



In fünf Schritten zum zielführenden Evaluationsdesign

Eine Handreichung für
Bildungsinitiativen im MINT-Bereich

Reinhold Nickolaus, Svitlana Mokhonko (Hrsg.)

acatech MATERIALIEN

Autoren/Herausgeber:

Prof. Reinhold Nickolaus
Universität Stuttgart
Institut für Erziehungswissenschaft, Abteilung Berufs-, Wirtschafts- und Technikpädagogik
Geschwister-Scholl-Straße 24 D
70174 Stuttgart

Dr. Svitlana Mokhonko
Universität Stuttgart
Institut für Erziehungswissenschaft, Abteilung Berufs-, Wirtschafts- und Technikpädagogik
Geschwister-Scholl-Straße 24 D
70174 Stuttgart

Projekt:

acatech Arbeitskreis Bildung

Reihenherausgeber:

acatech – DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN, 2016

Geschäftsstelle
Karolinenplatz 4
80333 München

Hauptstadtbüro
Pariser Platz 4a
10117 Berlin

Brüssel-Büro
Rue d'Egmont/Egmontstraat 13
1000 Brüssel
Belgien

T +49(0)89/5 20 30 90
F +49(0)89/5 20 30 99 00

T +49(0)30/2 06 30 96 0
F +49(0)30/2 06 30 96 11

T +32(0)2/2 13 81 80
F +32(0)2/2 13 81 89

E-Mail: info@acatech.de
Internet: www.acatech.de

Empfohlene Zitierweise:

Nickolaus, R./Mokhonko, S. (Hrsg.): *In fünf Schritten zum zielführenden Evaluationsdesign. Eine Handreichung für Bildungsinitiativen im MINT-Bereich* (acatech MATERIALIEN), München 2016.

ISSN: 2191-8481

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.d-nb.de> abrufbar.

© acatech – DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN, 2016

Koordination: Dr. Thomas Lange, Susanne Schröder

Redaktion: Susanne Schröder

LayoutKonzeption: acatech

Konvertierung und Satz: Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Sankt Augustin

> DIE REIHE acatech MATERIALIEN

In dieser Reihe erscheinen Diskussionspapiere, Vorträge und Vorstudien, die im Rahmen der acatech Projektarbeit entstanden sind. Die Bände dieser Reihe liegen in der inhaltlichen Verantwortung der jeweiligen Herausgeber und Autoren.

Alle bisher erschienenen acatech Publikationen stehen unter www.acatech.de/publikationen zur Verfügung.

Der vorliegende Materialienband ist ein Impuls des acatech Arbeitskreis Bildung und aus den dort geführten Diskussionen zum Thema Qualitätssicherung und Wirkungsmessung von MINT-Initiativen hervorgegangen. In der Auseinandersetzung mit diesen Themen zeigte sich, dass viele Förderer und Bildungsanbieter ihre Arbeit evaluieren möchten, häufig jedoch nicht wissen, wie die Wirksamkeit der eigenen Arbeit zielführend gemessen werden kann. Es besteht ein großer Bedarf an Orientierung und der Wunsch nach Hilfestellung zur Identifikation und Durchführung von geeigneten Evaluationen. Die Auswahl eines, zur Maßnahme passenden, Evaluationsdesigns ist dabei ein erster wichtiger Schritt. Die dafür notwendigen Informationen stellt diese Handreichung zur Verfügung.

Sie soll zudem als ein Input aus der Wissenschaft zur aktuellen Diskussion um die Möglichkeiten der Wirkungsmessung im zivilgesellschaftlichen Sektor verstanden werden. Intensiv diskutiert wird dieses Thema auch im Nationalen MINT Forum in der Arbeitsgruppe Qualitätssicherung/Evaluation unter der Leitung von Prof. Dr. Reinhold Nickolaus, der Autor dieser Handreichung und Mitglied des acatech Arbeitskreis Bildung ist. Der kontinuierliche Austausch zwischen acatech, dem Nationalen MINT Forum und der Arbeitsgruppe Qualitätssicherung/Evaluation zu diesem wichtigen Thema wird im Rahmen der Zusammenarbeit in zukünftigen gemeinsamen Projekten fortgesetzt.

Wir danken Herrn Professor Nickolaus und seiner Mitarbeiterin Frau Dr. Svitlana Mokhonko für die Erstellung der Handreichung.

INHALT

1	GUT GEMEINT HEISST NICHT GUT GEMACHT	6
	Warum evaluieren wir überhaupt? – Grundlegende Funktionen von Evaluationen	6
2	WAHL DES EVALUATIONSDESIGNS – WIE FINDE ICH DAS RICHTIGE DESIGN FÜR MEINE ZWECKE?	7
	Schritt 1 – Ziele der Evaluation bestimmen	7
	Schritt 2 – Mögliche Wirkungszusammenhänge vorab klären	9
	Schritt 3 – Erfolgskriterien der untersuchten Maßnahmen bestimmen	11
	Schritt 4 – Fragestellung präzisieren	11
	Schritt 5 – Design festlegen	12
3	FAZIT UND CHECKLISTE	13
	Checkliste zur Prüfung eines Evaluationsvorschlags	14
4	BEISPIELE FÜR EVALUATIONEN	15
	Beispiele für Ergebnis-Evaluationen	15
	Beispiele für Prozess-Evaluationen	15
	Weitere Checklisten und Handreichungen	16
	LITERATUR	17

1 GUT GEMEINT HEISST NICHT GUT GEMACHT

Die Themen **Qualitätssicherung** und **Wirkungsmessung** haben im zivilgesellschaftlichen Sektor stark an Bedeutung gewonnen. Viele Initiativen, Bildungsanbieter und Fördereinrichtungen möchten die Wirkung ihrer Arbeit messen und durchgeführte Maßnahmen und Projekte evaluieren.

Gut gemeint ist dabei aber noch nicht gut gemacht. Mangelnde Sorgfalt zu Beginn eines Evaluationsprozesses konterkariert jede gute Absicht der Qualitätssicherung – denn selbst die ausgefeiltesten Analyseverfahren im Rahmen eines Evaluierungsverfahrens können Defizite des zugrunde liegenden Evaluationsdesigns später nicht mehr ausgleichen.

Diese Handreichung für Praktikerinnen und Praktiker soll Organisationen auf Basis wissenschaftlicher Erkenntnisse **in der frühen Phase der Entscheidung zur Einführung von Evaluationsmaßnahmen und deren Umsetzung** bei MINT-Initiativen unterstützen, um deren Wirksamkeit zu erhöhen.

WARUM EVALUIEREN WIR ÜBERHAUPT? – GRUNDLEGENDE FUNKTIONEN VON EVALUATIONEN

Evaluationen können mit unterschiedlichen Zielperspektiven durchgeführt werden. Je nach Zielsetzung muss ein passender Evaluationszuschnitt gewählt werden.

Zum Beispiel kann es einer Initiative darum gehen, eine (neu entwickelte) Maßnahme zu optimieren oder auch die Effekte einer Maßnahme zu überprüfen, um eine Entscheidungsgrundlage für die Weiterführung oder Weiterfinanzierung zu erhalten. Abhängig von diesen Zielsetzungen bedarf es jeweils spezifischer Evaluationsansätze.

Evaluationen beziehen sich immer auf ganz spezifische Maßnahmen und müssen am Ende verwertbare Informationen zu deren Beurteilung bereitstellen. Dies setzt voraus, dass die Evaluation explizit auf den Zweck der jeweiligen

Maßnahme ausgerichtet ist. Da sich die Maßnahmen nicht nur in ihren Zielsetzungen, sondern ebenso in den dafür eingesetzten Ressourcen erheblich unterscheiden können, muss zu Beginn geklärt werden, welcher Aufwand (vor und während der Evaluation) welchem Ertrag (Ergebnisse/Informationen nach der Evaluation) gegenüber steht.

Bei **kleineren und zeitlich befristeten Fördermaßnahmen** lassen sich größere Evaluationsaufwendungen nur rechtfertigen, wenn Interesse an grundlegenden Klärungen besteht und die Evaluation der Maßnahme geeignet scheint, solch elementare Fragen anzugehen. Das wird in der Praxis eher die Ausnahme sein.

Bei **umfangreichen und länger andauernden Maßnahmen**, in die erhebliche Ressourcen investiert werden, sind aufwändigere Evaluationen eher vertretbar. Die in Kapitel 2 skizzierten Evaluationszugänge können auch als „Baukasten“ genutzt werden, um für die jeweilige Maßnahme oder das Projekt einen geeigneten Ansatz auszuwählen beziehungsweise über Evaluationsangebote zu entscheiden.

Die Auswahl des Evaluationszuschnitts bestimmt die Aussagekraft der Evaluation – ebenso wie die handwerklich saubere Umsetzung.

Dazu müssen bei quantitativen Evaluationszugängen die **Gütekriterien empirischer Untersuchungen** berücksichtigt werden¹: Objektivität (das Ergebnis der Untersuchung ist unabhängig von der durchführenden Person), Reliabilität (die Genauigkeit beziehungsweise Zuverlässigkeit eines Verfahrens ist auch bei wiederholter Messung gegeben) und Validität (die Messung ist gültig und der Test misst, was gemessen werden soll). Darüber hinaus muss sichergestellt sein, dass die **Personen, die die Evaluation durchführen, entsprechend geschult sind**. Zudem sollte sich jede Organisation die Frage stellen, ob eine interne Evaluation ausreicht oder eine externe Evaluation nötig erscheint, um die gewünschte **Objektivität** zu gewährleisten.

¹ Bortz, Jürgen/Döring, Nicola 2006.

2 WAHL DES EVALUATIONSDESIGNS – WIE FINDE ICH DAS RICHTIGE DESIGN FÜR MEINE ZWECKE?

Die Wahl eines Evaluationsdesigns ist von verschiedenen Bedingungen abhängig. Um ein passendes Evaluationsdesign auszuwählen, ist es hilfreich, schrittweise vorzugehen und folgende Punkte organisationsintern zu klären:

1. **Ziele der Evaluation bestimmen**
Zum Beispiel Optimierung von Prozessen, Kontrolle/Überprüfen der Wirkung einer Maßnahme.
2. **Mögliche Wirkungszusammenhänge vorab klären**
Zum Beispiel strukturelle Zusammenhänge zwischen unterschiedlichen Einflussfaktoren und Zielgrößen.
3. **Kriterien bestimmen, die zur Beurteilung herangezogen werden sollen**
Zum Beispiel scheint Optimierungsbedarf zu bestehen, wenn 50 Prozent aller Teilnehmenden die Maßnahme abbrechen.
4. **Informationsbedarf identifizieren und Fragestellungen präzisieren**
Welche Informationen liegen bereits vor, welche müssen gewonnen werden, um eine ausreichende Beurteilungsbasis zu schaffen?

5. Passendes Design festlegen

Zum Beispiel Querschnitts- oder Längsschnittbefragung, Kontrollgruppe/-n.

6. Prüfen, ob das Design mit den verfügbaren Mitteln umgesetzt werden kann und gegebenenfalls Anpassungen vornehmen.

Die einzelnen Schritte werden auf den nächsten Seiten im Detail erläutert.

SCHRITT 1 – ZIELE DER EVALUATION BESTIMMEN

Bei empirischen Evaluationen geht es nicht allein um Erkenntnisgewinn, sondern vor allem auch um die Nutzung der gewonnenen Informationen. Das heißt: Bereits in der Vorbereitung der Evaluation müssen die Bewertungskriterien für den Erfolg einer Maßnahme erarbeitet werden.

Unterschiedliche Zielperspektiven von Evaluationen

Evaluationen können mit unterschiedlichen Zielperspektiven durchgeführt werden. Unterscheiden Sie zunächst nach der vordringlichsten Funktion, die die Evaluation erfüllen soll (in Anlehnung an Wesseler 1999):

Tabelle 1: Funktionsbeschreibungen von Evaluationen in Anlehnung an Wesseler

FUNKTION DER EVALUATION	ERKLÄRUNG
Legitimation	Im Vorfeld der Evaluation wurde bereits eine Entscheidung für die Beibehaltung der Maßnahme getroffen, und es wird der Versuch unternommen, diese Entscheidung zu unterstützen. Dieser Entscheidungsprozess ist in den anderen Fällen (siehe linke Spalte) noch mehr oder weniger offen.
Vergewisserung	Die Evaluation dient der Vergewisserung, dass zum Beispiel das Ziel der Interessenförderung im MINT-Bereich tatsächlich erreicht wird. Bei negativem Ausgang kann es aber auch zu einer grundsätzlichen Infragestellung der Maßnahme kommen. Gegenstand der Vergewisserung sind in der Regel die Zielsetzungen der Maßnahmen.
Verbesserung	Kriterien, die bei Evaluationen mit Verbesserungsfunktion bei Bildungsmaßnahmen relevant sein könnten: Inputmerkmale (wird zum Beispiel die gewünschte Zielgruppe erreicht?) – Prozessmerkmale (sind beispielsweise jene Prozessmerkmale, für die auf Basis des Erkenntnisstands unterstellt werden kann, dass sie den Output beeinflussen, positiv ausgeprägt? Oder sind die Prozessmerkmale optimierungsbedürftig?) – Outputmerkmale (werden beispielsweise wünschenswerte Kompetenzzuwächse oder Orientierungsveränderungen tatsächlich erreicht?)
Kontrolle und Steuerung	Bei Evaluationen mit dieser Funktion geht es in der Regel um die Zielerreichung (Output) oder in einer längerfristigen Perspektive um den Outcome (werden beispielsweise die beruflichen Pläne, die in einer Maßnahme positiv stimuliert werden konnten, auch tatsächlich umgesetzt?).

Tabelle 2: Aussagemöglichkeiten und Grenzen der Anwendung von Ergebnis-Evaluationen

ANSÄTZE	AUSSAGEMÖGLICHKEITEN UND GRENZEN	ANWENDUNGSBEISPIELE
Querschnitts- untersuchungen	<p>Daten werden zu einem bestimmten Zeitpunkt gewonnen.</p> <p>Aussagemöglichkeiten:</p> <ul style="list-style-type: none"> – Momentaufnahmen – Ausprägungen von interessierenden Merkmalen – Zusammenhangsanalysen <p>Grenzen:</p> <ul style="list-style-type: none"> – Es sind keine Entwicklungsaussagen möglich. – Zusammenhangsanalysen (zum Beispiel zwischen Kompetenzausprägungen und zeitgleich retrospektiv erhobenen Qualitätsmerkmalen der Maßnahmen) sind nicht als kausale Zusammenhänge interpretierbar. 	<p>Querschnittsuntersuchungen bieten sich an, um Informationen zu (wünschenswerten) Merkmalsausprägungen zu einem bestimmten Zeitpunkt zu gewinnen. Beispielsweise wenn überprüft werden soll, ob die anvisierte angestrebte Zielgruppe tatsächlich erreicht wird, oder – bei Kompetenzmessungen – um festzustellen, ob es einen dringenden Handlungsbedarf gibt.</p>
Pseudolängsschnitte	<p>Zu einem Messzeitpunkt werden Daten beispielsweise in unterschiedlichen Altersgruppen und Klassenstufen erhoben.</p> <p>Aussagemöglichkeiten:</p> <ul style="list-style-type: none"> – Erste Abschätzungen, inwieweit sich Merkmalsausprägungen in einer Altersspanne ändern. <p>Grenzen:</p> <ul style="list-style-type: none"> – Es kann nicht ausgeschlossen werden, dass unterschiedliche Ausprägungen auf Stichproben- oder Selektionseffekten beruhen. – Bei großen Zufallsstichproben ist dies seltener der Fall; in Evaluationsvorhaben zu Maßnahmen im MINT-Bereich sind solche Verzerrungen eher wahrscheinlich. 	<p>Pseudolängsschnitte werden im MINT-Bereich zum Beispiel bei breit angelegten Interessentstudien oder auch in Studien zu Unterschieden von Kompetenzausprägungen in unterschiedlichen Klassenstufen eingesetzt.</p> <p>Für die Evaluation von Bildungsmaßnahmen mit eher kleineren Stichproben sind sie nicht geeignet.</p>
Echte Längsschnitte	<p>In der gleichen Untersuchungsgruppe werden mindestens zu zwei Zeitpunkten Daten gewonnen.</p> <p>Aussagemöglichkeiten:</p> <ul style="list-style-type: none"> – Um Entwicklungen einzelner Merkmale beschreiben zu können, müssen diese mehrfach erhoben werden. – Um die Nachhaltigkeit von Maßnahmen beurteilen zu können, sind nach deren Abschluss in einem angemessenen Abstand Nacherhebungen erforderlich. – Um Aussagen über die Effekte einer Maßnahme treffen zu können, sind Experimental-Kontrollgruppen-Designs erforderlich. Andernfalls kann nicht abgeschätzt werden, ob sich die Veränderungen auch ohne die Maßnahme, zum Beispiel durch den Regelunterricht, einstellen. – Sofern die Bedingungen, unter denen die Maßnahmen durchgeführt werden, variieren oder systematisch modifiziert werden, sind auch Aussagen zu Wirkmechanismen möglich. <p>Grenzen:</p> <ul style="list-style-type: none"> – Längsschnitte sind aufwändig, teuer und zum Teil anspruchsvoll in der Umsetzung. 	<p>Dieser Evaluationsansatz sollte zum Einsatz kommen, wenn Entwicklungsaussagen bereitgestellt werden sollen. Das gilt zum Beispiel für die Evaluation von Laborangeboten, in welchen Interessenentwicklungen oder Selbstkonzepte stimuliert werden sollen. Bei neu entwickelten Maßnahmen ist es auch denkbar, längsschnittliche Ergebnis-Evaluationen erst dann anzusetzen, wenn der Ansatz bereits erprobt und gegebenenfalls optimiert wurde.</p> <p>Optimierungsbedarfe können sich zum Teil jedoch auch erst beim Einsatz solcher Evaluationsansätze zeigen.</p>

Designoptionen – eine Auswahlhilfe

Sobald Outputmerkmale beziehungsweise die Wirkung von Maßnahmen im Rahmen einer Evaluierung erfasst werden sollen, werden in der Regel **Ergebnis-Evaluationen** durchgeführt (Fachleute sprechen von **summativen Evaluationen**); **Prozess-Evaluationen (formative Evaluationen)** kommen in erster Linie zum Einsatz, wenn Hinweise zu Prozessoptimierungen gewonnen werden sollen.

Um kausale Zusammenhänge zwischen Maßnahmen und beobachteten Ergebnissen erkennen zu können, sind experimentelle Evaluationen notwendig. In der Praxis wird häufig versucht, diesem Anspruch durch den Vergleich einer Experimental- und einer Kontrollgruppe gerecht zu werden. Damit lässt sich weitgehend ausschließen, dass beobachtete Effekte auf andere Ursachen als die Maßnahme zurückzuführen sind.

Entscheidungen, welche Evaluationsdesigns im jeweiligen Falle angemessen sind, müssen vor dem Hintergrund der Evaluationsfunktionen, der relevanten Evaluationskriterien und der verfügbaren Mittel getroffen werden.

Wann ist die Anwendung von Ergebnis-Evaluationen sinnvoll und welche Methoden stehen mir zur Verfügung?

Ergebnis-Evaluationen werden unterschieden in Querschnittsuntersuchungen, Pseudolängsschnitte und echte Längsschnittuntersuchungen. Je nach Zielsetzung können diese Ansätze mehr oder weniger geeignet sein. In Tabelle 2 sind die mit Ergebnis-Evaluationen prinzipiell verbundenen Ausagemöglichkeiten dargestellt und mit Beispielen illustriert.

Wann ist die Anwendung von Prozess-Evaluationen sinnvoll?

Prozess-Evaluationen sind notwendig, um sich ein genaueres Bild von Prozessen machen zu können. Sie werden in der Regel eingesetzt, um Optimierungsbedarfe zu identifizieren. Vor allem bei Neuentwicklungen oder bei vorliegenden Hinweisen auf suboptimale Prozesse sind Prozess-Evaluationen häufig hilfreich, um erste Informationen über Optimierungsmöglichkeiten zu erhalten.

Eingesetzt werden dazu **verschiedene Instrumente**, zum Beispiel **Interviews und Workshops**, in denen die unterschiedlichen Perspektiven der Betroffenen berücksichtigt werden können. **Prozess-Evaluationen lassen in der Regel keine belastbaren Aussagen zur Wirksamkeit von Maßnahmen zu.** Werden subjektive Einschätzungen zur Wirksamkeit von Maßnahmen durch die Beteiligten erfasst, so sind in der Regel systematische Überschätzungen zu erwarten.

Bei kleineren Projekten, deren Zuschnitt aufwändige Ergebnis-Evaluationen kaum rechtfertigt, sind Prozess-Evaluationen häufig die einzige verbleibende Möglichkeit, um Informationen zur Bewertung der Maßnahme zu gewinnen. Umso wichtiger ist es, die Evaluation so anzulegen, dass ihre Stärken genutzt werden können. **Das bedeutet, den Schwerpunkt auf Informationen zu legen, die für weitere Optimierungsschritte genutzt werden können, und auf zweifelhafte Wirksamkeitseinschätzungen zu verzichten.**

SCHRITT 2 – MÖGLICHE WIRKUNGSZUSAMMENHÄNGE VORAB KLÄREN

In aller Regel wird die Zielerreichung einer Maßnahme von vielen unterschiedlichen Faktoren beeinflusst. Vor Beginn der Evaluation ist es wichtig, sich die entsprechenden Zusammenhänge so weit wie möglich bewusst zu machen. Die folgende Abbildung zeigt beispielhaft die strukturellen Zusammenhänge von **Einfluss- und Zielgrößen** unter Berücksichtigung außerschulischer und schulischer Rahmenbedingungen.

Außerschulische Fördermaßnahmen stellen ergänzende Lerngelegenheiten zur Schulbildung dar. Die entsprechenden Angebote sind lediglich ein Faktor, der für die Entwicklung der Lern- und Leistungsergebnisse von Kindern und Jugendlichen relevant ist. Welches Gewicht die einzelnen Faktoren haben, wurde in der Gesamtkonstellation bislang nur unzureichend untersucht. Deutlich wird anhand dieses Strukturmodells auch, dass es eine erhebliche Herausforderung darstellt, in Evaluationsstudien alle denkbaren Wirkmechanismen in Summe zu berücksichtigen.

In der Praxis beschränkt man sich vor diesem Hintergrund in aller Regel auf die Untersuchung von Ausschnitten beziehungsweise einfachen Wirkungsketten.

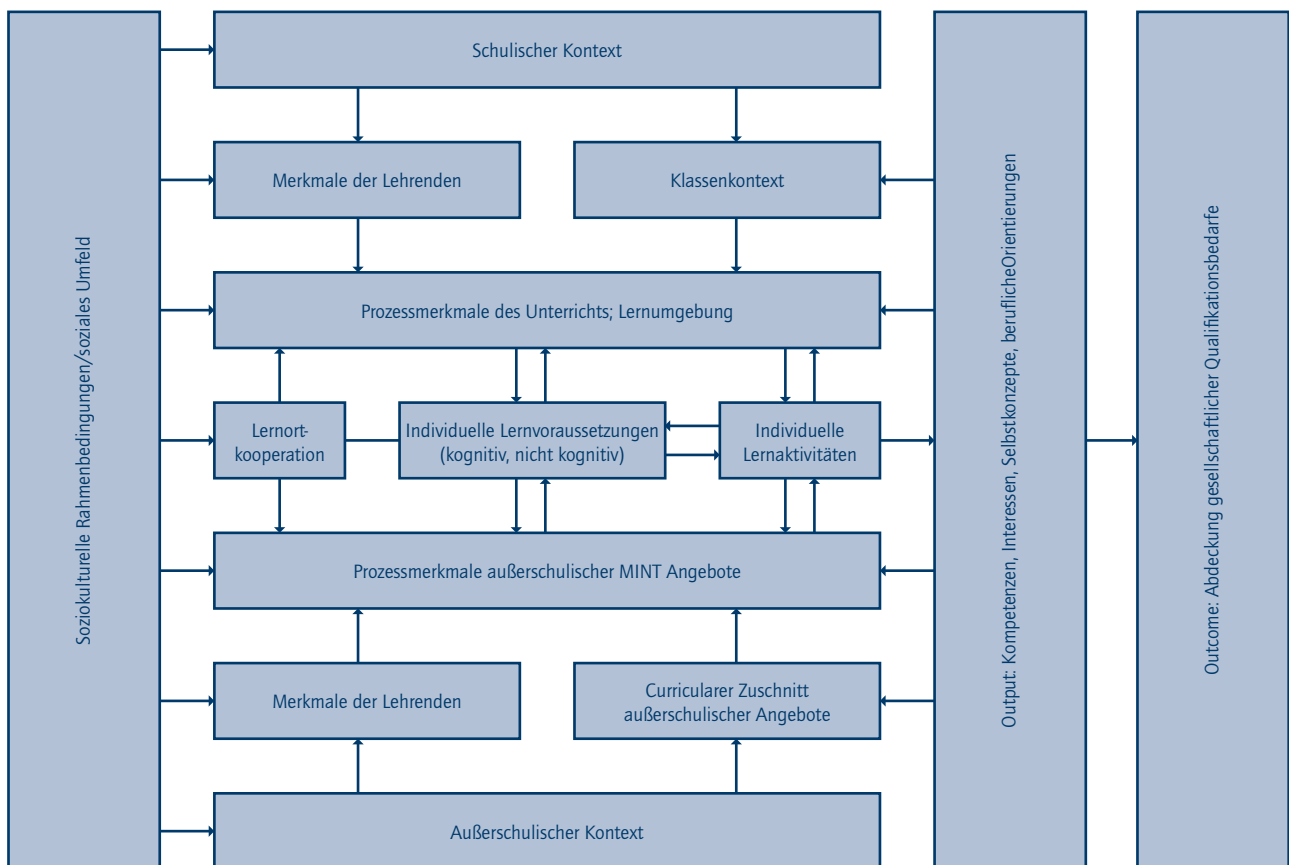
Beispiel für eine Wirkungskette:

Kompetenz der Lehrkraft → Qualität der Lehr-Lern-Prozessgestaltung → individuelle Entwicklungen

Beim obigen Modell bleiben andere Einflüsse wie etwa das elterliche Erziehungs- und Unterstützungsverhalten unberücksichtigt, obgleich sie ebenfalls Auswirkungen haben können. Die Evaluierenden sollten zumindest ein

Bewusstsein für potenziell verstärkende oder gegenläufige Effekte haben, die im Zusammenhang mit den Maßnahmen auftreten können, die sie bewerten möchten. Andernfalls ziehen sie unter Umständen falsche Schlüsse aus ihren Analysen. Wünschenswert ist eine Kontrolle möglicher Einflüsse; das heißt, dass deren Ausprägung ebenfalls erfasst und daraufhin geprüft wird, ob sich bei ihrer Berücksichtigung andere Ergebnisse zeigen als ohne ihren Einbezug.

Abbildung 1: Einflüsse außerschulischer MINT-Angebote auf Kompetenzen und Orientierungen mit Rückkopplungen



SCHRITT 3 – ERFOLGSKRITERIEN DER UNTERSUCHTEN MASSNAHMEN BESTIMMEN

Ist der Zweck der Evaluation definiert, wird das **Ziel der zu untersuchenden Maßnahme** ins Auge gefasst. Darauf aufbauend muss geklärt werden, woran erkennbar wird, ob die Maßnahme zum Erfolg führt beziehungsweise ob der Output der Maßnahme wie geplant ausfällt und die Prozesse qualitativ befriedigend sind.

Belastbare Aussagen hängen von den verfügbaren Messinstrumenten ab

Eine notwendige Bedingung für belastbare Aussagen sind in allen hier angesprochenen Varianten die **Verfügbarkeit und der Einsatz adäquater Messinstrumente**.

Bei Längsschnittuntersuchungen kommt zu den wissenschaftlichen Gütekriterien der Objektivität (Unabhängigkeit), Reliabilität (Zuverlässigkeit/Genauigkeit) und Validität (Gültigkeit einer Messung) hinzu, dass die Instrumente geeignet sein müssen, um Entwicklungen auch im Detail, sozusagen „unter der Lupe“, abbilden zu können.

Zur Planung einer Evaluation – sowohl bei eigenen Evaluationen als auch bei externen Angeboten – sollten zunächst stets folgende Fragen zu den Messinstrumenten beantwortet werden:

- Was soll mit der Maßnahme (dem Input, dem Prozess, der Maßnahme selbst) erreicht werden? Wie würde sich dies zeigen?
- Wie lassen sich diese Kriterien erfassen?
- Erfüllen die gewählten Messinstrumente die oben genannten Gütekriterien?
- Kommen Selbsteinschätzungen oder objektivere Verfahren zum Einsatz? Sind diese für die gewünschten Aussagen sinnvoll?
- Kann das Messinstrument – falls nötig – Entwicklungen abbilden?

Achtung: Selbsteinschätzungen als „Messinstrument“ zu Kompetenzausprägungen sind für Ergebnis-Evaluationen weitgehend ungeeignet. Auch Fremdeinschätzungen (zum Beispiel von Lehrkräften) erweisen sich nicht als verlässliche Datenbasis².

SCHRITT 4 – FRAGESTELLUNG PRÄZISIEREN

Wenn geklärt ist, welche Kriterien mit welchen Messinstrumenten erfasst werden sollen, muss überlegt werden, wie mögliche Ergebnisse interpretiert werden:

Wie muss das Ergebnis ausfallen, damit vom Erfolg einer Maßnahme gesprochen werden kann? Welche Ergebnisse könnten Anlass sein, um über eine Veränderung von Prozessen nachzudenken?

Zur Planung einer Evaluation sollten folgende Fragen bezüglich der Interpretation möglicher Ergebnisse beantwortet werden:

- Woran ist erkennbar, ob das Ziel erreicht wurde?
- Ab welchen Werten wird eine angestrebte Interessenänderung als hinreichend erachtet, um die Maßnahme weiterzuführen?
- Genügt es schon, wenn das Interesse an technisch-naturwissenschaftlichen Sachverhalten im Gegensatz zur Kontrollgruppe nicht absinkt und stabil bleibt?
- Lässt sich dies erfassen (messen, beobachten, erfragen)?
- Ab welchen Werten wäre das Ziel erreicht? Was wird erwartet? Welche Kriterien sprechen für Erfolg?
- Ist als Kriterium für den Erfolg bereits eine stark positive Erlebnisqualität („hat Spaß gemacht“) ausreichend oder sind zusätzlich härtere Outputkriterien heranzuziehen?

² Nickolaus/Seeber 2013.

SCHRITT 5 – DESIGN FESTLEGEN

Die Wahl des Evaluationsdesigns hängt vom Ziel der Evaluation ab. Soll die Wirkung einer Maßnahme überprüft werden, muss immer ein längsschnittliches Design gewählt werden. Eine Kontrollgruppe erlaubt Aussagen darüber, ob die Entwicklung dank der gewählten Maßnahme tatsächlich positiver ist als ohne.

In **Experimental-Kontrollgruppendesigns** ist darauf zu achten, dass

- (1) die Eingangsbedingungen der Gruppen vergleichbar sind
- (2) nicht nur in der Experimentalgruppe, sondern auch in der Kontrollgruppe Informationen zu den Lehr-Lernqualitäten erfasst werden. So können Aussagemöglichkeiten zu den Ursachen eventueller Effekte bereitgestellt werden.

Wenn Evaluationsergebnisse einer anderen/externen Initiative auf die eigene Maßnahme übertragen werden sollen, muss ebenfalls geprüft werden, ob die Eingangsbedingungen der Gruppen und die Förderkonzepte und deren Umsetzungsqualität vergleichbar sind.

In diesem Zusammenhang ist auch zu klären, ob die Evaluation intern durchgeführt werden kann oder ob eine externe Einrichtung beauftragt werden muss, um eine qualitativ hochwertige Umsetzung sicherzustellen.

Für den Fall einer externen Beauftragung können folgende Aussagen zur Beurteilung von Angeboten zur Durchführung von Evaluationen herangezogen werden:

- Sind im Angebot Zweck, Ziele, Beurteilungskriterien angemessen beschrieben?
- Sind die Ausführungen zu den notwendigerweise zu gewinnenden Informationen nachvollziehbar?
- Sind bei Wirkungsuntersuchungen längsschnittliche Experimental-Kontrollgruppendesigns vorgesehen?
- Werden ergänzende Ausführungen zu den notwendigen Stichprobengestaltungen eingebracht und nachvollziehbar begründet?
- Werden bei Prozess-Evaluationen die relevanten Kriterien berücksichtigt und deren Einbezug begründet?

Wichtig ist, dass sich die beauftragende Organisation darüber im Klaren ist, dass Evaluationen häufig in Konstellationen vorgenommen werden, in welchen besonders günstige Ausgangsbedingungen herrschen (zum Beispiel hoch engagierte und überdurchschnittlich kompetente Akteurinnen und Akteure, besonders günstige materielle Rahmenbedingungen).

Achtung: Effekte von Maßnahmen unter „Normalbedingungen“ können deshalb deutlich von denen abweichen, die unter besonders günstigen Bedingungen erzielt wurden. Dieser Sachverhalt ist auch bei den Evaluationszuschnitten und den einzubeziehenden Untersuchungsgruppen sowie der Interpretation der Ergebnisse zu berücksichtigen.

Zur Planung einer Evaluation sollten folgende Fragen bezüglich der Wahl des Evaluationsdesigns beantwortet werden:

- Ist ein Längsschnitt notwendig oder ist der Querschnitt hinreichend?
- Ist das Kontrollgruppendesign umsetzbar? Falls ja: Sind die Experimental- und die Kontrollgruppe in zentralen Merkmalen vergleichbar?
- Soll eine interne oder eine externe Evaluation durchgeführt werden?

3 FAZIT UND CHECKLISTE

Die Sicherung einer hohen Evaluationsqualität beginnt immer mit der Entscheidung für ein angemessenes Evaluationsdesign, das heißt einer Definition von Zweck und Zielen der Evaluation, der Bestimmung von Beurteilungskriterien, der Präzisierung von Fragestellungen und Informationsbedarfe sowie der Festlegung des Evaluationsdesigns.

Hinweis für Expertinnen und Experten: Um ein angemessenes Evaluationsdesign zu entwickeln, genügt es in der Regel nicht, lediglich einen der oben skizzierten Evaluationsansätze auszuwählen. Notwendig sind zusätzlich theoretische Orientierungen, die sicherstellen, dass auch andere potenzielle Einflussfaktoren berücksichtigt werden. So wäre beispielsweise theoretisch zu unterstellen, dass bei Laborangeboten, die sich über einen längeren Zeitraum erstrecken, nicht nur die Einflüsse aus den Laboren, sondern auch jene aus dem schulischen oder familiären Kontext bedeutsam für die Interessenentwicklung sind (siehe auch Abbildung 1). Die zusätzliche Erhebung solcher potenziellen Einflussfaktoren stellt eine Herausforderung dar. Diese kann auch in aufwändigeren

Evaluationsstudien meist nur bedingt bewältigt werden, da der Aufwand zur Erfassung der theoretisch relevanten Einflussfaktoren sehr groß werden kann. Das gilt teilweise auch für Studien, in welchen zum Teil ernüchternde Evaluationsergebnisse zu außerschulischen Laborangeboten zur Förderung naturwissenschaftlicher und technischer Interessen sowie einschlägigen beruflichen Orientierungen dokumentiert wurden³. Darin wird weitgehend das Ausbleiben von Effekten beziehungsweise deren mangelnde Nachhaltigkeit festgestellt, die unterstellten gegensätzlich wirkenden Einflussfaktoren werden jedoch nur begrenzt berücksichtigt. Auf dieser Basis wird zwar deutlich, dass die zunächst erzielten Effekte nicht nachhaltig sind (was mit hoher Wahrscheinlichkeit auf Erfahrungen im schulischen oder privaten Bereich zurückzuführen ist), es wird jedoch nicht hinreichend geklärt, wie die schulischen und privaten Erfahrungen mit denen in den außerschulischen Fördermaßnahmen zusammenwirken. Das wäre jedoch sehr hilfreich, um zielführende Handlungsempfehlungen aus den Evaluationsergebnissen ableiten zu können.

³ Mokhonko 2015; Mokhonko et al. 2014.

Checkliste zur Prüfung eines Evaluationsvorschlags

Tabelle 3: Checkliste zur Prüfung eines Evaluationsvorschlags

VERFAHRENSCHRITTE	ZU KLÄRENDE FRAGEN	ANTWORTEN
Ziel der Evaluation festlegen	Soll eine Maßnahme optimiert werden? Wenn ja: Input oder Prozesse? → Prozess-Evaluation (formative Evaluation)	
	Soll eine Maßnahme auf ihre Effekte (Output) überprüft werden? → Ergebnis-Evaluation (summative Evaluation)	
Erfolgskriterien für Maßnahmen festlegen	Sind Kriterien festgelegt, wie sich der Erfolg einer Maßnahme im Hinblick auf Input, Prozesse oder Output zeigen sollte?	
	Lassen sich diese Kriterien erfassen?	
	Erfüllen die gewählten Messinstrumente die Gütekriterien?	
	Kommen Selbsteinschätzungen oder objektivere Verfahren zum Einsatz? Sind diese für die gewünschten Aussagen sinnvoll?	
	Kann das Messinstrument, falls nötig, zeitliche Entwicklungen abbilden?	
Fragestellung präzisieren	Woran ist erkennbar, ob das Ziel erreicht wurde?	
	Ab welchen Werten wäre das Ziel erreicht? Was wird erwartet? Welche Kriterien sprechen für Erfolg?	
Design festlegen	Ist ein Längsschnitt notwendig oder reicht ein Querschnitt?	
	Ist ein Kontrollgruppendesign umsetzbar? Falls ja: Sind die Experimental- und die Kontrollgruppe in zentralen Merkmalen vergleichbar?	
	Erfolgt die Evaluation intern oder extern?	
Theoretische Basis dokumentieren	Ist beschrieben, welche theoretischen Überlegungen leitend sind?	
	Ist beschrieben, welche Zusammenhänge unterstellt werden und welche der unterstellten Einflüsse im Evaluationsansatz (nicht) berücksichtigt werden?	
	Können trotz einer partiellen Berücksichtigung potenzieller Einflüsse belastbare Aussagen erwartet werden?	

4 BEISPIELE FÜR EVALUATIONEN

Beispiele für Ergebnis-Evaluationen

Ein Überblick über vorliegende Evaluationen, in den rund tausend Modellprojekte Eingang gefunden haben, wurde von einer acatech Projektgruppe um Prof. Dr. Ortwin Renn erstellt und unter dem Titel **Monitoring von Motivationskonzepten für den Techniknachwuchs** als acatech Publikation „acatech BERICHTET und EMPFIEHLT“ (Nr. 5, 2011) der Öffentlichkeit zur Verfügung gestellt. Vorgestellt werden im Folgenden einige ausgewählte Studien mit relativ anspruchsvollen Untersuchungszuschnitten, die als Anschauungsbeispiele für Evaluationen dienen können. Die Arbeiten sind überwiegend über folgenden Link abrufbar: <http://www.lernortlabor.de/publications.php?tl=15>

- a) Weßnik, S.: *Kooperatives Arbeiten an industrienahen außerschulischen Lernorten* (Dissertation), Universität Kiel 2013.

Untersucht wurden unter anderem die Effekte eines industrienahen Schülerlabors auf das Image sowie das Fähigkeitsselbstkonzept zu Chemie, Physik und naturwissenschaftlichen Berufsorientierungen. Durch Teams mit unterschiedlichen Aufgabenzuschnitten war es zudem möglich zu prüfen, ob solche unterschiedlichen Aufgabenzuschnitte ergebnisrelevant werden. Berichtet werden relativ deutliche und zum Teil auch nachhaltige Effekte, was relativ selten ist.

- b) Guderian, P.: *Wirksamkeit außerschulischer Lernorte. Der Einfluss mehrmaliger Besuche eines Schülerlabors auf die Entwicklung des Interesses an Physik* (Dissertation), Humboldt-Universität Berlin 2006.

Durch das gewählte Untersuchungsdesign können nicht nur Aussagen zur Interessenentwicklung getroffen, sondern auch Hinweise zu den Effekten einer engen Abstimmung der außerschulischen Aktivitäten mit dem schulischen Unterricht eingebracht werden. Insgesamt dokumentiert die Studie eher schwache und nicht stabile Effekte. Die Anbindung der Laborangebote an den

Unterricht scheint sich günstig auszuwirken, der Effekt wird jedoch nicht signifikant, was durch die relativ kleine Untersuchungsgruppe begünstigt sein dürfte.

- c) Mokhonko, S.: *Nachwuchsförderung im MINT-Bereich. Aktuelle Entwicklungen, Fördermaßnahmen und ihre Effekte*, Stuttgart: Steiner Verlag 2015.

Unter Einbezug einer großen Stichprobe wurden in dieser Studie Entwicklungen von Interessen, Selbstkonzepten und Berufsorientierungen in einem Experimental- Kontrollgruppendesign mit drei Messzeitpunkten untersucht. Einbezogen waren mehrere Standorte und thematische Zuschnitte. Außerdem wurden die Maßnahmendauer und die Häufigkeit des Laborbesuchs variiert. Trotz sehr guter Beurteilung der Seminare zeigen sich überwiegend keine nachhaltigen Effekte.

Beispiele für Prozess-Evaluationen

- a) Weßnik, S.: *Kooperatives Arbeiten an industrienahen außerschulischen Lernorten* (Dissertation), Universität Kiel 2013.

In dieser Studie wurden auch qualitative Analysen der Teamprozesse vorgenommen, sodass auch der Frage nachgegangen werden konnte, inwieweit unter den jeweiligen Bedingungen stärkere oder schwächere Effekte beobachtet werden können.

- b) Engeln, K.: *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*, Berlin: Logos Verlag 2004.

Die Untersuchung liefert neben Aussagen zur Interessenentwicklung auch Ergebnisse zu den Wahrnehmungen der Laborangebote, wobei unterschiedliche Gestaltungsformen der Labore auch für die Wahrnehmungen und Verarbeitungen bedeutsam werden.

Weitere Checklisten und Handreichungen

Die beiden folgenden Handreichungen liefern einen allgemeinen Überblick zum Thema Qualitätssicherung und bieten einen guten Einstieg in die Thematik. Für MINT-Initiativen, die sich detaillierter mit der Evaluierung und der konkreten Wirksamkeitsmessung ihrer Projekte auseinandersetzen möchten, ist die vorliegende Handreichung gedacht.

- Arbeitsgruppe Qualitätssicherung/Evaluation des Nationalen MINT Forums (Hrsg.): *Leitfaden 1.0 für die Qualitätssicherung von MINT-Initiativen*

Dieser Leitfaden fasst die zentralen Punkte zusammen, die bei der Qualitätssicherung für MINT-Bildungsangebote berücksichtigt werden müssen. Er bietet MINT-Initiativen und Organisationen Anregungen, Unterstützung und Vorschläge zur Dokumentation für die längerfristige

Qualitätsentwicklung und -sicherung von Angeboten im MINT-Bereich.

http://www.nationalesmintforum.de/fileadmin/user_upload/gerke/NMF/Leitfaden_1.0_Webversion.pdf

- MINT Zukunft schaffen e. V. (Hrsg.): *Leitfaden zur Selbstüberprüfung und Sicherung der Qualität von MINT-Initiativen*

Dieser Leitfaden ist als Orientierungshilfe gedacht und richtet sich an alle potenziellen und aktiven MINT-Initiatoren. Er soll Hilfestellung und Unterstützung bei Aufbau und Umsetzung eines neuen oder bei der Weiterentwicklung eines existierenden MINT-Förderangebotes bieten, um diese möglichst effektiv und beständig zu gestalten.

<https://www.mintzukunftschaefen.de/leitfaden.html>

LITERATUR

Bortz/Döring 2006

Bortz, J./Döring, N.: *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4. Auflage, Heidelberg: Springer 2006.

Engeln 2004

Engeln, K.: *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*, Berlin: Logos Verlag 2004.

Guderian 2006

Guderian, P.: *Wirksamkeit außerschulischer Lernorte. Der Einfluss mehrmaliger Besuche eines Schülerlabors auf die Entwicklung des Interesses an Physik* (Dissertation), Humboldt-Universität Berlin 2006.

Kromrey 2005

Kromrey, H.: „Evaluation – ein Überblick“. In: Schöch, H. (Hrsg.): *Was ist Qualität. Die Entzauberung eines Mythos*, Berlin: Wissenschaftlicher Verlag, 2005, S. 31–85.

Mokhonko 2015

Mokhonko, S.: *Nachwuchsförderung im MINT-Bereich. Aktuelle Entwicklungen, Fördermaßnahmen und ihre Effekte*, Stuttgart: Steiner Verlag 2015.

Mokhonko et al. 2014

Mokhonko, S./Nickolaus, R./Windaus, A.: „Förderung von Mädchen in Naturwissenschaften: Schülerlabore und ihre Effekte“. In: *Zeitschrift für Didaktik der Naturwissenschaften*, (1), 2014, S. 143–159.

Nickolaus/Seeber 2013

Nickolaus, R./Seeber, S.: „Berufliche Kompetenzen: Modellierungen und diagnostische Verfahren“. In: Frey, A./Lissmann, U./Schwarz, B. (Hrsg.): *Handbuch Berufspädagogische Diagnostik*, Weinheim: Beltz 2013, S. 166–195.

Sanders 2006

Sanders, J. R.: *Handbuch der Evaluationsstandards: Die Standards des „Joint Committee on Standards for Educational Evaluation“*, Wiesbaden: VS Verlag 2006.

Wesseler 1999

Wesseler, M.: „Evaluation und Evaluationsforschung“. In: Tippelt, R. (Hrsg.): *Handbuch der Erwachsenenbildung/Weiterbildung*, Opladen 1999, S. 736–752.

Weßnik 2013

Weßnik, S.: *Kooperatives Arbeiten an industrienahen außerschulischen Lernorten* (Dissertation), Universität Kiel 2013.

Wottawa/Thierau 2003

Wottawa, H./Thierau, H.: *Lehrbuch Evaluation*, Bern: Verlag Hans Huber 2003.