Hierfür werden die in den einzelnen Evaluationen identifizierten Erfolgsfaktoren daraufhin abgeglichen, wie häufig sie aufgetreten sind. Es wird versucht herauszufinden, ob es sich nur um individuelle Erfolgsfaktoren handelt, die nur einmal im spezifischen Programm- oder Länderkontext auftreten, oder ob es sich um Faktoren handelt, die stets auftreten, wenn es sich um ein erfolgreiches Projekt handelt. In diesem Fall wurde gefolgert, dass es sich unabhängig vom jeweiligen Programmtypus und unabhängig vom jeweiligen Programmkontext um einen generalisierbaren, d.h. zwingend notwendigen Erfolgsfaktor handelt.

Für die Analyse wurden sieben in den Jahren 1992 bis 2017 durchgeführte Studien herangezogen, in denen 83 Projekte und Programme der Berufsbildungszusammenarbeit weltweit, mit dem gleichen Konzept, dem CEval- Evaluationsansatz, untersucht wurden.

Da die Berufsbildungsförderung zum Urgestein deutscher Entwicklungspolitik gehört, also schon über eine 50-jährige Historie verfügt, ist es nicht verwunderlich, dass die Förderziele und die Maßnahmen, um diese zu erreichen, in den vergangenen Dekaden einem deutlichen Wandel unterlagen. Waren die 1960er und 1970er Jahre zunächst noch durch die Förderung vollschulischer Ausbildungsmodelle geprägt, wurden in den 1980er und 1990er Jahren Systemreformen mit dem Ziel, duale oder kooperative Ausbildungsformen einzuführen, unterstützt. Spätestens um die Jahrtausendwende wurde die Berufsbildung stärker in den Fokus von Wirtschafts- und Beschäftigungsförderung gestellt und am Leitbild einer sozialen und ökologischen Marktwirtschaft ausgerichtet. Zusätzlich sollte die Berufsbildung aber auch direkt zur Armutsbekämpfung beitragen.

Klassifiziert man die untersuchten Projekte, lassen sich drei, allerdings unterschiedlich starke Gruppen ausmachen. Am umfangreichsten sind die Institutionenförderprojekte, die ihre Blütezeit zwischen 1960 und 1980 hatten und die danach zwar von systemreformerischen Projekten abgelöst wurden, ohne allerdings ihre wichtige Bedeutung bis heute zu verlieren. Projekte für marginalisierte Gruppen kamen vermehrt erst in neuerer Zeit auf und spielen insgesamt im Berufsbildungsportfolio nach wie vor eine marginale Rolle.

Schaubild 3: Klassifikation der untersuchten Projekte

- ✓ Institutionenförderprojekte:
 - Singuläre Projekte wurden unterstützt, um deren Leistungsangebot zu verbessern
- ✓ „Kooperative“ Ausbildungs-Projekte:
 - Erhöhung des Praxisanteils in der beruflichen Ausbildung und Verbreitung des Ausbildungstyps durch Systemreformen
- ✓ Marginalisierte Gruppen
 - Projekte im Zusammenhang mit friedensschaffenden, konfliktreduzierenden Maßnahmen
 - Erwachsenenbildung

Insgesamt lässt sich feststellen, dass die Ansprüche an Berufsbildungsvorhaben im Zeitverlauf immer höhergeschraubt wurden, indem immer mehr Anforderungen an sie gestellt wurden. So sollen sie einerseits dazu beitragen, in den Partnerländern soziale und ökologische Marktwirtschaften aufzubauen, andererseits aber auch Armut zu reduzieren. Diese Ziele sind jedoch nur schwer miteinander kompatibel, da im ersten Fall vor allem systemreformerische Ansätze auf makrogesellschaftlicher Ebene gefragt sind, für die zweite Zielvorgabe jedoch vor allem Projekte im informellen Sektor und auf mikrogesellschaftlicher Ebene vonnöten sind.

Die im Zeitverlauf zu beobachtende zunehmende Komplexität von Programmen und Projekten hin zu einem vielfältigen, ausdifferenzierten Zieletableau hat sich nicht unbedingt ausgezahlt, denn die Evaluationsstudien zeigen, dass weniger ambitionierte Projekte, wie zum Beispiel die Förderung ausgewählter Berufsbildungseinrichtungen (Institutionen-Förderprojekte) ohne Breitenwirksamkeits- und Systemveränderungsanspruch und Projekte, die ohne einen komplexen Mehrebenenansatz implementiert werden konnten, im Hinblick auf das Nachhaltigkeitskriterium, am besten abschneiden.

Dieser Befund sollte natürlich nicht dahingehend interpretiert werden, möglichst 'einfache' Projekte und Programme durchzuführen, da diese am erfolgversprechendsten sind, sondern komplexe Projekte, also zum Beispiel solche mit einem Systemveränderungsanspruch, vor Beginn besonders genau auf seine Erfolgchancen, zum Beispiel mithilfe einer Ex-Ante-Evaluation, zu prüfen.

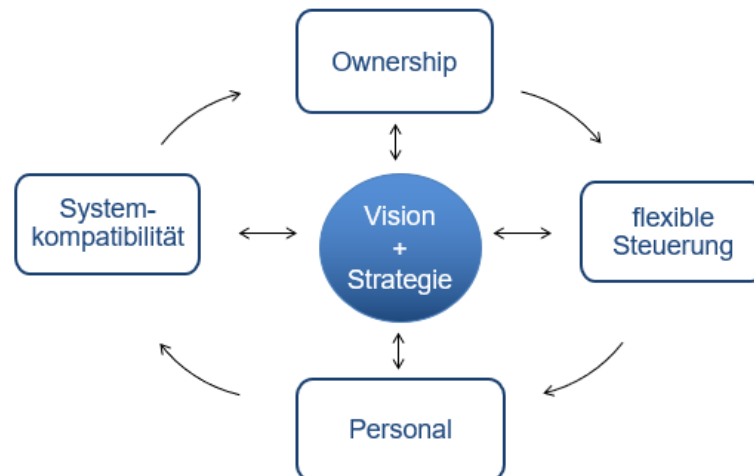
Dabei sollten die in unserer Systematic Evaluation Analysis herausgearbeiteten Erfolgsfaktoren beachtet werden. Hierfür wurde die Vielfalt der in den einzelnen Projekt- und Programmevaluationen ermittelten Erfolgsgrößen in einem, wie eingangs dargestellten, vergleichenden Verfahrens auf vier Kerngrößen verdichtet² (vgl. Schaubild 4). Diese vier ‚Schlüsselgrößen‘ sind:

1. Systemkompatibilität: die im Rahmen der Berufsbildungszusammenarbeit eingeführten Maßnahmen, Institutionen und Organisationen müssen mit dem bestehenden Berufsbildungssystem vereinbar sein oder dieses entsprechend einer besseren Passung verändern;
2. Flexible Steuerung: die Vorhaben der Berufsbildungszusammenarbeit müssen auf veränderte Rahmenbedingungen, technische Innovationen, wirtschaftliche Krisen, neue Kompetenzanforderungen und/oder veränderte Bedürfnisse der Beteiligten schnell und angemessen reagieren und die implementierten Maßnahmen, Institutionen und Organisationen entsprechend den veränderten Anforderungen anpassen;
3. Personal: die mit der Durchführung betrauten organisatorischen Einheiten müssen entsprechend der Anforderungen mit angemessenen Ressourcen ausgestattet sein, worunter in allererster Linie geeignetes Personal mit den benötigten Kompetenzen zu verstehen ist;

² Dieses Vorgehen orientiert sich am Modell der Balanced Scorecard („ausgewogener Berichtsbogen“). Hierbei handelt es sich um ein Konzept zur Messung, Dokumentation und Steuerung von Unternehmensaktivitäten. Es dient als Führungsinstrument zur Steuerung von Organisationen anhand strategischer Ziele (vgl. KA-PLAN/NORTON 1997, 2001; FRIEDAG/SCHMIDT 2011, GLEICH 2012.).

4. Ownership: die Trägerorganisationen von Programmen und Projekten müssen sich die Vorhaben zu eigen machen und diese nicht als „von außen“ (insbesondere von den ausländischen Kooperationspartnern) aufgesetzt begreifen, damit sie dauerhaft – auch nach Ende der Vorhabenlaufzeit – mit den notwendigen Ressourcen ausgestattet sind und die zur Erzielung von Wirkungen erforderlichen Entscheidungen getroffen werden.

Schaubild 4: Kernschlüsselgrößen

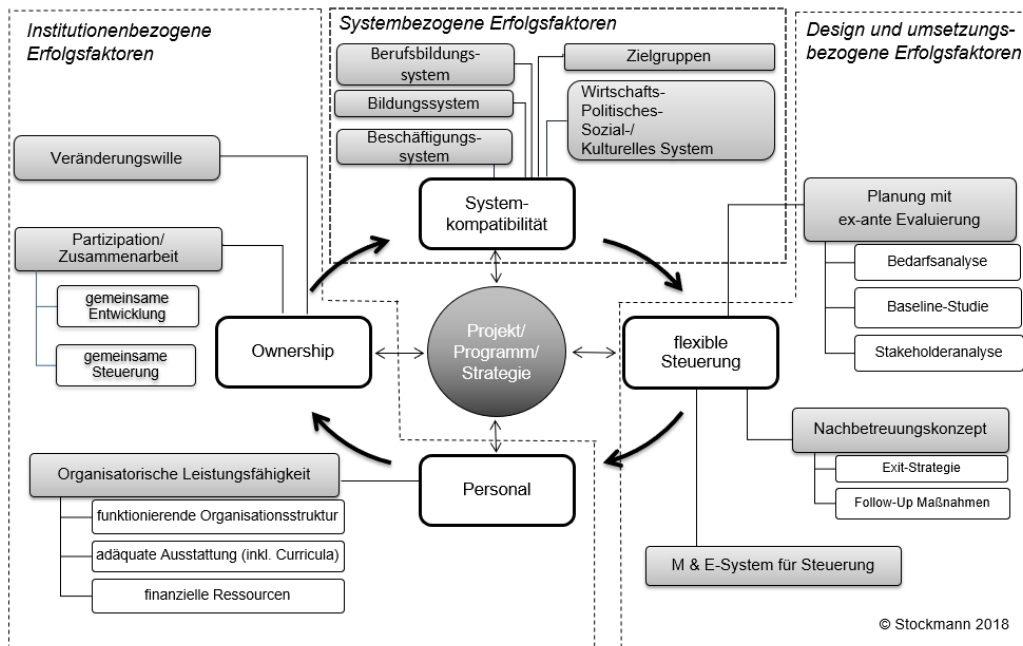


Diese Kerngrößen werden von drei unterschiedlichen Bündeln (Clustern) von Einflussfaktoren geprägt, wobei die Effektstärke zwischen unterschiedlichen Vorhabenarten variieren kann. Diese Cluster sind (vgl. Schaubild 5):

1. Systembezogene Erfolgsfaktoren: alle Einflüsse, die aus dem Berufsbildungssystem sowie den mit ihm verbundenen anderen gesellschaftlichen Systemen (speziell dem Beschäftigungs-, dem Bildungs-, dem politischen und dem Wirtschaftssystem) resultieren. Hier handelt es sich z.B. um kulturelle Eigenarten, gesetzliche Vorgaben, Curricula, gesellschaftliche Regeln, Verwaltungsmaßnahmen, wirtschaftliche und politische Entwicklungen.
2. Design und umsetzungsbezogene Einflussfaktoren: alle Einflüsse, die sich aus der Planung, Konzeption, Durchführung und Steuerung der Berufsbildungsvorhaben ergeben. Hier handelt es sich z.B. um die bereitgestellten Ressourcen, die Laufzeit, die implementierten Entscheidungsstrukturen, das Monitoring- und Evaluationssystem, Führungsstil und Führungskultur innerhalb der implementierten Strukturen.
3. Institutionenbezogene Erfolgsfaktoren: alle Einflüsse, die aus den Strukturen, Prozessen, Regeln und Verhaltensweisen der beteiligten Organisationen und Netzwerken abzuleiten sind. Unter diesen Faktoren sind die Leistungsfähigkeit der Trägerorganisationen, die formale und informelle Kooperation in der bilateralen Zusammenarbeit und zwischen allen

beteiligten Stakeholdern, sowie deren Einstellungen gegenüber dem Vorhaben und den implementierten Maßnahmen.

Schaubild 5: Modell der Schlüsselgrößen nachhaltiger BBZ



Aus diesem sehr allgemeinen und aus einer Vielzahl unterschiedlicher Vorhaben der Berufsbildungszusammenarbeit beruhenden Schema lassen sich Thesen und Erwartungen über die Wirkungen und die fördernden bzw. behindernden Faktoren konkreter Einzelbeispiele ableiten. An dieser Stelle seien als Beispiel nur einige zentrale Thesen des Modells angesprochen:

- Von den design- und umsetzungsbezogenen Faktoren (C) ist eine flexible Steuerung (K) am wichtigsten, die auf neue Informationen und sich verändernde Umweltbedingungen schnell und angemessen reagiert. Die hierfür benötigten Gestaltungsspielräume und schlanken Entscheidungsstrukturen sind für den Erfolg wichtiger als eine sorgfältige, umfassende und detaillierte Planung. Planungsfehler können durch eine flexible Steuerung eher aufgefangen werden als umgekehrt gut geplante, aber an den realen Erfordernissen vorbei gehende Maßnahmen, die entsprechend des Plans akribisch abgearbeitet werden. In der Durchführungsphase ist deshalb vor allem eine rechtzeitige Bereitstellung relevanter und korrekter Monitoring- und Evaluationsergebnisse für die notwendigen Entscheidungen wichtig, durch die Steuerungsfehler vermieden oder zumindest begrenzt werden können. Bei Auslaufen von Fördermitteln und einem entsprechenden Übergang ist die rechtzeitige Implementierung angemessener Überleitungsmaßnahmen entscheidend.

- Der zentrale Erfolgsfaktor jeder Berufsbildungsmaßnahme ist deren Systemkompatibilität (K), also eine hohe Vereinbarkeit mit bzw. eine Anschlussfähigkeit an bestehende Strukturen und Abläufe. Als kritischer Aspekt bei der Etablierung einer praxisnahen beruflichen Qualifizierung hat sich vor allem die Balance im Verhältnis Arbeitgeber, Arbeitnehmer und Staat bezüglich der Aus- und Weiterbildungsverantwortung herausgestellt. Hier helfen gut etablierte Vertretungsstrukturen und Institutionen zur Regulierung des Interessenausgleichs weiter, die allerdings eine entsprechende Vertretungs- und Verhandlungskultur voraussetzen. Dies lässt sich nicht einfach durch staatliche Maßnahmen von oben „verordnen“, sondern erfordert einen langwierigen und in der Regel schwierigen Einstellungs- und Kulturwandel, welcher durch entsprechende, den Diskurs fördernde Maßnahmen unterstützt werden muss.
- Die Umsetzung von Maßnahmen innerhalb eines Vorhabens der Berufsbildungszusammenarbeit erfordert schließlich den Aufbau eines entsprechenden institutionellen Rahmens, der vor allem durch das gegenseitige Vertrauen in die Zusammenarbeit sowie die Bereitschaft des Partners geprägt ist, die im Rahmen der Kooperation entwickelten Maßnahmen und Institutionen eigenständig auch nach Auslaufen der Förderung fortzuführen. Häufig bleiben Projekte und Programme „Fremdkörper“ innerhalb der Trägerorganisationen und können sich im bestehenden institutionellen Geflecht nicht etablieren. Dies führt zu fehlender „Ownership“ (K) des Trägers, wodurch die erreichten Erfolge nach Beendigung eines Vorhabens weitgehend verloren gehen. Aus diesem Grund spielt gerade in der Berufsbildungszusammenarbeit der menschliche Faktor eine besondere Rolle: ohne eine kritische „Masse“ von qualifiziertem und engagiertem Personal (K) ist weder eine überzeugende Performanz innerhalb des Vorhabens (z.B. durch kompetente Projektsteuerung und Maßnahmenumsetzung) noch eine diplomatisch geschickte Einbindung anderer Organisationseinheiten und/oder Netzwerkteilnehmer (und damit eine dauerhafte institutionelle Verankerung) möglich. In der Regel wird die Bereitstellung entsprechender personeller Ressourcen von Dritten auch als Signal für die Bedeutung angesehen, welche die Trägerorganisation dem Vorhaben beimisst und wertet dieses entsprechend in der Außendarstellung auf.

Die im Modell dargestellten Erfolgsfaktoren beruhen auf einer Vielzahl von Berufsbildungsevaluationen aus den letzten 40 Jahren, in denen ein enormer konzeptioneller Wandel stattfand. Sie umfassen eine Vielfalt unterschiedlicher Projekte und Programme und wurden in den verschiedensten sozio-kulturellen Kontexten und politischen Systemen durchgeführt. Dies gibt Anlass zu der Vermutung, dass diese Kerngrößen eine hohe zeitliche und regionale Stabilität aufweisen.

Es wäre deshalb wünschenswert, wenn bei der Planung und Implementation neuer Projekte und Programme der Berufsbildungszusammenarbeit dieses Schlüsselgrößenmodell Beachtung finden würde.

4. Literaturverzeichnis

- BOLAND, Angela; CHERRY, Gemma M.; DICKSON, Rumona: Carrying out a systematic review as a master's thesis. In: BOLAND, Angela; CHERRY, Gemma M.; DICKSON, Rumona (Hrsg.): Doing a systematic review. A Student's Guide. 2. Auflage. London 2017, S. 1-21.
- CASPARI, Alexandra: Lernen aus Evaluierungen. Meta-Evaluation & Evaluationssynthese von InWEnt-Abschlussevaluierungen 2009. Frankfurt 2009.
- COOPER, Harris: Research synthesis and meta-analysis. A step-by-step approach. 5. Auflage. Duke University 2017.
- COOPER, Harris; HEDGES, Larry V.: Research synthesis as a scientific process. In: COOPER, Harris; HEDGES, Larry V.; VALENTINE, Jeffrey C. (Hrsg.): The handbook of research synthesis and meta-analysis. 2. Auflage. New York 2009. S. 3-17.
- EULER, Dieter: Gemessenes und Angemessenes – Berufsbildungsforschung auf der Suche nach einem Profil... In: WEIß, Reinhold; SEVERING, Eckart (Hrsg.): Multidisziplinär – praxisorientiert – evidenzbasiert: Berufsbildungsforschung im Kontext unterschiedlicher Anforderungen. Bonn 2018, S. 30-53.
- FRIEDAG, Herwig R.; SCHMIDT, Walter: Balanced Scorecard. 4. Auflage. Haufe 2011.
- GLEICH, Ronald (Hrsg.): Balanced Scorecard. Best-Practice-Lösungen für die Unternehmenssteuerung. Freiburg: Haufe 2012.
- KAPLAN, Robert S.; NORTON, David P.: Balanced Scorecard – Strategien erfolgreich umsetzen. Stuttgart: Schäffer-Poeschel 1997.
- KAPLAN, Robert S.; NORTON, David P.: Die strategiefokussierte Organisation: Führen mit der Balanced Scorecard. Stuttgart: Schäffer-Poeschel 2001.
- LERNER, David: Modernization. Social Aspects. In: International Encyclopaedia of the Social Sciences, (1968) Vol. 10.
- MANTEN, Arie A.: Scientific Literature Review, Scholarly Publishing. (1973) Vol. 5, S.75-89.
- MERTENS, Donna: Research methods in education and psychology: integrating diversity with quantitative & qualitative approaches. Sage Publications 1998.

ORATA, Pedro T.: Evaluating Evaluation. In: Journal of educational research, (1940) Vol. 33(9), S. 641-661.

SCRIVEN, Michel: 1991: Evaluation Thesaurus. 4. Auflage. Newbury Park, London, New Delhi 1991.

SEVERING, Eckart; WEIß, Reinhold: Qualitätsentwicklung in der Berufsbildungsforschung: Plädoyer für eine neue Diskussion. In: SEVERING, Eckart; WEIß, Reinhold (Hrsg.): Qualitätsentwicklung in der Berufsbildungsforschung. Bonn 2012, S. 5-14.

STOCKMANN, Reinhard: Die Nachhaltigkeit von Entwicklungsprojekten. Eine Methode zur Evaluierung am Beispiel von Berufsbildungsprojekten. Wiesbaden 1992.

STOCKMANN, Reinhard: Die Wirksamkeit der Entwicklungshilfe. Eine Evaluation der Nachhaltigkeit von Programmen und Projekten der Berufsbildung. Opladen 1996.

STOCKMANN, Reinhard: Handbuch zur Evaluation. Eine praktische Handlungsanleitung. Münster 2007.

STOCKMANN, Reinhard; KOHLMANN, Uwe: Transferierbarkeit des Dualen Systems. Eine Evaluation dualer Ausbildungsprojekte in Entwicklungsländern. Berlin 1998.

STOCKMANN, Reinhard; LEICHT, René: Implementationsbedingungen eines kooperativen Ausbildungssystems in Ägypten. Berlin 1997.

STOCKMANN, Reinhard; MEYER, Wolfgang (Hrsg.): Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven. Münster 2017.

STOCKMANN, Reinhard; MEYER, Wolfgang: Chinas Berufsbildung im Wandel: 30 Jahre Entwicklungszusammenarbeit mit der Hanns-Seidel-Stiftung. Münster 2017.

STOCKMANN, Reinhard; MEYER, Wolfgang: Evaluation. Eine Einführung. 2. Auflage. Barbara Budrich 2014.

STOCKMANN, Reinhard; MEYER, Wolfgang; KRAPP, Stefanie; KÖHNE, Gerhard: Wirksamkeit deutscher Berufsbildungszusammenarbeit. Ein Vergleich staatlicher und nicht-staatlicher Programme in der Volksrepublik China. Wiesbaden 2000.

STOCKMANN, Reinhard; Silvestrini, Stefan (Hrsg.): Metaevaluierung Berufsbildung. Ziele, Wirkungen und Erfolgsfaktoren der deutschen Entwicklungszusammenarbeit. Münster 2013.

UMAN, Lindsay S.: Systematic reviews and meta-analyses. In: Journal of the Canadian Academy of Child and Adolescent Psychiatry, (2011) Vol. 20(1), S. 57-59.

WIDMER, Thomas: Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Haupt 1996.