



KONZEPT FÜR EVALUATIONEN VON PROGRAMMEN UND ANGEBOTEN

IMPACT

PRAXIS

OUTCOME

OUTPUT

WISSENSCHAFT

Herausgeber: Marie Meierhofer Institut für das Kind
Autorin: Dr. phil. Patricia Lannen
Umschlag und Illustrationen: Anna-Lea Guarisco
Layout/Gestaltung: Claudius Natsch
Druck: Borer Druck AG, Laufen
Vertrieb: Marie Meierhofer Institut für das Kind
Pfungstweidstrasse 16, 8005 Zürich
Tel.: 044 205 52 20 / Fax.: 044 205 52 22
info@mmi.ch / www.mmi.ch

© mmi, 2018



Konzept für Evaluationen von Programmen und Angeboten

Marie Meierhofer Institut für das Kind

Patricia Lannen, August 2018





Inhaltsverzeichnis

1	Einleitung	5
2	Ziel und Zweck des Konzepts	5
3	Bausteine des Konzepts	6
3.1	Partizipativ - mit und für die Praxis	6
3.2	Das Kind im Zentrum	10
3.3	Multi- und Transdisziplinarität	10
3.4	Multimethodaler Ansatz	11
4	Wirkungsmodelle	12
4.1	Ein erweitertes Modell	12
4.2	Erarbeitung des Modelles	15
4.3	Stolpersteine	17
5	Wahl des Evaluationsdesigns	20
6	Wissensmanagement	26
7	Anhang	28



1. Einleitung

Der Evaluationsansatz des MMI setzt eine systematische Vorgehensweise um und widerspiegelt Kernelemente des MMI Leitbildes (Anhang) sowie des MMI Forschungskonzeptes (Anhang). Es setzt diese in Verbindung mit einer systematischen Analyse und Dokumentation des Feldes und stellt eine Synthese mit Erfahrungen und Expertise aus vergangenen Evaluationen dar.

Das Konzept ist aus mehreren Bausteinen zusammengesetzt und illustriert Werkzeuge, welche im Rahmen der MMI- Evaluationen eingesetzt werden.

Im letzten Kapitel wird illustriert, wie die Nutzung der Erkenntnisse aus den Evaluationen im Sinne von Wissensmanagement optimiert wird.

Der Ansatz stellt hohe Anforderungen an die Forschenden sowohl in Bezug auf deren Fachwissen und die Verfügbarkeit eines breiten Werkzeugkoffers an sozialwissenschaftlichen Methoden als auch auf ihre Erfahrung mit Programmentwicklung und -umsetzung.

2. Ziel und Zweck des Konzepts

Das Evaluationskonzept verfolgt Ziele auf verschiedenen Ebenen:

MMI intern

Das Evaluationskonzept macht das kollektive Wissen und gemeinsame Verständnis zur Umsetzung von Evaluationen explizit, vereinheitlicht so den Wissensstand innerhalb der Organisation und macht diese Basis für alle (neuen) Fachpersonen verfügbar. Es ist ein Qualitätssicherungsinstrument und kann im Sinne von Wissenstransfer und -management kontinuierlich weiterentwickelt werden.

Praxis

Das Konzept schärft die Wahrnehmung für eine zu besetzende Nische im Feld und stärkt die Stellung des MMIs als Kompetenzzentrum für Evaluationen im Transmissionsriemen Forschung, Praxis, Gesell-

schaft. Es bildet eine Grundlage dafür, sich proaktiv in den Diskurs zu Wirkungsorientierung und Wirkungsmessung von Angeboten einzubringen. Das Evaluationskonzept ist ein Werkzeug für Capacity Strengthening hin zu evidenz-basierter Praxis und das Fundament für entsprechende Aufträge, Projekte, Workshops, Schulungen und Publikationen.

Wissenschaft

Das Evaluationskonzept bildet den Ausgangspunkt, sich aktiv an der Stärkung der Evidenzbasis und des Wissens um wirksame Interventionselemente zu beteiligen. Zudem soll es den Austausch über sozialwissenschaftliche Disziplinen hinweg anregen und die Relevanz praxisbezogener Forschung positionieren.

3. Bausteine des Konzepts

Die nachfolgenden „Bausteine“ widerspiegeln die Kernelemente des MMI Leitbildes wie auch des MMI Forschungskonzeptes und beschreiben deren Anwendung auf die Evaluationstätigkeit des MMI. Sie bilden den Kern des Konzeptes.

3.1 Partizipativ - mit und für die Praxis

In diesem Baustein spiegelt sich die partizipative Grundhaltung des MMI. Er veranschaulicht zudem die Arbeitsweise des MMI im Transmissionsriemen zwischen Forschung und Praxis.

Ein partizipatives Vorgehen und eine starke, funktionierende Partnerschaft zwischen dem Praxisanbieter („implementer“) und den Forschenden am MMI, welche von Vertrauen getragen ist, stehen im Zentrum des Evaluationsansatzes. Dafür werden die Ziele der Evalua-

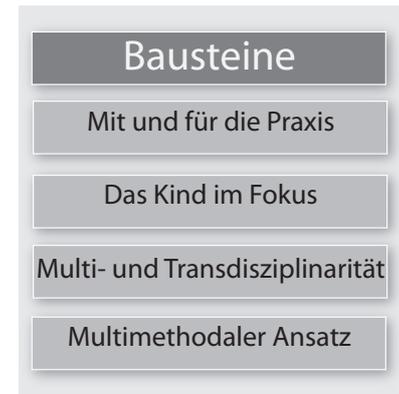


Abbildung 1:
Bausteine des Evaluationskonzeptes

tion, die Kernfragestellungen und das Evaluationsdesign bereits gemeinsam in einem Workshop im Rahmen der Erstellung einer Offerte erarbeitet. Damit wird eine optimale Passung zwischen dem Anliegen der Praxis (Umsetzer und Auftraggeber) und der Evaluation angestrebt, um die Durchführung der Evaluation zu sichern und deren Aussagekraft zu maximieren, Missverständnissen vorzubeugen und die Möglichkeiten und Grenzen verschiedener Evaluationsdesigns transparent zu machen (siehe auch Kapitel 4 zu Wirkungsmodellen und Kapitel 5 zu Wahl der Evaluationsdesigns). Angesprochen wird auch, wie die Unabhängigkeit der Forschung gewährleistet wird, und, was allfällige unerwartete oder gar unerwünschte Ergebnisse für den Praxispartner bedeuten und wie er damit umzugehen gedenkt.

In diesem ersten Workshop werden zudem mögliche Bedenken zur Evaluation besprochen (siehe Box 1). Die MMI-Wissenschaftlerin übernimmt dabei die Rolle eines „Research Brokers“. Sie/er ist jemand, der zwischen der wissenschaftlichen und der praktischen Sprache und Herangehensweise übersetzen kann und damit den Transmissionsriemen zwischen Forschung und Praxis in Schwung bringt.

Weiter wird in diesem ersten Workshop das Wirkungsverständnis des Praxispartners bezüglich dem fraglichen Angebot erfragt, damit die passenden Werkzeuge gewählt und später eingesetzt werden können (siehe Kapitel 4 zu Wirkungsmodellen).

Die Resultate der Evaluation werden gemeinsam mit dem Praxispartner vor Fertigstellung des Schlussberichtes kontextualisiert und wo, angezeigt, werden entsprechende Empfehlungen diskutiert. Bei grösseren Evaluationsvorhaben (über mehrere Jahre) werden Zwischenworkshops zur Vorstellung und Diskussion vorläufiger Resultate eingeschoben.

Wissenschaftliche Unabhängigkeit und Integrität sind in jeder noch so partnerschaftlich aufgegleisten Evaluation von grösster Wichtigkeit. Sie werden geschützt und, falls nötig, erklärt. Es werden entsprechende Massnahmen getroffen. Interaktionen und informeller Austausch während der Forschungsphase werden unterlassen und strikt unterbunden. Ein offener Ausgang mit verschiedenen möglichen und ggf. unerwarteten Resultaten muss zu jedem Zeitpunkt gewährleistet sein.

Häufige Einwände und Bedenken aus der Praxis und Möglichkeiten, darauf einzugehen:

„Ich weiss, dass mein Angebot wirkt“

Dieser Eindruck mag entstehen, a) da die Teilnehmenden positiv auf das Angebot reagieren, b) aufgrund spezifischer Beispiele von Teilnehmenden, bei denen das Angebot geholfen hat, c) weil die Intuition und professionelle Erfahrung besagen, dass das Angebot gut ist. Forschungsarbeiten konnten jedoch zeigen, dass die intuitive Beurteilung bezüglich Wirksamkeit oft trügt. Es mag sein, dass vor allem Rückmeldungen aus den Angeboten verzerrt sind und vorwiegend von Personen kommen, welche das Angebot mögen. Es ist auch möglich, dass es eine „Vermischung“ gibt und es Ausdruck davon ist, dass die Personen es schätzen, am Angebot teilzunehmen, und nicht unbedingt, dass es wirkt. Ein Beispiel hierfür ist DARE, ein Drogenpräventionsprogramm in den USA, welches über viele Jahre hinweg umgesetzt, aber erst nach vielen Jahren evaluiert wurde. Es zeigte sich, dass das Programm nicht wirksam ist, und Abschreckung nicht als Wirkfaktor für eine erfolgreiche Prävention ausreicht. Es zeigte sich, dass vielmehr Peer-Einflüsse ausschlaggebend sind und in einer Intervention berücksichtigt werden müssen.

Eine Evaluation kann die intuitive Einschätzung aber auch bestätigen und untermauern, was ein starkes Argument für die Weiterführung und ggf. die entsprechende Finanzierung ist.

„Evaluation ist teuer und nimmt viel Zeit in Anspruch. Sollten die Ressourcen nicht lieber den Kindern direkt zugute kommen?“

Dieses Argument ist sehr gut nachvollziehbar: es ist schwer, Geld zu finden für Evaluationen, und es ist noch schwerer, wenn Gelder zur Umsetzung des Angebotes für diese Kinder und Familien selber auch limitiert sind. Gleichzeitig ist es problematisch und unter Umständen unethisch, eine Dienstleistung anzubieten, von der man nicht weiss, ob sie den gewünschten Effekt überhaupt hat oder im Gegenteil vielleicht sogar Schaden anrichtet.

Wieder ist das Beispiel DARE illustrativ: es wurden über eine Milliarde USD in das Programm investiert, bevor zu Tage kam, dass das Programm keine Wirkung zeigt, d.h. die Adressat/innen ohne passende Unterstützung blieben.

„Was wir tun, ist nicht linear und nicht messbar“

Es ist in der Tat so, dass das Messen von menschlichen und zwischenmenschlichen Phänomenen durch die Sozialwissenschaft oft nicht so offensichtlich und geradlinig ist wie teilweise in anderen Disziplinen. So kann man zum Beispiel in der Medizin häufig mit relativ einfachen Tests das Vorhandensein einer Krankheit definitiv im Blut nachweisen (ja, die Krankheit ist vorhanden oder nein, sie ist nicht vorhanden). Oft ist dies in der Medizin und erst recht in den Sozialwissenschaften nicht ganz so eindeutig und schwarz- Weiss. Ein Beispiel ist die Erfassung von Gewalt: Welche Handlungen gelten als Gewalt und welche nicht? In welche Dynamik sind sie eingebettet? Verzerrt die soziale Erwünschtheit eventuell Antworten?.

Die Sozialwissenschaften haben in den letzten Jahren enorme Fortschritte hin zu valider und reliabler Messung von Phänomenen gemacht. Ein Evaluationsansatz, der qualitative und quantitative Methoden verbindet, ermöglicht die gleichzeitige Erfassung des "big picture" wie auch der Reichhaltigkeit der Erfahrungen. Sie ermöglicht einen Überblick und spezifische Einblicke in die Tiefe.

„Was mache ich, wenn die Evaluation zeigt, dass mein Angebot nicht funktioniert?“

Evaluationsergebnisse sind in den seltensten Fällen einseitig positiv oder negativ. Oft zeigt sich, dass eine Intervention für gewisse angestrebte Outcomes wirksam ist und für andere weniger. Oft liefert die Evaluation wertvolle Hinweise zur Stärkung und Verbesserung der Intervention. Und wenn die Intervention nicht wirkt oder teilweise am Ziel vorbei geht, würde man das nicht wissen wollen?

„Es ist unethisch, wenn nicht alle vom Programm profitieren können“

Diese Bedenken werden manchmal für Evaluationen mit Kontrollgruppe (quasi-experimentelle Studien oder randomisierte Kontrollstudien) geäußert und sind in gewissen Fällen gerechtfertigt.

Es ist allerdings zu bedenken, dass oft so oder so nicht genügend Ressourcen vorhanden sind, um alle potentiellen Teilnehmer im Programm einzuschließen. Die Erhebung von Informationen bei Personen, die nicht am Programm teilnehmen können, ist dann eine Möglichkeit, die letztlich auch der Programmausweitung zugute kommen kann.

Wenn eine Staffelung der Aufnahme in ein Programm nötig und möglich ist, stellt die Einführung von Warte-Kontrollgruppen eine sehr sinnvolle Option dar. Zudem ist zu bedenken, dass das Anbieten einer Dienstleistung, von der wir nicht wissen, ob sie wirkt oder gar negative Effekte hat, ethisch mindestens ebenso bedenklich ist wie der beschränkte Zugang zum Programm.

„Ich habe die Expertise nicht, eine Evaluation durchzuführen“

Dies ist häufig wahr. Die Ausbildung und die Fertigkeiten eines guten Praktikers/einer guten Praktikerin sind nicht dieselben wie die eines guten Evaluators/einer guten Evaluatorsin. Und je komplexer und im wissenschaftlichen Sinne strenger die Evaluation ("rigour") ist, desto spezialisierter ist das notwendige Wissen. Eine starke Partnerschaft zwischen ForscherIn und PraktikerIn ist deshalb das A und O, also der entscheidende Gelingensfaktor einer erfolgreichen Evaluation.

Quelle: Lannen, P., Jones, L., (in prep). Beyond Evaluation: Using Science to Accompany Program Development and Implementation. Chapter 2: Concerns about engaging in evaluation.

3.2 Das Kind im Zentrum

Das Kind und seine Lebenswelten stehen in der Arbeit des Marie Meierhofer Institutes im Zentrum jeder Tätigkeit. Thematisch werden Evaluationen priorisiert, welche auf die frühe Kindheit oder kinderrechtliche Fragen (vor allem auch im Sinne von Unterstützung von Kindern aus vulnerablen Umständen) fokussieren.

Die Grundhaltung des MMI, „vom Kind her“ und seiner Erlebniswelt zu denken, ist für die Evaluationen des MMI zentral (vgl. Modell des MMI zur Partizipation junger Kinder: Marie Meierhofer Institut für das Kind (2016), Modell kindlicher Partizipation aus Kinder- und aus Erwachsenenperspektive, Poster, Beilage und Kinder 98). Um die Perspektive der Kinder direkt abzuholen, werden Gefässe wie Kinderworkshops und passende Methoden der Kommunikation mit Kindern eingesetzt. Wo immer möglich und zielführend, werden Outcomes auf Ebene Kinder als zentrale Indikatoren eingeschlossen.

3.3 Multi-und Transdisziplinarität

Das MMI beherbergt unter einem Dach Forschungs- und Praxisexpertise (Transdisziplinarität) und zeichnet sich durch eine multidisziplinäre Arbeitsweise mit Fachpersonen aus der Entwicklungspsychologie, Frühpädagogik und Familiensoziologie aus. Zudem arbeitet das Team regelmässig mit Kolleg/innen der Entwicklungspädiatrie und des Familien- und Kinderrechts zusammen. Dies ermöglicht einen ganzheitlichen Blick auf das Kind sowie das niederschwellige Hinzuziehen eines breiten Spektrums an Expertise, einschliesslich erweitertem Grundlagen- und Praxiswissen in den verschiedenen Disziplinen, Expertise in therapeutischen und pädagogischen Interventionen sowie Organisationsentwicklung. Dies wird für jede Evaluation nach Bedarf entweder informell oder im Rahmen von bereichsübergreifenden Werkstätten hinzugezogen.

3.4 Multimethodaler Ansatz

Der Evaluationsansatz des MMI zeichnet sich durch ein systematisches Vorgehen und durch den Einsatz von Methoden gemäss höchsten sozialwissenschaftlichen Standards aus. Die Mitglieder des Forschungsteams sind in einer der relevanten Methodenrichtungen vertieft geschult (siehe Abschnitt 3.3) und verfügen über eine breit angelegte Tool-box an sozialwissenschaftlichen Methoden. Dies ermöglicht es, je nach Fragestellung das passende Design einzusetzen (siehe Kapitel 5 Wahl des Evaluationsdesigns).

Das Team arbeitet explizit mit qualitativen wie auch quantitativen Forschungsmethoden, wenn immer möglich, ergänzend in jeder Evaluation. Daten werden mit standardisierten und wo, immer möglich, validierten Erhebungsinstrumenten erhoben. Bei massgeschneiderten und innovativen Instrumenten wird hoher Wert auf systematische, reflektierte Anwendung gelegt. Es wird, wie üblich, unterschieden zwischen hypothesen-testendem und explorativem Vorgehen und es werden mehrere Perspektiven zu einer Fragestellung eingeholt (multinformant). Die Wahl eines Evaluationsdesigns soll explizit nicht nach den (qualitativen vs. quantitativen) Vorlieben des Wissenschaftlers/der Wissenschaftlerin gesteuert, sondern von der Programmreife und den Fragestellungen abhängig gemacht werden (siehe Kapitel 5 Wahl des Evaluationsdesigns).



4. Wirkungsmodelle

4.1 Ein erweitertes Modell

Ein zentrales Werkzeug im Rahmen der MMI Evaluationen ist die gemeinsame Erarbeitung eines Wirkungsmodells ("program theory", "theory of change", "logic model"). Ein Wirkungsmodell verdeutlicht, was die Programmverantwortlichen im Rahmen des Angebotes als notwendig und sinnvoll erachten, und erläutert in logischen Schritten, wie das Angebot das soziale Problem angeht, wie es zu Veränderung führt und was es genau erreichen will. Es benützt bestehende Evidenz, um die einzelnen Schritte zu untermauern, und benennt die zu erreichenden Ziele als klar definierte Outcomes. Es ermöglicht die Identifikation von «blinden Flecken» und Verbesserungspotenzialen und stärkt so bereits das Angebot in Richtung Wirkungsorientierung. Zudem ist ein Wirkungsmodell ein äusserst nützliches Werkzeug, um prägnant zu kommunizieren, worum es im Angebot geht. Es ist die Grundlage jeder MMI-Evaluation.

In vielen Angeboten ist das Wirkungsverständnis entweder intuitiv, nicht explizit zugänglich, hat logische Lücken oder es ist nicht mit der entsprechenden Evidenz untermauert.

Des Weiteren sind Outputs und Outcomes oft vage formuliert, sodass eine Wahl der zu messenden Ziele erschwert ist oder zu Missverständnissen führen kann.

Viele der klassischen Ansätze der Erläuterung von Wirkungsmodellen beginnen direkt bei den Elementen des Angebotes und machen explizit, wie diese zu spezifischen Zielen (outputs und outcomes führen).

Hier zwei Beispiele:

Beispiel 1:

Quelle: Stiftung ZEWO

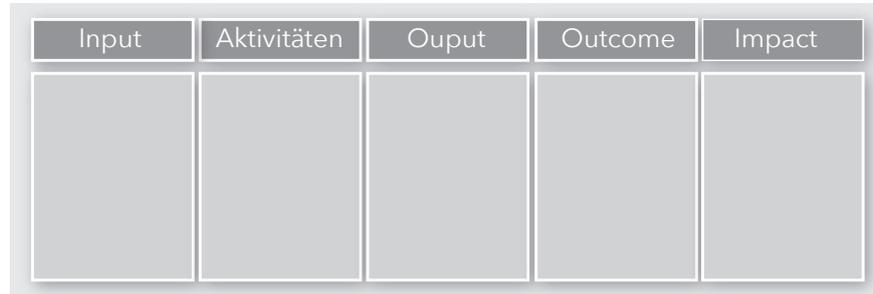


Abbildung 2: Beispiel Wirkungsmodell 1

Beispiel 2:

Quelle: Jones, L. (2014). Improving efforts to prevent children's exposure to violence- a handbook to support child maltreatment prevention programs. World Health Organization.



Abbildung 3: Beispiel Wirkungsmodell 2

Diese berücksichtigen aber meist nicht, welches „soziale Problem“ das Angebot zu lösen versucht („weshalb es das Angebot braucht“) und welche Evidenz zum Problem vorhanden ist (Prävalenzen, Ursachen, Risikofaktoren etc). Dabei sind eine logische Verkettung sowie eine Verlinkung zwischen der Problemstellung und den Angebotselementen sowie wiederum eine Einbettung der Verlinkung mit bestehender Evidenz essentiell.

Das MMI-Konzept erweitert daher das Modell und fügt diese Elemente hinzu: Was soll konkret mit dem Angebot angegangen werden und wie steht es mit der Passung der Programmelemente in Bezug auf das Adressieren des sozialen Problems? Welche Evidenz haben wir zu den Gründen, die zum sozialen Problem führen? Sind sie ausreichend im Angebot berücksichtigt? Was sind die Hindernisse, dass das Problem nicht schon lange gelöst ist (z.B. Zugang zu Zielgruppe), und wie wird dies im Angebot überwunden?

So ergibt sich das folgende Modell: Abbildung 4

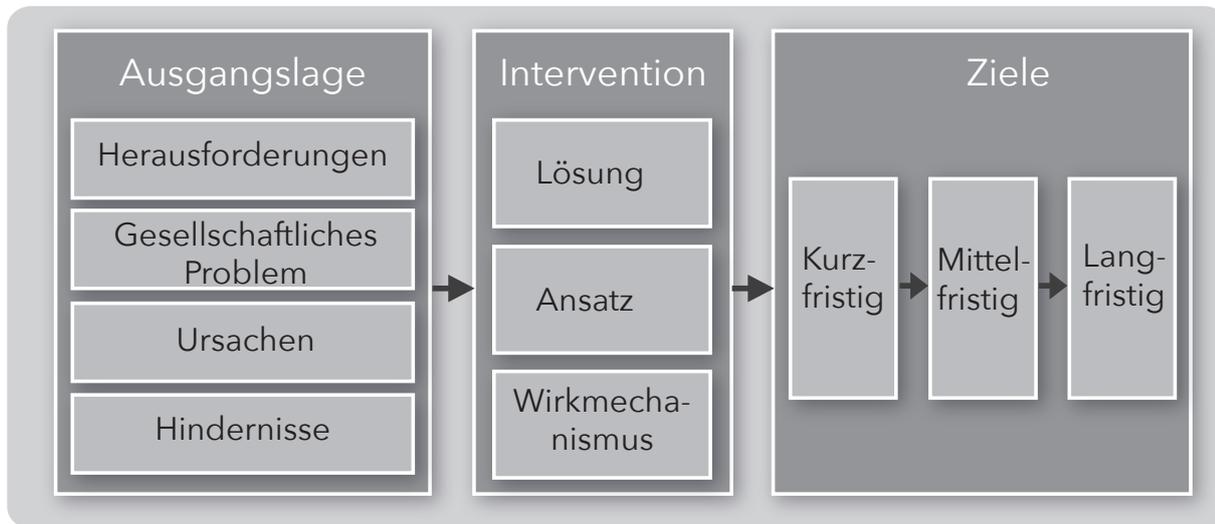


Abbildung 4: MMI- Wirkungsmodell

Erläuterungen:

Ausgangslage: Sie beschreibt das gesellschaftliche Problem, welches das Angebot anzugehen versucht, sowie dessen Ursachen. Sie dokumentiert Hindernisse (warum wurde das Problem nicht schon lange gelöst?) oder spezielle Herausforderungen bezüglich Umsetzbarkeit eines Angebotes (zum Beispiel Erreichbarkeit der Zielgruppe).

Intervention: Die Beschreibung der Intervention soll eine Lösung des identifizierten Problems sein. Der Ansatz soll die Hindernisse und Herausforderungen lösen (z.B. Niederschwelligkeit, um Erreichbarkeit zu verbessern) und den Wirkmechanismus, wie das Problem gelöst wird, explizit machen.

Ziele: Ziele sollen als kurz-, mittel- und langfristige Ziele formuliert und direkt als Lösung für das initial definierte „Problem“ erkennbar sein.

4.2 Erarbeitung des Modelles

Im Rahmen des ersten Workshops wird das Wirkungsverständnis des Praxispartners erfragt (siehe 3.1.). Je nach dessen Differenziertheit und Reifegrad, wird im Rahmen der Evaluation mehr oder weniger Zeit auf dieses Element verwendet. Bei nur rudimentär ausgereiften Wirkungsmodellen wird in einem bis zwei weiteren Workshops gemeinsam und im Dialog ein Wirkungsmodell erarbeitet.

Dazu werden unter anderem die folgenden Leitfragen eingesetzt, um die einzelnen Elemente des Wirkungsmodelles einzufüllen und in logischen Schritten schematisch abzufassen:

- Welche Inhalte werden angeboten?
- Weshalb braucht es das Angebot /was ist das Problem?
- Warum wurde dieses Problem nicht schon lange gelöst (Hindernisse)?
- Wie geht das Angebot das Problem an?
- Wer ist die Zielgruppe des Angebotes und warum?
- Welche Ziele gibt es kurz-, mittel- und langfristig?
- Welche Wirkungen hat das Angebot bei der Zielgruppe?

- Wie werden mit dem Angebot die Ziele erreicht?

Dabei ist die Notwendigkeit des Individualisierungsgrades der Unterstützung meist hoch. Oft sind Praktiker sehr in die operative Arbeit vertieft und es fällt ihnen schwer, ohne Strukturierungshilfe von aussen das Angebot schematisch abzufassen. Ein selbstständiges Erarbeiten (zum Beispiel mit Leitfragen ohne Dialog) hat sich nicht als erfolgreich herausgestellt.

Wichtig anzumerken ist, dass die im Wirkungsmodell formulierten Ziele nicht unbedingt vollständig deckungsgleich sein müssen mit dem, was dann in der Evaluation erhoben und erfasst wird. Je nach Rahmenbedingungen und Ressourcen der Evaluation, ist es meist nicht möglich, alle Ziele zu messen. Das Wirkungsmodell hilft aber, explizit zu machen, was innerhalb der erwarteten Wirkung und der angestrebten Angebotsziele im Rahmen einer konkreten Evaluation tatsächlich gemessen wird und was nicht. In der Evaluation nicht gemessene Ziele sollen, wenn immer möglich, durch bestehende Forschungsergebnisse aus anderen Studien belegt werden.

Neben der Wichtigkeit des Wirkungsmodells für das Angebot und als Basis für die Evaluation selber, dient die gemeinsame Erarbeitung des Modells als Instrument, um eine Arbeitsbeziehung aufzubauen, die Prioritäten der anderen Partei zu verstehen und Vertrauen zu schaffen, dass wirklich das gemessen wird, was im Zentrum des Angebots und der Fragen der Praktiker steht. Es ist eine Gelegenheit für den Forscher, das Wirkungsverständnis des Praxispartners zu erfassen, zu schärfen und explizit zu machen. Nur so kann die Logik der Impact Kette vom Problem zur Lösung überprüft werden und man kann schauen, ob die Logik der Kette dem, was wir heute aus der Forschung wissen, standhält.

Die Erarbeitung von Wirkungsmodellen setzt die Fähigkeit der Forscherin voraus, hochstrukturiert zu denken, und bedarf einer hohen Abstraktionsfähigkeit, um ein Angebot schematisch darstellen zu können. Zudem braucht es substantielles Fachwissen, um die inhaltlich logischen Schritte nachzuvollziehen und sie mit relevanter Forschungsliteratur zu untermauern. Die Forscherin muss den Praxispartner flexibel abholen und sich auf seine Gedankengänge und Erfahrung einlassen können (niederschwellig und hoch professionell).

4.3 Stolpersteine

Im Folgenden werden drei der häufigsten Stolpersteine bei der Erarbeitung eines Wirkungsmodells mit einem möglichem Umgang damit im Rahmen von MMI-Evaluationen aufgezeigt.

Outputs vs. Outcomes

Eine der grossen Herausforderungen ist das Festlegen von zu erreichenden Outcomes und deren Abgrenzung von Outputs. Oft wird die Wirksamkeit oder der Erfolg eines Projektes in Outputs ausgedrückt, die automatisch dem Outcome gleichgesetzt werden. Das Erreichen von 50 Kindern und deren Teilnahme am Angebot heisst zum Beispiel noch nicht unbedingt, dass sich ihre Entwicklung als Folge des Angebots auch wirklich verbessern wird.

Outputs: sind Projektaktivitäten und -leistungen wie Anzahl Durchführungen, Kinder, Sitzungen/Workshops etc.

Es ist essentiell zu beschreiben, wie viele Personen erreicht werden, was deren Merkmale sind und welche Leistungen im Rahmen des Angebotes erbracht werden. Bereits dies stellt Organisationen oft vor grosse Herausforderungen. Im Rahmen einer MMI-Evaluation wird, wenn nötig, immer auch besprochen, wie die wichtigsten Outputs im Rahmen eines Monitoringkonzepts definiert, sinnvoll festgehalten und verwendet werden können.

Outcomes: sind die Veränderungen, welche in der Zielgruppe angestrebt werden.

In Zusammenarbeit der Autorin mit der University of New Hampshire und der UBS Optimus Foundation wurden drei kurze Videos entwickelt, welche sich sehr gut als Werkzeuge eignen, mit Praktikern in Dialog zu kommen und einige dieser Punkte wieder aufzunehmen.

Why Evaluation: http://www.unh.edu/ccrc/evaluation-videos/Chapter_1_Master.mp4

Program Theory: http://www.unh.edu/ccrc/evaluation-videos/Chapter_2_Master.mp4

Measuring Outcomes: http://www.unh.edu/ccrc/evaluation-videos/Chapter_3_Master.mp4

Schwierigkeiten bereitet in sehr vielen Fällen die sinnvolle Definition von Outcomes. Oft sind sie vage und nicht oder nur sehr schwer überprüfbar (z.B. „Kinder sind sicherer in der Gesellschaft“; besser wäre es, was gemeint ist, präzise auszuformulieren: „Kinder erleben weniger Gewalt“; „Kinder haben weniger Unfälle“).

Ein häufiges Beispiel einer zu kurzichtigen, nicht zu Ende gedachten Wirkung und vernachlässigter Evidenz ist es von verbessertem Wissen auf eine angestrebte Wirkung auf der Verhaltensebene zu schliessen. Der Transfer von Wissen zu Verhaltensänderungen ist jedoch oft nicht ohne weitere Unterstützung gegeben.

Die klare Definition und ein gemeinsames Verständnis von Outcomes sind eine der wichtigsten Grundlagen für eine erfolgreiche Evaluation. Nur so kann sichergestellt werden, dass im Rahmen der Evaluation auch wirklich das gemessen wird, was das Angebot erreichen soll, und nicht etwas anderes. Ansonsten ist es nicht möglich, fehlende Wirkung (Null-Effekte) sinnvoll zu erklären. In diesem Fall misst der Forscher etwas, das gar nicht vom Angebot intendiert wird. Wenn durch die Messung dann keine Veränderung angezeigt wird, das Angebot also bezüglich dieses Outcomes „nicht wirkt“, ist dies ein Resultat, das keine schlüssige Interpretation ermöglicht.

Leider ist es so, dass dies immer wieder vorkommt und vor allem auf Seiten der Praktiker zu grosser Frustration, Verunsicherung und möglichem Reputationsverlust führen kann.

Logische Lücken

Logische Lücken im Wirkungsmodell zeigen sich meist auf eine dieser zwei Arten:

1. Es zeigt sich, dass die Intervention nicht die effizienteste/beste Art und Weise ist, das Problem anzugehen, oder dass es Forschungsergebnisse gibt, welche besagen, dass zusätzliche Angebots-elemente notwendig sind.
2. Es zeigt sich, dass es nicht möglich ist, mit den zur Zeit umgesetzten Angebotselementen die angestrebten Ziele zu erreichen.

Manchmal stellt sich sogar bei diesem frühen Evaluationsschritt heraus, dass die Diskrepanz entweder



zum skizzierten Problem oder den zu erreichenden Zielen so gross ist, dass es sich lohnt, zuerst das Angebot zu überarbeiten, um die Passung zwischen Problem und Erreichung der Ziele zu optimieren, bevor weitere Schritte der Evaluation umgesetzt werden.

Organisationsentwicklung

Manchmal stellt sich auch heraus, dass das Hauptanliegen einer Organisation nicht programmatischer Art ist, sondern dass Organisationsentwicklungselemente prioritär anzugehen sind. Dies können Unklarheiten in operativen Prozessen sowie bzgl. Verantwortung und Steuerung (Governance) oder mangelnde Finanzierungs- bzw. Fundraisingexpertise sein. Oft sind kleinere Organisationen von Fachpersonen geführt, welche inhaltlich reichhaltige Expertise mitbringen, aber nur minimal geschult sind in Organisationsführung. Bei erfolgreichen Programmen und somit wachsenden Organisationen ist dieses Manko besonders kritisch.

Damit sich die Investition in eine Evaluation lohnt und deren Ergebnisse nicht durch Mängel der Organisationsführung verzerrt werden, kommt es auch in solchen Fällen vor, dass die Forscherin des MMI vorschlägt, zuerst oder zumindest parallel auf einen Organisationsentwicklungsprozess zu fokussieren, bevor die Evaluation des Angebots zu weit fortschreitet. Zur Unterstützung kann das MMI hier auf die interne Organisationsentwicklungsexpertise des pädagogischen Teams und auf ein Netz von externen Berater/innen zurückgreifen.

5. Wahl des Evaluationsdesigns

Die Wahl des passenden Evaluationsdesigns steht neben der Investition in eine solide Partnerschaft ebenfalls im Zentrum von MMI-Evaluationen. Wie unter 3.4. bereits erläutert, werden Evaluationsdesigns nicht aufgrund von Fertigkeiten und Vorlieben des Evaluators (eher qualitativ vs. quantitativ, eher formativ vs. summativ, eher explorative vs. Wirkungsevaluation) gewählt, sondern in einem systematischen Prozess erarbeitet, der verschiedene Faktoren berücksichtigt.

Ziele sind dabei, die Stärke („rigour“) der Evidenz, die aus der Evaluation gewonnen werden kann, zu maximieren und gleichzeitig den Nutzen für das Programm zu optimieren. Das Design wird auf die gemeinsam definierten Ziele und die Kernfragestellungen der Evaluation abgestimmt. So bedürfen explorative und formative Zugänge neben einer guten Basisdokumentation der Outputs eher qualitativer Methoden, während die Stärke der Evidenz signifikant verbessert werden kann, wenn für hypothesentestende Fragestellungen sowie für Wirkungsevaluationen, wo immer möglich, ein quantitatives Vorgehen gewählt wird.

Zudem fließen die Ausgereiftheit des Wirkungsverständnisses (siehe Kapitel 4) sowie die Phase, in der das Angebot umgesetzt wird (Konzeptphase, Pilotphase, Umsetzungsphase, Scale up Phase), wesentlich in die Entscheidung mit ein.

Nicht zuletzt ist selbstverständlich der verfügbare Ressourcenrahmen (finanzielle Mittel und Zeithorizont der Evaluation) ausschlaggebend.

Im Idealfall wird in der Entwicklung eines Angebots bereits in der Konzeptphase ein logisches, explizites Wirkungsmodell erarbeitet, welches mit wissenschaftlichen Erkenntnissen untermauert ist. In einer Pilot- und anschließenden Umsetzungsphase werden fortlaufend systematisch formative Daten erhoben, die Praxis reflektiert und die Umsetzung gefestigt. Zudem wird eine Abfolge zunehmend strengerer Wirksamkeitsevaluationen durchgeführt. Erst wenn die Machbarkeit und Wirksamkeit eines Angebotes etabliert sind, ist es sinnvoll und vertretbar, unterstützt mit weiteren formativen Elementen, eine Verbreitung (Scale up) dieses nun evidenz-basierten Angebots anzustreben.

Ein kontinuierlicher und aktualisierter Einbezug von Forschungsergebnissen und eine auf den Umsetzungsstand angepasste Evaluation des Angebots sind in jeder Phase wertvoll und wichtig. Im Rahmen der Evaluationsarbeit begleitet das MMI Angebote in jeder Phase der Angebotsentwicklung. Intern ist das MMI bestrebt, im Austausch zwischen dem Fach- und dem Forschungsteam die eigenen Angebote auf diese Weise zu entwickeln und kontinuierlich zu evaluieren. Dies im Bewusstsein, dass sich damit je nach Fragestellung und Art des Angebots zu gegebener Zeit eine externe Evaluation nicht erübrigt.

Viele Angebote und Programme werden allerdings nicht in einem solch engen Austausch zwischen Praxis und Forschung entwickelt. Die Realität der meisten Evaluationsaufträge, welche an das MMI gelangen, sieht daher etwas anders aus. Viele Praktiker haben keinen oder limitierten Zugang zu wissenschaftlichem Vorgehen und/oder werden während der Programmentwicklung nicht von einem wissenschaftlichen Partner begleitet. Deshalb gelangen Anfragen für einen externen Evaluationsauftrag in jeder Angebotsphase und mit jeder Art von Ausreifung der Nutzung von Evidenz an uns. Der MMI-Evaluationsansatz schätzt zur Bestimmung des passenden Evaluationsdesigns systematisch einerseits die Projektphase und die „Reife“ des Programms oder Angebots und andererseits die bereits bestehende Evidenzebene ein. Es ergeben sich die folgenden möglichen „Stadien“ von „Angebotszuständen“ (Tabelle 1). Die Bedeu-

		Projektphase			
		Konzept	Pilot	Umsetzung	Scale Up
Evidence of impact	Keine	1	2	3	4
	Evidence- informed	5	6	7	8
	Evidence- based	9	10	11	12

Tabelle 1: Angebote

Evidence-informed (auf reflektierte Erkenntnisse oder auf Forschungsergebnisse aus anderen Studien gestützt) bedeutet, dass das Projekt in ein Wirkungsmodell eingebettet ist, auf bestehenden Forschungserkenntnissen aufbaut und eine systematische Umsetzung von Praxisreflexion und/oder formativer Evaluation besteht.

Evidence-based (auf belegter Wirksamkeit basierende Evidenz) bedeutet, dass Erkenntnisse zur Wirkung bestehen, welche in einer formalen Wirksamkeitsevaluation durchgeführt wurden, idealerweise in einem Design mit Vergleichsgruppe.

Die Begriffe *evidence-informed* und *evidence-based* werden in Box 2 erläutert.

Es ist nicht zu erwarten und auch gar nicht möglich, dass ein Angebot in der Konzeptphase bereits Wirksamkeit nachweisen kann, da noch keine Umsetzung stattgefunden hat (Feld 9). Weiter ist es eher selten und nicht zwingend notwendig, dass ein Angebot in der Pilotphase bereits Wirksamkeit auf einem hohen Niveau (oder überhaupt) nachweisen kann (Feld 10). Spätestens in der Umsetzungsphase sollte aber in die Wirksamkeitsevaluation in Richtung *Evidence of impact* investiert werden. Es gibt aber Angebote, die bereits weit verbreitet sind und grossflächig umgesetzt werden (*Scale up*), jedoch in keiner Weise evidenzbasiert sind und deren Wirkung oder potentielle Risiken oder unerwünschte Effekte unbekannt sind (Feld 4). Im Rahmen einer MMI-Evaluation wird zu Beginn der Evaluation besprochen, wie schrittweise die entsprechenden Stufen „aufgeholt“ und die Wirkungsorientierung und Evidenz des Angebotes gestärkt werden können. Um ein Angebot „*evidence-informed*“ zu machen, werden Wirkungsmodelle erarbeitet und sie in die bestehenden wissenschaftlichen Erkenntnisse eingebettet, formative Evaluationen durchgeführt und ein Konzept für reflektierte Praxis eingeführt. Anschliessend werden in zunehmender Stärke (*rigour*) der Evaluationen Erkenntnisse zur Wirkung etabliert.

Das MMI setzt explizit keine strengen Evaluationen wie randomisierte Kontrollstudien mit schlecht etablierten und konzeptuell mangelhaft eingebetteten Angeboten um. Ein zu früh eingesetztes, strenges Evaluationsdesign – wenn dies auch für den auf wissenschaftliche Publikationen angewiesenen Forscher interessant ist – schwächt die Chancen auf eine erfolgreiche Durchführung einer Evaluation beträchtlich, schmälert den möglichen Erkenntnisgewinn und birgt damit ein substantielles Risiko für den Praxispartner.

Es ist in der Arbeit mit dem Praxispartner sorgfältig zu beachten, dass oft grosse Verwirrung bezüglich Möglichkeiten und Grenzen eines Evaluationsdesigns besteht: ein Angebot, welches sich einer formativen Evaluation unterzogen hat, kann keine Aussagen zur Wirksamkeit des Angebotes machen und es kann sich unter Umständen „evidence -informed“, aber nicht „evidenz-basiert“ nennen. Aus diesem Grund wird im Rahmen des Workshops zur Wahl des Designs mit den Praktikern im Detail diskutiert, was die Vor- und Nachteile eines Designs sind, welche Art von Fragen mit den unterschiedlichen Herangehensweisen beantwortet werden und mit welcher „Sicherheit“ entsprechende Aussagen zu Wirksamkeit gemacht werden können (siehe Box 3).

Prozess- und Wirksamkeitsevaluation:

Unterschiedliche Evaluationsansätze sind zu unterschiedlichen Zeitpunkten und Phasen eines Angebotes nützlich. Jeder Ansatz beantwortet unterschiedliche Fragen und liefert andere Informationen.

Prozessevaluation

Umsetzung

Das wichtigste Ziel bei einer Prozessevaluation ist das Sammeln von Information, welche in die Angebotsumsetzung gespiesen werden kann. Zum Beispiel:

- Wie gut wird das Angebot umgesetzt? Wird so umgesetzt wie es geplant war?
- Wie gut wird die Zielpopulation erreicht?
- Welche Knacknüsse zeigen sich?
- Welche möglichen Lösungen wurden ausprobiert?
- Welche Elemente bewähren sich / werden nachgefragt, welche nicht?

Wirksamkeitsevaluationen

Prä-Post

Hier wird die Ausprägung der Kernoutcomes zu Beginn und zum Schluss einer Intervention gemessen.

sen. In manchen Fällen ist ein Prä-Post Design ausreichend, manchmal braucht es aber eine Vergleichsgruppe, um zu bestimmen, ob die gemessene Veränderung zwischen den Zeitpunkten tatsächlich auf das Angebot zurückzuführen ist. Ein Beispiel dafür sind Indikatoren zur Entwicklung des Kindes: Ein Kind reift und entwickelt sich natürlich über die Zeit. Es ist anzunehmen, dass entsprechende Werte zu einem zweiten, späteren Zeitpunkt höher ausfallen als in einer ersten Messung. Dies geschieht sehr wahrscheinlich auch ohne Intervention.

Um zu entscheiden, ob und zu welchem Ausmass dies aber auf die Intervention zurückzuführen ist, braucht es eine Gruppe, welche nicht am Angebot teilnimmt. Möglicherweise zeigen Vergleiche, dass sich der Rückstand zu einer unbelasteten Gruppe verkleinert oder durch das Angebot ein Abstand zu einer Gruppe ohne Intervention entsteht (beides z.B. beim Erwerb der Umgebungssprache von Kindern aus zugewanderten Familien durch den Besuch einer qualitativ guten Spielgruppe).

Mit Vergleichsgruppe

Quasi-experimentell:

Die Strenge der Evaluation und deren Aussagekraft verbessern sich also, wenn eine Vergleichsgruppe zur Verfügung steht. Manchmal kann es "einfach" eine andere Gruppe von Personen sein, in einem regulären Angebot, aus einer anderen Gemeinde. Aber auch dies kann zu Problemen führen, da die Teilnehmenden sich unter Umständen bereits zu Beginn unterscheiden und die Resultate somit verzerrt sind. Deshalb braucht es eine Ausgangsmessung als sogenannte Baseline, also eine Prä- und Postmessung.

Randomisierte Kontrollstudien (RCT):

Dafür gilt es eine Vergleichsgruppe zu finden, die der Gruppe, welche das Angebot erhält, so ähnlich wie möglich ist. Die Auswahl, wer das Angebot erhält und wer nicht, soll per Zufall entschieden werden. Diese Art von Evaluation wird in der Forschung als Goldstandard gehandelt, ist aber nicht ganz einfach durchzuführen. Sie ist sehr ressourcenintensiv.

6. Wissensmanagement

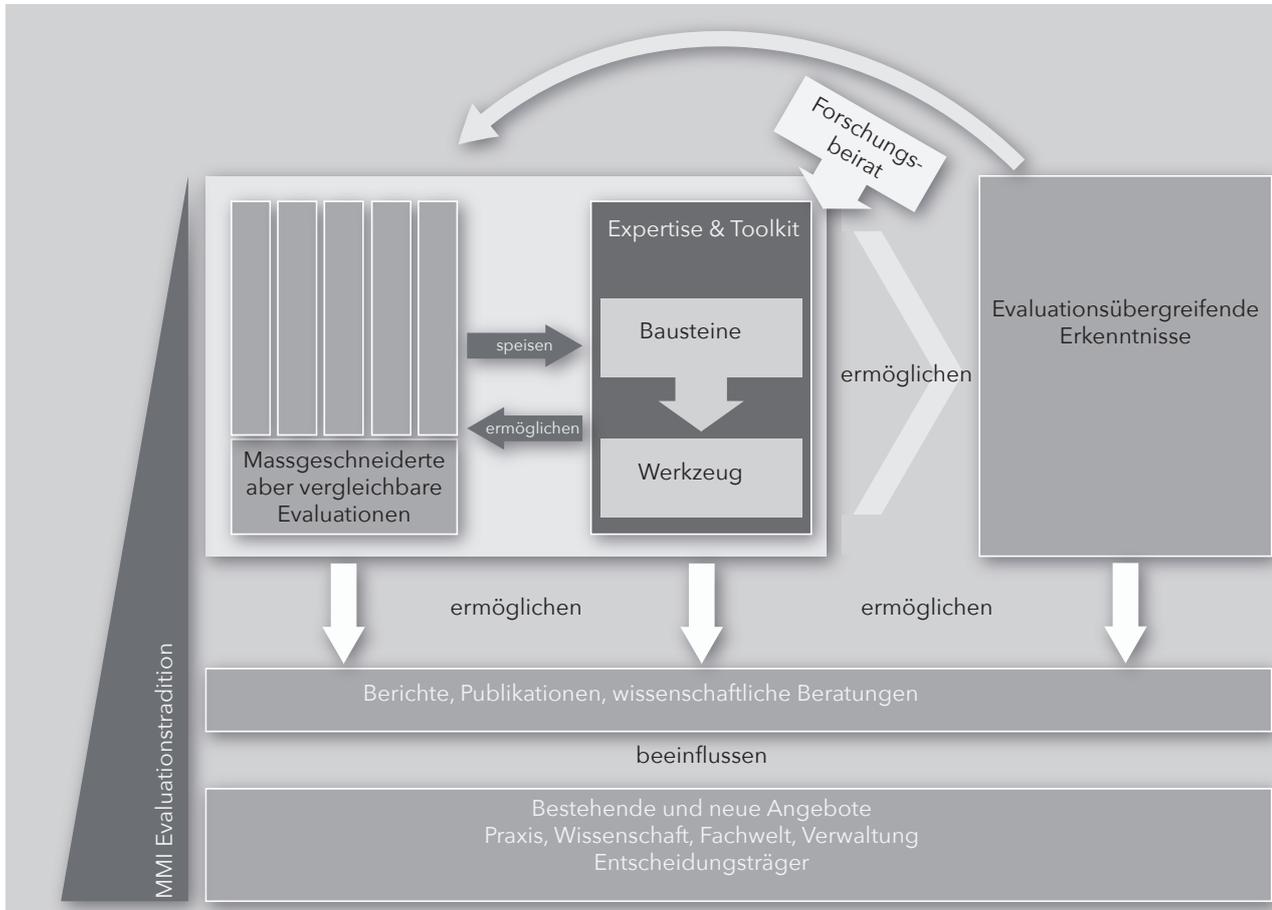
Im Zentrum von MMI-Evaluationen steht schliesslich auch der Anspruch, dass die Resultate der individuellen Evaluationen stets auf mehreren Ebenen einen Erkenntnisgewinn bringen sollen. Um evaluationsübergreifende Erkenntnisse ableiten zu können, ist ein systematisches Vorgehen unabdingbar und bedarf des in diesem Konzept beschriebenen massgeschneiderten und gleichzeitig vergleichbaren und deshalb standardisierten Vorgehens (Abbildung 5).

Im Fokus steht selbstverständlich der Erkenntnisgewinn zu Gunsten des evaluierten Angebots selber. Dieses soll sich anhand der Resultate weiterentwickeln können und zu einem wirkungsorientierten Ansatz gestärkt werden. Dazu werden die Rohresultate gemeinsam mit dem Praxispartner in einem Workshop besprochen und in einem wissenschaftlich soliden, jedoch in erster Linie für Praktiker verständlichen Bericht abgefasst (gray paper).

Die Resultate sollen aber auch anderen, ähnlichen Angeboten zur Verfügung stehen, um kollektives Lernen im Felde zu stärken und die Entwicklung von *good practice* zu unterstützen. Wenn immer möglich, verfasst das Team des MMI Publikationen, in denen diese angebotsübergreifend relevanten Resultate synthetisiert sind. Auf diese Weise sollen die einzelnen Evaluationen wie auch die Evaluationsarbeit in ihrer Summe Relevanz für soziale Veränderungen sowie politische und fachliche Diskussionen gewährleisten.

Wenn immer möglich, sollen die Resultate aus der Evaluation des Angebotes auch in wissenschaftlichen, peer-reviewed Artikeln abgefasst werden, um die Evidenzbasis zu stärken und die Resultate der wissenschaftlichen Gemeinschaft zugänglich zu machen. Ergeben sich Erkenntnisse aus einer Serie von Evaluationen, seien diese inhaltlicher oder methodischer Art, werden diese als Meta-Ergebnisse ebenfalls in wissenschaftlichen Artikeln abgefasst.

Intern werden die Erkenntnisse gebündelt und die Bausteine und Werkzeuge des Evaluationskonzeptes bei Bedarf entsprechend ergänzt oder abgeändert.



7. Anhang

Leitbild Forschung Marie Meierhofer Institut für das Kind

Analog zum Leitbild des Marie Meierhofer Instituts für das Kind beschäftigt sich die Forschung am MMI mit der ganzheitlichen und gesunden Entwicklung des Kindes. Der Schwerpunkt liegt dabei auf der frühen Kindheit. Die Aktivitäten bewegen sich stets am Transmissionsriemen von Forschung, Praxis, Verwaltung und Gesellschaft und verfolgen ein duales Ziel, indem parallel sowohl Forschungs-impact wie auch sozialer Impact angestrebt werden. So wird, um die Relevanz der Erkenntnisse zu gewährleisten, die umgesetzte Forschung massgeblich von Impulsen aus der Praxis, Verwaltung und Gesellschaft gespiesen. Die Resultate werden entsprechend kontextualisiert.

Um das Kind und seine Lebenswelten möglichst ganzheitlich zu erfassen, werden ausgewählte Themen durch ein multi-disziplinäres Fenster beleuchtet. Vertreten sind im Team des MMI namentlich die folgenden Disziplinen: Frühpädagogik (Erziehung und Bildung von Kindern unter sechs Jahren), Entwicklungspsychologie (zeitlich überdauernde und aufeinander aufbauende Veränderungen des kindlichen Erlebens und Verhaltens), Entwicklungspsychopathologie (abweichende Entwicklung und deren Ursachen und Verläufe bei Kleinkindern) und Familien- und Bildungssoziologie (Sozialisation des Kindes in der Familie und den weiteren Lebenswelten). Ferner arbeitet das Team des MMI regelmässig mit Expert/innen der Entwicklungspädiatrie und des Familien- und Kinderrechts zusammen.

In den einzelnen Disziplinen und übergreifend werden Schwerpunkte gesetzt, innerhalb derer dann spezifische Fragestellungen bearbeitet werden.

Ebenfalls wird ein multi-methodaler Ansatz verfolgt. Dies bedeutet, dass sowohl quer-, wie auch längsschnittliche Studien mit qualitativen und / oder quantitativen Methoden umgesetzt werden. Zudem werden auch theoretisch-konzeptionelle Arbeiten durchgeführt.

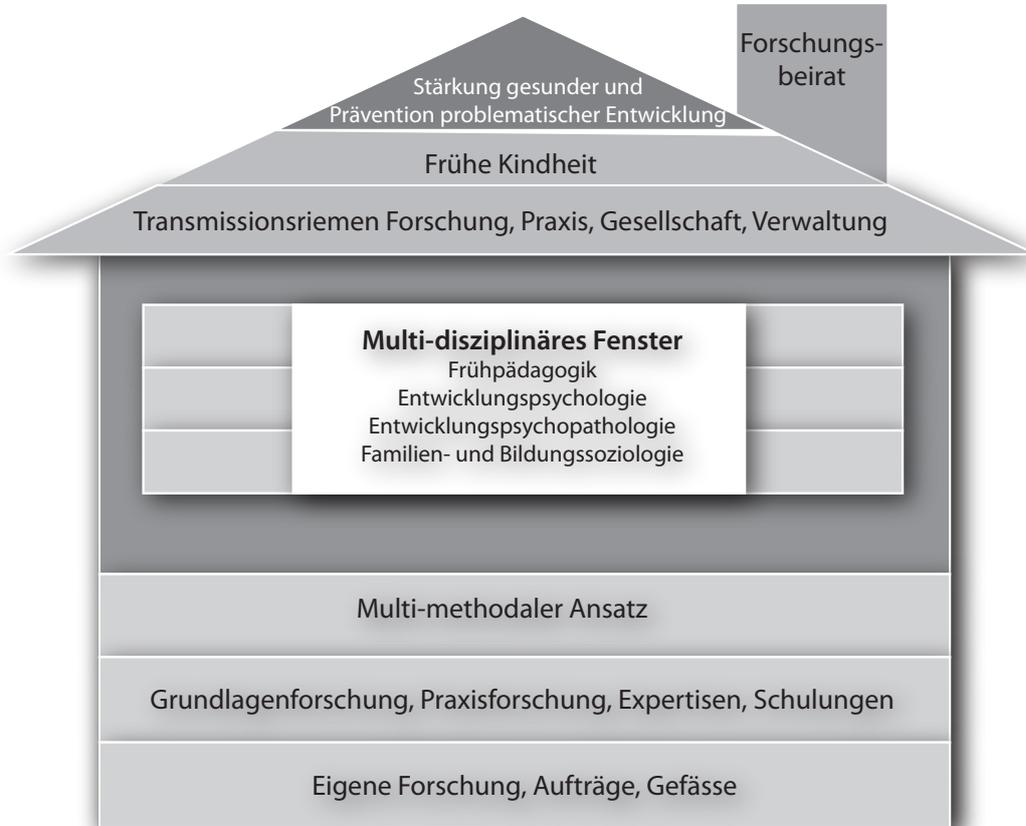


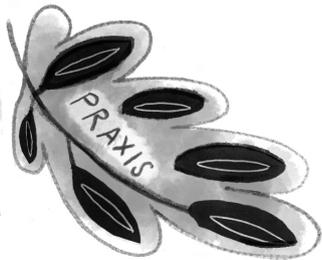
Abbildung 6: Leitbild Forschung

Umgesetzt wird dieser bewusst breit gesetzte Ansatz anhand von strategischer Suche und Anstellung von Personen mit komplementärer methodischer Expertise und unterschiedlichem disziplinären Hintergrund.

Neben Grundlagenforschung, in der erkenntnisorientiertes Elementarwissen erarbeitet wird, hat auch die Praxisforschung einen wichtigen Stellenwert. Diese beinhaltet Evaluationen, Angebots- und Konzeptentwicklung mit wissenschaftlicher Begleitung und die Erarbeitung von Wirkungsmodellen. Es werden sowohl eigene Forschungsideen konzipiert und in Projekten umgesetzt, wie auch Forschungsaufträge Dritter bearbeitet. Des Weiteren bietet das MMI Expertisen und Schulungen als Dienstleistungen an.

Zudem führt das MMI regelmässig Veranstaltungen in einem Spektrum von internen, halb-offenen (mit selektiver Einladung) und offenen (öffentlich ausgeschrieben) Gefässen durch. Es sind dies Werkstätten, Kolloquien und die jährliche wissenschaftliche Tagung zu frühkindlicher Bildungs- und Entwicklungsforschung (in Kooperation mit der Universität Zürich). Die Gefässe dienen dazu, wissenschaftliche Fragen gemeinsam zu diskutieren, gemeinsame Lerngelegenheiten anhand konkreter Projekte zu schaffen und den Transfer von Fragen aus der Praxis in die Forschung und von wissenschaftlichen Erkenntnissen in die Praxis zu gewährleisten.

Die Forschung des MMIs wird durch einen Forschungsbeirat unterstützt. Dessen Funktion ist es, die Forschungsstrategie und die konkreten Forschungstätigkeiten des MMI zu reflektieren, zu diskutieren und beratend zu begleiten.





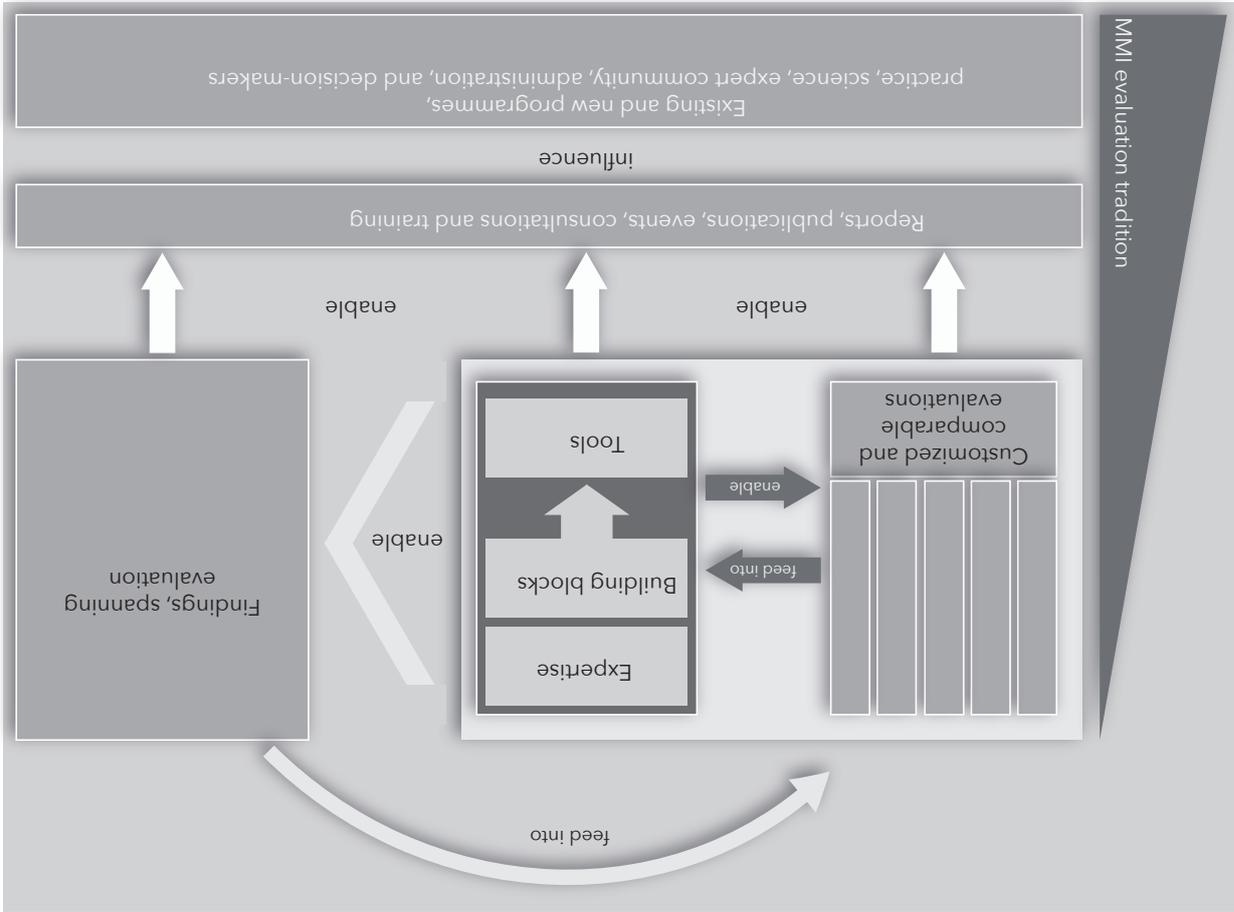


Figure 5: Knowledge management

programme. The decision of who participates and who does not should be determined by chance. This type of evaluation is considered the gold standard, but is not easy to carry out. It is very resource-intensive, and can also be logistically and ethically complicated.

6. Knowledge Management

Finally, a core aim of MMI evaluations is to ensure that the results of individual evaluations provide insights on a variety of levels. A systematic approach is indispensable to derive multi-level findings, ideally following the customized and simultaneously comparable, standardized procedure described in this concept (Figure 5). Of course, the main objective is to generate knowledge that benefits the evaluated programme itself. The evaluation results should enable the programme to develop further towards an impact-oriented approach. To this end, the raw findings from the evaluation are discussed in a workshop with the practice partner, and written up in a scientifically sound report that is nevertheless tailored primarily for practitioner audiences (grey paper).

However, the results should also be made available to other similar programmes so as to strengthen collective learning in the field and support development of *good practice*. Whenever possible, the MMI team authors publications that synthesize relevant results spanning multiple programmes. In this way, individual evaluations and the evaluation work as a whole are made relevant for social change as well as political and professional dialogue.

Whenever possible, programme evaluation results should also be taken up in peer-reviewed scientific articles, so as to strengthen the evidence base and make the results accessible to the scientific community. If insights emerge from a series of evaluations, whether of a content-based or methodological nature, these meta-results may also be taken up in scientific articles. Internally, the evaluation findings are bundled together and, where applicable, the building blocks and tools of our evaluation concept are correspondingly enhanced or revised.

already widespread and implemented at a large scale (scale up), which are not evidence-based and lack knowledge of impacts, potential risks, or undesirable effects (field 4).

As part of an MML evaluation, we discuss from the beginning how corresponding programme stages can be "caught up" step by step, strengthening the programme's impact-orientation and evidence basis. To make a programme "evidence-informed", impact models are developed and embedded in existing scientific knowledge, formative evaluations are carried out, and a concept for reflective practice is introduced. Subsequently, insights into programme effectiveness are established in evaluations of increasing strength (rigour). The MML expressly avoids applying rigorous evaluations, such as randomized controlled trials, to poorly established programmes that lack a solid conceptual framework. Premature application of a strict evaluation design – even if interesting for researchers concerned with scientific publications – considerably weakens the chance of successfully completing an evaluation, narrows the scope of possible insights, and bears substantial risks for the practice partner.

When working with practice partners, it should be carefully noted that there is often a lot of confusion about the possibilities and limits of an evaluation design: A programme that has undergone a formative evaluation cannot make any conclusive statements about its effectiveness. It may refer to itself as "evidence-informed", but not "evidence-based": For this reason, the workshop on choice of evaluation design must include detailed discussion with practitioners about the advantages and disadvantages of each design, what kinds of questions are answered by different approaches, and with what level of "certainty" corresponding statements about effectiveness may be made (see Box 3).

Process and impact evaluations:

Different evaluation approaches are useful at different times and phases of a programme. Every approach answers different questions and provides different information.

Process evaluations:

Implementation:
The most important goal of a process evaluation is to collect information that can be fed into programme implementation. For example:

It should not be expected, nor is it even possible, for a programme in the concept phase to demonstrate effectiveness, since implementation has yet to occur (field 9). It is also relative uncommon, and not absolutely necessary, for a programme in the pilot phase to demonstrate high (or any) effective-ness (field 10). However, in the implementation phase at the very latest, investments should be made into identifying and evaluating evidence of impacts. Unfortunately, there are programmes that are

Box2:

Evidence-informed (underpinned by considered insights or research results from other studies) means that a project is embedded in an impact model, builds on existing research results, and comprises a systematic implementation of practical reflection and/or formative evaluation.
 Evidence-based (underpinned by evidence of proven effectiveness) means that evidence of efficacy exists which was derived in a formal impact evaluation, ideally including a comparison group in its design.

Project phase	Evidence of impact			
	None	Evidence-informed	Evidence-based	
Concept	1	5	9	
Pilot	2	6	10	
Implementation	3	7	11	
Scale Up	4	8	12	

Table 1: Interventions

on 4) as well as the phase in which the programme is implemented (concept phase, pilot phase, implementation phase, scale-up phase).

Last but not least, the level of available resources (financial means, time horizon of the evaluation) are, of course, also decisive.

Ideally, the development of a programme will include establishment of a logical, explicit impact model that is underpinned by scientific findings. In a pilot phase and subsequent implementation phase, systematic formative data are collected on an ongoing basis, reflecting practice and strengthening implementation. In addition, a series of increasingly rigorous impact evaluations will be carried out. Only once a programme's feasibility and effectiveness have been established is it sensible and justifiable, supported by further formative elements, to strive for dissemination (scale up) of the now evidence-based programme.

Continuous, updated inclusion of research results and an evaluation of the programme, adapted to its implementation status, are valuable and important in every phase. In the course of its evaluation work, the MML accompanies programmes in every phase of their development. Internally, the MML strives to develop and continually evaluate its own programmes in the same manner, based on ongoing exchange between its technical and research teams. This is done with full awareness that, depending on the relevant issues and programme type, an external evaluation may also be necessary in due course.

However, many programmes and interventions are not developed in such a close exchange between practice and research. Indeed, the reality behind most evaluation assignments MML receives is different. Many practitioners have little or no access to scientific procedures and/or are not supported by a scientific partner during programme development. Thus, we receive requests for external evaluation of programmes in many different phases, displaying various levels of sophistication regarding evidence use.

To determine the appropriate evaluation design, MML's evaluation approach systematically assesses the project phase and the "maturity" of the intervention or programme, on the one hand, and the existing evidence base, on the other.

This results in the possible "stages" of "programme status" shown below (Table 1). The meaning of the terms *evidence-informed* and *evidence-based* are explained in Box 2.

5. Choice of Evaluation Design

Other times, it is revealed that the main concern of an organization is not programmatic in nature, but rather has to do with elements of organizational development that should be addressed as a matter of priority. These may involve ambiguous operational processes, responsibility, and direction (governance), or lack of financing and fundraising expertise. Smaller organizations, in particular, are often run by specialist staff who have a wealth of expertise, but have little training in organizational management. This shortcoming can be especially critical with respect to successful programmes and growing organizations.

To ensure that the investment in an evaluation is worthwhile and its results are not distorted by deficiencies in organizational management, the MML evaluators may recommend focusing first, or at least in parallel, on processes of organizational development, prior to advancing too far with the programme evaluation. For support in such cases, the MML can draw on its pedagogical team's expertise in organizational development as well as its network of external consultants.

In addition to cultivation of a solid partnership, the choice of an appropriate evaluation design is at the heart of MML evaluations. As described under 3.4, evaluation designs are not selected based on the skills or preferences of the evaluator (qualitative vs. quantitative, formative, explorative vs. impact evaluative), but are instead selected in a systematic process considering various factors. The aim is to maximize the strength ("rigour") of the evidence generated by the evaluation as well as to optimize its benefits for the programme itself. The design is adapted to the jointly defined goals and the core questions of the evaluation. In addition to good basic documentation of outputs, explorative and formative approaches require more qualitative methods, whereas the strength of the evidence basis can be improved significantly by using quantitative methods, whenever possible, to test hypotheses and assess impacts.

Also key to the evaluation design are the level of sophistication of their understanding of impact (see Secti-

meant, it may be better to formulate it more specifically, e.g. "Children experience less violence"; "Children have fewer accidents".

One common example of short-sighted, insufficient impact thinking and lack of evidence is drawing conclusions about desired *behavioral* impacts based solely on improved *knowledge* among a target group. However, the transformation of knowledge into behavioural changes often does not occur without further support.

Clearly defined, joint understanding of outcomes is one of the most important foundations of a successful evaluation. It is the only way to ensure that the evaluation measures what was targeted by the programme and not something else. Without this joint understanding, it is not possible to explain missing impacts (zero effects) in a meaningful way. In this case, the researcher tries to measure something that was never intended by the programme. When the resulting measurement indicates no change, i.e. that the programme "does not work" regarding a particular outcome, virtually nothing is revealed that can be interpreted conclusively.

Unfortunately, this is a frequent occurrence and one that can lead, especially among practitioners, to major frustration, uncertainty, and possible loss of reputation.

Logical gaps

Logical gaps in the impact model typically appear in one of two ways:

1) It is revealed that the intervention is not the most efficient/best way of addressing the problem, or research results suggest that additional programme elements are needed.

2) It is revealed that the targeted goals cannot be achieved using the currently implemented programme elements.

Sometimes, this initial evaluation step reveals a big discrepancy regarding the problem outlined, or the targeted goals, that it appears better to refine the programme, optimizing the problem-goal fit, before going further with subsequent evaluation steps.

4.3 Stumbling blocks

In the following, we discuss three of the most common stumbling blocks when developing an impact model, and how they may be dealt with in the context of MML evaluations.

Outputs vs. outcomes

One of the biggest challenges is defining targeted outcomes and distinguishing them from outputs. Often, the effectiveness or success of a given project is expressed in terms of *outputs*, which are automatically (and incorrectly) equated with *outcomes*. For example, reaching 50 children and securing their participation in a programme does not necessarily mean that their development will improve as a result of the programme.

Outputs are project activities and products such as the number of implementations, participating children, meetings/workshops, etc.

It is essential to describe the number of people reached, their characteristics, and what services have been provided in the context of a programme. This alone often presents major challenges to organizations. In an MML evaluation, we discuss as needed how the most important outputs can be defined, sensibly recorded, and applied in the framework of a monitoring concept.

Outcomes are the changes striven for in the target group.

In collaboration with Dr. Lisa Jones, University of New Hampshire, Dr. Sonja Schenkel and the UBS Optimus Foundation, the author developed three short videos that serve as useful tools for initiating dialogue with practitioners and discussing key points:

Why Evaluation: http://www.unh.edu/ccrc/evaluation-videos/Chapter_1_Master.mp4

Programme Theory: http://www.unh.edu/ccrc/evaluation-videos/Chapter_2_Master.mp4

Outcomes: http://www.unh.edu/ccrc/evaluation-videos/Chapter_3_Master.mp4

In numerous cases, defining meaningful outcomes proves challenging. They are often left vague and difficult or impossible to assess (e.g. "Children are safer in society" – depending on what is actually

At the same time, it is usually very important to tailor and individualize support to the project partners. Practitioners are often deeply involved in operational activities and have difficulty schematically formulating programmes without external support in structuring. Facilitation of independent model development (e.g. provision of guiding questions, without discussion) has not proven to be effective. It is important to note that the goals formulated in the impact model need not overlap entirely with what is ultimately measured and recorded in the evaluation. Depending on the framework and resources of the evaluation, it is usually not possible to measure every goal. However, the impact model helps make explicit what is actually measured or not among the expected effects and programme goals, in the context of a concrete evaluation. Whenever possible, goals that are not measured in the evaluation should be substantiated based on existing research from other studies.

In addition to the importance of the impact model for the programme itself and as a basis for evaluation, the process of jointly developing an impact model can serve as an instrument for building a working relationship, understanding the priorities of the other party, and establishing trust that the evaluation will focus on what is most relevant to the programme and the aims of the practitioners. It provides an opportunity for the researchers to ascertain, sharpen, and make explicit the understanding of impact of the practice partners. This is the only way to examine the logic of the impact chain, from problem to solution, and see whether it holds up vis-à-vis what we know from current research.

The development of impact models requires skills in highly structured thinking on the part of researchers, as well as a high degree of competence in abstract conceptualization and schematic presentation. In addition, it requires significant expert knowledge to grasp the logical steps involved and reinforce them with relevant research literature. The researchers must be able to respond flexibly to the needs of practice partners and engage with their thought processes and experience.

Explanation: Initial situation: Describes the societal problem and the root causes that the programme seeks to address. It documents obstacles (why hasn't the problem already been solved?) or particular challenges to implementation of a programme (e.g. accessibility of target group).
 Intervention: The description of the intervention should present a solution to the identified problem. The approach should overcome the obstacles and challenges (e.g. easy access to improve accessibility) and explicitly describe the mechanism of action that serves to solve the problem.
 Goals: These should be formulated as short-, medium-, and long-term, and should be directly recognizable as solutions to the initially defined "problem".

4.2 Development of the model

In the first workshop, the implementation partners' understanding of impact is discussed (see 3.1). Depending on its level of differentiation and degree of development, more or less time will be spent elaborating this element. In cases of only rudimentary impact models, one or two additional workshops will be held to facilitate joint discussion and development of the impact model. Among others, the following key questions may be used to fill in each element of the impact model, formulating them in logical steps:

- What contents will be offered?
- Why is the programme needed? What is the problem?
- Why wasn't the problem solved a long time ago (obstacles)?
- How does the programme address the problem?
- Who is the target group of the programme and why?
- What are the short-, medium-, and long-term goals?
- What effects does the programme have on the target group?
- How does the programme achieve its goals?

However, these descriptions often do not specify what "social problem" the programme seeks to address ("why the programme is needed") or what evidence there is of the problem (prevalence, causes, risk factors, etc.). Yet it is essential to articulate a logical chain and linkage between the problem and programme elements, as well as how these are embedded in existing evidence. Thus, the MMI concept expands upon such models, adding elements reflecting the following questions: What concrete social problem does the programme seek to address and how are the individual programme elements suited to addressing it? What evidence do we have of causes leading to the social problem? Are they adequately considered within the programme? What obstacles have prevented the problem from being solved in the past (e.g. access to the target group) and how does the programme overcome them?

This gives rise to the following model (Figure 4):

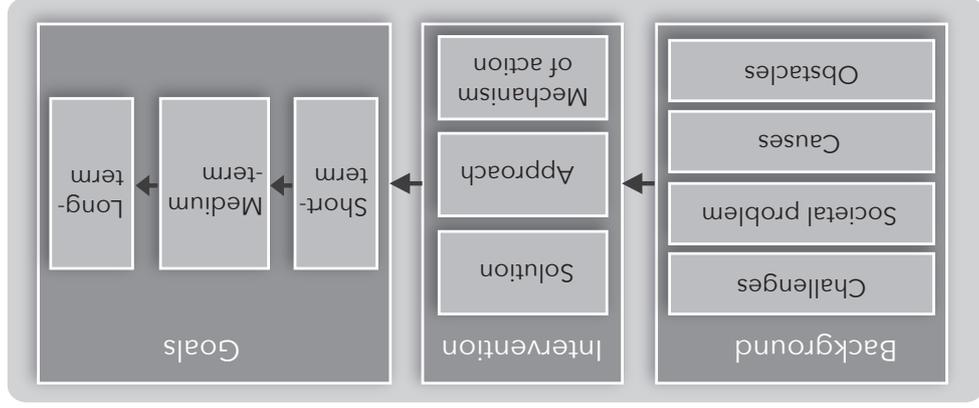


Figure 4: MMI impact model

Classical approaches to explanation of impact models often begin directly with individual programme elements, explicitly describing how they lead to specific goals (outputs and outcomes).

Below are two examples:

Example 1:

Source: Zewo Foundation



Figure 2: Impact model example 1

Example 2:

Source: Jones, L. (2014). *Improving efforts to prevent children's exposure to violence: A handbook to support child maltreatment prevention programs*. World Health Organization.

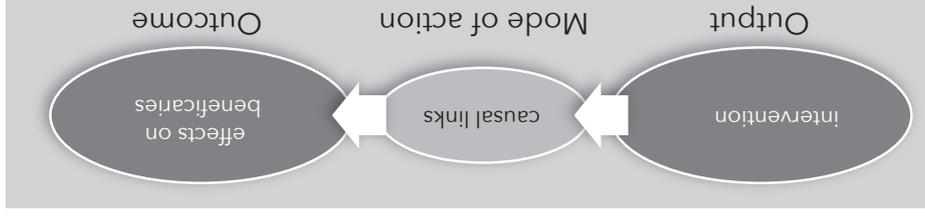


Figure 3: Impact model example 2

The team works explicitly with both qualitative and quantitative research methods, whenever possible, in a complementary way in each evaluation. Data are collected using standardized and, wherever possible, validated survey instruments. In cases of customized and innovative instruments, great importance is attached to systematic, reflective application. As is standard, a distinction is made between hypothesis-testing and explorative approaches, and multiple perspectives are obtained regarding the question at hand (multi-informant). The choice of evaluation design should expressly *not* be made according to the (qualitative vs. quantitative) preferences of the researchers, but rather according to the programme maturity and questions posed (see Section 5 Choice of Evaluation Design).

4. Impact Models

4.1 An expanded model

A central tool in MMI evaluations is the joint development of an impact model ("programme theory", "theory of change", "logic model"). An impact model clarifies what those responsible for the programme consider necessary and useful in the programme framework and explains in logical steps how the programme addresses a given social problem, how it leads to change, and what it aims to achieve. It uses existing evidence to substantiate the individual steps and names the goals to be achieved as clearly defined outcomes. It enables the identification of "blind spots" and potential improvements, strengthening the programme in terms of impact orientation. In addition, an impact model is a highly useful tool for succinctly communicating what the programme involves. It is the foundation of every MMI evaluation.

In many programmes, the understanding of impact is either intuitive, not made explicit, displays logical gaps, or is not substantiated with corresponding evidence. In addition, outputs and outcomes are often vaguely formulated, complicating the selection of measurable goals or causing misunderstandings.

MMI's evaluation concept is characterized by a systematic approach and use of methods reflecting the highest standards of social science. The members of our research team have received in-depth training in at least one of several relevant methodologies (see Section 3.3) and have a broad toolbox of social science methods at their disposal. This makes it possible to implement the right design depending on the issue at hand (see Section 5 Choosing the Evaluation Design).

3.4 Multi-method approach

The MMI houses expertise in both research and practice (transdisciplinarity), and is characterized by a multidisciplinary approach with team members specialized in development psychology, early childhood education, and family sociology. In addition, the MMI team works regularly with colleagues in developmental paediatrics and family and children's law. This enables a holistic view of children and easy access to a wide range of expertise, including extensive theoretical and practical/clinical knowledge in various disciplines, proficiency in therapeutic and pedagogical interventions, and know-how in organizational development. These resources are drawn upon as needed for every evaluation, either informally or in the framework of cross-disciplinary workshops.

3.3 Multi- and transdisciplinary

In our work at the Marie Meierhofer Institute, we put children and their worlds at the centre of every undertaking. Thematically, we prioritize evaluations focussed on early childhood or issues of child rights (especially in terms of supporting children in vulnerable circumstances). MMI's core stance – i.e. viewing issues from the perspective of the child and her/his world – is central to all MMI evaluations (cf. *MMI model on participation of young children: Marie Meierhofer Institut für das Kind (2016): Modell kindlicher Partizipation aus Kinder- und aus Erwachsenensperspektive, Poster, Beilage und Kinder 98*). In order to acquire children's perspectives more directly, we use vehicles like children's workshops and implement child-appropriate methods of communication. Whenever feasible and sensible, child-level outcomes are included as central indicators.

3.2 Putting children at the centre of our work

vely simple blood test ("Yes", the disease is present or, "No", it is not present). Other times, things are not so obvious, especially in the social sciences. One example is the documentation of violence: Which actions count as violence and which do not? What dynamics are they embedded in? Are answers distorted by social desirability?

In recent years, the social sciences have made enormous progress towards more reliable, valid measurement of phenomena. An evaluation approach that unites qualitative and quantitative methods makes it possible to grasp the "big picture" and specific experiences simultaneously. It enables an overview and specific in-depth findings.

"What do I do if the evaluation shows that my programme doesn't work?"

Evaluation results are seldom just positive or negative. They often show that a programme is effective regarding certain desired outcomes and less so regarding others. Evaluations often provide valuable insights that can be used to strengthen and improve the programme. And if a programme doesn't work or partly misses its mark, isn't that something you would want to know?

"It's not ethical if some people are excluded from the programme"

Such concerns are occasionally expressed regarding evaluations with control groups (quasi-experimental studies or randomized control studies). They may be justified in certain cases.

However, there are often not enough resources to include all possible participants either way. Meanwhile, collecting information from non-participants (as a control) may eventually help to expand the programme at a later date.

If it is possible to stagger the uptake of programme participants, use of a waitlisted control group may represent a sensible option. Finally, it is worth noting that providing a service with unknown effectiveness, or possibly negative effects, is at least as ethically questionable as limiting people's access to a programme.

"I don't have the expertise to carry out an evaluation"

This is often true. The training and skills of a good practitioner are not the same as those of a good evaluator. And the more complex and scientifically rigorous an evaluation is, the more specialized are the knowledge requirements. A strong partnership between researchers and practitioners is thus the decisive factor behind a successful evaluation.

Source: Lannen, F., Jones, L., (in prep). Beyond Evaluation: Using Science to Accompany Program Development and Implementation. Chapter 2: Concerns about engaging in evaluation.

Common objections and concerns from practice and ways of addressing them:

“I know my programme works”

This impression may arise a) because participants react positively to a programme, b) based on specific cases of participants who were helped by a programme, c) because intuition and professional experience suggest that a programme is good. Research has shown, however, that intuitive assessment of effectiveness is often mistaken. It may be that programme feedback is distorted, as it comes mainly from people who like the programme. There may also be a “mix-up” between people’s enthusiasm to participate in a programme versus actual effectiveness.

One example of this is DARE, a drug prevention programme in the USA, which was implemented for many years before being evaluated. The evaluation showed that the programme was ineffective, with deterrence proving to be an insufficient factor for successful prevention. Instead, peer influences were shown to be much more decisive and necessary to incorporate in the programme. However, an evaluation may also serve to confirm and underpin one’s intuitive assessment, providing a strong argument for continuation and corresponding financing.

“Evaluation is expensive and takes a lot of time. Wouldn’t it be better for the resources to benefit the children directly?”

This argument is very easy to understand: it’s difficult to find money for evaluations, and it is even harder when there are already funding constraints for programme implementation to benefit children and families. At the same time, it appears problematic – perhaps even unethical – to provide a service without knowing whether it has the desired effect, or might even do the opposite and cause harm. Again, the example of DARE is illustrative: Billions of dollars were invested in the programme before it was discovered to be ineffective, that is, participants remained without suitable support.

“What we do is not linear and not measurable”

Measuring human and interpersonal phenomena via social science is frequently less clear and straightforward when compared with other disciplines. In medicine, for example, one can often clearly determine the presence of a disease with a rela-

between the aims of practitioners (implementers and commissioning agencies) and evaluators, secures implementation of the evaluation and maximizes its usefulness, aids in preventing misunderstandings, and assists in clarifying the potential and limitations of different evaluation designs (see also Section 4 on Impact Models and Section 5 on Choice of Evaluation Design). Other key topics of discussion include how research independence can be ensured, what unexpected – or even undesired – results could mean for the practice partner, and how the latter should be dealt with.

In addition, possible reservations about evaluation will be discussed in the first workshop (see Box 1). Here, an MII researcher will assume the role of “Research Broker”; She/he will be someone who can translate between the languages and approaches of science versus practice, thereby setting in motion the intersection between research and practice.

Further, the first workshop will also address the project partner’s understanding of impact with regard to the relevant programme, so that appropriate tools can be selected and eventually implemented (see Section 4 on Impact Models). Before completing the final report, the results of the evaluation will be contextualized together with the practice partner and, where appropriate, corresponding recommendations will be discussed. In the case of larger evaluation projects (lasting several years), interim workshops will be scheduled for presentation and discussion of ongoing results. Scientific independence and integrity are of the utmost importance in every evaluation carried out in partnership. They must be protected and, where necessary, clarified. During the research phase, interactions and informal exchanges will be avoided and strictly prohibited. An open outcome including various possible or unexpected results must be guaranteed at all times.

proactive contribution to the discourse on impact orientation and impact measurement of programmes. The evaluation concept is a tool for capacity strengthening towards an evidence-based practice and provides a foundation for corresponding assignments, projects, workshops, trainings, and publications.

Science

The evaluation concept forms the starting point for active participation in strengthening the evidence base and our knowledge of effective intervention components. In addition, it should stimulate exchange across social science disciplines and contribute to positioning the importance of practice-oriented research and implementation science.

3. Concept Building Blocks

The following "building blocks" reflect the core elements of MMi's mission and research concept, describing their application to MMi's evaluation activities. They represent the heart of the concept.

3.1 Participatory: with and for practice

This building block reflects MMi's fundamental participatory stance. Further, it illustrates MMi's function at the intersection of research and practice. Central to the evaluation concept are a participatory approach and a strong, functioning, trust-based partnership between the practitioner (implementer) and the researchers at MMi. To this end, evaluation goals, core questions, and the evaluation design are collaboratively developed in a proposal-preparation workshop. This helps to ensure an optimal fit

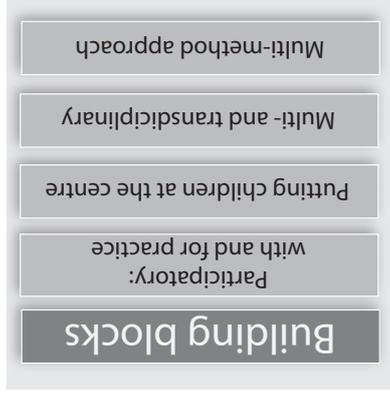


Figure 1:
Building blocks of the evaluation concept

1. Introduction

The MMI evaluation concept follows a systematic approach and reflects core elements of the MMI mission statement (annex) and the MMI research concept (annex). It combines these with systematic analysis and documentation of the field, and synthesizes experiences and expertise from previous evaluations.

The concept is comprised of individual building blocks and describes tools that are implemented in the framework of MMI evaluations.

The last section illustrates how use of learnings from evaluation are optimized as part of proactive knowledge management.

The approach places significant demands on researchers in terms of expertise, access to a broad toolbox of social science methods, and experience in programme development and implementation.

2. Purpose of the Concept

The evaluation concept pursues goals on different levels:

MMI internal

The evaluation concept makes explicit the collective knowledge and mutual understanding related to the implementation of evaluations, unifies the level of knowledge within the organization, and makes this knowledge basis available to all (new) professionals. It is a quality assurance instrument and may be continuously developed within a framework of knowledge transfer and knowledge management.

Practice

The concept focuses on a key niche in the field and strengthens MMI's position as a centre of excellence for evaluations at the intersection of science, practice, and society. It provides a foundation for

Contents

1	Introduction	2
2	Purpose of the Concept	3
3	Concept Building Blocks	4
3.1	Participatory: with and for practice	4
3.2	Putting children at the centre of our work	6
3.3	Multi- and transdisciplinary	6
3.4	Multi-method approach	6
4	Impact Models	6
4.1	An expanded model	6
4.2	Development of the model	8
4.3	Stumbling blocks	9
5	Choice of Evaluation Design	11
6	Knowledge Management	14

Patricia Lannen, August 2018

Marie Meierhofer Children's Institute

Framework for programme evaluations

Published by:

Marie Meierhofer Children's Institute

Author:

Dr. phil. Patricia Lannen

Cover and illustrations:

Anna-Lea Guarisco

Layout/Design:

Claudius Natsch

Print:

Borer Druck AG, Laufen

Distribution:

Marie Meierhofer Children's Institute

Phone:

+41 44 205 52 20

Fax::

+41 44 205 52 22

info@mmi.ch / www.mmi.ch

© mmi, 2018



SCIENCE

OUTPUT

OUTCOME

PRACTICE

IMPACT

FRAMEWORK FOR
PROGRAM EVALUATIONS