



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Wirtschaft, Bildung und Forschung WBF
Staatssekretariat für Wirtschaft SECO

Die Wirkung von arbeitsmarktlichen Massnahmen

Eine Analyse bisheriger Evaluationen

*mit einer deutsch- und französischsprachigen
Zusammenfassung*

Studie im Auftrag der
Aufsichtskommission für den
Ausgleichsfonds
der Arbeitslosenversicherung

B,S,S. Volkswirtschaftliche Beratung

**Michael Morlok
David Liechi
Nathanael Moser
Mirjam Suri**

Wissenschaftliche Begleitung:

**Patrick Arni
Aderonke Osikominu**

B , S , S .

VOLKSWIRTSCHAFTLICHE BERATUNG

Die Wirkung von arbeitsmarktlichen Massnahmen

Eine Analyse bisheriger Evaluationen

B,S,S. Volkswirtschaftliche Beratung

in Kooperation mit

Prof. Dr. Aderonke Osikominu

Prof. Dr. Patrick Arni

Basel, den 24.8.2018

Die Wirkung von arbeitsmarktlichen Massnahmen

Eine Analyse bisheriger Evaluationen

zuhanden des Staatssekretariats für Wirtschaft, Direktion für Wirtschaftspolitik

Projektteam seitens Auftragnehmer: Michael Morlok, David Liechti, Nathanael Moser, Mirjam Suri

Wissenschaftliche Begleitung: Patrick Arni, Aderonke Osikomunu

Verantwortlich seitens Auftraggeber: Katharina Degen

Begleitgruppe: Mario Ceppi (SECO), Katharina Degen (SECO), Gabriel Fischer (Travailsuisse), Martin Karrer (AWA Zürich), Dieter Kläy (SGV), Daniel Lampart (SGB), Stefan Leist (SECO), Daniella Lützelschwab (SAV), Simon Röthlisberger (SECO)

B,S,S. Volkswirtschaftliche Beratung AG, Aeschengraben 9, CH-4051 Basel

Tel: 061-262 05 55, Fax: 061-262 05 57, E-Mail: contact@bss-basel.ch

Inhaltsverzeichnis

Abkürzungsverzeichnis	ii
1. Einleitung.....	1
2. Methodik.....	3
3. Analyse der AMM-Wirkung.....	8
3.1. Deskriptive Untersuchung: Inhaltliche Aspekte	11
3.2. Deskriptive Untersuchung: Methodische Aspekte.....	19
3.3. Regressionsanalyse	33
3.4. Erkenntnisse aus weiteren Studien.....	37
3.5. Empfehlungen.....	41
3.6. Wirkungskanäle	43
4. Fazit.....	48
4.1. Inhaltliche Erkenntnisse.....	48
4.2. Methodische Erkenntnisse	50
4.3. Empfehlungen.....	56
5. Anhang.....	58
5.1. Literaturverzeichnis	58
5.2. Auszug Erfassungsraster.....	67
5.3. Workshop Liste der Teilnehmenden.....	73

Abkürzungsverzeichnis

AHV	Alters- und Hinterlassenenversicherung
AK ALV	Aufsichtskommission für den Ausgleichsfonds der Arbeitslosenversicherung
ALK	Arbeitslosenkasse
AMA	Amt für den Arbeitsmarkt
AMM	Arbeitsmarktliche Massnahmen
APO	Andere persönlichkeitsorientierte Kurse
ASAL	Auszahlungssysteme der Arbeitslosenkassen
AVAM	Informationssystem für die Arbeitsvermittlung und die Arbeitsmarktstatistik
AWA	Amt für Wirtschaft und Arbeit
DIHA	Dienststelle für Industrie, Handel und Arbeit
EAZ	Einarbeitungs- und Ausbildungszuschuss
EFK	Eidgenössische Finanzkontrolle
IT	Informationstechnik
IV	Instrumentalvariablen
k.A.	Keine Angabe
KIGA	Kantonales Amt für Industrie, Gewerbe und Arbeit
LAM	Logistikstellen für arbeitsmarktliche Massnahmen
LAMDA	Informationssystem des SECO für die Analyse von Arbeitsmarktdaten
ORP	Office régional de placement
PB	Personalberater/in
PvB	Programme zur vorübergehenden Beschäftigung
SAV	Schweizerischer Arbeitgeberverband
RAV	Regionales Arbeitsvermittlungszentrum
SDL	Sezione del lavoro
SECO	Staatssekretariat für Wirtschaft
SEVAL	Schweizerische Evaluationsgesellschaft
TN	Teilnehmende
VSAA	Verband Schweizerischer Arbeitsmarktbehörden
WiAMM	Wirkungsmessung Arbeitsmarktliche Massnahmen
ZV	Zwischenverdienst

Management Summary

Die vorliegende Studie verfolgt drei Ziele: Erstens sollen alle bisherigen Evaluationen von arbeitsmarktlichen Massnahmen (AMM) identifiziert werden. Zweitens sollen die Ergebnisse der Evaluationen zusammengefasst und drittens Empfehlungen hinsichtlich der AMM sowie der Evaluationsmethoden formuliert werden.

Der Fokus der statistischen Metaanalyse liegt auf den 23 Evaluationen, die die Wirkung der AMM mit einer Vergleichsgruppe oder -situation bestimmen. Die Analyse zeigt, dass zwei Drittel dieser Evaluationen die von ihnen untersuchten AMM positiv bewertet haben. Das weitere Drittel ermittelte entweder keine Wirkung oder kam zum Schluss, dass die AMM die Stellensuchdauer verlängert. Weitere 33 Evaluationen nutzen keine Vergleichsgruppe oder -situation oder zumindest keine, welche die Situation ohne AMM abbildet. Diese Studien weisen häufig keine Bewertung der Gesamtwirkung der AMM auf; insgesamt ist aber ein höherer Anteil positiver Bewertungen unter diesen Studien zu finden.

Der Wunsch, aus den bisherigen Evaluationen generelle Handlungsanweisungen für das Design und die Anwendung von AMM ableiten zu können, kann durch die vorliegende Studie nur teilweise erfüllt werden: Je nach Ausgestaltungsform, Teilnehmerschaft, und Verfügungszeitpunkt kann ein konkretes Angebot mehr oder weniger wirkungsvoll sein. Der Umstand, dass für alle mehrfach evaluierten AMM-Typen sowohl positive wie auch negative Evaluationsresultate vorliegen, signalisiert eine grosse Heterogenität in der Wirkung.

Um zu ermitteln, welche AMM für welche Person in welcher Situation wirksam sind, müsste diese Heterogenität besser verstanden werden. Die Studie kann somit nicht als Entscheidungsgrundlage für die Intensivierung oder Reduzierung der Nutzung bestimmter AMM-Typen dienen. Die Resultate der vorliegenden Studie können aber von den Spezialisten der kantonalen Arbeitsmarktbehörden als Diskussionsgrundlage genutzt werden, um die AMM-Palette im Kanton sowie deren Nutzung genauer zu untersuchen. Die Analyse zeigt beispielsweise, dass Beschäftigungsprogramme, Coachingangebote und der Zwischenverdienst in Evaluationen überdurchschnittlich gut abschneiden. Bei den Basisprogrammen und Fachkursen hingegen weist jede zweite Evaluation ein negatives Evaluationsresultat auf.

Die Evaluationen haben weiter gezeigt, dass AMM bei Stellensuchenden mit schlechten Arbeitsmarktchancen bessere Resultate aufzeigen als bei solchen, die ein hohes Wiedereingliederungspotenzial haben. Eine zusätzliche Gemeinsamkeit ist, dass Evaluationen wiederholt zu erhöhter Individualisierung, zu mehr Kommunikation mit Stakeholdern sowie zu mehr Arbeitsmarktnähe geraten haben.

Hinsichtlich der Evaluationsinstrumente ist festzuhalten, dass sich je nach Situation und Erkenntnisinteresse andere Instrumente anbieten. Folgende drei Instrumente eignen sich unseres Erachtens besonders gut für die regelmässige Anwendung durch die kantonalen Arbeitsmarktbehörden: Monitoring der Abmeldequoten (auf Ebene Angebote), Monitoring der Kompetenzentwicklung (auf Ebene Teilnehmende) sowie die Wirkungsmessung neu pilotierter AMM durch eine zufällige Zuteilung der Teilnehmenden.

Auf Basis der ermittelten Informationen empfiehlt das Forschungsteam:

- AMM, die sich an Stellensuchende mit schlechten Arbeitsmarktchancen richten, zu priorisieren
- Den Individualisierungsgrad, die Kommunikation mit Stakeholdern sowie die Arbeitsmarktnähe der AMM regelmässig zu überprüfen
- Die Bereitstellung und Anwendung von AMM-Typen, deren Evaluationen viele negative Resultate beinhalten, die teuer sind oder häufig angewandt werden, besonders sorgfältig zu prüfen
- In künftigen Evaluationen stärker auf Wirkungskanäle zu fokussieren
- Evaluationen nach Möglichkeit so auszugestalten, dass ihre Ergebnisse mit jenen anderer Studien vergleichbar sind (u.a. wenn immer möglich eine formelle Vergleichsgruppe resp. -situation zu verwenden)
- Evaluationsresultate zwischen Kantonen / Bund stärker miteinander zu teilen, zu kommunizieren und gemeinsam zu interpretieren
- Evaluationen so zu koordinieren und zu gestalten, dass neue Erkenntnisse gewonnen werden. Dazu könnte eine gemeinsame AMM-Evaluationsagenda formuliert werden. In deren Erarbeitung könnte auch geprüft werden, ob eine erhöhte Frequenz oder Bearbeitungstiefe von Evaluationen zu beschleunigten und detaillierteren Erkenntnisgewinnen führen.

Synthèse

La présente étude poursuit trois objectifs. En premier lieu, il s'agit d'identifier toutes les évaluations qui ont été réalisées à ce jour concernant les mesures relatives au marché du travail (MMT). Deuxièmement, le but de l'étude est de résumer les résultats de ces évaluations. Enfin, l'étude vise à formuler des recommandations au sujet des MMT et des méthodes d'évaluation.

La méta-analyse statistique se concentre sur les 23 évaluations qui démontrent l'effet des MMT à l'aide d'un groupe ou d'une situation de comparaison. L'analyse montre que deux tiers de ces évaluations ont jugé les MMT qu'elles ont examinées comme étant positives. Le tiers restant a conclu que les MMT soit n'ont pas eu d'effet, soit ont prolongé la durée de la recherche d'un emploi. Trente-trois autres évaluations n'ont utilisé aucun groupe ou aucune situation de comparaison, ou du moins aucun groupe ou situation ne montrant quel aurait été le résultat sans recours à une MMT. Ces études ne présentent souvent aucune évaluation de l'effet global des MMT ; néanmoins, elles sont dans l'ensemble plus nombreuses à juger les MMT positivement.

La présente étude ne comble qu'en partie le souhait de pouvoir déduire, à partir des évaluations effectuées jusqu'à présent, des instructions générales sur la conception et l'utilisation des MMT. En effet, une offre peut être plus ou moins efficace selon sa forme, ses participants ou sa disponibilité. Le fait que les évaluations aient abouti à des résultats tantôt positifs, tantôt négatifs pour tous les types de MMT examinés plusieurs fois montre que l'effet des MMT est très hétérogène.

Afin de déterminer quelles MMT sont efficaces pour quelle personne dans quelle situation, il faudrait mieux comprendre cette hétérogénéité. La présente étude ne peut donc pas constituer une base de décision pour intensifier ou réduire le recours à certains types de MMT. Mais les résultats de l'étude peuvent servir de base de discussion aux spécialistes des autorités cantonales du marché du travail afin d'examiner plus en détail la palette des MMT proposées dans leur canton et leur utilisation. L'analyse montre par exemple que les programmes d'emploi, les offres de coaching et le gain intermédiaire sont jugés de manière particulièrement positive dans les évaluations. En revanche, une évaluation sur deux se conclut par des résultats négatifs en ce qui concerne les programmes de base et les cours spécialisés.

Les évaluations entreprises ont aussi montré que lorsqu'elles sont suivies par des demandeurs d'emploi dont les chances sur le marché du travail sont faibles, les MMT ont des résultats meilleurs que lorsqu'elles s'adressent à des demandeurs d'emploi présentant un fort potentiel de réinsertion. Un autre point commun entre les évaluations est le fait qu'elles ont préconisé de façon répétée une plus grande

individualisation, davantage de communication avec les parties prenantes et une plus grande proximité avec le marché du travail.

En ce qui concerne les instruments d'évaluation, il convient de relever que des instruments différents s'imposent en fonction de la situation et du résultat visé. À notre avis, les trois instruments suivants sont particulièrement bien adaptés pour une utilisation régulière par les autorités cantonales du marché du travail : le monitoring des taux de désinscription (niveau des offres), le monitoring de l'évolution des compétences (niveau des participants) et la mesure des résultats de MMT soumises à un nouveau pilote avec assignation au hasard des participants.

Sur la base des informations récoltées, l'équipe de recherche recommande :

- de mettre la priorité sur les MMT qui s'adressent aux demandeurs d'emploi qui ont de moins bonnes chances de réinsertion dans le marché du travail ;
- d'examiner régulièrement les MMT sous les angles de leur degré d'individualisation, de la communication avec les parties prenantes et de leur proximité avec le marché du travail ;
- d'examiner minutieusement la mise à disposition et l'utilisation des types de MMT dont les évaluations ont conclu à des résultats négatifs, qui sont onéreux ou qui sont utilisés souvent ;
- de mettre davantage l'accent sur les canaux d'influence dans les prochaines évaluations ;
- de concevoir autant que possible les évaluations de manière à ce que leurs résultats soient comparables avec ceux d'autres études (notamment en utilisant si possible un groupe ou une situation formel/-le de comparaison) ;
- d'envisager une plus grande collaboration entre les cantons et la Confédération pour partager, communiquer et interpréter les résultats des évaluations ;
- de coordonner et concevoir les évaluations de manière à obtenir de nouveaux constats. On pourrait formuler un agenda commun pour l'évaluation des MMT. Lors de son élaboration, on pourrait aussi examiner si une augmentation de la fréquence ou du niveau de traitement des évaluations entraînerait des résultats plus rapides et plus détaillés.

1. Einleitung

Die arbeitsmarktlichen Massnahmen (AMM) wurden 1996 eingeführt. Wenig später (2000/2001) wurde ihre Wirkung in einer ersten vom SECO in Auftrag gegebenen „Evaluationswelle“ analysiert. Weitere Wellen folgten 2004-2006 und 2012-2014. Das SECO erstellte Syntheseberichte, welche die wichtigsten Erkenntnisse aus den Studien zusammenfassten.¹ Es liegen zudem verschiedene internationale Übersichtsstudien vor, in die auch Studien aus der Schweiz eingeflossen sind.² Was hingegen fehlt, ist eine vollständige Übersicht zu den Schweizer Studien, welche auch Evaluationen der Kantone oder weiterer Auftraggeber berücksichtigt.

Ziel der vorliegenden Studie ist es, diese Lücke zu schliessen oder zumindest zu reduzieren. Die Studie soll alle Evaluationen identifizieren und die ermittelten Resultate systematisch erfassen. Anschliessend sollen Schlussfolgerungen und Empfehlungen zu den AMM sowie den Evaluationsmethoden formuliert werden.

Das Netz wurde dazu möglichst weit ausgeworfen: Das Projektteam kontaktierte die zuständigen Abteilungen der kantonalen Arbeitsmarktbehörden, Evaluationsteams, Forschungsinstitute an Universitäten und Fachhochschulen sowie Stiftungen. Insgesamt konnten so rund 56 Evaluationen und Studien identifiziert werden; deutlich mehr als vermutet. 23 dieser Berichte beinhalteten Schätzungen zur AMM Wirkung, die miteinander vergleichbar sind und so für eine Metaevaluation genutzt werden können.

Einschränkend soll zu Beginn vermerkt sein: AMM, AMM-Teilnehmende, der Arbeitsmarkt und die Evaluationsmethoden haben sich über die letzten 20 Jahre teilweise merklich verändert. Die Zahl der Schweizer Evaluationen ist hingegen zu tief, als dass man nur auf die neusten Studien fokussieren könnte; entsprechend flossen auch die älteren Studien in die Analyse ein. Weiter enthalten die Evaluationen und Studien vielfältige Erkenntnisse und Informationen, die in der vorliegenden Übersichtsanalyse nicht oder jedenfalls nicht abschliessend verarbeitet werden können. Die vorliegende Studie fokussiert auf die Resultate zur Wirkung, auf Wirkungskanäle und die in den Evaluationen präsentierten Empfehlungen.

Im Verlaufe der Erarbeitung der Studie zeigte sich die grosse Heterogenität der identifizierten Evaluationen und ihren Ergebnissen. Durch diese Heterogenität war es nicht möglich, eindeutige Aussagen zur Wirkung einzelner AMM-Typen zu erzie-

¹ SECO 2015, Ragni 2007.

² Neben Card et al. 2015 und 2010 siehe auch Vooren et al. 2017, Kluge 2016 sowie Özge 2015.

len. Die Studie kann jedoch erste Hinweise geben, welche AMM-Typen resp. Verfügungssituationen weiter geprüft werden müssten. Zudem bietet sie erstmalig einen Überblick über alle Wirkungsmessungen von AMM in der Schweiz und kann damit als Übersichtswerk und zur Orientierung bei der Entwicklung von AMM resp. Evaluationen im AMM-Bereich verwendet werden. Auch wenn mit der vorliegenden Studie nicht alle offenen Fragen beantwortet werden können, erlaubt sie doch eine bessere Übersicht darüber, welche Fragestellungen noch offen bleiben und wie sie allenfalls in Zukunft beantwortet werden könnten.

Der Bericht ist wie folgt aufgebaut: Kapitel 2 erläutert das methodische Vorgehen. Die Resultate der Analysen werden in Kapitel 3 dargestellt und diskutiert. Die Schlussfolgerungen und Empfehlungen werden in Kapitel 4 besprochen.

2. Methodik

Die vorliegende Studie beinhaltet drei hauptsächliche Arbeitsschritte, die im Folgenden kurz beschrieben werden. Das Vorgehen orientiert sich an der Metaevaluation von Card et al. 2015, wobei deren Methodik für den Schweizer Kontext adaptiert wurde: Insbesondere wurden die Selektionskriterien geöffnet, um mehr Evaluationen berücksichtigen zu können.

1. Identifikation und Erfassung der Studien

Um eine möglichst vollständige Übersicht aller Evaluationen zu schaffen, wurden folgende Stellen resp. Quellen konsultiert:

- Kantonale Ämter für Wirtschaft und Arbeit resp. derer mit der Organisation der AMM beauftragten LAM-Stellen
- Die an den bisherigen Evaluationen beteiligten Forschungsteams
- Die in den entsprechenden Bereichen forschenden Abteilungen der Universitäten und Fachhochschulen³
- Ausgewählte Stiftungen⁴
- Literaturverzeichnisse bereits bekannter Studien

Die so identifizierten Studien wurden gesichtet und thematisch abgegrenzt. Ausgewählt wurden alle Evaluationen, die a) von der Arbeitslosenversicherung finanzierte Massnahmen untersuchen und b) sich schwerpunktmässig mit deren Wirkung auseinandersetzen. Daraus resultierte eine Liste mit 56 Studien.

Für die statistischen Analysen (Kapitel 3.1 bis 3.3) müssen die Studien zusätzliche Anforderungen erfüllen (siehe unten). Von den 56 Studien erfüllen 23 diese Anforderungen. Sie stellen die Basis für die vertieften statistischen Analysen dar. Die restlichen 33 Studien wurden ebenfalls gesichtet, aber nur hinsichtlich ausgewählter Aspekte (Empfehlungen, Wirkungskanäle; siehe zudem Abschnitt 3.4. zu den Erkenntnissen aus diesen weiteren Studien) ausgewertet.

³ Fachbereiche: Wirtschaftswissenschaften, Soziologie, Psychologie, Politikwissenschaften, Soziale Arbeit. Zusätzlich dazu wurden die Pädagogischen Hochschulen kontaktiert.

⁴ Migros Kulturprozent, Paul Schiller Stiftung, Stiftung Mercator, Volkart Stiftung.

Anforderungen für die Metaevaluation

Folgende Bedingungen müssen die Evaluationen / Studien erfüllen, um in die Metaevaluation einzufließen:

- Es wird die beobachtbare Situation mit AMM-Besuch mit der nicht beobachtbaren, hypothetischen Situation ohne AMM-Besuch verglichen (sog. *Counterfactual*). So werden beispielsweise Studien, die die Vermittlungsquote der Teilnehmenden ohne eine Vergleichsgruppe untersuchen oder ausschliesslich mit den Teilnehmenden anderer Massnahmen vergleichen, ausgeschlossen.
- Die Wirkungsmessung basiert auf Indikatoren, die sowohl positive wie negative Werte annehmen können. So werden beispielsweise Evaluationen, die ausschliesslich auf Fragen wie: „Fanden Sie AMM x nützlich? 0: nicht nützlich, 5: sehr nützlich“ oder „Wurde das Ziel der AMM erreicht?“ basieren, aus der statistischen Analyse ausgeschlossen.
- Es wurden nur Evaluationen aufgenommen, bei denen nachvollzogen werden kann, ob die beiden zuerst genannten Kriterien erfüllt sind (z.B. summarische Bewertungen werden nicht aufgenommen).

Diese Ausschlusskriterien stellen keine Bewertung der Studien dar: Auch Studien ohne die oben genannten Eigenschaften sind in vielen Situationen wertvoll und erfüllen das Erkenntnis- und Verwertungsinteresse der jeweiligen Auftraggeber/innen. Diese Studien eignen sich aber nicht für einen direkten Vergleich. Insbesondere ist nicht eindeutig bestimmbar, was einen akzeptablen (positiven) und nicht akzeptablen (negativen) Wert darstellt; je nach AMM und Teilnehmerpopulation verschieben sich diese Werte.

Für die Erfassung der Studien wurde das nachfolgend abgebildete Raster verwendet. Ein Auszug aus dem Erfassungsraster ist im Anhang 6.2 zu finden.

Tabelle 1 Übersicht Erfassungsraster

Studie	AMM	Methodik	Wirkung	Reliabilität	Kontext	Bewertung
Autor/in	Typ	Datenerhebung	Wirkung	Daten	Zeitraum	Empfehlungen
Auftraggeber/in	Merkmale TN	Indikatoren	Wirkungskanäle	Signifikanz	Region	
Jahr	Kosten	Auswertungsverfahren	Ergebnisse Subgruppen	Stichprobengrösse		
			Interaktions-/ Nebeneffekte			

Die im Raster dokumentierten Merkmale bilden die Basis für die weiteren Auswertungen. Die Merkmale Kosten und Interaktions- resp. Nebeneffekte wurden derart selten erfasst, dass sie im Bericht nicht weiter diskutiert werden.

2. Auswertung

Im Zentrum der statistischen Analyse stehen die in den Studien ausgewiesenen Schätzungen zur Wirkung der AMM. Wurden in einer Studie mehrere Schätzungen ausgewiesen (z.B. für mehrere AMM-Typen oder Subgruppen der Teilnehmenden) wurden diese einzeln erfasst. Für die Erfassung der Schätzungen wurden zunächst die relativen Effektgrößen (Wirkung in %) erfasst. Diese wurden entweder explizit in den Studien dokumentiert oder durch das Studienteam berechnet.⁵

Um die verschiedenen Angaben zur Wirkung vergleichbar zu machen, wurde in Anlehnung an Card et al. 2015 folgende Kategorisierung vorgenommen: Das Vorzeichen wurde in drei Kategorien „positiv“, „negativ“ und „kein Effekt“ eingeteilt, wobei kein Effekt einer Wirkung von weniger als ± 0.5 % entspricht.⁶ Diese Vereinfachung erlaubt eine Vergleichbarkeit über alle aufgenommenen Studien. Eine zweite Kategorisierung berücksichtigt zudem die statistische Signifikanz.⁷

Zusätzlich zu den deskriptiven Analysen wurde eine Ordered Probit Regression geschätzt. Dieses Auswertungswertungsverfahren eignet sich bei abhängigen Variablen, die wie im vorliegenden Fall ordinalskaliert sind (positive Wirkung, keine Wirkung, negative Wirkung). Mit der Regression lässt sich bestimmen, wie sich die Wahrscheinlichkeit, dass in der Evaluation eine positive AMM-Wirkung beobachtet wird, ändert, wenn ein bestimmtes Merkmal der AMM oder der Studie vorliegt. Aufgrund der tiefen Fallzahlen und der wenigen signifikanten Ergebnisse sind die Resultate mit Vorsicht zu interpretieren.

Alle 56 Studien wurden schliesslich auf ihre Empfehlungen und die diskutierten Wirkungskanäle untersucht. Die Diskussion der Methoden(elemente) (Kapitel 3 und 4) basiert auf der wissenschaftlichen Literatur zu Wirkungsmessungsmethoden, der

⁵ Dabei haben wir die relative Wirkung als Verhältnis der absoluten Wirkung zum bei der Kontrollgruppe oder in der Vergleichssituation beobachteten Indikatorwert berechnet (oder uns basierend auf den vorhandenen Informationen bestmöglich diesem Verhältnis genähert).

⁶ Es wäre möglich gewesen, diese Schwelle höher anzusetzen (z.B. bei 3 % oder 5 % relativer Wirkung). Diese Einteilung wäre aber arbiträrer als eine simple Rundung gewesen.

⁷ Eine Analyse der (standardisierten) Effektstärken, wie sie ebenfalls von Card et al. 2015 vorgeschlagen wird, liess sich für die Schweiz aufgrund einer zu kleinen Anzahl verwertbarer Studien nicht sinnvoll durchführen.

Bewertung der jeweiligen Studienautor/innen selbst (sofern vorhanden), der Erfahrung des Projektteams sowie auf Inputs aus dem Validierungsworkshop (siehe nächstes Teilkapitel).

Studien mit mehreren Schätzungen

Damit Studien mit vielen Schätzungen in der Auswertung nicht mehr Gewicht erhalten als solche mit wenigen, wurde anhand der folgenden Kriterien (in abnehmender Wichtigkeit) für jede Studie eine primäre Schätzung ausgewählt:

- Die umfassendste Schätzung (Gesamtergebnisse über alle AMM / Gruppen, sofern Ergebnisse für mehrere Angebote / Teilnehmendengruppen dokumentiert wurden)
- Der langfristigste Indikator (sofern mehrere Indikatoren untersucht wurden, z.B. Anteil Teilnehmende mit einer Beschäftigung nach 6 anstatt nach 3 Monaten)
- Das von den jeweiligen Studienautoren präferierte methodische Verfahren (sofern mehrere Verfahren genutzt wurden und eines klar bevorzugt wurde)

Falls in einer Studie mehr als eine Schätzung diese Kriterien erfüllen, wurde so gewichtet, dass jede Studie insgesamt mit dem gleichen Gewicht in die Auswertung einfließt. Wenn eine Studie beispielsweise Schätzungen für mehrere Massnahmen ausweist (ohne eine durchschnittliche Wirkung darzustellen) fließen alle Schätzungen einzeln, jedoch gewichtet in die Analyse ein.⁸ Es wurde keine weitere Gewichtung (beispielsweise Stichprobengrösse, wahrgenommene Qualität der Studie etc.) vorgenommen. Dies stellt einerseits sicher, dass die Analyse auf möglichst vielen Studien beruht, und andererseits, dass die Analyse nicht durch Präferenzen des Forschungsteams beeinflusst wird.

3. Validierung und Interpretation

Im Rahmen eines Workshops wurden die Resultate der Analysen vorgestellt und gemeinsam interpretiert. Am Workshop nahmen die Begleitgruppe des Projekts, Vertreter/innen der RAV und LAM-Stellen sowie des SECO, des VSAA und Arbeitsintegration Schweiz teil (siehe Anhang 6.3 für die Liste der Teilnehmenden).⁹ Die Rückmeldungen aus dem Workshop wurden dokumentiert und fließen in den Bericht an ausgewählten, hervorgehobenen Stellen ein.

⁸ Das Gewicht der Schätzungen wurde folgendermassen berechnet: $1 / \text{Zahl der Schätzungen der Studie}$. Dies stellt sicher, dass jede Studie insgesamt ein Gewicht von 1 erhält. Wenn eine Studie in mehreren Auswertungskategorien (z.B. mehrere Indikatoren bei der Auswertung nach Indikatoren) vorkommt wird die Gewichtung so angepasst, dass die Studie in jeder Auswertungskategorie das Gewicht von 1 erlangt.

⁹ Wir danken dem VSAA für organisatorische und logistische Unterstützung und den Teilnehmenden für die engagierten und spannenden Diskussionen.

4. Grenzen der Methodik

Die Zahl der untersuchten Studien ist insgesamt gering. Sobald die Auswertungen differenziert nach einzelnen Merkmalen der Studien (z.B. Untersuchungsperiode) oder der Schätzungen (z.B. AMM-Typen) vorgenommen werden, fließen bei einigen Ausprägungen dieser Merkmale nur noch wenige Schätzungen oder sogar nur eine einzige Schätzung in die Auswertung ein. Entsprechend sind sämtliche nachfolgende Ergebnisse mit Vorsicht zu interpretieren, resp. die jeweils zugrunde liegende Fallzahl zu beachten (diese ist jeweils direkt in den Grafiken abgebildet).

Um möglichst robuste Resultate zu besprechen, wurde weitgehend darauf verzichtet, Merkmale kombiniert auszuwerten (z.B. AMM-Typ kombiniert mit Teilnehmergruppen): Solche Kombinationen könnten nur basierend auf Einzelstudien diskutiert werden. Ebenso ist der Detaillierungsgrad, nach dem Studien und AMM-Typen klassifiziert werden aufgrund der erwähnten Restriktionen beschränkt.

Es besteht insgesamt ein Spannungsfeld zwischen dem Interesse, möglichst detaillierte Informationen über die Umstände der Wirkung von AMM wiederzugeben und dem Wunsch, eine Übersicht über sämtliche gesammelten Studien zu schaffen. Aufgrund der Tatsache, dass die Zahl der verfügbaren Studien (mit Vergleichsgruppe) klein ist und sich die untersuchten AMM hinsichtlich vieler Aspekte und Umstände unterscheiden, können detailliertere Eigenschaften von AMM nicht studienübergreifend untersucht werden. So kann nicht untersucht werden, welche Aussagen zur Wirkung für eine Ausgestaltungsform oder für eine spezifische Teilnehmerschaft gelten. Um eine möglichst grosse Datenbasis zu erhalten, wurden zudem alle Studien seit 2000 aufgenommen; die meisten Auswertungen berücksichtigen den Umstand nicht, dass die AMM im Verlauf der Zeit laufend überarbeitet und verbessert wurden (eine Ausnahme stellt die Auswertung nach Untersuchungsperiode dar). Schliesslich erschwert auch die Heterogenität in der beobachteten AMM Wirkung, übergreifende Erkenntnisse zu identifizieren. Im Rahmen der vorliegenden Metaanalyse fokussieren wir auf das Ziel, über alle Studien hinweg Aussagen machen zu können. Neben der Wirksamkeit einer AMM wären auch ihre Nutzung sowie die entstehenden Kosten zu beachten, um eine Aussage über ihre gesamthafte Zweckmässigkeit treffen zu können.

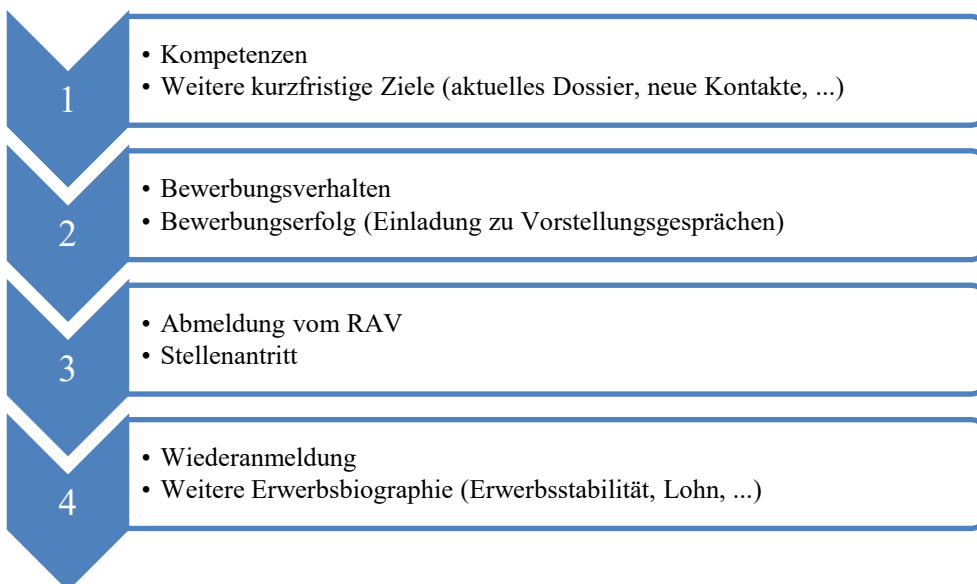
Aufgrund der erläuterten Herausforderungen ist es nicht möglich, mit den vorliegenden Angaben klare Empfehlungen zur schwächeren oder stärkeren Nutzung von einzelnen AMM-Typen zu formulieren. Die Studie ermöglicht aber einen Überblick über die Evaluationen, gibt diverse Hinweise zur Wirkung von AMM und liefert damit Orientierungshilfen für die AMM-Praxis und künftige Evaluationen.

3. Analyse der AMM-Wirkung

Im vorliegenden Kapitel wird die in den Studien und Evaluationen ermittelte AMM-Wirkung zunächst insgesamt und dann unterteilt nach verschiedenen Merkmalen der Studien (Untersuchungsperiode und regionale Abdeckung) sowie nach Merkmalen der Schätzungen (AMM-Typen, Teilnehmergruppen, Indikatoren, Informationsgrundlage, Stichprobengrösse, Vergleichsebene und Identifikationsstrategie) untersucht.

Dabei ist darauf hinzuweisen, dass nicht alle Studien dieselbe Ebene von AMM-Wirkung messen. Dies kann anhand einer einfachen, idealtypischen Wirkungskette illustriert werden (siehe Abbildung 1). AMM werden aus unterschiedlichen Gründen genutzt. Letztendlich sollen die AMM die teilnehmenden Stellensuchenden dabei unterstützen, sich rasch und dauerhaft (erneut) im Arbeitsmarkt zu integrieren. Personalberatende verfolgen mit einer AMM Verfügung allerdings immer auch kurzfristige Zwischenziele: Der Lebenslauf soll aktualisiert, das Suchfeld vergrössert oder die Sprachkenntnisse verbessert werden.

Abbildung 1 Wirkungskette AMM¹⁰



AMM können also darauf abzielen, die Kompetenzen der Stellensuchenden zu beeinflussen. Dies wiederum kann zu Veränderungen im Such- und Bewerbungsverhalten sowie bei den Bewerbungschancen führen. Kurz- bis mittelfristig kommt es

¹⁰ Darstellung übernommen aus Morlok und Liechti 2018.

somit zur Abmeldung vom RAV und längerfristig sollen bessere und stabilere Beschäftigungen erzielt werden. Die untersuchten Studien untersuchen unterschiedliche Glieder dieser Wirkungskette und verwenden dazu unterschiedliche Indikatoren. In den Auswertungen wird diese Unterscheidung der verschiedenen Wirkungsebenen vernachlässigt und die Indikatoren werden gemeinsam besprochen (ausser in der Auswertung der Indikatoren).¹¹

Die in den 23 statistisch erfassten Studien dargestellte AMM-Wirkung wurde auf zwei unterschiedliche Weisen codiert:

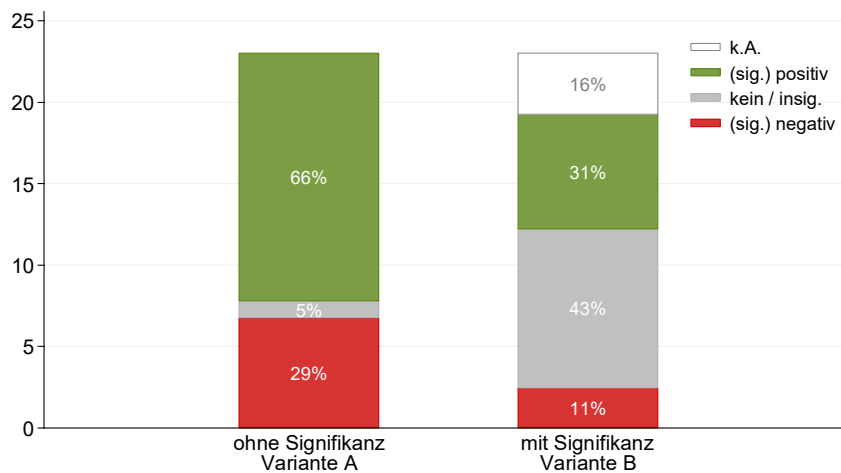
- *Variante A*: Einteilung in „negativ“, „positiv“ oder „kein Effekt“, je nach Vorzeichen des geschätzten Effekts. Werte unter ± 0.5 % wurden als „kein Effekt“ klassifiziert.¹²
- *Variante B*: Einteilung in „signifikant negativ“, „signifikant positiv“, „insignifikant“ oder „keine Angabe“ auf Grundlage der Signifikanz der dokumentierten Effekte (mindestens auf dem 90 %-Niveau).¹³ Unter „keine Angabe“ fallen die Studien, welche keine Signifikanz ausweisen.

¹¹ Das Wirkungsglied „Abmeldung vom RAV / Stellenantritt“ steht häufig im Zentrum der Diskussionen über die AMM-Wirkung. Auch die anderen Glieder haben aber eine wichtige Rolle in der Analyse. Beispielsweise kann mit den Wirkungsgliedern Kompetenzen und Bewerbungsverhalten resp. -erfolg besser verstanden werden, wie die Gesamtwirkung der AMM zustande kommt. Auf dieser Ebene ist die Wirkung auch unmittelbarer resp. weniger von anderen Dienstleistungen der RAV (Beratung etc.) sowie Einflüssen des Arbeitsmarkts abhängig; dies vereinfacht die Wirkungsmessung. Das letzte Wirkungsglied (Wiederanmeldung, weitere Erwerbsbiographie) schliesslich bildet die längerfristige Wirkung ab. Siehe auch Diskussion in Kapitel 3.2, „Bewertung Indikatoren“.

¹² „Positiv“ bezieht sich nicht auf das mathematische Vorzeichen, sondern auf die gewünschte Richtung des Effekts, z.B. Erhöhung Stellenantrittsrates, Verkürzung Stellensuchdauer.

¹³ Die statistische Signifikanz ist ein Mass für die Robustheit der Angabe. Ein Effekt wird als „statistisch signifikant“ bezeichnet, wenn es sehr unwahrscheinlich ist, dass die geschätzte Effektgrösse sich per Zufall merklich von Null unterscheidet.

Abbildung 2 Gesamteffekt



Hinweis: Die Auswertung basiert auf 23 Studien.

Bei Variante A liegen zwei Drittel (66 %) der Schätzungen im positiven Bereich, während ein knappes Drittel (29 %) nicht nur keinen sondern gar einen negativen Effekt ermittelte. Das Verhältnis von positiven zu negativen Schätzungen ist bei Variante B sehr ähnlich wie bei Variante A. Allerdings liegen bei Variante B fast zwei Drittel der Schätzungen in den Kategorien „insignifikant“ (43 %) oder „keine Angabe“ (16 %, siehe Abbildung 2). Um die Anteile positiver / negativer Resultate basierend auf möglichst vielen Studien zu berechnen, werden die weiteren Auswertungen mit Variante A durchgeführt.

Insgesamt kommen folglich zwei Drittel der statistisch ausgewerteten Studien zum Schluss, dass die Wirkung der AMM positiv ist.¹⁴ Es gibt aber auch eine nicht vernachlässigbare Zahl an Studien, die nicht nur keine Wirkung konstatieren, sondern sogar einen negativen Effekt ausweisen. Ein negativer Effekt wird meistens mit dem sogenannten Lock-in erklärt, d.h. es wird argumentiert, dass während einer AMM weniger oder weniger gute Bewerbungen verfasst werden (siehe weitere Ausführung in Kapitel 3.6).

Dieses Gesamtergebnis wirft weitere Fragen auf: Welche AMM-Typen konnten zu einem positiven Resultat beitragen und welche nicht? Bei welcher Teilnehmergruppe erzielt die AMM zu welchem Zeitpunkt eine gute Wirkung? Die unterschied-

¹⁴ In den Studien, welche nicht in die statistische Analyse miteinbezogen werden konnten, wurden mehrheitlich positive Resultate ermittelt (siehe Kapitel 3.4).

lichen Ausgangslagen müssen bei der Beurteilung miteinbezogen werden. Die folgenden Teilkapitel behandeln verschiedene Dimensionen der Studien und Schätzungen und stellen die Resultate differenziert nach diesen Dimensionen dar. Dies mit dem Ziel – soweit möglich – herauszufinden, welche Massnahme sich unter welchen Umständen in Bezug auf die untersuchte Zielgrösse als besonders wirksam erwiesen hat.

3.1. Deskriptive Untersuchung: Inhaltliche Aspekte

Erscheinungsjahr

Die Auswertung nach Erscheinungsjahr zeigt, dass pro Jahr maximal drei Evaluationen durchgeführt wurden (Abbildung 3). In mehreren Jahren wurde kein Bericht verfasst, oder zumindest keiner, welcher die Aufnahmekriterien für die statistischen Analysen erfüllte (siehe Kapitel 2).

Abbildung 3 *Erscheinungsjahr*



Hinweis: Die Auswertung basiert auf 23 Studien.

Lesebeispiel: Im Jahr 2002 erschien eine Studie (Gerfin und Lechner 2002). Dieser Studie wurden sieben Schätzungen entnommen, wobei vier negativ sind (57 %), zwei als ineffektiv eingestuft wurden (29 %) und eine zu einem positiven Ergebnis kam (14 %).

Werden die Studien in eine Periode vor 2007 und eine nach 2007 aufgeteilt (was der ersten und zweiten Dekade nach Einführung der AMM entspricht), zeigt sich, dass die Zahl der Studien weitgehend konstant ist: Vor 2007 wurden 13 Evaluationen erstellt, nach 2007 sind es 12.

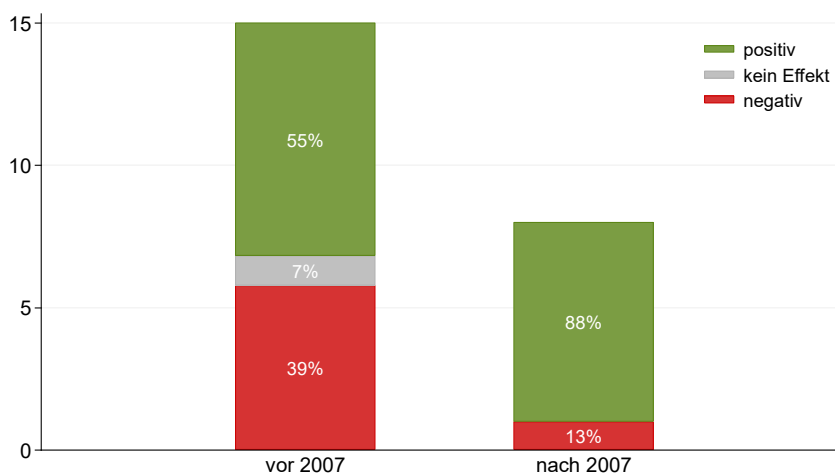
Der Anteil an positiven Schätzungen steigt über die Zeit. Studien mit einem Erscheinungsjahr nach 2007 kommen häufiger (9.3 von 12 Studien) zu einem positiven Resultat als Studien, die vor 2007 veröffentlicht wurden (6.0 aus 11 Studien). Zu beachten ist, dass das Erscheinungsjahr nur begrenzt darüber Aufschluss gibt, welche Beobachtungsperiode abgebildet wird. Wichtiger für eine Aussage dazu, ob die heutigen AMM wirksamer sind als die früheren, ist eine Analyse der Untersuchungsjahre.

Untersuchungsperiode

Es stellt sich die Frage, ob Massnahmen, welche kurz nach der Einführung der AMM eingesetzt wurden, andere Resultate erzielen als jene Massnahmen, die in der jüngeren Vergangenheit durchgeführt wurden. Konnten die AMM und die Verfügungspraxis über die Zeit verbessert werden?

Dazu wird das mittlere Untersuchungsjahr berechnet, als Mittelwert vom Beginn und Ende der Untersuchungsperiode. Erneut wurden die Studien in zwei Gruppen mit dem mittleren Untersuchungsjahr vor resp. nach 2007 eingeteilt (siehe Abbildung 4), also in die erste und die zweite Dekade nach Einführung der AMM im Jahr 1996.

Abbildung 4 *Untersuchungsperiode*



Hinweis: Die Auswertung basiert auf 23 Studien.

Es fällt zunächst auf, dass zwei Drittel der Schätzungen Daten verwenden, deren mittlere Untersuchungsperiode vor 2007 liegt.¹⁵ Evaluationen, bei denen das mittlere Untersuchungsjahr nach 2007 liegt, schneiden besser ab. Man könnte dies so interpretieren, dass die angebotenen AMM über die Zeit ihre Wirkung steigern konnten, oder dass das AMM-Portfolio optimiert wurde, indem wenig wirksame Massnahmen eingestellt wurden. Hier ist zu erwähnen, dass dieses Resultat durch die Regression, welche verschiedene weitere Faktoren berücksichtigt (z.B. Verschiebungen in den untersuchten AMM oder verwendeten Methoden), nicht bestätigt wird (siehe Kapitel 3.3). Die Ergebnisse der Regression deuten auf eine negative Entwicklung hin, wobei dieses Resultat nicht statistisch signifikant ist und daher zufällig zustande gekommen sein mag. Zu erwähnen ist zudem, dass weitere Einflussfaktoren, insbesondere die Veränderung der konjunkturellen Lage, einen Einfluss auf die Ergebnisse unterschiedlicher Untersuchungsperioden haben können.

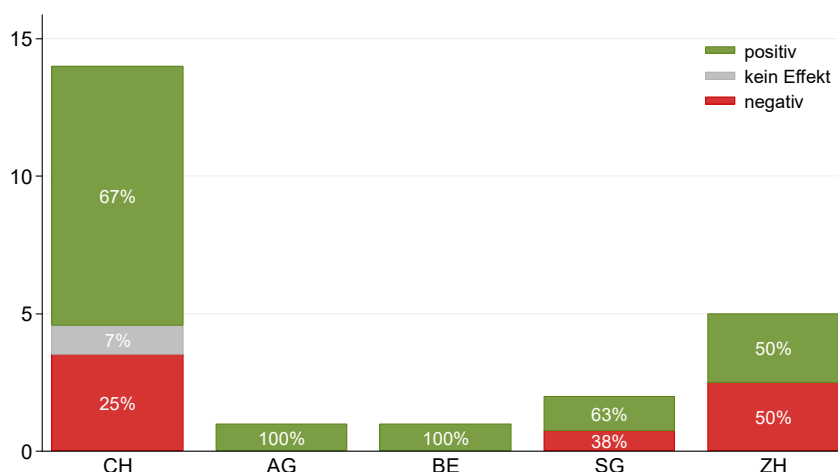
Regionale Abdeckung

In einem nächsten Schritt wird die Verteilung der Studien auf die Regionen/Kantone untersucht. Dies einerseits um zu ermitteln, ob sich die beschriebenen Erkenntnisse auf die ganze Schweiz übertragen lassen und andererseits um allfällige regionale Unterschiede in der Effektivität bei der Durchführung von AMM aufzuspüren.

Studien, die nicht explizit als gesamtschweizerisch deklariert waren, jedoch mindestens fünf Kantone in mindestens zwei Sprachregionen abdeckten, wurden auch als gesamtschweizerisch eingeteilt.

¹⁵ Die durchschnittliche Differenz zwischen mittlerem Untersuchungsjahr und Erscheinungsjahr beträgt 4.5 Jahre. Dies kann mit der Aktualität der Daten aber auch mit der Beobachtungsdauer erklärt werden: Wenn eine 2017 publizierte Evaluation Daten aus der Periode 2006 - 2016 verwendet, beträgt die entsprechende Differenz zum mittleren Untersuchungsjahr bereits sechs Jahre.

Abbildung 5 Regionale Abdeckung



Hinweis: Die Auswertung basiert auf 23 Studien.

Es fällt auf, dass sich die kantonalen Untersuchungen auf die Deutschschweiz beschränken. Hier möchten wir nochmals darauf hinweisen, dass sich die Analyse ausschliesslich auf Studien bezieht, die die Anforderungen der statistischen Analyse erfüllen (siehe Kapitel 2). Unter den Studien, welche die Selektionskriterien nicht erfüllen, sind auch solche, die die Wirkung von AMM in der Westschweiz untersuchen.¹⁶

Es bestehen gewisse Unterschiede zwischen den Kantonen resp. der Gesamtschweiz hinsichtlich der Verteilung positiver / negativer Evaluationsresultate; aufgrund der tiefen Fallzahlen können diese aber nicht als robuste Differenzen interpretiert werden.

AMM-Typen

Zentral ist die Frage, ob bestimmte AMM-Typen besser oder weniger gut wirken als andere. Es gibt nur wenige Schätzungen zur Gesamtwirkung über alle AMM-Typen; die meisten Studien nehmen sich eines AMM-Typs an oder untersuchen mehrere Typen, aber weisen deren Wirkung getrennt aus (in Abbildung 6 sind sowohl Schätzungen zur Gesamtwirkung als auch zu einzelnen AMM-Typen dargestellt). Beschäftigungsprogramme, Fachkurse und Basisprogramme wurden am häufigsten

¹⁶ Beispielsweise wurde in der Westschweiz mittels eines Experiments die unterschiedliche Wirkung eines privaten Angebots zur Unterstützung bei der Stellensuche mit dem entsprechenden Angebot der öffentlichen Hand verglichen (siehe Ausführungen in Kapitel 3.4). Eine weitere Studie prüfte die Zufriedenheit der Teilnehmenden mit einer Massnahme zur Förderung der Selbstständigkeit.

evaluiert. Dabei spielen vermutlich einerseits die Häufigkeit der Nutzung und andererseits die Kosten eine Rolle: Basis- und Beschäftigungsprogramme gehören zu den meist genutzten AMM. Beschäftigungsprogramme sind zudem aufgrund ihrer Dauer und Intensität besonders teure Angebote.

Klassifizierung Massnahmentypen

In der Analyse werden neun Typen unterschieden:

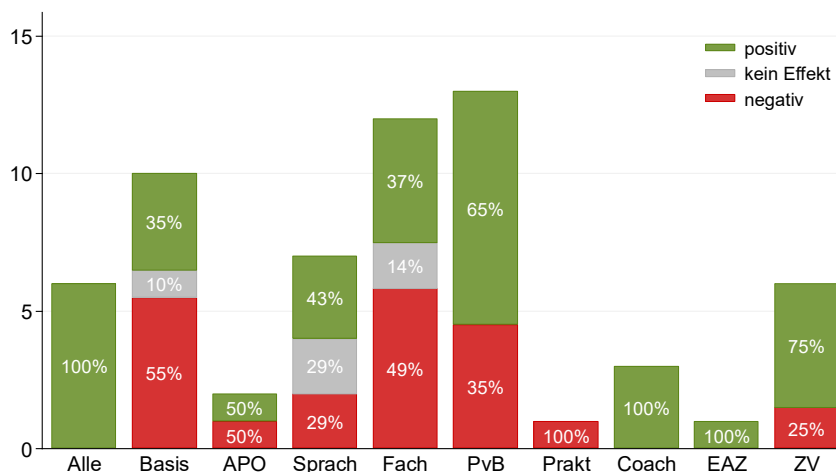
- *Basisprogramme* fokussieren auf berufliche Standortbestimmung sowie Bewerbungstraining.
- *APO (Andere persönlichkeitsorientierte Kurse)* beinhalten ähnliche Komponenten wie Basisprogramme, dauern aber im Durchschnitt deutlich länger.
- *Sprachkurse* zielen auf den Erwerb angemessener Kenntnisse der Landessprache für die Verständigung auf Stellensuche und in beruflichen Situationen.
- *Fachkurse* sollen die fachlichen Qualifikationen aktualisieren und allfällige Bildungslücken, welche die berufliche Wiedereingliederung erschweren, schliessen.
- *Praktika* umfassen Ausbildungspraktika (Vertiefung berufliche Kenntnisse) sowie Berufspraktika (Erwerb praktischer Arbeitserfahrung nach Abschluss Ausbildung).
- *PvB (Programme zur vorübergehenden Beschäftigung)* bieten zeitlich begrenzte und fachlich begleitete Arbeitsgelegenheiten an. Die Programme beinhalten Bildungs- und Coachingelemente für eine arbeitsplatznahe Qualifizierung.
- *Coaching* deckt Beratungsformen ab, die als gesonderte Dienstleistung angeboten werden (d.h. nicht als Teil der oben genannten AMM oder der allg. RAV Beratung)
- *EAZ (Einarbeitungs- und Ausbildungszuschuss)* sind Zahlungen an Arbeitgeber für besondere Leistungen bei der Einarbeitung oder Ausbildung neuer Mitarbeitender.
- *Zwischenverdienst* kompensiert Stellensuchende, die eine vorübergehende Stelle annehmen und dabei einen Lohn erhalten, der tiefer als der versicherte Verdienst ist.

Hinweis: Die Klassifizierung stellt eine Vereinfachung dar: Viele AMM kombinieren inhaltliche Elemente und sind nicht eindeutig zuteilbar. Die Ausgestaltung der Angebote innerhalb einer Kategorie kann sich merklich unterscheiden. Schliesslich wurden die AMM-Typen laufend weiterentwickelt; in der Auswertung können diese Differenzierungen oder zeitlichen Entwicklungen nicht berücksichtigt werden. Schliesslich ist darauf hinzuweisen, dass in der Analyse auch Evaluationen von Zwischenverdiensten aufgenommen wurden. Der Zwischenverdienst ist keine AMM. Indem er vorübergehende (und tiefer entlohnte) Beschäftigungen ermöglicht, stellt er aber hinsichtlich seiner Zielsetzung ein mit dem Beschäftigungsprogramm verwandtes Instrument dar.

Auffällig ist, dass die sechs Studien, die eine Gesamtwirkung über das ganze AMM-Angebot schätzen, alle zu einem positiven Resultat kommen. Unter jenen Massnahmen, die mindestens zweifach evaluiert wurden, weisen Coachingangebote, Beschäftigungsprogramme und Zwischenverdienste den höchsten Anteil positiver Resultate auf. Basisprogramme, APO (Andere persönlichkeitsorientierten Kurse) und Fachkurse schliessen mit 49 % und mehr negativen Schätzungen am schlechtesten

ab. Die Sprachkurse weisen einen vergleichsweise tiefen Anteil an Studien mit einem negativen Effekt auf; allerdings konnten mehrere Schätzungen gar keinen Effekt nachweisen. Zu beachten ist, dass vom Verhältnis der positiven und negativen Wirkungsmessungen eines AMM-Typs nicht auf alle Angebote dieser Kategorie geschlossen werden kann. So können auch bei den Praktika, für die nur ein negatives Resultat vorliegt, Angebote bestehen, welche eine positive Wirkung erzielen.

Abbildung 6 Massnahmentypen



Hinweis: Die Auswertung basiert auf 22 Studien, wobei eine Studie in mehrere Kategorien einfließen kann. Nicht dargestellt ist eine Studie, welche ein Massnahmenbündel (Kombination aus verschiedenen Massnahmen) untersuchte.

Diverse Gründe können einen Einfluss darauf haben, ob ein AMM-Typ einen positiven Effekt erzielt: Die genaue Ausgestaltungsform, der Inhalt, die Dauer und der Zeitpunkt der Teilnahme aber auch die Frage, wie gut das Angebot zum betreffenden Stellensuchenden passt. Die diesbezüglichen Diskussionen im Validierungsworkshop werden im nachfolgenden Abschnitt wiedergegeben. Verschiedene Einflussfaktoren werden zudem in den weiteren Teilkapiteln besprochen, u.a. in der Beschreibung der Teilnehmendengruppen (nächste Seite), der Empfehlungen der Evaluationen (Kapitel 3.5. 3.6. Die Erkenntnisse aus den Evaluationen können aufgrund deren tiefen Anzahl nicht getrennt für verschiedene AMM-Typen besprochen werden. Folgende Interpretation bietet sich dennoch an: Ein hoher Individualisierungsgrad (der als positiver Faktor bewertet wird) mag mitunter den Umstand erklären, dass Coachingangebote überdurchschnittlich und Basisprogramme unterdurchschnittlich positiv abschneiden. Die durchgezogenen Evaluationsresultate bei den

Basisprogrammen mögen auch mit dem potentiell stärkeren Lock-in Effekt zu Beginn der Stellensuche zusammenhängen (allerdings sind Basisprogramme kurz, was den Lock-in Effekt reduzieren sollte).

Interpretation Resultate im Validierungsworkshop

Im Rahmen des Validierungswshops wurde in drei Gruppen folgende Frage diskutiert: Was erklärt das unterschiedliche Abschneiden der AMM-Typen?

Die erste Gruppe begründete den hohen Anteil negativer Resultate bei den Basisprogrammen mit deren mangelnder Individualisierung. Es wurde weiter angemerkt, dass die unterdurchschnittlichen Resultate bei den Fachkursen möglicherweise darauf zurückzuführen sind, dass deren Wirkung erst langfristig zum Tragen komme. Schliesslich wurde diskutiert, ob der hohe Anteil positiver Evaluationsresultate bei den Beschäftigungsprogrammen auf deren Droheffekt zurückgeführt werden könnte.

Die zweite Gruppe machte hinsichtlich der verhältnismässig schlechten Wirkung der Fachkurse darauf aufmerksam, dass diese in Kombination mit weiteren AMM eingesetzt werden; dies mag die Wirkungsmessung erschweren. Grundsätzlich ist der „Mix“ der AMM für die Wirkung verantwortlich: So ist es denkbar, dass ein AMM-Typ für sich alleine betrachtet keine oder sogar eine negative Wirkung erzielt, aber die Wirkung einer nachfolgenden AMM in einer positiven Weise beeinflusst. Dies könnte beispielsweise auf die Basisprogramme zutreffen, die oft auch dazu genutzt werden, mehr Informationen über die Stellensuchenden zu erhalten, um sie dann einer passenden AMM zuzuweisen. Die unterdurchschnittliche Wirkung der Basisprogramme wurde insbesondere durch die Häufigkeit der Anwendung erklärt („Giesskannenprinzip“). Weitere Gründe könnten aus Sicht der Gruppe sein, dass die Grundvoraussetzungen für eine Teilnahme an einem Basisprogramm nicht bei allen Stellensuchenden erfüllt sind, oder wiederum Schwierigkeiten bei der Wirkungsmessung.

In der dritten Gruppe wurden die Unterschiede auf die je nach AMM unterschiedlichen Erwartungen und Zielsetzungen seitens Stellensuchenden und PB zurückgeführt sowie auf Unterschiede zwischen Angeboten, die sich an geringqualifizierte resp. hochqualifizierte Stellensuchende richten. Es wurde darauf hingewiesen, dass die Wirkung jeweils durch das Angebot selbst sowie die Verfügungspraxis zustande kommt; dies kann die Wirkungsmessung erschweren.

Teilnehmergruppen

Weiter wird untersucht werden, ob die Massnahmen bei verschiedenen Gruppen von Teilnehmenden unterschiedlich positive oder negative Wirkung entfalten. Die Resultate sind einerseits interessant für das Design von AMM seitens der LAM-Stellen, aber auch für die Verfügungspraxis seitens der RAV. Die Zahl der Studien, die verschiedene Gruppen von Teilnehmenden hinsichtlich der AMM-Wirkung

vergleichen, ist relativ tief. Dennoch gibt es einige interessante Resultate, die in Tabelle 2 dokumentiert sind.¹⁷

Tabelle 2 Teilnehmergruppen

Grössere Wirkung	Tiefere Wirkung	Studien
Frauen	Männer	4
Pessimisten	Optimisten	1
Motivierte	Nicht motivierte	1
Geringqualifizierte	Gutqualifizierte	1
Hilfsfunktion	Kader	1
Schlecht vermittelbar	Gut vermittelbar	2
Tiefes Einkommen vor AL	Hohes Einkommen vor AL	1
Ausländer/innen	Schweizer/innen	2 vs. 1

Insbesondere wurde in mehreren Studien beobachtet, dass AMM eine grössere Wirkung bei Frauen als bei Männern zeigen. Eine in diesem Zusammenhang diskutierte These ist, dass Frauen eine höhere Flexibilität, mehr Wille zur Veränderung und eine erhöhte Aufnahmebereitschaft haben, was sich positiv auf die AMM-Teilnahme auswirkt. Das Resultat könnte aber auch durch ein unterschiedliches Zuweisungsmuster in AMM zustande kommen, oder durch mit dem Geschlecht korrelierte Faktoren wie die Verteilung der Stellensuchenden auf Berufe, Branchen und Beschäftigungsgrad (Vollzeit / Teilzeit).¹⁸ Es bräuchte mehr Analyse, um dieses Phänomen genauer zu verstehen.

Weitere Studien zeigten, dass Stellensuchende mit schlechten Chancen auf dem Arbeitsmarkt mehr von einer AMM profitieren als andere: So wurde bei Geringqualifizierten, Stellensuchenden mit einem tiefen versicherten Verdienst sowie Personen mit einer Hilfsfunktion jeweils eine überdurchschnittliche Wirkung beobachtet. Dies

¹⁷ Vier Studien untersuchen die Unterschiede zwischen Männern und Frauen (Martinovits et al. 2000, Lalive und Zweimüller 2000, Morlok et al. 2014 und Knaus et al. 2017), drei davon zudem schweizerische und ausländische Stellensuchende (Martinovits et al. 2000, Lalive und Zweimüller 2000 und Morlok et al. 2014). Des Weiteren untersuchen Arni und Wunsch 2014 Optimisten und Pessimisten und Lalive et al. 2011 Stellensuchende, die (nicht) motiviert an der AMM teilnehmen. Knaus et al. 2017 schliesslich untersuchten die Rolle der Qualifikation, beruflichen Stellung und das Einkommen der Teilnehmenden.

¹⁸ Der Unterschied könnte an sich auch entstehen, wenn nicht Frauen, sondern überdurchschnittlich von Frauen ausgeübte Berufsprofile stärker von AMM profitieren.

mag damit zu tun haben, dass Stellensuchende mit guten Chancen verhältnismässig rasch auch selbst eine Stelle finden. Die Unterstützung durch eine AMM könnte unnötig oder sogar hinderlich sein (Lock-in Effekt, negativer Signaleffekt).

Im Rahmen der Analyse wurde auch untersucht, ob bei AMM, die auf eine spezifische Personengruppe fokussieren, eine bessere Wirkung nachgewiesen wurde. Lediglich drei Studien, die in diese statistische Analyse aufgenommen wurden, untersuchen eine derartige AMM.¹⁹ Diese kommen alle zu einem positiven Ergebnis. Die drei untersuchten Massnahmen grenzen ihre Teilnehmerschaft in erster Linie über das Alter ein: Eine Studie untersucht eine Intensivberatung für ältere Stellensuchende, eine weitere ein Coaching für junge Erwachsene (v.a. nach der Lehre) und die dritte ein niederschwelliges Angebot für Jugendliche und junge Erwachsene (Motivationssemester).

3.2. Deskriptive Untersuchung: Methodische Aspekte

In einer Metaanalyse können neben inhaltlichen Dimensionen (AMM-Typ, Untersuchungsperiode etc.) auch methodische Aspekte der Evaluationen untersucht werden. So kann insbesondere analysiert werden, ob die Wahl des Forschungsdesigns einen Einfluss auf die Evaluationsresultate hat: Ermitteln beispielsweise Studien, welche eigene Daten erheben, andere Resultate als jene, die bestehende Informationsgrundlagen für ihre Analysen verwenden? Beeinflusst die Wahl des Indikators die ermittelten Ergebnisse? Wir ergänzen die Diskussion in diesem Abschnitt mit einer Bewertung, in welchen Evaluationssituationen sich die Methodenelemente eignen.

Indikator

Wir beginnen mit den Wirkungsindikatoren, die methodische und inhaltliche Elemente vereinen: Indikatoren sind einerseits ein zentraler Aspekt des methodischen Evaluationsdesigns. Andererseits wird mit den Indikatoren festgelegt, welche inhaltlichen Wirkungsdimensionen untersucht werden und welche nicht. Es wird auch die Fristigkeit festgelegt: Gewisse Indikatoren nehmen eine kurzfristige, andere eine mittel- und langfristige Perspektive ein. Die folgende Analyse orientiert sich an der bereits zu Beginn des Kapitels 3 erläuterten idealtypischen Wirkungskette.²⁰

¹⁹ Arni 2011, Rüfenacht und Neuenschwander 2013 sowie Morlok et al. 2017.

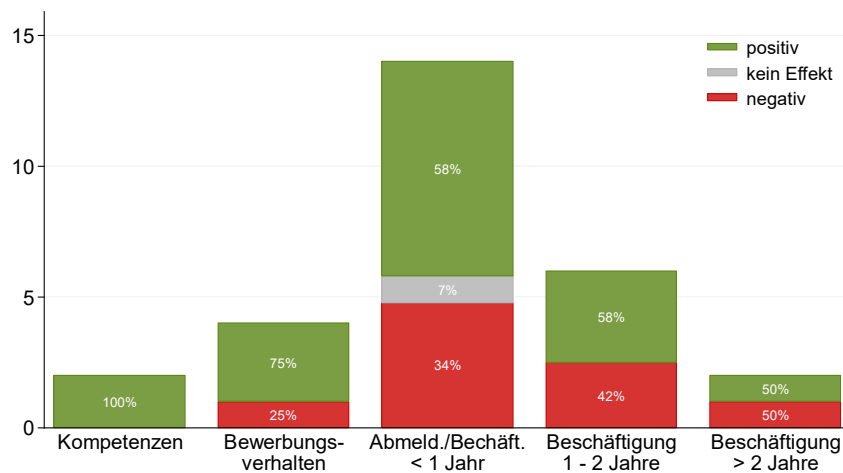
²⁰ Die meisten AMM zielen darauf ab, die Kompetenzen der Stellensuchenden zu beeinflussen. Die gesteigerten Kompetenzen können einerseits zu Veränderungen im Such- und Bewerbungsverhalten

Klassifizierung Indikatoren

- *Kompetenzen*: In dieser Kategorie wurden beispielsweise untersucht: Auftritt, Selbstvertrauen, Motivation, Deutschkenntnisse.
- *Bewerbungsverhalten und -chancen*: Diese Studien untersuchten Anzahl Bewerbungen pro Monat, Anzahl Vorstellungsgespräche pro Monat und Wahrscheinlichkeit eines Vorstellungsgesprächs
- *Abmeldung / Beschäftigung weniger als 1 Jahr nach AMM*: Beispielsweise Anteil Erwerbstätiger nach 9 Monaten oder Veränderung der Abmelderate direkt nach AMM Teilnahme
- *Abmeldung / Beschäftigung 1 bis 2 Jahre nach AMM*: Beispielsweise „Veränderung der Wahrscheinlichkeit am Tag 461 in einer Beschäftigung zu sein“
- *Abmeldung / Beschäftigung mehr als 2 Jahre nach AMM*: Beispielsweise „Veränderung der Anzahl Monate mit einer Anstellung seit Programmbeginn während den ersten 31 Monate“

Die letzten drei Kategorien entsprechen der Einteilung von Card et al. 2015. Um den in der Schweiz ermittelten Studien Rechnung zu tragen, wurden die Kategorien weiter ergänzt.

Abbildung 7 Indikator



Hinweis: Die Auswertung basiert auf 23 Studien, wobei eine Studie in mehrere Kategorien einfließen kann.

und andererseits zu Veränderungen bei den Bewerbungschancen führen, was sich wiederum in mehr Vorstellungsgesprächen und einer schnelleren Abmeldung vom RAV niederschlägt. Längerfristig sollten die AMM zu besseren (z.B. mehr Lohn, bessere Beschäftigungsbedingungen, mehr Jobzufriedenheit) und stabileren Beschäftigungen führen.

Etwas mehr als die Hälfte der Evaluationen fokussiert auf das Abmeldeverhalten bis zu einem Jahr nach der AMM. Das erste (Kompetenzen) und das letzte Glied (Beschäftigung nach mehr als zwei Jahren) in der Wirkungskette wurden hingegen nur selten abgebildet. Zu berücksichtigen ist einmal mehr, dass die Analyse verschiedene Selektionskriterien verwendete (siehe Kapitel 2). Dadurch sind überdurchschnittlich viele Studien weggefallen, die auf Kompetenzen fokussierten (siehe Kapitel 3.4. zu den Erkenntnissen aus u.a. diesen Studien).

Mit den kurzfristigen Indikatoren, die die Kompetenzen und das Bewerbungsverhalten messen, wurde praktisch immer ein positives Resultat erzielt (wobei verschiedenste AMM-Typen untersucht resp. miteinander verglichen wurden). Die einzige Ausnahme bildet die Studie von Falk et al. 2004, die sich mit der Reaktion von Arbeitgebern auf IT-Zertifikate, die von AMM-Anbietern ausgestellt werden, befasste (Signaleffekt). Bei jenen Studien, die die Abmeldung vom RAV bzw. die Beschäftigung untersuchen, ist eine leicht schlechtere Wirkung in der längeren Frist zu beobachten. Zu berücksichtigen ist, dass in der Regression (siehe Kapitel 3.3) sich dieses Resultat nicht bestätigt: Der Unterschied ist nicht statistisch signifikant.

Auffällig ist zudem, dass die Zahl der Studien, die die längerfristige Wirkung von AMM auf die Beschäftigung untersuchen, sehr klein ist. Diese Erkenntnislücke könnte vor allem das Bild der humankapitalorientierten AMM (v.a. Fachkurse) negativ verzerren, da sich deren Wirkung häufig erst über die Zeit positiv entwickelt, wie internationale Evidenz zeigt (z.B. Card et al. 2015 als Übersicht).

Bewertung Indikatoren

Die Indikatoren weisen verschiedene Stärken und Schwächen auf. Wir besprechen diese anhand der folgenden Eignungskriterien, die teilweise aus dem Validierungsworkshop (siehe Kapitel 4) stammen und anschliessend vom Projektteam weiter ergänzt wurden:

- *Wirkung auf rasche/dauerhafte Reintegration*: Ein Teil der Indikatoren bildet die Ziele der öffentlichen Arbeitsvermittlung direkt ab; andere nur indirekt. Bei letzteren (insb. Kompetenzen) muss zusätzlich gezeigt werden, warum eine Steigerung die Chancen auf eine Reintegration fördert.
- *Verständnis Wirkungsmechanismen*: Häufig interessiert nicht nur die Wirkung, sondern auch wie diese zustande kommt. Diese Eigenschaft erfüllen Indikatoren, die den Inhalt der AMM abbilden (insb. Kompetenzen), besser.
- *Differenzierungsgrad*: Je mehr Faktoren ausserhalb der AMM den Indikator beeinflussen, desto komplexer wird es, die Wirkung aufzuzeigen. Je nach Indikator sind deshalb mehr Beobachtungen notwendig, um differenzierte Aussagen über verschiedene Teilgruppen (z.B. AMM-Typen oder Teilnehmergruppen) machen zu können.

- *Mehrfache Messung*: Indikatoren, die eine mehrfache Messung zulassen, können bei der gleichen Person vor, während und nach der AMM verglichen werden. Das reduziert den Selektionseffekt und vereinfacht die Wirkungsmessung.
- *Validität / Reliabilität*: Die Indikatoren unterscheiden sich auch danach, wie eindeutig und objektiv sie sich bestimmen lassen.
- *Geringer Aufwand Datenerhebung*: Für bestimmte Indikatoren liegen Daten vor (AVAM/ASAL), für andere müssen die Daten erst noch erhoben werden.
- *Einbindung Stakeholder*: Es kann prozedurale Vorteile haben, z.B. RAV-PB in die Erhebung miteinzubeziehen; dies generiert aber auch Aufwand.
- *Aktualität*: Wie rasch nach AMM kann ein Indikator aufbereitet werden?

Die folgende Tabelle fasst die Bewertung der Indikatoren anhand der Eignungskriterien zusammen:

Tabelle 3 Vor- und Nachteile der Indikatoren

Eignungskriterien	Kompetenzen	Bewerbungsverhalten und -chancen	Abmeldung / Beschäftigung
Wirkung Reintegration	*	**	***
Verständnis Mechanismen	***	**	*
Differenzierungsgrad	***	**	*
Mehrfache Messung	***	***	*
Validität / Reliabilität	*	***	***
Geringer Aufwand	*	*	***
Einbindung Stakeholder	***	*	*
Aktualität	***	**	* / **

** in tiefem Ausmass erfüllt; ** in mittlerem Ausmass erfüllt; *** in hohem Ausmass erfüllt*

Hinweis: Die Tabelle unterscheidet nicht zwischen kurz-, mittel- und längerfristiger Abmeldung/Beschäftigung; die Bewertung ist weitgehend gleich. Längerfristige Indikatoren können die Dauerhaftigkeit/Nachhaltigkeit besser erfassen, erlauben aber ein geringeres Verständnis der Kausalität, einen tieferen Differenzierungsgrad und eine geringere Aktualität.

Informationsgrundlage

Die in den ausgewerteten Studien genutzten Informationsgrundlagen wurden in vier Kategorien eingeteilt. Die grosse Mehrheit der Evaluationen nutzte die von der öffentlichen Arbeitsvermittlung erhobenen Registerdaten, d.h. die AVAM- und ASAL-Daten. Das überrascht nicht: Die Daten sind reichhaltig, decken alle Stellensuchenden ab und sind forschungsfreundlich aufbereitet resp. dokumentiert. Zudem lohnt es sich aus forschungswirtschaftlichen Gründen, soweit möglich bestehende Daten zu nutzen.

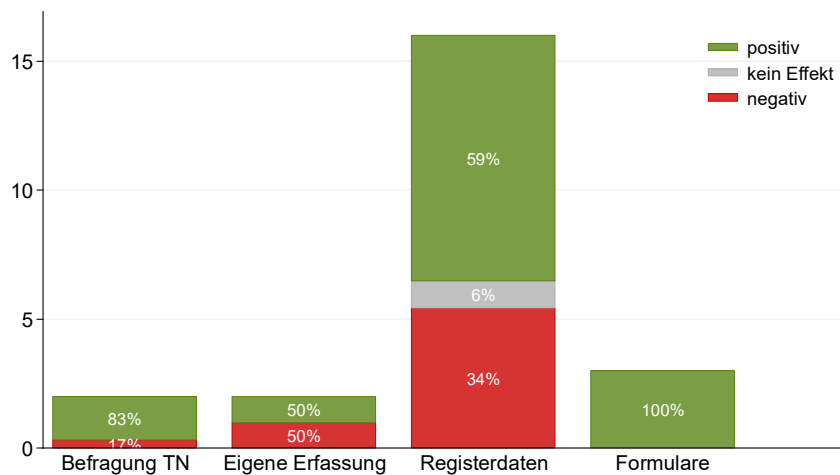
Nur vier der Evaluationen erhoben eigene Primärdaten im Rahmen einer Befragung oder mit Hilfe eigens erstellter Erhebungsinstrumente. Weitere drei Studien erfass-ten Informationen aus bestehenden Formularen (die manuell oder mittels Bilderken-nungssoftware eingelesen wurden).

Die bereits wiederholt erwähnten Selektionskriterien für die statistische Analyse führten dazu, dass verschiedene Informationsgrundlagen, darunter fast alle Befra-gungen von Teilnehmenden, Personalberatenden oder Arbeitgebern, ausgeschlossen wurden. Das war insbesondere dann der Fall, wenn die Befragungen die Wirkung der AMM mit einer Frage im Sinne von „War die besuchte AMM nützlich?“ abbil-deten und keine negativen Antwortkategorien zur Verfügung stellten (d.h. die tiefste Antwortkategorie war „kein Nutzen“). In diesem Fall resultiert per Definition ein Wert, der zwischen 0 und 100 % liegt, was dazu führt, dass sich diese Evaluationen in der statistischen Analyse schlecht mit anderen erfassten Studien vergleichen las-sen.

Klassifizierung Informationsgrundlage

- *Befragung Teilnehmende*: Alle Typen von Befragungen (persönlich, telefonisch oder schriftlich/online).
- *Eigene Erfassung*: Evaluationen, die eigene Erfassungsinstrumente entwickelten und einsetzten, z.B. zur Erfassung von Kompetenzen, Selbsteinschätzungen oder die Messung des Bewerbungserfolges mit einem IT Diplom in der Studie von Falk et al. 2005.
- *Registerdaten*: Daten der Arbeitslosenstatistik (AVAM/ASAL), teilweise verknüpft mit den AHV-Daten
- *Analyse Formulare*: Hierzu zählen wir die Analyse von bereits existierenden For-mularen, wobei es sich in allen drei Studien um das Formular „Nachweis persön-liche Arbeitsbemühungen“ handelt, auf welchem die Stellensuchenden ihre Be-werbungen dokumentieren.

Abbildung 8 Informationsgrundlage



Hinweis: Die Auswertung basiert auf 23 Studien, wobei eine Studie in mehrere Kategorien einfließen kann.

Die Kategorie „Formulare“ weist mit 100 % den höchsten Anteil an positiven Resultaten auf²¹, während die Ergebnisse der Befragungen mehrheitlich positiv ausfielen. Bei den anderen beiden Kategorien mündeten ähnlich viele Evaluationsresultate in positiven wie in negativen Ergebnissen.

Bewertung Informationsgrundlage

Für die Bewertung werden die gleichen Eignungskriterien wie bei den Indikatoren verwendet, wobei ein zusätzliches Kriterium verwendet wird:

- *Historisierte Daten:* Sind die Daten auch rückblickend beziehbar?

Die Bewertung hängt häufig davon ab, was genau erfasst resp. ausgewertet wird. Beispielsweise enthalten die Registerdaten viele Variablen, die vollständig erfasst sind, sowie einige Datenfelder, die fehlende Werte aufweisen. Die meisten Variablen sind objektiv bestimmbar, so ist z.B. die Abmeldung vom RAV ein eindeutig feststellbares Ereignis. Andere Variablen hingegen enthalten subjektive Einschätzungen (z.B. die Vermittlungschancen, welche der zuständige Personalberater für den jeweiligen Stellensuchenden einschätzt). Wir geben daher Bandbreiten an.

Tabelle 4 Vor- und Nachteile der Informationsgrundlagen

Eignungskriterien	Befragung Teilnehmende	Eigene Erfassung	Registerdaten	Analyse Formulare
Wirkung Reintegration	*/**/**	*/**/**	*/**/**	*/**/**

²¹ Die Evaluationen in dieser Kategorie wurden von den gleichen Autoren (Rafael Lalive, Josef Zweimüller und Michael Morlok), teilweise mit weiteren Mitautor/innen, verfasst.

Verständnis Mechanismen	***	*/**/***	*/**	*/**
Differenzierungsgrad	*/**/***	*/**	***	**/***
Mehrfache Messung	*/**	*/**	*/**2	*/**/***
Validität / Reliabilität	*/**1	*/**/***	**/***	**/***
Geringer Aufwand	*	*	***	**
Einbindung Stakeholder	***	**	*	*
Aktualität	***	***	*/**	*/**
Historisierte Daten	*	*	***	**

** in tiefem Ausmass erfüllt; ** in mittlerem Ausmass erfüllt; *** in hohem Ausmass erfüllt*

¹Subjektive Rückmeldungen, teilweise schwierig, von der Situation Stellenverlust zu abstrahieren.

²Die meisten Variablen werden über die Dauer der Stellensuche nicht angepasst.

Stichprobengrösse

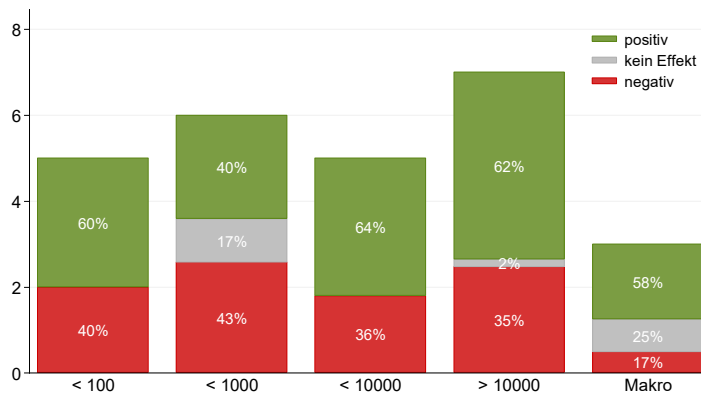
Weiter wurde der Einfluss der Stichprobengrösse untersucht. Es wurde dabei die Anzahl der AMM Teilnehmenden betrachtet; die Beobachtungen in einer allfälligen Kontrollgruppe wurden nicht mitgezählt. Die verwendeten Kategorien kommen in etwa gleich häufig vor, zwischen fünf und sieben Mal.

Klassifizierung Stichprobengrösse

Die Stichprobengrösse wurde wie folgt unterteilt:

- Weniger als 100 AMM-Teilnehmende
- Zwischen 100 und 1'000 AMM-Teilnehmende
- Zwischen 1'000 und 10'000 AMM-Teilnehmende
- Mehr als 10'000 AMM-Teilnehmende
- Makro betrifft Studien, die die Wirkung von AMM nicht auf Individualebene, sondern auf der Makroebene untersuchen, beispielsweise auf der Ebene der RAV oder Kantone; der Vergleich findet zwischen RAV resp. Kantonen statt, die viele AMM einsetzen und solchen, die dies wenig tun. Damit wird die Zahl der Beobachtungen stark reduziert, auch wenn die Beobachtungseinheiten selbst die Daten von zahlreichen Stellensuchenden zusammenfassen.

Abbildung 9 Stichprobengrösse



Hinweis: Die Auswertung basiert auf 23 Studien, wobei eine Studie in mehrere Kategorien einfließen kann.

Auch hier unterscheiden sich die Resultate relativ wenig. Der Unterschied liegt insbesondere darin, dass die Kategorien „< 1‘000“ und „Makro“ mehr Schätzungen enthalten, die keinen Effekt ausweisen.

Bewertung Stichprobengrößen

Grundsätzlich gilt für die Analyse: Je höher die Fallzahlen, desto besser. Aus forschungsökonomischen Gründen sind die Beobachtungszahlen insbesondere bei Primärerhebungen häufig begrenzt; weitere Gründe für eine Einschränkung mag die Zahl der Teilnehmenden an der interessierenden AMM darstellen.

Werden quantitative Auswertungsverfahren verwendet, wird häufig angestrebt, statistisch signifikante Effekte aufzeigen zu können. Ob das möglich ist, hängt von der Grösse des Effekts ab, von der Varianz der Wirkung in den Beobachtungen sowie der Zahl der Beobachtungen. Es gibt daher keine einfachen Regeln für die Bestimmung der notwendigen Fallzahlen. Es besteht aber die Möglichkeit, sog. statistische Power-Berechnungen im Vorfeld einer Studie durchzuführen. Card et al. (2011) zeigen verschiedene Beispiele. Angenommen die Erwerbstätigenquote beträgt 50 Prozent und der erwartete Effekt der AMM wird auf 2.5 Prozentpunkte geschätzt, dann sind 6‘354 Personen im Sample notwendig, um den Effekt auf einem 5-Prozent Signifikanzniveau nachzuweisen. Beträgt der erwartete Effekt hingegen 15 Prozentpunkte, dann sind nur 183 Personen notwendig. Die Berechnungen zeigen, dass insbesondere dann, wenn der erwartete Effekt klein ist, sehr viele Beobachtungen notwendig sind, damit ein signifikanter Effekt geschätzt werden kann. In der Realität werden die Anzahl der AMM-Teilnehmenden durch die AMM und weniger durch die Anforderungen der Evaluation gesteuert. Eine mögliche Stellenschraube ist aber der Zeitpunkt, an welchem die Evaluation durchgeführt wird, oder die Dauer der Evaluation, da die Anzahl der beobachtbaren Teilnehmenden wächst, je länger eine AMM resp. die Evaluation durchgeführt wird. Wird eine Evaluation mit weniger als 100 Teilnehmenden durchgeführt, muss man sich bei der Interpretation bewusst sein,

dass die Resultate von Zufälligkeiten geprägt sein können. Entsprechend müssen die Werte vorsichtig diskutiert und interpretiert werden.²²

Vergleichsebene

Schliesslich wurden die Evaluationen hinsichtlich ihrer Vergleichsebene unterschieden (Abbildung 10). Es geht hier darum, wie sich die Studien der unbeobachtete Vergleichssituation („Counterfactual“) annähern: Was wäre ohne AMM-Teilnahme geschehen?

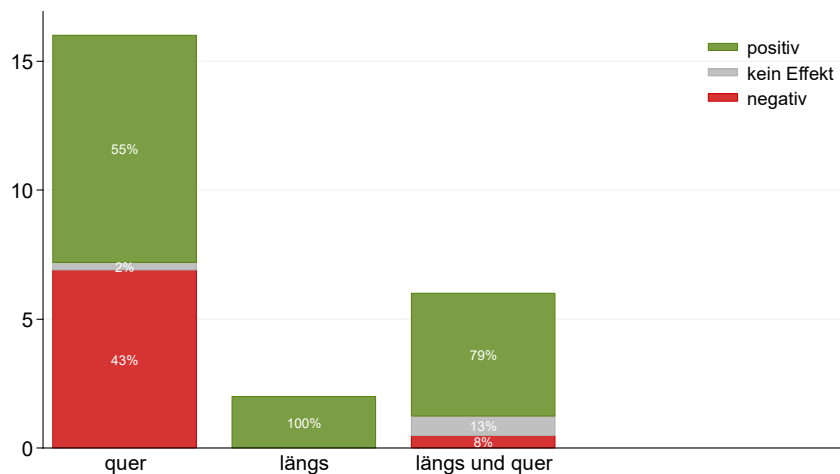
Klassifizierung Vergleichsebenen

Die Vergleichsebenen wurden wie folgt definiert:

- *Quervergleich*: Vergleich zwischen Stellensuchenden, zu einem Zeitpunkt
- *Längsvergleich*: Vergleich zwischen mehreren Zeitpunkten, bei den gleichen Stellensuchenden bzw. gleichen übergeordneten Einheiten
- *Längs und quer*: Kombination aus beidem

Dabei gab nicht etwa die Datenstruktur oder die Aufbereitung der Resultate Ausschlag: Wenn beispielsweise bei einem Matchingverfahren die Wirkung für verschiedene Zeitpunkte ausgewiesen wird (Beschäftigung nach 6, 12 und 18 Monaten), handelt es sich dennoch um einen Quervergleich: Die Wirkung wird eingeschätzt, indem ein Vergleich mit möglichst ähnlichen Nichtteilnehmenden durchgeführt wird.

Abbildung 10 Vergleichsebene



Hinweis: Die Auswertung basiert auf 23 Studien, wobei eine Studie in mehrere Kategorien einfließen kann.

²² Die Ausführungen sind eng angelehnt an Stalder, Lindenmeyer, Morlok und Frey 2013.

Die Analysen zeigen auf, dass es mit grosser Wahrscheinlichkeit eine Rolle spielt, wie der Vergleich durchgeführt wird: Die beiden Längsvergleiche waren positiv, während bei den reinen Quervergleichen fast die Hälfte der Studien zu einem negativen Resultat kam. Bei den Längs- und Quervergleichen waren rund drei Viertel positiv.

Bewertung Vergleichsebenen

Hinsichtlich der Vor- und Nachteile der unterschiedlichen Vergleichsebenen lassen sich folgende Überlegungen festhalten:

- *Quervergleich:* Bei einem reinen Quervergleich besteht das Risiko, dass Unterschiede zwischen Teilnehmenden und Nichtteilnehmenden nicht ausreichend berücksichtigt werden. Die Registerdaten, die dazu meist genutzt wurden, sind zwar sehr umfangreich, decken aber verschiedene zentrale Eigenschaften, die einen Einfluss auf die Arbeitsmarktfähigkeit ausüben, nicht ab (beispielsweise die psychische und physische Gesundheit, Motivation oder das Auftreten).
- *Längsvergleich:* Hier sind zwei Fälle zu unterscheiden. Bei einem Vorher-Nachher-Vergleich über die Zeit – etwa bei der Einführung einer neuen AMM – besteht die Gefahr, dass andere Veränderungen über die Zeit zu wenig berücksichtigt werden. Dies betrifft insbesondere die Verbesserung der Konjunktur und damit die Nachfrage auf dem Arbeitsmarkt. Ein zweiter Fall ist der Vorher-Nachher-Vergleich der Situation eines Stellensuchenden innerhalb der Arbeitslosigkeit (z.B. vor und nach einer AMM). Hier kann sich die Schwierigkeit ergeben, dass sich unbeobachtete Eigenschaften der Person verändern (etwa Kompetenzen, Flexibilität, Motivation oder Selbstvertrauen) und damit Vergleiche verzerren.
- *Längs- und quer:* Grundsätzlich ist es daher ein Vorteil, wenn die Analyse auf verschiedenen Beobachtungspunkten bei der gleichen Person oder Personengruppe beruht und gleichzeitig Teilnehmer und Nichteilnehmer miteinbezieht („längs und quer“). Zu berücksichtigen ist, dass diese allgemeine Aussage nur bei ansonsten gleichen Bedingungen gilt. Häufig werden für „längs und quer“ Vergleiche bei einer geringeren Anzahl von Stellensuchenden eigene Daten erhoben, während viele „quer“ Vergleiche auf Registerdaten beruhen und entsprechend die Daten von sehr vielen Personen verarbeiten können.

Es ist im Einzelfall sorgfältig zu prüfen, welcher Vergleich geeignet resp. überhaupt möglich ist und welche Annahmen beim Vergleich getroffen werden.

Identifikationsstrategie

In den Evaluationen wurden unterschiedliche Strategien angewandt, um den kausalen Effekt zu identifizieren. Die Klassifizierung dieser Strategien kann als weitere Differenzierung der Vergleichsebene verstanden werden und berücksichtigt sowohl

Aspekte der Auswertung wie auch des Designs (z.B. zufällige Zuweisung der Teilnehmenden zur AMM in einem Experiment). Die Klassifizierung erwies sich als besonders komplex, da verschiedene Mischformen zu beobachten sind: Beispielsweise kann ein Experiment mit einem Matching und einem Doppelvergleich kombiniert werden. Diese Unschärfe und die gewählte mechanische Zuteilungsregel ist bei der Interpretation zu berücksichtigen: So gibt es besonders viele Studien in der Kategorie „Matching“, weil alle Studien, die unter anderem ein Matchingverfahren genutzt haben, dieser Kategorie zugeteilt wurden (siehe untenstehende Ausführungen).

Klassifizierung Identifikationsstrategie

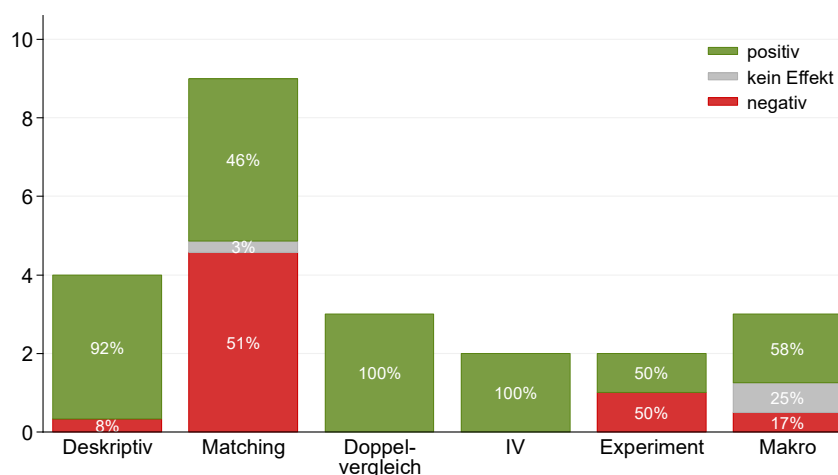
- *Deskriptiv*: Dieser Kategorie wurden (nicht-experimentelle) Studien zugeteilt, die die Wirkung durch einen direkten Vergleich zwischen Teilnehmenden und Nichtteilnehmenden vornehmen, oder einen Vorher – Nachher Vergleich bei den gleichen Personen durchführen, ohne Kontrollvariablen einzusetzen
- *Matching*: Diese Kategorie enthält Studien, die ein Matchingverfahren zur Bildung der Kontrollgruppe nutzen: Beim Matching werden Teilnehmende mit Nichtteilnehmenden verglichen, die möglichst ähnliche Merkmale aufweisen resp. deren Wahrscheinlichkeit an der AMM teilzunehmen gleich hoch ist (sog. Propensity Score Matching). Das Ziel besteht darin, die unterschiedlichen Merkmale der Stellensuchenden, welche zur Selektion in die AMM führten, in der Schätzung möglichst gut zu berücksichtigen, um so eine „künstliche Kontrollgruppe“ zu konstruieren.
- *Difference-in-differences (Doppelvergleich)*: Vorher-Nachher Vergleich sowohl bei AMM-Teilnehmenden wie auch bei einer Vergleichsgruppe. Die beiden Vorher-Nachher Vergleiche werden miteinander verglichen (daher Doppelvergleich). Das Ziel besteht darin, zeitunabhängige Unterschiede zwischen Teilnehmenden und Nichtteilnehmenden herauszufiltern und damit möglichst viele Selektionseffekte zu eliminieren. Die Bereinigung dient somit dazu, allgemeine Entwicklungen über die Zeit, welche nicht mit der AMM in Verbindung stehen, herauszufiltern.²³

²³ Die doppelten Vorher-Nachher-Vergleiche können entweder über die Kalenderzeit geschehen, etwa bei einer neuen Einführung (oder Anpassung) einer AMM. Eine andere Herangehensweise wäre, dieselben Stellensuchenden vor und nach der AMM (innerhalb derselben Stellensuchperiode) zu untersuchen und bei der Vergleichsgruppe bei einem analogen Zeitpunkt den Vorher-Nachher-Vergleich anzusetzen. Diese verschiedenen Ansätze sind mit unterschiedlichen methodischen Herausforderungen verbunden – siehe auch die Ausführungen zu den Vergleichsebenen.

- *IV*: Instrumentalvariablen (IV) setzen Variablen ein, welche mit der AMM-Teilnahme korrelieren, nicht aber mit allfälligen unbeobachteten Merkmalen, welche sich auf interessierenden Zielgrössen (Beschäftigung etc.) auswirken. Gelingt dies, kann die Problematik des Selektionseffekts umgangen werden.
- *Experiment*: Die potentiellen Teilnehmenden werden zufällig in die Gruppe der AMM-Teilnehmenden und Nichtteilnehmenden eingeteilt. Durch die zufällige Auswahl der AMM-Teilnehmer wird sichergestellt, dass sich die Gruppe der AMM-Teilnehmenden nicht systematisch von den Nichtteilnehmenden unterscheidet; letzteres würde die Ergebnisse verfälschen. Dieser Kategorie haben wir auch die Studie von Falk et al. 2005 zugeteilt, in welcher einer zufälligen Auswahl von fiktiven Bewerbungen das Diplom aus einer AMM beigelegt wurde.
- *Makro*: In diese Kategorie fallen Studien, die die Wirkung von AMM auf einer höheren Aggregatsebene (RAV oder Kanton) untersuchen (siehe Erklärung im Abschnitt „Stichprobengrösse“). Um die Wirkung von AMM zu vergleichen, wird die unterschiedliche Nutzungsintensität zwischen den Kantonen ausgenutzt.

Bei der Einteilung haben wir einfache, mechanische Regeln verwendet, die den speziellen Charakter einer Studie hervorheben sollen. Hervorgehoben wurden insbesondere selten eingesetzte und terminologisch klar abgegrenzte Verfahren: Wenn eine Studie eine Instrumentenvariable nutzte, einen experimentellen oder einen aggregierten Ansatz („Makro“) wählte, wurden sie diesen Kategorien zugeteilt. Die restlichen Studien wurden als Matching klassifiziert, sofern ein Matchingverfahren Teil des Vorgehens vor, ansonsten als Doppelvergleich (sofern ein Vorher-Nachher Vergleich umgesetzt wurde) oder als deskriptiv.

Abbildung 11 Identifikationsstrategie



Hinweis: Die Auswertung basiert auf 22 Studien, wobei eine Studie in mehrere Kategorien einfließen kann. Eine weitere Studie konnte nicht zugeteilt werden.

Neben den neun Evaluationen mit einem Matchingverfahren verwendeten vier Studien eine rein deskriptive Analyse und je drei einen Doppelvergleich resp. einen Makro-Ansatz. Angesichts ihrer Prominenz in der Forschung in anderen Bereichen wurden experimentelle Ansätze vergleichsweise sehr selten genutzt (zwei Evaluationen).²⁴

Die Matchingverfahren führten zu einem besonders hohen Anteil negativer Evaluationsresultate, zusammen mit den Experimenten (zwei Studien) und „Andere“ (eine Studie, nicht dargestellt). Andererseits waren die Resultate der deskriptiven Verfahren, Studien mit einem Doppelvergleich oder einem Instrumentenvariablenansatz fast ausschliesslich positiv.

Die Informationsgrundlage und Identifikationsstrategie hängen eng zusammen: Alle Studien, die ein Matchingverfahren anwenden, basieren auf bestehenden Registerdaten, die reichhaltige Daten und sehr viele Beobachtungen zur Verfügung stellen (beide stellen Bedingungen für eine erfolgreiche Anwendung dar). Die Doppelvergleiche basieren alle auf bestehenden Formularen (Nachweis persönlicher Arbeitsbemühungen).

Bewertung Identifikationsstrategie

Der folgende Abschnitt diskutiert die grundlegenden Stärken und Schwächen der Ansätze. Zu berücksichtigen ist, dass diese bei Mischformen oder in speziellen Anwendungssituationen nur reduziert oder gar nicht zutreffen mögen.

- *Deskriptiv*: Die deskriptive Auswertung ist grundsätzlich die einfachste Form der Wirkungsmessung, sowohl in der Umsetzung wie auch in der Kommunikation. Zentral ist, dass eine adäquate Vergleichssituation („Counterfactual“) gefunden wird; dies ist in der Praxis häufig schwierig. Ansonsten bleibt unklar, ob die ermittelte Wirkung auch ohne Massnahme hätte erreicht werden können. Wird als Vergleichssituation eine andere AMM gewählt, kann die Wirkung nur relativ ermittelt werden, das heisst die absolute Wirkung der Massnahme ist nicht bekannt.
- *Matching*: Matchingverfahren können auch verwendet werden, wenn sich AMM-Teilnehmende und Nichtteilnehmende systematisch unterscheiden (was fast immer der Fall ist). Ein Vorteil des Matching ist, dass das Verfahren mit einem einzigen Messzeitpunkt pro Person und Indikator auskommt. Es können so Indikatoren verwendet werden, die einmalig pro Person vorliegen (z.B. Stellensuchdauer) und in

²⁴ Ein drittes Experiment wird aktuell abgeschlossen (Morlok und Liechti 2018), aber die Resultate lagen zum Zeitpunkt der Auswertung noch nicht vor.

den Registerdaten abgebildet sind. Dies reduziert den Aufwand. Um genügend Personen matchen zu können, braucht es hohe Fallzahlen, was wiederum bedeutet, dass die Untersuchungsperiode ausgedehnt werden muss (verringerte Aktualität). Hauptkritik stellt dar, dass für das Matching zentrale Charakteristika in den Daten nicht abgebildet sind und daher der Selektionseffekt nicht vollständig berücksichtigt wird. Es muss in der Anwendung folglich aufgezeigt werden, warum dies gelang.

- *Difference-in-differences (Doppelvergleich)*: Dieser Ansatz berücksichtigt sowohl Querschnitt- und Längsschnittinformationen und ist dadurch robuster (siehe Bewertung Vergleichsebene). Er ist in der Umsetzung und Kommunikation vergleichsweise einfach. Ein Nachteil besteht darin, dass in der Grundvariante nur jene Differenzen zwischen Gruppen (z.B. AMM-Teilnehmende und Nichtteilnehmende) berücksichtigt werden können, die sich über die Zeit nicht ändern.
- *IV*: Eine geeignete Instrumentalvariable (IV) muss zwei Bedingungen erfüllen, die Exogenität (siehe Beschreibung bei „Klassifizierung Identifikationsstrategie“) und die Relevanz: Die IV muss einen signifikanten Einfluss auf die Teilnahmewahrscheinlichkeit eines relevanten Personenkreises haben. In der Praxis ist es schwierig, eine solche Variable zu identifizieren. Wenn es nicht gelingt, dann sind die Ergebnisse nicht valide oder relevant. Wenn es gelingt, dann kann die kausale Wirkung mit verhältnismässig geringem Aufwand dargelegt werden (da meist Sekundärdaten verwendet werden). Allerdings ist der Ansatz komplex und schwierig zu vermitteln (da wenig intuitiv).
- *Experiment*: Ein Experiment stellt sicher, dass eine saubere Vergleichsgruppe besteht (sofern die beteiligten Personen das experimentelle Protokoll befolgen). Die Auswertung reduziert sich in der Regel auf eine einfache qualitative oder quantitative Analyse mit wenigen Annahmen. Zudem sind dabei i.d.R. auch mit tieferen Fallzahlen robuste Auswertungen möglich. Demgegenüber steht als Negativpunkt die teilweise geringe Akzeptanz der Methode, da eine Ungleichbehandlung als unfair angesehen werden kann.²⁵ Experimentelle Evaluationen eignen sich besonders für Pilotprojekte, in denen beispielsweise die Wirksamkeit neuer AMM vor einer flächendeckenden Einführung evaluiert werden soll.
- *Makro*: Anhand von Studien mit aggregierten Massen (Nutzungsintensitäten u.ä.) können beispielsweise Substitutionseffekte (jemand findet eine Anstellung dank einer AMM, ohne diese AMM hätte aber eine andere stellensuchende Person diese Anstellung erhalten) oder Regimeeffekte (Wirkung von mehr oder weniger intensiven Unterstützungsstrategien der RAV oder PB auf alle Stellensuchenden) untersucht werden. Allerdings ist es häufig schwierig, solche breit wirkenden Effekte auf

²⁵ Dem könnte allerdings entgegengehalten werden, dass die ungleiche Behandlung nicht unfair ist, solange die Wirkung des Programms unbekannt ist, oder dass die Anwendung eines Programms mit unbekannter Wirkung auf alle Personen ebenfalls ethische Fragen aufwirft. Zudem führen Pilotprojekte, die nicht alle Klienten umfassen, per se zu einer Ungleichbehandlung. Vgl. Arni 2012.

einzelne AMM oder weitere Elemente zurückzuführen. Oft bedarf es weiterer methodischer Modellierungen, um solche Effekte genauer zu identifizieren.

3.3. Regressionsanalyse

In Kapitel 3.2. wurden die Studien nach einzelnen Dimensionen separat ausgewertet. Diese deskriptiven Analysen ermöglichen einen Überblick über die Häufigkeiten der verschiedenen Ausprägungen und geben erste Hinweise, wie sich die Wirkung von AMM zwischen den analysierten Dimensionen unterscheiden kann. In den deskriptiven Analysen werden allerdings potentielle Zusammenhänge zwischen den einzelnen Dimensionen nicht berücksichtigt. So ist es möglich, dass es auf den ersten Blick so wirkt, als würde ein Zusammenhang zwischen einem Merkmal und dem Anteil positiver Evaluationsresultate bestehen. Bei genauerer Analyse stellt sich aber heraus, dass dieser Zusammenhang daher rührt, dass eine dritte Variable im Zusammenhang mit dem Merkmal und dem Evaluationsresultat steht. So wäre es beispielsweise denkbar, dass Basisprogramme weniger gute Evaluationsresultate aufweisen als Beschäftigungsprogramme, weil sie systematisch mit anderen Methoden ausgewertet wurden. Umgekehrt könnten Zusammenhänge in den deskriptiven Analysen nicht sichtbar sein, weil sie von anderen Zusammenhängen überdeckt werden.

Um hierzu Klärung zu verschaffen, werden Regressionen durchgeführt.²⁶ Im vorliegenden Fall wird in Anlehnung an Card et al. 2015 eine sogenannte Ordered Probit Regression geschätzt. Dieser Regressionstyp eignet sich, wenn die zu erklärende Variable wie in der vorliegenden Analyse eine geordnete Ausprägungen aufweist („positiv“, „kein Effekt“ und „negativ“). Dabei werden verschiedene Faktoren überprüft, wobei jeweils eine Ausprägung pro Faktor als „Referenzkategorie“ gilt, mit welcher die anderen Ausprägungen verglichen werden.

Aufgrund der tiefen Fallzahlen mussten verschiedene in der deskriptiven Analyse gesondert ausgewertete Kategorien aggregiert werden. Zusammengelegt wurden zudem Kategorien, die ausschliesslich positive oder negative Werte generierten; solche Kategorien können ansonsten nicht ausgewertet werden.

Die Koeffizienten einer Ordered Probit Regression sind nicht direkt interpretierbar. Aus diesem Grund werden im Folgenden die durchschnittlichen partiellen Effekte dargestellt: Diese geben den durchschnittlichen Einfluss einer isolierten Änderung

²⁶ Einschränkung ist zu erwähnen, dass nur Scheinkorrelationen ermittelt werden, welche durch Merkmale hervorgerufen werden, die in den Daten beobachtet werden können.

eines Merkmals der Evaluation resp. der untersuchten AMM auf die Wahrscheinlichkeit an, dass die Evaluation ein positives Resultat aufweist, unter Berücksichtigung der Ausprägungen bei den anderen beobachteten Faktoren.²⁷

Die Resultate der Regression (Tabelle 5) sollten mit Vorsicht interpretiert werden, da sie auf einer kleinen Beobachtungszahl basieren.

²⁷ Um die durchschnittlichen Effekte zu berechnen wird jede Schätzung einmal mit und einmal ohne die interessierende Ausprägung berechnet und die Differenz gebildet: $P(y = pos | Basis\ und\ APO = 1, x) - P(y = pos | Basis\ und\ APO = 0, x)$
 $Diff_i = \Phi(\beta_0 + \beta_1 * 1 + \beta_2 * x_{2i} + \dots) - \Phi(\beta_0 + \beta_1 * 0 + \beta_2 * x_{2i} + \dots)$
Darüber wird dann der Durchschnitt gebildet.

Tabelle 5 Wahrscheinlichkeit eines positiven Evaluationsresultats

Merkmal Evaluation	Ausprägung		Effekt
AMM-Typ (relativ zu PvB / Praktika)	Basisprogramme und APO	<input type="checkbox"/>	- 8 %
	Sprachkurse	<input type="checkbox"/>	- 0 %
	Fachkurse	<input type="checkbox"/>	- 15 %
	ZV und EAZ	<input checked="" type="checkbox"/>	+ 34 % *
	Coaching und Andere	<input type="checkbox"/>	+ 23 %
Indikator (relativ zu Indikatoren > 1 J.)	Indikatoren < 1 Jahr	<input type="checkbox"/>	- 8 %
Vergleich (relativ zu Quervergleich)	Längsvergleich oder Längs- und Quervergleich	<input checked="" type="checkbox"/>	+ 56 % ***
Analysetyp (relativ zu Matching)	Deskriptiv	<input checked="" type="checkbox"/>	+ 39 % *
	Doppelvergleich	<input type="checkbox"/>	- 12 %
	Andere	<input checked="" type="checkbox"/>	- 36 % **
Untersuchungsperiode (relativ zu: vor 2007)	Mittleres Untersuchungsjahr nach 2007	<input type="checkbox"/>	- 23 %

Beobachtungen: 71 Schätzungen aus 23 Studien

Signifikanz: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ (nach Studien geclusterte Standardfehler). Grüne Kästchen markieren signifikant positive Resultate, rote Kästchen signifikant negative, weisse Kästchen insignifikante.

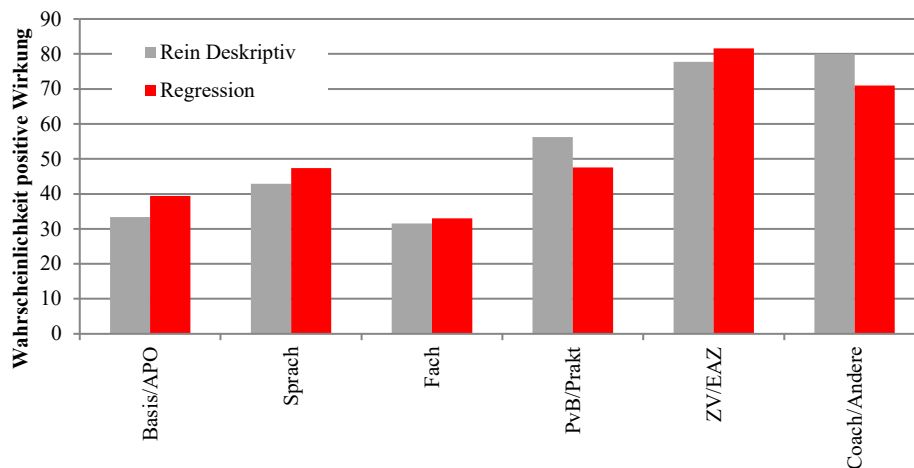
Die Regressionsresultate bestätigen die in der deskriptiven Analyse gezeigten Resultate bei den AMM-Typen: Beispielsweise schneiden PvB/Praktika besser ab als Basisprogramme/APO und Fachkurse (allerdings sind die Differenzen nicht statistisch signifikant), hingegen schlechter als EAZ/Zwischenverdienst und Coaching. Die in der deskriptiven Analyse gezeigten Grössenverhältnisse sind entsprechend keine Scheinkorrelationen; sie bleiben bestehen, wenn für zusätzliche Merkmale der Evaluationen (insb. Methodik und Informationsgrundlage) kontrolliert wird.²⁸ Mit anderen Worten: Zwischenverdienste und EAZ schneiden besser ab als Basisprogramme, und dies nicht, weil sie mit „freundlicheren“ Methoden evaluiert wurden.

In Abbildung 12 werden die Resultate nach AMM-Typen graphisch miteinander verglichen, wobei hier nur der Anteil positiver Resultate abgebildet wird. Laut der deskriptiven Analyse weisen beispielsweise 36 % der Evaluationen von Basisprogrammen / APO ein positives Evaluationsresultat auf. Werden in der Regression andere Faktoren konstant gehalten, erhöht sich dieser Anteil auf 39 %. Auch bei den

²⁸ Diese Aussage stimmt für die Vorzeichen der Koeffizienten. Die Regression zeigt aber auch auf, dass ein Teil der Differenzen statistisch nicht signifikant ist.

anderen AMM-Typen verschieben sich die Resultate relativ geringfügig resp. nicht ausreichend, als dass sich damit das Ranking zwischen den AMM-Typen verändern würde.

Abbildung 12 AMM-Typen: Vergleich Resultate Regression und deskriptive Analyse



Bei den *Indikatoren* hat sich die deskriptive Beobachtung, dass kurzfristige Indikatoren (Auswirkungen auf Kompetenzen, Bewerbungsverhalten sowie Beschäftigung innerhalb eines Jahres) positivere Resultate aufweisen, in der Regression nicht bestätigt. Tatsächlich führen neu längerfristige Indikatoren (Beschäftigung nach mehr als einem Jahr) zu einem höheren Anteil positiver Evaluationsresultate, wobei der betreffende Koeffizient nicht statistisch signifikant ist. Die Veränderung mag damit zu tun haben, dass die Indikatoren neu in zwei grössere Gruppen aufgeteilt wurden. Zuvor hatten die (wenigen) Studien mit den separat ausgewiesenen kurzfristigsten Indikatoren (Kompetenzen und Bewerbungsverhalten) am positivsten abgeschnitten.

Die Ergebnisse der *Vergleichsebene* werden durch die Regression betätigt: Reine Quervergleiche weisen weit häufiger negative Resultate auf als die anderen Kategorien (zusammengenommen). Der Unterschied ist hoch signifikant.

Bei den *Identifikationsstrategien* ist das Bild uneinheitlich: Die Kategorie „deskriptive Analyse“ führt auch in der Regressionsanalyse zu einem hohen Anteil positiver Evaluationsresultate. Während in der deskriptiven Darstellung die Kategorie Matchingverfahren besonders häufig zu negativen Resultaten führte, weisen jetzt sowohl

die Doppelvergleiche wie auch die Kategorie „Andere“ noch häufiger negative Evaluationsresultate auf.²⁹

Auch das Ergebnis zur *Untersuchungsperiode* widerspiegelt nicht das deskriptive Bild. Unter Konstanthaltung aller beobachteten Faktoren hat das mittlere Untersuchungsjahr in der Regressionsanalyse den umgekehrten Einfluss, als die deskriptive Analyse vermuten liess. Evaluationen mit aktuelleren Daten schneiden schlechter ab als solche mit älteren Daten. Der Unterschied ist nicht signifikant.

Insgesamt sind viele der deskriptiven Resultate auch in der Regressionsanalyse beobachtbar. Insbesondere sind die für die Nutzung der AMM besonders wichtigen Resultate zu den Typen stabil. Es gibt hingegen einige wichtige Verschiebungen bei den Indikatoren, den Auswertungsverfahren und der Untersuchungsperiode.

3.4. Erkenntnisse aus weiteren Studien

Wie bereits erwähnt, wurden von den 56 identifizierten Studien 23 in die statistische Analyse aufgenommen. Die nicht aufgenommenen Studien sind (vielleicht noch stärker als jene in der statistischen Analyse) sehr heterogen und entsprechend schwer zusammenzufassen. Bereits die Gründe für die Nichtaufnahme in die statistische Analyse sind durchaus unterschiedlich: Bei fünf Studien lag der Grund darin, dass ihre Resultate in einer anderen Studie dargestellt und für die vorliegende Analyse erfasst wurden. Alle anderen Studien und Evaluationen weisen keine Vergleichssituation auf (resp. keine, die die Situation ohne AMM abbildet) oder aber verwenden Indikatoren, die so spezifiziert sind, dass sie ausschliesslich in positiven Werten (z.B. Anteilen zwischen 0 und 100 %) resultieren.

Die Resultate dieser Studien werden im Folgenden kurz zusammengefasst, unterteilt in 7 Studien mit einem ökonometrischem Ansatz und 21 Studien, welche in erster Linie qualitative und / oder deskriptive Auswertungen enthalten.³⁰

²⁹ Der Unterschied bei den Doppelvergleichen ist besonders markant; diese Kategorie hat in der deskriptiven Analyse zu ausschliesslich positiven Resultaten geführt. Der Unterschied mag daher rühren, dass bei der deskriptiven Analyse nur die jeweiligen Hauptwirkungen betrachtet wurden und in der Regression die programmspezifischen Schätzungen mitberücksichtigt wurden. Dieser Umstand erklärt auch, warum der Doppelvergleich nicht einen „perfect predictor“ darstellt: Während alle Resultate auf Ebene der Hauptergebnisse positiv waren (siehe deskriptive Auswertung), ist ein Teil der Schätzungen auf der Programmebene negativ. Es liegen also leicht unterschiedliche Untersuchungseinheiten zugrunde. Erneut wurden die Strategien zudem zu Gruppen aggregiert.

³⁰ In diesem Abschnitt werden ausschliesslich Inhalte von Studien dargestellt, welche nicht in die statistische Analyse eingeflossen sind. Es ist denkbar, dass Studien, deren Ergebnisse in die statistische Auswertung einfließen, noch zusätzliche Beobachtungen enthalten (z.B. durch eine ergänzende Befragung). Diese sind hier nicht berücksichtigt.

Ökonometrische Untersuchungen

Mehrere Studien untersuchten die Wirkung von speziellen oder neuen AMM im Vergleich zu den Angeboten, welche Stellensuchenden im Regelfall zur Verfügung stehen.³¹ Dazu gehören u.a. zwei Experimente, welche private mit öffentlichen Angeboten zur beruflichen Wiedereingliederung von Langzeitarbeitslosen (Unterstützung bei der Stellensuche und -vermittlung) verglichen. Dabei zeigte sich ein kurzfristig positiver Effekt des privaten Angebotes gegenüber dem öffentlichen Angebot, welcher sich jedoch laufend verringert und nach zwei Jahren gar negativ wird.³²

In einer Studie wurde die Variation der Signalwirkung von AMM (Sprachkurse und PvB) in Abhängigkeit von persönlichen Eigenschaften der Stellensuchenden untersucht. Dabei wurde eine positive Signalwirkung für jene Stellensuchenden aufgezeigt, welche unterdurchschnittliche Chancen aufweisen, eine neue Stelle zu finden. Für besser vermittelbare Stellensuchende hingegen hat die AMM-Teilnahme eine negative Signalwirkung, da sie stigmatisierend wirkt (siehe auch Diskussion in Kapitel 3.6).³³

In einer weiteren Studie wurde bei der Untersuchung zwar eine Vergleichsgruppe gebildet, allerdings fanden sich so wenige Vergleichspersonen, dass die Autoren der Studie sich entschieden, keine Aussagen über die Kausalität zu machen.³⁴

Qualitative oder rein deskriptive Untersuchungen

In elf Studien werden deskriptive Auswertungen zur Tätigkeit der AMM-Teilnehmenden nach Programmabschluss vorgenommen. Untersucht wurden dabei ganz unterschiedliche AMM: Fachkurse, PvB, Coaching, EAZ und Vergleiche zwischen verschiedenen AMM-Typen. Ausgewiesen werden entweder eine Stellenantrittsquote oder aber alle Anschlusslösungen. Die ausgewiesenen Quoten können keine direkte Aussage zur Wirkung machen, da keine Vergleichsangabe zur Verfügung steht, d.h. wir wissen nicht, ob diese Teilnehmenden auch ohne die Massnahme eine Stelle gefunden hätten und ob z.B. eine Stellenantrittsquote von 50 % für die

³¹ Diese Studien wurden ausgeschlossen, da das Counterfactual sich nicht auf die Nicht-Teilnahme an der AMM, sondern auf die Teilnahme an einer anderen AMM bezog.

³² Cottier et al. 2017 sowie Cottier et al. 2015.

³³ Liechti et al. 2017.

³⁴ Guggisberg et al. 2013. Konkret wurde ein Pilotprojekt untersucht, bei welchem schwer vermittelbare Stellensuchende längerfristig in einer Sozialfirma beschäftigt wurden. Rund ein Drittel der Teilnehmenden konnte innerhalb des Untersuchungszeitraums eine Stelle im ersten Arbeitsmarkt antreten. Vorgesehen war, eine Vergleichsgruppe mit Personen zu bilden, welche die Teilnahme am Projekt ablehnten, wofür aber zu wenig passende Personen identifiziert werden konnten.

spezifische Zielgruppe hoch oder tief ist. In drei Studien wurden Stellenantrittsquoten von 44 % bis 90 % ermittelt und als positiv bewertet. Dabei handelte es sich um AMM im Bereich Praktika, PvB sowie ein Coaching spezifisch für junge Menschen. In fünf weiteren Studien wurden zwar Stellenantrittsquoten (im Bereich zwischen 32 % und 86 %) angegeben aber nicht bewertet. In drei Fällen wurde festgehalten, dass der vorab definierte Zielwert des Projekts nicht erreicht wurde, wobei in einer Studie Fachkurse untersucht wurden, in einer weiteren EAZ und in der dritten verschiedene AMM-Typen erfasst wurden.

Eine in zwei Kantonen durchgeführte Machbarkeitsstudie überprüfte, ob die von PB und Stellensuchenden vereinbarten Zielsetzungen für die AMM-Teilnahme erreicht wurden. Dies war grossmehrheitlich der Fall, und zwar in allen untersuchten AMM-Typen (je nach Kanton alle kollektiven AMM resp. Basisprogramme, Deutschkurse und Beschäftigungsprogramme).³⁵

In 15 Studien wurden Befragungen der Teilnehmenden durchgeführt und deskriptiv oder qualitativ ausgewertet. Untersucht wurden dabei unterschiedlichste AMM-Typen: Am häufigsten PvB, aber ebenso Fachkurse, EAZ, Coaching. Einige Studien deckten zudem mehrere AMM-Typen ab. Auffällig ist, dass sich etwas mehr als die Hälfte der untersuchten AMM auf eine spezifische Zielgruppe fokussiert: Auf eine bestimmte Altersgruppe (junge oder ältere Stellensuchende), Personen mit Suchtmittelproblematiken oder anderen Mehrfachproblematiken oder aber gut qualifizierte Stellensuchende. Häufig wurde bei den Befragungen der Teilnehmenden erfragt, ob die Teilnehmenden Verbesserungen bei ihrer Stellensuche wahrnehmen oder ob sie den Eindruck haben, das Programm sei nützlich für ihren beruflichen Werdegang. Auch wenn stets Aspekte hervorgehoben wurden, welche noch optimiert werden könnten / sollten, war der Grundtenor bei allen Untersuchungen durchwegs positiv. Auch Aspekte der Arbeitsfähigkeit, der Stabilität und der Motivation wurden teilweise angesprochen und als positiv empfunden. Einzig bei der Evaluation einer AMM, die auf die Vermittlung von Stellen und Coaching fokussiert, sahen die Teilnehmenden zwar einen persönlichen Nutzen in der Teilnahme (soziale Kontakte etc.), nicht aber Auswirkungen auf die Stellensuche: Von den im Rahmen des untersuchten Projekts vermittelten und subventionierten Arbeitsverhältnissen für ältere Erwerbslose wurde ein Grossteil bereits nach kurzer Zeit wieder aufgelöst. Als

³⁵ Morlok und Liechti 2018. In einem dritten Kanton wurde nicht die Zielerreichung, sondern die Steigerung ausgewählter Kompetenzen auf einer 10er Skala untersucht (bei einem Basisprogramm). Am häufigsten wurde eine Steigerung um 1-2 Punkte beobachtet (gemeinsame Einschätzung PB und Stellensuchende). Für eine ausführlichere Besprechung der pilotierten Methodik siehe Kapitel 4.2.

Gründe dafür wurden ein zu wenig abgestimmtes Anforderungsprofil, schwache Motivation sowie zu hohe Lohnforderungen identifiziert.

Befragungen der Stellensuchenden zur Kundenzufriedenheit beinhalten üblicherweise auch Fragen zu den AMM. 2017 wurden 20'000 Stellensuchende zu ihrer Zufriedenheit mit den Dienstleistungen der RAV und der ALK befragt. Die Aussage „Die Massnahme war für die Stellensuche nützlich.“ wurde auf einer Skala von 1 (stimme gar nicht zu) bis 6 (stimme voll zu) mit durchschnittlich 4.4 bewertet, was die Autoren als mittelmässig positives Feedback bewerten.³⁶

Vereinzelt wurden in Studien zudem weitere Personengruppen befragt, wie ehemalige Teilnehmende der Massnahmen, Arbeitgebende sowie RAV-Mitarbeitende. Dabei zeigte sich beispielsweise in einer Studie, dass die Arbeitgebenden, welche in einem Projekt involviert waren, das intensive und langfristige Coaching der Teilnehmenden sehr schätzen.³⁷ Zudem hervorzuheben ist eine Befragung, die nicht auf eine bestimmte Teilnehmergruppe fokussierte, sondern bei Stellensuchenden im Alter von mehr als 50 Jahren ihre Einschätzung der Angebote für diese Altersgruppe ermittelte. Für knapp zwei Drittel der Befragten (61 %) entspricht das vorhandene Angebot (d.h. alle zur Verfügung stehenden AMM) den Erwartungen; für ein gutes Drittel der Befragten war das Angebot ungenügend.³⁸ Eine weitere Befragung widmete sich ebenfalls dieser Altersgruppe, wobei hier eine Expertenbefragung Kernstück der Untersuchung war.³⁹ Dabei wird festgehalten, dass es für diese Altersgruppe besonders entscheidend ist, eine intensive Beratung zu erhalten und – sofern gute Qualifikationen vorhanden sind – möglichst rasch arbeitsmarktnahe AMM besuchen zu können. Zudem kommt die Studie zum Schluss, dass ein Coaching die realistische Selbstwahrnehmung sowie das Selbstwertgefühl positiv beeinflussen kann, was sich wiederum positiv auf die Stellensuche auswirkt.

Fazit Erkenntnisse aus anderen Studien

Die vorangehenden Ausführungen verdeutlichen die grosse Heterogenität der Evaluationen, welche nicht in die statistische Analyse miteinbezogen wurden: Angesichts der unterschiedlichen Untersuchungsgegenstände resp. Forschungsfragen ist es schwierig, über die Einzelstudien hinausführende allgemeingültige Erkenntnisse

³⁶ Empiricon 2017.

³⁷ Stremlo et al. 2013.

³⁸ Notter 2011.

³⁹ Schmutge et al. 2007.

zu gewinnen. Dies gilt in besonderem Masse für die ökonometrischen Untersuchungen. Bei den qualitativen und deskriptiven Untersuchungen kann zusammengefasst werden, dass die Mehrheit dieser Studien die AMM positiv bewertet haben. Bei den elf Auswertungen der Abgangsraten halten sich die durch die Autor/innen als positiv und negativ beurteilten Resultate zwar die Waage (je drei, zusätzlich zu fünf Analysen, die keine Bewertung abgaben). Die 15 Befragungen hingegen kamen mit einer Ausnahme alle zu einem positiven Resultat. Die Untersuchung der Zielerreichung in ausgewählten AMM kam ebenfalls zu einem positiven Resultat.

3.5. Empfehlungen

In 19 der 56 Studien wurden explizit Empfehlungen formuliert.⁴⁰ Nachfolgend werden jene Aspekte aufgeführt, welche in mehreren Studien (mind. drei) genannt wurden:

- *Zielgruppengerechte und bedarfsorientierte Ausgestaltung (6 Studien)*: Dieser Typ Empfehlung basiert auf der in den Evaluationen getätigten Beobachtung, dass die AMM-Teilnehmenden unterschiedliche Bedürfnisse und Möglichkeiten aufweisen, und dass diesem Umstand noch zu wenig Rechnung getragen wird. Entsprechend kommen die Evaluationen zum Schluss, dass es sinnvoll wäre, beispielsweise je nach Altersgruppe oder beruflichem Hintergrund unterschiedliche Angebote bereitzustellen.
- *Verbessertes Monitoring (insb. Abbruchgründe dokumentieren) (4)*: In mehreren Evaluationen wurde moniert, dass die Datengrundlage für die Steuerungs- und / oder Lernprozesse nicht ausreicht. Entsprechend wurde empfohlen, zusätzliche Daten zu erheben oder diese systematischer aufzubereiten und zu nutzen.
- *Arbeitsmarktnähe erhöhen (3)*: Einige Evaluationen kritisierten die fehlende Ausrichtung der AMM an den Anforderungen des Arbeitsmarkts, oder die fehlende Praxisanbindung durch Einsätze im ersten Arbeitsmarkt. Die Arbeitsmarktnähe kann z.B. durch die Vermittlung von Einsatzplätzen bei Unternehmen oder den Einbezug von Arbeitgeber/innen (z.B. Austausch Teilnehmende mit HR-Fachpersonen etc.) erreicht werden. Die Arbeitsmarktnähe steigert die Motivation der Teilnehmenden (da sie die Massnahme als sinnvoller empfinden) und fördert die positive Signalwirkung, da Arbeitgeber die vorhandenen und neu erlangten Fähigkeiten der Stellensuchenden besser einschätzen können.

⁴⁰ Dokumentiert sind Empfehlungen, welche explizit als solche formuliert wurden.

- *Flexibilität und Spielraum für Individualisierung bei der Ausgestaltung (3)*: Durch die Möglichkeit, innerhalb eines Angebots flexibel auf die Bedürfnisse der Stellensuchenden eingehen zu können, könnte gemäss den Evaluationen eine bessere Wirkung bei den untersuchten AMM erzielt werden.
- *Fundierte Abklärung vor AMM und angepasste Zuweisung (3)*: Indem die Bedürfnisse und Fähigkeiten der Stellensuchenden besser eingeschätzt werden, können passende(re) Massnahmen zugewiesen werden.
- *Klare Kommunikation / Zielformulierung gegenüber Teilnehmenden (3)*: Drei Evaluationen forderten eine klarere Kommunikation mit den Stellensuchenden, um so ein gemeinsames Verständnis der zu erreichenden Ziele sicherzustellen, was sich wiederum positiv auf die Motivation der Teilnehmenden sowie ihren Nutzen aus der AMM auswirken kann.
- *Verbesserte Zusammenarbeit / Kommunikation zwischen involvierten Stellen (3)*: Bei dieser Empfehlung steht die Kommunikation zwischen PB und Anbieter bezüglich der Ziele und Zielerreichung im Vordergrund. Wenn weitere institutionelle Akteure involviert sind (z.B. IV oder Sozialhilfe) kann sich die Empfehlung auch auf die klare Aufgabenteilung und Kommunikation zwischen diesen zusätzlichen Beteiligten beziehen, um Missverständnisse, Doppelspurigkeiten sowie fehlende Zielorientierung zu vermeiden.
- *Intensive Beratung / Coaching ausbauen (3)*: Drei Evaluationen schlugen vor, eine individuellere und engere Begleitung zu ermöglichen bzw. weiter zu fördern. Es handelt es sich bei der Beratung / Coaching um einen individualisierten und flexiblen Bestandteil der AMM, welche es erlaubt, spezifisch auf die jeweiligen Bedürfnisse der Person einzugehen (siehe Empfehlung zur Individualisierung).

Weiter wurden in verschiedenen Studien Empfehlungen dazu formuliert, ob die jeweils evaluierte AMM weitergeführt werden soll oder nicht. Diese Empfehlungen lassen sich nicht über die spezifisch untersuchte AMM hinaus verallgemeinern und werden hier nicht weiter diskutiert.

3.6. Wirkungskanäle

Eine kleinere Anzahl Studien untersuchte, über welche Wirkungskanäle AMM eine Wirkung entfalten. Dabei wurden verschiedene Hypothesen überprüft. Davon werden nachfolgend die zentralen Thesen dargestellt.⁴¹ Es wird auf jene Studien fokussiert, die die Wirkungskanäle empirisch untersuchen. Würde man auch jene Studien inkludieren, welche die Wirkungskanäle zur Erklärung ihres Gesamtergebnisses herbeiziehen, würde sich eine höhere Anzahl Studien ergeben: So wurde beispielsweise der Lock-in Effekt in all jenen Evaluationen thematisiert, die eine negative AMM-Wirkung zeigten.

Antizipationseffekte: Droheffekt / Anziehungseffekt

Der Droheffekt bezeichnet die These, dass Stellensuchende ihren Bewerbungseffort vor der AMM erhöhen, um noch vor deren Beginn eine Stelle zu finden, da sie dem als unangenehm empfundenen AMM-Besuch entgehen möchten.⁴² Als Gegenstück dazu bezeichnet der Anziehungseffekt die Reduktion des Bewerbungsefforts, da Stellensuchende gerne an einer bestimmten AMM teilnehmen möchten.

In fünf Studien wurde der Droh- resp. der Anziehungseffekt empirisch untersucht: Zwei dieser Studien beobachteten einen Droheffekt, zwei einen Anziehungseffekt. In einer weiteren Studie wurde der Droheffekt nicht beobachtet.⁴³ Der Befund ist folglich nicht eindeutig, wobei zu beachten ist, dass die Effekte je nach untersuchter AMM sehr unterschiedlich ausfallen können und folglich unterschiedliche Resultate nicht im Widerspruch stehen müssen. Unter jenen drei Studien, die diesbezüglich alle AMM untersucht haben, ist die Verteilung aber weiterhin gleich (eine Studie identifizierte einen Droheffekt, eine keinen Effekt, und eine einen Anziehungseffekt).

Lock-in Effekt

Unter dem Begriff „Lock-in“ wird der Effekt verstanden, dass Teilnehmende während der AMM ihren Bewerbungseffort reduzieren und entweder weniger oder aber weniger qualitativ hochstehende Bewerbungen schreiben. Dies, da die AMM zeitin-

⁴¹ Bei den 56 gesichteten Studien wurden die Zusammenfassungen, Abstracts resp. Einleitungen daraufhin geprüft, ob explizit auf einen Wirkungskanal Bezug genommen wurde.

⁴² Oder die Stellensuchenden melden sich ohne Stelle vom RAV ab.

⁴³ Droheffekt beobachtet: Martinovits et al. 2000, Lalive et al. 2000; Droheffekt nicht beobachtet: Morlok et al. 2014; Anziehungseffekt beobachtet: Arni 2011; Lalive et al. 2006.

tensiv ist, als nützlich empfunden wird (und der / die Teilnehmende die AMM vollständig besuchen will) oder da die Teilnehmenden während dem Besuch keine Notwendigkeit für die Bewerbung sehen. Der Lock-in Effekt gilt als zentrale Erklärung für (kurzfristig) negative AMM-Effekte.

Sieben Studien untersuchten den Lock-in Effekt empirisch. Fünf davon bestätigten den Effekt während er in zwei Untersuchungen nicht beobachtet werden konnte.⁴⁴ Eine Studie beobachtete beispielsweise eine statistisch signifikante Reduktion der Stellenantrittsrate während der Teilnahme an Fachkursen, d.h. die Kurse wurden selten abgebrochen um eine neue Stellen anzutreten.⁴⁵ Eine andere Studie zeigte, dass der Lock-in Effekt für die Teilnehmenden von Basisprogrammen stärker auftritt als beim Zwischenverdienst.⁴⁶ Auch der zeitliche Einsatz der AMM mag eine wichtige Rolle spielen: Eine Studie beobachtete, dass im dritten und vierten Monat der Stellensuche die Abgangsrate in die Beschäftigung bei den Nichtteilnehmenden vergleichsweise hoch ist und hier der Lock-in Effekt für die Teilnehmenden besonders relevant ist. Hingegen ist die Abgangsrate im ersten Monat der Stellensuche am niedrigsten, hier sollte ein kurzes Basisprogramm zur Abklärung keine nennenswerte Lock-in Effekte verursachen.⁴⁷ Eine der beiden Studien, die keinen Lock-in Effekt feststellte, beobachtete nicht nur keine Reduktion der Bewerbungsbemühungen, sondern stellte stattdessen eine geringe Steigerung bei der Anzahl Bewerbungen und eine starke Steigerung bei der Anzahl Vorstellungsgespräche (als Indikator für die Qualität der Bewerbungen) während der AMM fest.

Festzuhalten ist, dass diverse Faktoren die Frage beeinflussen, ob in einem individuellen Fall ein Lock-in Effekt zustande kommt oder nicht: Nicht zuletzt spielt auch die Haltung und die Kommunikation durch die PB und Anbieter sowie die persönliche Einstellung der Stellensuchenden eine Rolle.

Humankapitaleffekt / Arbeitsmarktfähigkeit

Gemäss dem Humankapitaleffekt entfaltet eine AMM ihre Wirkung dadurch, indem sie die Kompetenzen der Teilnehmenden steigert und damit ihre Chancen auf dem Stellenmarkt erhöht (Steigerung der Arbeitsmarktfähigkeit). In Bezug auf die Steigerung der Fachkompetenzen kann an dieser Stelle auf die Auswertungen zu den

⁴⁴ Lock-in Effekt beobachtet: Arni und Wunsch 2014, Lalive et al. 2008, Lalive et al. 2000, Baumann et al. 2015, Arni 2010; Lock-in Effekt nicht beobachtet: Morlok et al. 2014, Guggisberg et al. 2013.

⁴⁵ Lalive et al. 2000

⁴⁶ Arni und Wunsch 2014.

⁴⁷ Lalive et al. 2000

Fach- und Sprachkursen in Kapitel 3.1 verwiesen werden: Überall wo empirisch positive Effekte dieser Kurse belegt werden konnten, zeigt sich der Humankapitaleffekt.⁴⁸

Ein weiterer Aspekt des Humankapitals resp. der Arbeitsmarktfähigkeit sind persönliche Grundkompetenzen wie Pünktlichkeit, Verlässlichkeit, Teamfähigkeit etc. Eine Wirkung kann dadurch zustande kommen, dass die AMM diese Grundkompetenzen der Teilnehmenden steigert oder zumindest aufrechterhält. Dieser Effekt wurde in vier Studien explizit untersucht und in allen bestätigt. So konnten durch die AMM-Teilnahme Ängste vermindert und soziale Beziehungen gestärkt werden.⁴⁹

Verschiedentlich wurden auch die Auswirkungen auf die Erwartungshaltung diskutiert. So halten zwei Studien fest, dass die Integration in den Arbeitsmarkt besser verläuft, wenn die Stellensuchenden eine positive aber realistische Selbstwahrnehmung und Erwartungshaltung haben. Zwei Studien konnte zudem aufzeigen, dass durch den Besuch einer AMM eine klarere und realistischere Einschätzung der eigenen Jobchancen erreicht werden kann. In einem Fall wurde jedoch auch darauf hingewiesen, dass die Teilnahme an Abklärungs- und Bewerbungskursen sowie Informatikkursen Frustration und Motivationslosigkeit auslösen kann, wenn die Teilnahme als Zwang empfunden wird.⁵⁰

Signaleffekt

Der Signaleffekt bezieht sich als weiterer möglicher Wirkungskanal darauf, dass Arbeitgeber die Kompetenzen der Stellensuchenden aufgrund der AMM-Teilnahme als höher einstufen oder besser erkennen (z.B. durch ein Diplom / Zeugnis). Der Signaleffekt ist eng verknüpft mit dem Humankapitaleffekt, so dass üblicherweise in empirischen Untersuchungen beide Effekte zusammen betrachtet werden. Schliesslich kann die Signalwirkung auf den Arbeitgeber kaum losgelöst von den gesteigerten Kompetenzen durch die AMM-Teilnahme betrachtet werden, da von jeder Teilnahmebestätigung auch eine Aussage zu den Kompetenzen ausgeht.

⁴⁸ Internationale Evidenz (u.a. Card et al. 2015 für die Wirkung von Trainingskursen) weist darauf hin, dass sich der Humankapitaleffekt v.a. längerfristig positiv auswirkt. Kurzfristige Indikatoren mögen die Wirkung daher unterschätzen.

⁴⁹ Effekt auf Arbeitsmarktfähigkeit bestätigt: Guggisberg et al. 2013, Morlok et al. 2017, Rufenacht et al. 2013, Martinovits et al. 2000.

⁵⁰ Effekt auf die Erwartungshaltung: Arni und Wunsch 2014, Schmugge et al. 2007, Rufenacht et al. 2013, Arni 2011, Prey 2000.

In vier Studien wurde der Signaleffekt – resp. die Kombination von Signal- und Humankapitaleffekt – empirisch untersucht. Zwei davon ermittelten einen positiven Effekt: Die Zahl der Vorstellungsgespräche konnten gesteigert werden und die Arbeitgebenden beurteilten die Teilnahme positiv. Eine weitere Studie ermittelte für Kandidaten mit tendenziell schlechteren Vermittlungschancen positive Effekte, nicht aber bei gut Vermittelbaren, bei welchen die AMM-Teilnahme als Stigma wirkte. Die vierte Studie beobachtete für einen Computerfachkurs einen negativen Signaleffekt, wenn die entsprechenden Kompetenzen für die betreffende Stelle vorausgesetzt wurden.⁵¹

Eigenschaften der AMM / Verfügungspraxis

Verfügungszeitpunkt: Zwei Studien beobachteten, dass früh verfügte AMM eine bessere Wirkung erzielen als spät verfügte, wobei eine dieser Studien als Ausnahme die Fachkurse festhält, bei welchen der Verfügungszeitpunkt keine Rolle spielen würde. Eine weitere Studie ermittelte als optimalen Zeitraum für einen AMM-Antritt den Zeitraum zwischen dem 4. und 6. Monat der Stellensuche. Früher verfügte Massnahmen erzielten laut dieser Studie schlechtere Wirkungen. Als mögliche Erklärungsgründe wurde aufgeführt, dass die Personalberatenden die Stellensuchenden zu Beginn der Arbeitslosigkeit noch zu wenig kennen, als dass eine zielführende Zuweisung sichergestellt werden könnte, dass der Lock-in Effekt später in der Arbeitslosigkeit kleiner wird und dass es zu einem späteren Zeitpunkt auch einfacher wird, Stellensuchende mit unterdurchschnittlicher Arbeitsmarktfähigkeit zu identifizieren; diese wiederum weisen eine besonders gute AMM-Wirkung auf.⁵²

Passende Zuweisung: In vielen Studien wird darauf hingewiesen wie wichtig es ist, für einen Stellensuchenden die individuell passende Massnahme zu finden. In drei Studien wurde die Wichtigkeit der passenden Auswahl der Massnahmen für den jeweiligen Stellensuchenden explizit betont.⁵³

Inhalte: In zahlreichen Studien werden Erkenntnisse dazu dargestellt, welche Programminhalte und/oder Ausgestaltungsformen sich als besonders wirkungsvoll erwiesen. Diese Erkenntnisse beziehen sich jedoch stets auf das sehr spezifische Setting der jeweiligen Untersuchung und können entsprechend nicht verallgemeinert

⁵¹ Positiver Signaleffekt: Thurnherr et al. 2012, Morlok et al. 2014; gemischte Resultate: Liechti et al. 2017; Negativer Signaleffekt: Falk et al. 2004.

⁵² Früher Verfügungszeitpunkt positiv: Lalive et al. 2008, Morlok et al. 2011; Ideal zwischen 4. und 6. Monat der Stellensuche: Morlok et al. 2014.

⁵³ Lechner et al. 2004, Behncke et al. 2006, Schmutge et al. 2007.

werden. In einer Studie wurden die Effekte von unterschiedlichen Modalitäten resp. Inhalten massnahmenübergreifend untersucht.⁵⁴ Dabei erwiesen sich folgende Faktoren als besonders positiv: viel Bewerbungstraining, kleine Gruppengrösse, Coachingelement, Modularität, Mindestmass an Intensität/Dauer.

Die Analyse von Wirkungskanälen könnte durch weitere Aspekte ergänzt werden, etwa die Rolle von psychologischen Komponenten (Motivation, Selbstwahrnehmung etc.) oder die Rolle der Information (Information und Kommunikation über Inhalt und Ziele oder mögliche Wirkungen der AMM; dies prägt wiederum der Erwartungshaltungen der Stellensuchenden). Dazu bestehen bisher noch keine systematischen Evaluationen. Weitere Untersuchungen zu Wirkungskanälen könnten sich für die Gestaltung von künftigen AMM als sehr hilfreich erweisen.

⁵⁴ Morlok et al. 2014.

4. Fazit

Die vorliegende Studie verfolgt drei Ziele: Es soll erstens eine Übersicht über alle AMM Evaluationen in der Schweiz geschaffen werden. Zweitens sollen die Studien und ihre zentralen Resultate untersucht und zusammengefasst werden. Und drittens sollen Schlussfolgerungen und Empfehlungen auf Basis der Ergebnisse formuliert werden. Das vorliegende Kapitel adressiert nun das dritte Ziel, unterteilt in inhaltliche (4.1) und methodische Erkenntnisse (4.2) sowie Empfehlungen (4.3).

4.1. Inhaltliche Erkenntnisse

Die Analyse zeigt, dass Beschäftigungsprogramme, Coachingangebote und der Zwischenverdienst in Evaluationen überdurchschnittlich gut abschneiden. Bei den Basisprogrammen und Fachkursen hingegen weist jede zweite Evaluation ein negatives Evaluationsresultat auf. Unklar ist, welche Faktoren die unterschiedlichen Effekte verursachen. Im Validierungsworkshop wurde angesprochen, dass ein möglicher Grund für das schlechte Abschneiden der Basisprogramme sei, dass diese bei sehr vielen Stellensuchenden als Erstmassnahme angewendet werden („Giesskannenprinzip“) und entsprechend wenig bedarfsgerecht zugeschnitten seien. Bei den Fachkursen wurde darauf verwiesen, dass deren Wirkung schwieriger zu erfassen sei, da sie sich erst langfristig entfalte und Fachkurse häufig in Kombination mit anderen AMM eingesetzt werden. Hinsichtlich der Beschäftigungsprogramme wurde im Validierungsworkshop geäußert, dass das positive Ergebnis möglicherweise darauf zurückzuführen sei, dass hier der Droheffekt grösser sei als bei anderen AMM-Typen und dies einen Teil der Wirkung erklärt.

Der Umstand, dass für alle (mehrfach evaluierten) Typen sowohl positive wie auch negative Evaluationsresultate vorliegen, signalisiert eine grosse Heterogenität in der Wirkung. Diese ist einerseits eine Folge der unterschiedlichen Ausgestaltung der Angebote: Die Kategorie „Beschäftigungsprogramme“ beispielsweise fasst verschiedenartige Angebote zusammen. Andererseits spielt der unterschiedliche Einsatz der Massnahmen eine Rolle: Die Auswahl der Teilnehmer/innen und der Zeitpunkt der Teilnahme an einem bestimmten AMM-Typ sowie allfällige weitere vor- und nachgeschaltete AMM variieren zwischen den einzelnen Angeboten. Um zu ermitteln, welche AMM für welche Person in welcher Situation wirksam sind, müsste die Heterogenität der Evaluationsresultate resp. der Wirkung besser verstanden werden. Dazu ist aber die Anzahl der in der vorliegenden Metaevaluation untersuchten

Studien zu gering.⁵⁵ Zudem wurden Evaluationen seit 2000 miteinbezogen, während die Angebote laufend weiterentwickelt wurden.

Insgesamt kann aus der Metaevaluation folglich keine klare Empfehlung zur verstärkten oder schwächeren Nutzung von bestimmten AMM-Typen abgeleitet werden. Die Resultate könnten von den Spezialisten der kantonalen Arbeitsmarktbehörden hingegen als Grundlage genutzt werden, um die AMM-Palette im Kanton sowie deren Nutzung zu überprüfen. Zudem können sie dazu dienen, zukünftige Evaluationsschwerpunkte zu bestimmen: Eine vertiefte Diskussion und Analyse bietet sich insbesondere bei Typen an, die überdurchschnittlich viele negative Evaluationsergebnisse aufweisen. Auch die Häufigkeit und Kosten der Nutzung sollten dabei berücksichtigt werden: Gelingt es eine AMM leicht zu verbessern, welche sehr viele Stellensuchende nutzen, hat dies möglicherweise einen grösseren Effekt als ein stark erhöhter positiver Effekt bei einem Angebot für Einzelne. Unter Berücksichtigung dieser zusätzlichen Eigenschaften müsste der Fokus neben den Basisprogrammen (55 % negative Resultate, sehr hohe Nutzung, tiefe Kosten) und Fachkursen (49 % negative Resultate, mittlere Nutzung und Kosten) auch auf den besonders teuren Beschäftigungsprogrammen (35 % negative Evaluationsergebnisse, hohe Nutzung und Kosten) liegen.⁵⁶

Auch bei den Wirkungskanälen könnte die Heterogenität der AMM-Typen und Zuweisungssituationen die unterschiedlichen Resultate erklären. Immerhin besteht eine bestimmte Einigkeit unter den Autor/innen, dass Lock-in Effekte entstehen können und diese durch die angemessene Selektion der Teilnehmer/innen sowie des Zuweisungszeitpunktes verhindert oder zumindest reduziert werden können. Bei der konkreten Wahl des Zeitpunktes resp. der Selektion zeigt sich aber bereits wieder eine grosse Heterogenität bei den Studienergebnissen. Klar scheint einzig, dass der Lock-in Effekt bei Stellensuchenden, die schlechte Arbeitsmarktchancen aufweisen, geringer sind. Der Signaleffekt wiederum hängt vom Profil des jeweiligen Stellensuchenden und vom angestrebten Berufsfeld ab. Insgesamt haben nur wenige Stu-

⁵⁵ Weitere Gründe für die Heterogenität könnte die unterschiedliche Methodenwahl der Evaluationen oder unterschiedliche Validität der Resultate sein. Die Regressionsanalysen legen die Interpretation nahe, dass die Methodenwahl zwar wichtig ist, aber die Resultate zu den AMM-Typen weitgehend robust sind. Was die Validität betrifft: Die Selektionskriterien für die statistische Analyse stellten sicher, dass eine minimale Vergleichbarkeit zwischen den Studien besteht. Eine weiterführende Beurteilung der Validität auf Basis der Dokumentationen wäre komplex, aufwändig und subjektiv(er) als eine Gleichbehandlung aller Studien.

⁵⁶ Die durchschnittlichen Kosten für die Teilnahme an einem Beschäftigungsprogramm betragen 7'400 CHF, für die Teilnahme an einem Basisprogramm 1'400 CHF und für die Teilnahme an einem Fachkurs 2'500 CHF (Morlok et al. 2014).

dien und Evaluationen explizit Wirkungskanäle untersucht; angesichts der Bedeutung solcher Resultate für das Verständnis der Wirkung der AMM insgesamt sollte hier zukünftig ein stärkerer Fokus gelegt werden.

Am meisten Einigkeit scheint bei den Subgruppen zu bestehen: Mehrere Evaluationen haben gezeigt, dass AMM bei Stellensuchenden mit schlechten Arbeitsmarktchancen eine bessere Wirkung zeigen als bei solchen mit guten Chancen. Ebenso weisen Frauen systematisch bessere Resultate auf als Männer. Während bei den Gruppen mit unterschiedlichen Arbeitsmarktchancen die Lösung darin liegen könnte, vermehrt Angebote für diese Gruppe (und weniger Angebote für Personen mit intakten Chancen) anzubieten, scheint eine solche Fokussierung bei Männern und Frauen wenig fair und sinnvoll. Hier stellt sich für zukünftige Diskussionen und Untersuchungen die Frage, wie die Angebote verändert werden könnten, um die Bedürfnisse männlicher Stellensuchenden besser abzudecken.

Auch bei den Empfehlungen, die aus den Evaluationen stammen, ist eine Übereinstimmung auszumachen. So wurde mehrmals die Bedeutung einer bedarfsgerechten Ausgestaltung und Zuweisung hervorgehoben: Um eine bestmögliche Wirkung zu erzielen ist es wichtig, sorgfältig zu klären, welche AMM zu welchem Zeitpunkt für die/den jeweiligen Stellensuchenden geeignet ist. Ebenfalls als wichtig hervorgehoben wurde eine zielgruppenfokussierte und flexible Ausgestaltung der AMM, welche es erlaubt, auf die individuellen Bedürfnisse der Stellensuchenden einzugehen. Coachingangebote wurden positiv beurteilt, u.a. aufgrund der hohen Individualität dieser Massnahme. Weitere Empfehlungen betonen die Bedeutung einer klaren Kommunikation, sei es zwischen den involvierten behördlichen Akteuren oder aber mit den Stellensuchenden, insbesondere im Rahmen der Zielformulierung. Weiter wurden in den gesichteten Empfehlungen die Verbesserung des Monitorings und die Förderung der Arbeitsmarktnähe erwähnt.

4.2. Methodische Erkenntnisse

Im Folgenden diskutieren wir zunächst Erkenntnisse, die sich auf alle Evaluationen beziehen. Anschliessend adressieren wir die Frage, welche Methoden die kantonalen Arbeitsmarktbehörden mit verhältnismässig geringem Aufwand selbst anwenden könnten.

Evaluationen AMM generell

Im Rahmen des Workshops wurde mit der Begleitgruppe und Praxisexpert/innen der RAV und LAM-Stellen diskutiert, welche Anforderungen bei Evaluationen von

AMM erfüllt sein müssen. Es wurden zunächst im Plenum Erwartungen und Anforderungen gesammelt und diese anschliessend von den Teilnehmer/innen bewertet (Tabelle 6).⁵⁷ Diese Anforderungen können Orientierung bei der Auftragsverteilung und beim Entwerfen von Evaluationsdesigns geben, wobei die Anforderungen im jeweiligen Fall gemeinsam (Auftraggeber und -nehmer) zu klären und konkretisieren sind.

Tabelle 6 Anforderungen an Evaluationen aus Sicht Teilnehmende Workshop

Anforderung an Evaluationen	Punkte	in %
Konkrete Handlungsempfehlungen / Verwertung	14	19 %
Klare Definition Messung / Indikatoren	13	17 %
Ermitteln der Gründe, warum etwas (nicht) wirkt	11	15 %
Einbezug Stakeholder	11	15 %
Differenzierung	7	9 %
Abstimmung Evaluationen um Wissenslücken zu füllen	7	9 %
Wiederholt durchführbar	3	4 %
Wissenschaftlich korrekt, valide	3	4 %
Vorgabe Methoden / Indikatoren durch SECO (teilweise)	2	3 %
Berücksichtigung externer Faktoren	2	3 %
Aktuell	1	1 %
Effizient	1	1 %
Total	75	100 %

Quelle: Workshop mit Begleitgruppe und Praxisexpert/innen. 16 Personen erhielten 5 Punkte, wobei nicht alle Punkte verwendet wurden.

Viele der Anforderungen können grundsätzlich von allen in Kapitel 3 diskutierten Methodenelementen erfüllt werden und informieren daher nur bedingt, welche Elemente sich besonders eignen. Gewisse Unterschiede gibt es bei den Anforderungen „konkrete Handlungsempfehlung / Verwertung“, „Differenzierung“, „wiederholt durchführbar“, „aktuell“ und „effizient“ (siehe dazu die Bewertung der Methodenelemente in Kapitel 3).

⁵⁷ Aus Zeitgründen wurden die Anforderungen vor der Bewertung nicht weiter debattiert; es ist daher wahrscheinlich, dass die Anforderungen unterschiedlich interpretiert und gewichtet wurden. Es ist beispielsweise bei der Anforderung „Wissenschaftlich korrekt, valide“ davon auszugehen, dass alle Stakeholder es relevant finden, dass die Evaluationsresultate korrekt sind; möglicherweise wurde es aber als weniger wichtig erachtet, dass die Resultate „wissenschaftlich korrekt“ hergeleitet / dokumentiert werden, was die relativ tiefe Punktezahl erklären würde. Die Bewertung könnte auch davon abhängen, was bereits als gegeben erachtet wird. Insgesamt scheint die Gesamtliste der Anforderungen wichtiger als die Priorisierung der einzelnen Einträge.

Weitere Orientierung geben die von der Schweizerischen Evaluationsgesellschaft entwickelten SEVAL-Standards, wobei es eine klare Korrelation mit den von den Workshop-Teilnehmenden erwähnten Anforderungen gibt, beispielsweise „nützliche Empfehlungen“ (Standard C3), „nachvollziehbare Bewertung und begründete Folgerungen“ (Standard C2).⁵⁸

Indem Evaluationen stärker koordiniert werden, könnten zudem zusätzliche / vertiefte Erkenntnisse generiert werden („Abstimmung Evaluationen um Wissenslücken zu füllen“ in Tabelle 6). Hier könnte einerseits die thematische Fokussierung der Evaluationen eine Rolle spielen (siehe diesbezügliche Empfehlungen in Kapitel 4.1). Andererseits wäre auch denkbar, dass mehrere Evaluationen koordiniert werden oder eine grössere Evaluation die Angebote und die Nutzung in mehreren Kantonen/Regionen vergleicht (beispielsweise bei Basisprogrammen: flächendeckende Anwendung versus selektive Anwendung). Damit kann der Heterogenität möglicherweise besser begegnet werden: Es können die unterschiedlichen Angebote resp. unterschiedliche Nutzungssituationen direkt miteinander verglichen werden.⁵⁹ Umgekehrt wäre es auch wertvoll, verschiedene Methoden bei der gleichen AMM / Vergleichssituation anzuwenden und die Resultate miteinander zu vergleichen.⁶⁰

Es wäre aus unserer Sicht sinnvoll, wenn das SECO zusammen mit den Kantonen eine mehrjährige AMM-Evaluationsagenda definieren und in dieser künftige Evaluationsziele und -schwerpunkte festhalten würde. Basierend auf unseren Auswertungen fänden wir folgende Ausrichtung der Agenda sinnvoll:

- Fokus auf AMM, die in den bisherigen Evaluationen unterdurchschnittlich abgeschnitten haben, und/oder teurer sind und/oder häufig genutzt werden.
- Identifikation von Wirkungskanälen
- Vertiefte Analyse der Angebote und Nutzung einzelner AMM-Typen (d.h. Evaluation und Vergleich mehrerer AMM des gleichen AMM-Typs).

Um die Basis zu legen für künftige Analysen, welche die Erkenntnisse aus mehreren Studien gemeinsam untersuchen sollen, wäre es zudem wertvoll, dass wenn immer

⁵⁸ Siehe www.seval.ch/documents/Standards/SEVAL-Standards%202016_d_neue%20Version.pdf

⁵⁹ Besonders interessant wären Nutzungssituationen, die sich sehr ähnlich sind, aber sich in einem Element unterscheiden (z.B. frühere versus spätere Nutzung). Ansonsten besteht die Gefahr, dass zu viele Faktoren beim Vergleich berücksichtigt werden müssen.

⁶⁰ Uns sind nur zwei Studien in der Schweiz bekannt, die einen Vergleich verschiedener quantitativer Methoden anhand der gleichen Situation/Daten durchführen: Lechner 2001 vergleicht verschiedene Propensity Score (Matching) Verfahren und Lalive et al. 2008 vergleichen Matching mit Timing-of-Events Methoden. Andere Evaluationen beinhalteten zwar einen Methodenmix (z.B. Befragung und deskriptive Analyse), beinhalten aber keine weiterführende methodische Diskussion.

möglich die Situation des AMM-Besuchs der Situation ohne AMM gegenüberzustellen (Counterfactual). Erst dieser Vergleich erlaubt es, die Wirkung auf eine Art zu messen, welche später auch mit anderen Resultaten verglichen werden kann.

Teil der besseren Koordination sollte auch sein, stärker über die Evaluationen zu informieren (an Tagungen und Workshops) und die Resultate durchgeführter Studien möglichst proaktiv mit den LAM-Leiter/innen anderer Kantone zu teilen. Schliesslich sollte unseres Erachtens geprüft werden, ob eine Steigerung der Frequenz, in der Evaluationen durchgeführt werden, oder eine Erhöhung der Bearbeitungstiefe zu beschleunigten resp. detaillierteren Erkenntnisgewinnen führen; dazu müssten die Investitionen in Evaluationen erhöht werden.⁶¹ Sofern die Erkenntnisse es erlauben, die Wirkung der AMM weiter zu verbessern, dürften sich solche zusätzlichen Investitionen durch die erzielten Einsparungen rechtfertigen lassen.

Evaluationsinstrumente

Zum Schluss setzen wir uns mit der Frage auseinander, ob sich bestimmte Evaluationsmethoden stärker für die zukünftige Anwendung anbieten als andere. Grundsätzlich sind wir der Ansicht, dass es keine guten/schlechten resp. validen/nicht validen Methoden(elemente) gibt, sondern nur solche, die sich für eine bestimmte Evaluationssituation besonders gut eignen. Hier spielen insbesondere das Erkenntnis- und Verwertungsinteresse, zur Verfügung stehende Ressourcen sowie die zu untersuchende „AMM-Situation“ (u.a. Pilot ja/nein, einzelne AMM oder ganze Palette, ex-post oder ex-ante) eine Rolle. Es ist wertvoll, wenn verschiedene Evaluationsinstrumente angewandt werden. Diese Art von Triangulation erhöht die Robustheit der Resultate und verhindert „blinde Flecken“.

Im Folgenden skizzieren wir drei Instrumente, die sich unseres Erachtens für die regelmässige interne Anwendung der kantonalen Arbeitsmarktbehörden eignen.⁶²

1. Monitoring Zielerreichung auf Ebene Angebote

⁶¹ Aus Sicht des Projektteams sprechen folgende Gründe dafür, die Investitionen zu erhöhen: a) Die bisherigen Evaluationsresultate zeichnen ein gemischtes Bild. b) Die auf neuen Daten beruhenden Evaluationen enthalten nicht mehr positive Resultate als jene, die ältere Daten nutzen. c) Die Praxis unterscheidet sich stark zwischen den Kantonen, was darauf hinweisen mag, dass die Wirkung noch zu wenig verstanden wird. d) Die Massnahmen sind grossen Veränderungen unterworfen. e) Es scheint wahrscheinlich, dass Evaluationen zu Lerneffekten führen, welche die Wirkung der Massnahmen erhöhen.

⁶² Wir klammern Befragungen und qualitative Ansätze (Beobachtungen) für die interne Anwendung aus: Um verlässliche Informationen zu erheben, scheint es sinnvoll, wenn diese Ansätze von einer unabhängigen Stelle durchgeführt werden. Dies gilt insbesondere für Befragungen von Stellensuchenden, bei denen ein gewisses Abhängigkeitsverhältnis zum Anbieter aber auch zur Arbeitslosenversicherung generell besteht. Ebenfalls wenig geeignet scheinen uns komplexe quantitative Verfahren, die Statistiksoftware sowie spezialisierte Methodenkenntnisse voraussetzen.

2. Monitoring Zielerreichung auf Ebene Teilnehmende
3. Wirkungsmessung neu pilotierte AMM durch zufällige Zuteilung

Die drei Instrumente nehmen verschiedene Rollen wahr und komplementieren sich gegenseitig:

1. Monitoring Zielerreichung auf Ebene Angebote: Viele Kantone definieren Zielwerte in Bezug auf die Stellenantrittsrate nach AMM-Besuch (z.B. Abmeldung innert drei Monate). Diese Ziele werden den Anbietern kommuniziert und regelmässig überprüft. Es handelt sich hierbei in erster Linie um ein Steuerungsinstrument und weniger um eine Wirkungsmessung: Es wird nicht untersucht, was ohne AMM geschehen wäre. Das Instrument scheint sich in den Kantonen bewährt zu haben. Seine Stärken sind der geringe Aufwand (die Stellenantrittsraten können in AVAM/LAMDA ermittelt werden), die rasche Verfügbarkeit, die objektive Datengrundlage und der Umstand, dass das Konzept und die Messung einfach zu vermitteln sind. Nachteilig ist, dass es komplex ist, sinnvolle Zielwerte zu bestimmen: Diese müssten in Abhängigkeit der Teilnehmergruppen, Zeitpunkt der Nutzung sowie der Kosten für jede AMM unterschiedlich lauten. Ohne Vergleichswerte („Counterfactual“) kann nicht ermittelt werden, ob sich die AMM als Investition lohnt. Bei mehrfachen AMM-Teilnahmen ist schliesslich unklar, welcher Anteil der Wirkung welcher AMM zugeteilt werden soll.

2. Monitoring Zielerreichung auf Ebene AMM-Teilnehmende: Das SECO hat 2017 zusammen mit den Kantonen Aargau, Schwyz und Zürich ein neues Verfahren pilotiert;⁶³ aufgrund der positiven Resultate wird die Methodik ab 2019 schweizweit umgesetzt. Kernidee ist, dass RAV-PB und Stellensuchende vor der AMM die Zielsetzung der Teilnahme besprechen und im Anschluss der AMM gemeinsam diskutieren, ob die Ziele erreicht wurden. Die Zielerreichung kann anschliessend zwischen den AMM verglichen werden. Stärken des Verfahrens sind insbesondere der prozedurale Mehrwert (gemeinsame Zielverständnis, Steigerung der Wirkungsorientierung, bessere Verfügungspraxis).⁶⁴ Die Analyse zeigt auf, wo der Fokus der AMM liegt resp. liegen sollte und es können differenzierte Auswertungen (Einzelprodukte, Teilgruppen) durchgeführt werden. Schwächen sind, dass die Einschätzungen subjektiv sind und durch andere Eigenschaften (u.a. Sprachkenntnisse der

⁶³ Vgl. Machbarkeitsstudie Morlok et al. 2018.

⁶⁴ Damit wurden Empfehlungen aus früheren Studien, u.a. der EFK (siehe Baumann et al. 2015) aufgenommen.

Stellensuchenden) beeinflusst werden; bestimmte Kompetenzen (z.B. Selbstvertrauen) sind durch die PB schwierig einzuschätzen, insbesondere zu Beginn der Stellensuche. Die Erfassung und Kontrolle der Kompetenzziele benötigte im Pilot relativ viel Zeit; dies soll durch die Umsetzung in AVAM optimiert werden. Schliesslich bleibt unklar, ob die Steigerung der Kompetenzen auch tatsächlich zu einer schnelleren / dauerhafteren Abmeldung führt; dies muss separat geprüft werden.⁶⁵

3. Wirkungsmessung bei neu pilotierten AMM durch zufällige Zuteilung: Bei der Einführung von neuen AMM ist es besonders wichtig, deren Wirkung zu überprüfen. Hierzu eignet sich ein Pilotversuch mit ausgewählten Stellensuchenden. Idealerweise werden die Teilnehmenden zufällig aus dem Pool potentieller Teilnehmenden ausgewählt. So kann auf vergleichsweise einfache Weise die Wirkung aufgezeigt werden: Sofern die Zahl der Teilnehmenden resp. Nichtteilnehmenden genügend gross ist, sollten sich die beiden Gruppen in allen relevanten Eigenschaften, insbesondere in ihren Arbeitsmarktchancen, sehr ähnlich sein. Deshalb können die Indikatoren wie der Anteil der Abmeldungen innert drei oder sechs Monaten direkt zwischen den Gruppen verglichen werden. Stärken sind zudem der verhältnismässig geringe Aufwand (AVAM Daten). Als Schwäche wird häufig die ungleiche Behandlung von Stellensuchenden aufgeführt. Hier ist aber zu beachten, dass ein Pilotprojekt ohnehin zu einer Ungleichbehandlung führt (siehe Fussnote 25). Zudem mag es zu einem erhöhten Koordinationsaufwand für die kantonale Behörde kommen.

Die drei Ansätze stellen nicht Substitute dar, sondern ergänzen sich gegenseitig. Sie haben ihre jeweils eigenen Rollen resp. ihren eigenen Fokus: Steuerung Anbieter, Zieldiskussion mit Stellensuchenden, Wirkungsmessung bei neuen Angeboten. Als vierte Kategorie ist die Wirkungsmessung bei bestehenden AMM zu nennen.⁶⁶ Untersuchungen zur Wirkung bestehender AMM informieren über sinnvolle Zielwerte für die Anbieter (Ansatz 1), den Zusammenhang zwischen Kompetenzen und Abmeldung (Ansatz 2), über Wirkungskanäle und die Heterogenität der Zielerreichung (Ansatz 3; siehe zudem Ausführungen in Kapitel 4.1). Um klare Orientierungshilfen für die Bereitstellung und Anwendung zu schaffen, braucht es alle vier Ansätze.

⁶⁵ Bei der Pilotumsetzung hat sich gezeigt, dass viele AMM sehr gut abschneiden und wenig Varianz zwischen den AMM besteht. Es wird in der weiteren Umsetzung zu prüfen sein, wie eine kritischere und gleichsam standardisierte Beurteilung gefördert werden kann.

⁶⁶ Wie eingangs dieses Kapitels erwähnt, wäre es zu begrüssen, wenn in dieser Kategorie weiterhin sehr unterschiedliche methodische Instrumente genutzt werden.

4.3. Empfehlungen

Basierend auf den in den vorgehenden Teilkapiteln formulierten Erkenntnissen können folgende Empfehlungen abgeleitet werden:

- AMM, welche sich an Stellensuchende mit schlechten Arbeitsmarktchancen richten, sollten priorisiert werden.
- Es sollte regelmässig untersucht werden, ob a) die Nutzung bedarfsgerecht erfolgt und/oder die Inhalte eine ausreichende Individualisierung zulassen, b) die Ziele der AMM-Teilnahme allen Beteiligten genügend klar kommuniziert werden, und c) die Arbeitsmarktnähe sichergestellt ist.⁶⁷ *Hinweis:* Die vom SECO pilotierte und ab 2019 umgesetzte Überprüfung der Kompetenzziele (siehe 4.2) adressiert sowohl a) wie auch b).
- Die Bereitstellung und Anwendung von AMM-Typen, deren Evaluationen besonders häufig negative Resultate beinhalten, die hohe Kosten verursachen oder die von sehr vielen Stellensuchenden genutzt werden, sollten von den Arbeitsmarktbehörden kritisch geprüft werden (beim Design neuer Massnahmen, bei der Qualitätssicherung, bei Evaluationen). Mit diesem Fokus sollte zukünftig der Anteil negativer Evaluationsresultate reduziert werden können.⁶⁸
- Es sollte bei zukünftigen Evaluationen mehr darauf geachtet werden, die Wirkungskanäle zu verstehen. Deren Untersuchung ist für die weitere Diskussion besonders wertvoll. Mitunter können Faktoren identifiziert werden, die zur positiven Wirkung von Massnahmen beitragen. Um mehr Erkenntnisse zu Wirkungskanälen zu generieren, ist es häufig von grossem Nutzen, die Auswertung von quantitativen Daten mit Befragungen der Stellensuchenden und / oder PB zu kombinieren.
- Bei der Erarbeitung künftiger Evaluationen ist ein Augenmerk darauf zu legen, dass – sofern dies möglich ist – ein methodischer Ansatz gewählt wird, welcher eine spätere Vergleichbarkeit der Resultate mit anderen Ergebnissen sicherstellt. Dabei ist es zentral, dass die beobachtbare Situation mit AMM-Besuch mit der nicht beobachtbaren, hypothetischen Situation ohne AMM-Besuch verglichen wird (sog. *Counterfactual*).

⁶⁷ In diesem Zusammenhang könnte es nützlich sein, systematische Evaluationen der Verfügungspraxis von diversen Massnahmentypen durchzuführen.

⁶⁸ Es ist nochmals festzuhalten, dass auch bei AMM-Typen, die ausschliesslich negative Resultate ergaben, einzelne Angebote durchaus positive Wirkungen haben können.

-
- Es wird empfohlen, Evaluationsresultate zwischen Kantonen / Bund stärker miteinander zu teilen, zu kommunizieren und gemeinsam zu interpretieren.
 - Künftige Evaluationen sollten so koordiniert und gestaltet werden, dass neue Erkenntnisse gewonnen werden können, welche die Diskussion um die wirksamsten Angebote voranbringen. Dazu könnte eine gemeinsame AMM-Evaluationsagenda formuliert werden. Eine systematische Kombination sich ergänzender Evaluationsansätze und -aktivitäten fördert den Erkenntnisgewinn insgesamt. Es wäre zudem zu prüfen, ob eine erhöhte Frequenz oder Bearbeitungstiefe von Evaluationen zu beschleunigten resp. detaillierteren Erkenntnisgewinnen führen.

5. Anhang

5.1. Literaturverzeichnis

Studien, welche in die statistische Metaevaluation eingeflossen sind

- Arni Patrick (2010), Langzeitarbeitslosigkeit verhindern: Intensivberatung für ältere Stellensuchende, Schlussbericht Universität Lausanne.
- Arni Patrick und Conny Wunsch (2014), Die Rolle von Erwartungshaltungen in der Stellensuche und der RAV-Beratung; Teilprojekt 1: Wechselwirkungen zwischen Erwartungshaltungen der Stellensuchenden, Determinanten des Sucherfolgs und arbeitsmarktlichen Massnahmen, Arbeitsmarktpolitik No 39 (4. 2014), SECO Publikation.
- Falk Armin, Rafael Lalive und Josef Zweimüller (2005), The Success of Job Applications: A New Approach to Program Evaluation, Labour Economics 12 (2005) 739–748.
- Frölich Markus und Michael Lechner (2010), Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies, Journal of the American Statistical Association 105.491: 1014-1029.
- Frölich Markus und Michael Lechner (2014), Combining Matching and Nonparametric Instrumental Variable Estimation: Theory and Application to the Evaluation of Active Labour Market Policies, Journal of Applied Econometrics 30.5: 718-738.
- Gerfin Michael und Michael Lechner (2002), Microeconomic Evaluation of Active Labour Market Policies in Switzerland, The Review of Economics and Statistics, 84, 205-220.
- Gerfin Michael, Michael Lechner und Heidi Steiger (2005), Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes, Labour Economics 12 (2005), 807–835.
- Knaus Michael C., Michael Lechner und Anthony Strittmatter (2017), Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach, IZA DP No. 10961.
- Lalive Rafael und Josef Zweimüller (2000), Arbeitsmarktliche Massnahmen, Anspruch auf Arbeitslosenentschädigung und die Dauer der Arbeitslosigkeit, Publikation der Ausgleichsstelle für den Ausgleichsfonds der Arbeitslosenversicherung, SECO Bern.

- Lalive Rafael, Jan C. von Ours und Josef Zweimüller (2008), The Impact of Active Labour Market Programmes on the Duration of Unemployment in Switzerland, *The Economic Journal* 118.525: 235-257.
- Lalive Rafael, Michael Morlok und Josef Zweimüller (2011), Applying for Jobs: Does ALMP Participation help?, Working Paper Series, Department of Economics, University of Zurich.
- Lalive Rafael, Tanja Zehnder und Josef Zweimüller (2006), Makroökonomische Evaluation der Aktiven Arbeitsmarktpolitik der Schweiz, Arbeitsmarktpolitik No 19 (10. 2006), SECO Publikation.
- Lechner Michael (2001), Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies, *The Review of Economics and Statistics*, 84, 205-220.
- Lechner Michael, Markus Frölich und Heidi Steiger (2004), Mikroökonomische Evaluation aktiver Arbeitsmarktpolitik, Nationales Forschungsprogramm Bildung und Beschäftigung, *Synthesis* 16: 313-323.
- Marti Michael und Stephan Osterwald (2006), Wirkungen der arbeitsmarktlichen Massnahmen auf den schweizerischen Arbeitsmarkt, Arbeitsmarktpolitik No 15 (10. 2006), SECO Publikation.
- Martinovits Alex, André Rothenbühler und Dennis Ganzaroli (2000), Panelbefragungen bei Massnahmenteilnehmern - Wirkungsmessung bei Gastgewerbe- und Winword-Kursen sowie Einsatzprogrammen, im Auftrag des Ausgleichsfonds der Arbeitslosenversicherung (SECO).
- Morlok Michael und David Liechti (2011), Evaluation quick & smart Pilotphase 2011, im Auftrag des Amt für Arbeit St. Gallen.
- Morlok Michael, David Liechti, Rafael Lalive, Aderonke Osikominu und Josef Zweimüller (2014), Evaluation Arbeitsmarktliche Massnahmen. Wirkung auf Bewerbungsverhalten und -chancen, Arbeitsmarktpolitik No 41 (5. 2014), SECO Publikation.
- Morlok Michael, David Liechti, Rafael Lalive, Aderonke Osikominu und Josef Zweimüller (2016), Evaluation der Arbeitsmarktlichen Massnahmen Wirkung AMM auf Bewerbungsverhalten und -chancen (WiAMM 2015), im Auftrag der Volkswirtschaftsdirektion des Kanton Zürich.
- Morlok Michael, Harald Meier, Mirjam Suri, Luzia Zimmermann und David Liechti (2017), Evaluation SEMO Plus, im Auftrag des Mittelschul- und Berufsbildungsamts des Kantons Bern.

- Prey Hedwig (2000), Evaluationsprogramm Arbeitsmarktpolitik: Wirkungsanalyse zu Weiterbildungs- und Umschulungsmassnahmen in St. Gallen, Arbeitsmarktpolitik N 3 (2.2000), SECO Publikation.
- Rüfenacht Mascia und Markus P. Neuenschwander (2013), Praxisbeitrag Teil 2: Evaluation des Coaching Programms CT2, Neuenschwander, M.P. (Hrsg.). Selektion in Schule und Arbeitsmarkt (pp. 231–240). Zürich/Chur: Rüegger.
- Tucci Jürg (2006), Einfluss der arbeitsmarktlichen Massnahmen (AMM) auf die vier Wirkungsindikatoren des AVIG-Vollzugs, im Auftrag der Nationalen Fachgruppe Arbeitsmarktmassnahmen des Verbandes Schweizerischer Arbeitsämter (VSAA).

Studien, die bzgl. Wirkungskanäle und Empfehlungen ausgewertet wurden

- Amt für Wirtschaft und Arbeit Kanton AG (2017), Q-Bericht 2016 (internes Dokument), Kanton Aargau, Amt für Wirtschaft und Arbeit.
- Arni Patrick (2011), Evaluating the During and Post Unemployment Effects of Labor Market Policies, Thesis, University of Lausanne.
- Baumann Martin, Virginie Chevrier und Claudine Morier (2015), Die Programme zur vorübergehenden Beschäftigung und die Berufspraktika der Arbeitslosenversicherung, Eidgenössische Finanzkontrolle (EFK).
- Behncke Stephanie, Markus Frölich und Michael Lechner (2006), Aktive Arbeitsmarktpolitik in Deutschland und der Schweiz – Eine Gegenüberstellung, Vierteljahrshefte zur Wirtschaftsforschung 75/3, S. 118–154.
- Bieri Oliver, Helen Amberg und Flurina Dietrich (2016), Begleitete Evaluation zum Projekt Restaurant Libelle, SAH Zentralschweiz.
- Bonoli Giuliano (2017), Sind Arbeitsmarktprogramme immer nützlich?, kurz und bündig #7, Schweizerischer Nationalfonds.
- Cottier Lionel, Pierre Kempeneers, Yves Flückinger und Rafael Lalive (2017), Does Intensive Job Search Assistance Help Job Seekers Find and Keep Jobs?, working paper.
- Cottier Lionel, Pierre Kempeneers, Yves Flückinger und Rafael Lalive (2015), Does Outsourcing Job Search Assistance help Job Seekers Find and Keep Jobs, working paper.

- Egger Marcel, Urs Birchler, Thomas Dreher, Carlos Lenz, Véronique Merckx und Thomas Studer (2000), Evaluation des Einsatzes arbeitsmarktlicher Massnahmen, im Auftrag des SECO.
- Egger, Dreher und Partner AG (2015), Evaluation der Pforte Arbeitsmarkt in Menziken, im Auftrag des Kanton Aargau.
- Empiricon (2017), Kundenbefragung bei den Stellensuchenden zu den Dienstleistungen der Regionalen Arbeitsvermittlungszentren (RAV) und der Arbeitslosenkassen (ALK) 2017, im Auftrag des WBF und des SECO.
- Guggisberg Jürg, Beat Baumann und Eva Mey (2013), Evaluation Pilotversuch Arbeitsintegrationszuschuss (AIZ) Sozialfirma der DOCK AG, im Auftrag des SECO.
- Jurt Luzia und Claudia Daigler (2011), Evaluation zu Wirkungen des Programms Move On aus Sicht der Adressatinnen und Adressaten, Hochschule für Soziale Arbeit der Fachhochschule Nordwestschweiz.
- Känzig Marc (2011), Le « Soutien à l'Activité Indépendante » (SAI) en tant que mesure du marché du travail dans le canton de Fribourg : étude concernant les bénéficiaires entre 2006-2008, Etat de Fribourg, Service public de l'emploi SPE.
- Künzi Kilian, Susanne Schmutz und Silvia Strub (2007), Evaluation Pilotprojekt «Startinsel» (2005 – 2007): Ein Projekt zur Reintegration älterer gut qualifizierter Stellensuchender in Kooperation mit Neuunternehmen, im Auftrag des SECO.
- Lalive Rafael, Aderonke Osikominu, David Liechti, Michael Morlok, Jeremy Zuchuat und Joseph Zweimüller (2017), WiAMM - Income Effects, im Auftrag des SECO.
- Liechti Fabienne, Flavia Fossati, Giuliano Bonoli und Daniel Auer (2017), The Signalling Value of Labour Market Programmes, *European Sociological Review*, Vol. 33, No. 2, 257–274.
- Morger Mario, Melania Rudin, Heidi Stutz, Jolanda Jäggi und Victor Legler (2017), Zusatzevaluation des Pilotbetriebs Pforte Arbeitsmarkt Menziken, im Auftrag des Vereins Pforte Arbeitsmarkt, Menziken.
- Morlok Michael und David Liechti (2018a), Begleitung Machbarkeitsprüfung einer Wirkungsmessung bei Arbeitsmarktlichen Massnahmen, im Auftrag des SECO.

- Morlok Michael und David Liechti (2018b), Evaluation Stabe Stebe G ü50, im Auftrag des Kanton Solothurn.
- Mühlebach Christine (2016), Auswertung der Umfrage bei den Motivationssemestern der Schweiz, Hochschule Luzern Soziale Arbeit.
- Neuenschwander Markus P. (2014), Coaching und Schulung sind wirksam, Panorama 2, 30 – 31.
- Notter Daniel (2011), Stellensuchende 50+ in der Schweiz, Masterarbeit, Kalaidos Fachhochschule Schweiz.
- Oegerli Thomas (2010), Pilotprojekt Power 40+, im Auftrag des Amts für Arbeit St. Gallen.
- Ragni Thomas (2007), Die Wirksamkeit der öffentlichen Arbeitsvermittlung in der Schweiz, im Auftrag des SECO.
- Schaffhausen, Kantonales Arbeitsamt (2014), Fokus Zukunft.
- Schaffhausen, Kantonales Arbeitsamt (2013), Evaluationsbericht Durchführung Pilotstaffel Perspektive Generation plus.
- Schallberger Peter und Bettina Wyer (2009), Ermächtigung oder Entmutigung? Eine fallrekonstruktive Untersuchung von Programmen zur vorübergehenden Beschäftigung (PvB), im Auftrag des Amt für Wirtschaft und Arbeit (AWA) des Kantons Thurgau.
- Schläfli Katrin (2010), Evaluation Projekt PvB Südhang, beco Fachbereich LAM.
- Schmugge Susanne, Kilian Künzi und Jürg Guggisberg (2007), Wirkungsvolle Massnahmen für Erwerbslose über 50, im Auftrag des Amts für Arbeit St. Gallen.
- Stastny Hoffet Michaela und Barbara Zimmermann (2014), 10 Jahre IFBB, Zur beruflichen und persönlichen Entwicklung der Programmteilnehmerinnen, Masterarbeit ZHAW.
- StremLOW Jürgen und Barbara Trommsdorff (2013), Evaluation des Pilotprojekts Berufsintegrationsprogramm Transit Express der Arbeitsintegration Winterthur, im Auftrag der Arbeitsintegration Winterthur.
- Thurnherr Gregor und Werner Fuchs (2012), Evaluationsbericht 50+ Luzern Schlussbericht, im Auftrag der Dienststelle für Wirtschaft und Arbeit (wira) des Kanton Luzern.

Ergänzende Studien

- Arni Patrick (2012), Kausale Evaluation von Pilotprojekten: Die Nutzung von Randomisierung in der Praxis, *LeGes – Gesetzgebung und Evaluation*, 2012/3, 23 (3), 355-386.
- Arni Patrick und Amelie Schiprowski (2015), Die Rolle von Erwartungshaltungen in der Stellensuche und der RAV-Beratung: Der Jobchancen-Barometer (Teilprojekt 2), *Arbeitsmarktpolitik* No 43 (11. 2015), SECO Publikation.
- Arni Patrick und Amelie Schiprowski (2016), Evaluation der AVIG-Revision 2011 (Taggelder für Junge, Sanktionierungen) Analyse der Wirkung von Suchvorgaben der Arbeitsbemühungen, *Arbeitsmarktpolitik* No 44 (10. 2016), SECO Publikation.
- Arni Patrick, Giuliano Bonoli, Rafael Lalive, Daniel Oesch, Pierpaolo Parrotta, Nicolas Turtschi und Anna von Ow (2013), L’impact des réseaux sociaux sur le retour à l’emploi des chômeurs, *Publication du SECO, Politique du marché du travail* n°37 (9.2013).
- Arni Patrick, Rafael Lalive, Giuliano Bonoli, Daniel Oesch, Pierpaolo Parrotta, Anna von Ow und Nicolas Turtschi (in Bearbeitung), Are Social Networks an Effective Job Search Channel? A Field Experiment, University of Lausanne.
- Arni Patrick, Rafael Lalive, und Jan van Ours (2012), How Effective are Unemployment Benefit Sanctions? Looking beyond Unemployment Exit, *Journal of Applied Econometrics* 28: 1153–1178.
- Bachinger Karin, Andrea Dorr und Eva Heckl (2015), Maßnahmen zur Verhinderung von Jugendarbeitslosigkeit in der Schweiz, *Arbeitsmarktservice Österreich Report* 110.
- Balthasar Andreas (2007), Institutionelle Verankerung und Verwendung von Evaluationen. Praxis und Verwendung von Evaluationen in der schweizerischen Bundesverwaltung, Zürich/Chur, Rüegger.
- Behncke Stephanie, Markus Frölich und Michael Lechner (2009), Targeting Labour Market Programmes – Results from a Randomized Experiment, *Swiss Society of Economics and Statistics*, Vol. 145 (3): 221–268.
- Bieri Oliver, Franziska Müller und Andreas Balthasar (2004), Übersicht über die Professionalisierung der arbeitsmarktlichen Massnahmen (AMM) seit 1997, im Auftrag des SECO.

- Bieri Olivier und Ruth Bachmann (2006), RAV-Strategien zur Arbeitsmarktintegration und deren Wirksamkeit, Seco Publikation Arbeitsmarktpolitik No 17 (10. 2006).
- Bilgili Özge (2015), Evaluating Impact: Lessons Learned from Robust Evaluations of Labour Market Integration Policies, <http://mipex.eu/evaluating-impact-lessons-learned-robust-evaluations-labour-market-integration-policies>.
- Burri Andreas (2012), Chancen und Risiken der Wirkungsmessung von arbeitsmarktlichen Massnahmen, Zertifikatsarbeit der Universität Bern.
- Card David, Jochen Kluge und Andrea Weber (2010), Active Labour Market Policy Evaluations: A Meta-Analysis, *Economic Journal*, 120(548), F452-F477.
- Card David, Jochen Kluge und Andrea Weber (2015), What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations, IZA DP No. 9236.
- Card David, Pablo Ibararán und Juan Miguel Villa (2011), Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide, IZA Discussion Paper No. 6085.
- Flubacher Mi-Cha, Alexandre Duchêne und Renata Coray (2017), Language investment and employability: The Uneven Distribution of Resources in the Public Employment Service, Palgrave Macmillan, Cham.
- Frölich Markus, Michael Lechner, Stefanie Behncke, Stephan Hammer, Nicolas Schmidt, Sarah Menegale, Anette Lehmann und Rolf Iten (2007), Einfluss der RAV auf die Wiedereingliederung von Stellensuchenden, SECO Publikation Arbeitsmarktpolitik No 20 (1.2007).
- Gruber Marc, John Dencker, Melvin Haas und Peter Vogel (2011), Unternehmensgründungen aus der Arbeitslosigkeit, im Auftrag des Schweizerischen Nationalfonds und des SECO.
- Hofmann Claudia und Simone Schaub (2015), Berufliche Integration durch Ausbildung auf dem ersten Arbeitsmarkt und Gelingensbedingungen, interkantonale Hochschule für Heilpädagogik, Mitfinanziert durch das Eidgenössische Büro für die Gleichstellung von Menschen mit einer Behinderung und die „Stiftung La Capriola“.
- Huber Martin, Michael Lechner und Anthony Strittmatter (2017), Direct and Indirect Effects of Training Vouchers for the Unemployed, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181.2 (2018), 441-463.

- Huber Martin, Michael Lechner und Giovanni Mellace (2014), Why do Tougher Caseworkers Increase Employment? The Role of Programme Assignment as a Causal Mechanism, working paper.
- Kaltenborn Bruno und Petra Kaps (2013), Steuerung der öffentlichen Arbeitsvermittlung in der Schweiz, SECO Publikation Arbeitsmarktpolitik No 35 (9. 2013).
- Kaiser Boris, David Liechti und Michael Morlok (2018), Entwicklung der Wirkung der öffentlichen Arbeitsvermittlung in der Schweiz im Zeitraum 2003-2016, SECO Publikation Arbeitsmarktpolitik No 50 (2. 2018).
- Kluve Jochen (2016), A Review of the Effectiveness of Active Labour Market Programmes with a Focus on Latin America and the Caribbean, ILO Working Paper.
- Lalive Rafael, Josef Zweimüller und Jan C. van Ours (2002), The Effect of Benefit Sanctions on the Duration of Unemployment, working paper.
- Lechner Michael (2002), Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods, *The Review of Economics and Statistics*, 84, 205-220.
- Lechner Michael und Conny Wunsch (2009), Are Training Programs more Effective when Unemployment is High?, *Journal of Labor Economics*, 27, 653-692.
- Lechner Michael und Conny Wunsch (2013), Do Long-Term Unemployed Workers Benefit from Targeted Wage Subsidies?, *German Economic Review* 16(1): 43–64.
- Lechner Michael und Stephan Wiehler (2013), Does the Order and Timing of Active Labor Market Programs Matter?, *Oxford Bulletin of Economics and Statistics*, 72-2, 180-211.
- Luzzi Ferro Giovanni, Yves Flückinger, José V. Ramirez und Anatoli Vassiliev (2001), Performance Measurement and Determinants of Inefficiency of Regional Employment Offices: A non parametric Frontier Analysis for Switzerland, Department of Economics, University of Geneva.
- Martin Huber, Michael Lechner, und Conny Wunsch (2016), The Effect of Firms' Phased Retirement Policies on the Labour Market Outcomes of their Employees, *ILR Review* Vol 69, Issue 5, pp. 1216 – 1248.
- Morlok Michael, David Liechti, Rafael Lalive, Aderonke Osikominu und Josef Zweimüller (2015), Wirkung von Beraterinterventionen, Der Einfluss von

Zuweisungen, Sanktionen und Beraterwechsel auf das Bewerbungsverhalten von Stellensuchenden, Arbeitsmarktpolitik No 42 (1. 2015), SECO Publikation.

Oegerli Thomas (2009), Wiederholte Arbeitslosigkeit Situationen und Massnahmen, AMOSA Arbeitsmarktbeobachtung Ostschweiz, Aargau, Zug und Zürich.

Rieder Stefan (2011), Kosten von Evaluationen, LeGes – Gesetzgebung und Evaluation, 2011/1, S. 73-88.

SECO (2015), Synthesebericht Dritte Evaluationswelle der aktiven Arbeitsmarktpolitik 2011-2014. Bern.

Sheldon George (2008), Entwicklung der Performance der öffentlichen Stellenvermittlung der Schweiz im Zeitraum 1998 – 2007, im Auftrag des SECO.

Stalder Martin, Hannes Lindenmeyer, Michael Morlok und Miriam Frey (2013), Evaluationskonzept für das Projekt „Zusammenarbeit Arbeitslosenversicherung / Sozialhilfe: Konkrete Kooperationsvorhaben“, im Auftrag des SECO.

Vooren Melvin, Carla Haelermans, Wim Groot und Henriëtte Maassen van den Brink (2017), The Effectiveness Of Active Labor Market Policies: A Meta-Analysis, Tier Working Paper Series, Amsterdam.

Zweimüller Josef, Tanja Zehnder und Reto Foellmi (2014), Rückruf durch den ehemaligen Arbeitgeber, SECO Publikation Arbeitsmarktpolitik No 40 (4. 2014).

5.2. Auszug Erfassungsraster

Studien, welche in die statistische Metaevaluation eingeflossen sind

Autoren	Erscheinungs- jahr	Untersuchungs- periode	Regionale Abdeckung	AMM-Typ	Indika- tor ⁶⁹	Wirkung	Signifikanz
Arni P.	2011	2008 - 2009	AG	Coaching	4	positiv	sign.
					3	positiv	nicht sign.
Arni P. und C. Wunsch	2014	2007 - 2008	ZH	Basisprogramm	4	negativ	nicht sign.
				ZV	4	negativ	nicht sign.
Falk A., R. Lalive und J. Zweimüller	2005	1999 - 2000	ZH	Fachkurs	2	negativ	nicht sign.
Frölich M. und M. Lechner	2010	1998 - 2006	CH	Gesamtwirkung	3	positiv	sign.
Frölich M. und M. Lechner	2014	1998 - 2006	CH	Gesamtwirkung	3	positiv	sign.
				Gesamtwirkung	5	positiv	sign.
Gerfin M. und M. Lechner	2002	1996 - 1999	CH	Basisprogramm	3	negativ	sign.
				Sprachkurs	3	negativ	sign.
				Fachkurs	3	negativ	sign.
				Fachkurs	3	k.A.	nicht sign.
				Fachkurs	3	k.A.	nicht sign.

⁶⁹ Codierung Indikatoren: 1: Kompetenzen, 2: Bewerbungsverhalten, 3: Abmeldung / Beschäftigung weniger als 1 Jahr, 4: Beschäftigung 1 – 2 Jahre, 5: Beschäftigung mehr als 2 Jahre

				PvB	3	negativ	sign.
				ZV	3	positiv	sign.
Gerfin M., M. Lechner und H. Steiger	2005	1997 - 1999	CH	EAZ	4	positiv	sign.
				PvB	4	positiv	sign.
Knaus M. C., M. Lechner und A. Strittmatter	2017	2003 - 2004	CH	Basisprogramm	5	negativ	sign.
				Basisprogramm	3	negativ	sign.
				Basisprogramm	4	negativ	sign.
				Basisprogramm	3	negativ	keine Angabe
Lalive R. und J. Zweimüller	2000	1997 - 1998	CH	Sprachkurs	3	positiv	keine Angabe
				Fachkurs	3	positiv	keine Angabe
				PvB	3	positiv	keine Angabe
Lalive R., J. C. von Ours und J. Zweimüller	2008	1997 - 1998	CH	Fachkurs	3	negativ	sign.
				PvB	3	negativ	sign.
				Fachkurs	3	negativ	keine Angabe
				ZV	3	positiv	keine Angabe
				Basisprogramm	2	positiv	sign.
				APO	2	negativ	nicht sign.
Lalive R., M. Morlok und J. Zweimüller	2011	2007 - 2008	ZH	Fachkurs	2	positiv	nicht sign.
				Sprachkurs	2	negativ	nicht sign.
				PvB	2	negativ	nicht sign.
				Gesamtwirkung	2	positiv	nicht sign.
Lalive R., T. Zehnder und J. Zweimüller	2006	1998 - 2004	CH	PvB	3	positiv	keine Angabe
				ZV	3	negativ	keine Angabe

				Basisprogramm	3	positiv	keine Angabe
				Fachkurs	3	negativ	keine Angabe
				Basisprogramm	3	negativ	sign.
				ZV	3	positiv	sign.
				Fachkurs	3	negativ	nicht sign.
				PvB	3	negativ	sign.
				ZV	4	positiv	sign.
Lechner M.	2001	1997 - 1999	ZH	Andere	4	negativ	sign.
				Fachkurs	4	positiv	nicht sign.
				PvB	4	negativ	nicht sign.
				ZV	3	positiv	sign.
Lechner M., M. Frölich und H. Steiger	2004	1997 - 1999	CH	PvB	3	negativ	sign.
				ZV	4	positiv	sign.
				PvB	4	positiv	nicht sign.
				Basisprogramm	3	kein Effekt	keine Angabe
Marti M. und S. Osterwald	2006	2000 - 2004	CH	Sprachkurs	3	kein Effekt	keine Angabe
				Fachkurs	3	kein Effekt	keine Angabe
				PvB	3	positiv	keine Angabe
				Fachkurs	3	negativ	keine Angabe
Martinovits A., A. Rothenbühler und D. Ganzaroli	2000	1998 - 1999	CH	Fachkurs	3	positiv	keine Angabe
				PvB	3	positiv	keine Angabe
				Basisprogramm	2	positiv	nicht sign.
Morlok M., D. Liechti, R. Lalive, A. Osikominu und J. Zweimüller	2014	2012 - 2013	CH	Sprachkurs	2	positiv	sign.

				Fachkurs	2	positiv	nicht sign.
				PvB	2	positiv	sign.
				APO	2	positiv	sign.
				Praktika	2	negativ	nicht sign.
				Gesamtwirkung	2	Positiv	sign.
Morlok M., D. Liechti, R. Lalive, A. Osikominu und J. Zweimüller	2016	2014 - 2015	ZH	Basisprogramm	2	-	-
				Sprachkurs	2	-	-
				Fachkurs	2	-	-
				PvB	2	-	-
				Gesamtwirkung	2	-	-
Morlok M., H. Meier, M. Suri, L. Zimmermann und D. Liechti	2017	2015 - 2017	BE	PvB	1	positiv	keine Angabe
Morlok, M. und D. Liechti	2011	2011 - 2011	SG	Coaching	3	positiv	keine Angabe
				Basisprogramm	3	negativ	nicht sign.
Prey H.	2000	1996 - 1998	SG	Sprachkurs	3	positiv	sign.
				Fachkurs	3	negativ	nicht sign.
				Fachkurs	3	negativ	nicht sign.
Rüfenacht M. und M. P. Neuenschwander	2013	2010 - 2012	CH	Coaching	1	positiv	nicht sign.
Tucci J.	2006	2000 - 2004	CH	Gesamtwirkung	3	positiv	nicht sign.

Studien, die bzgl. Wirkungskanäle und Empfehlungen ausgewertet wurden

Autoren	Erscheinungs- jahr	Untersuchungs- periode	Regionale Abdeckung	AMM-Typ	Gruppe
Amt für Wirtschaft und Arbeit Kanton AG (<i>internes Do- kument</i>)	2017	-	AG	-	Qual. / Desk.
Arni P.	2011	1998 - 2003	CH	Diverse	Bereits abgedeckt
Baumann M., V. Chevrier und C. Morier	2015	2013	CH	PvB Praktika	Qual. / Desk.
Behncke S., M. Frölich und M. Lechner	2006	2005	CH	Diverse	Bereits abgedeckt
Bieri O., H. Amberg und F. Dietrich	2016	2014 - 2016	LU	PvB	Qual. / Desk.
Bonoli G.	2017	2015	CH	Fachkurs EAZ PvB	Bereits abgedeckt
Cottier L., P. Kempeneers, Y. Flückiger und R. Lalive	2017	2006 - 2007	GE	Andere	Ökonometrisch
Cottier L., P. Kempeneers, Y. Flückiger und R. Lalive	2015	2006 - 2007	GE	Andere	Ökonometrisch
Egger, Dreher und Partner AG	2015	2013 - 2014	AG	Diverse	Ökonometrisch
Egger M., U. Birchler, T. Dreher, C. Lenz, V. Merckx und T. Studer	2000	1999	CH	Diverse	Ökonometrisch
Empiricon	2017	2017	CH	Diverse	Qual. / Desk.
Guggisberg J., B. Baumann und E. Mey	2013	2011 - 2013	LU	EAZ	Ökonometrisch
Jurt L. und C. Daigler	2011	2007 - 2011	ZH	PvB	Qual. / Desk.
Känzig M.	2011	2006 - 2008	FR	Andere	Qual. / Desk.
Künzi K., S. Schmutz und S. Strub	2007	2005 - 2007	SG	EAZ / Coaching	Qual. / Desk.

Lalive R., A. Osikominu, D. Liechti, M. Morlok, J. Zuchuat und J. Zweimüller	2017	2012 - 2014	CH	Diverse	Ökonometrisch
				Fachkurs	
Liechti F., F. Fossati, G. Bonoli und D. Auer	2017	2015	CH	EAZ	Ökonometrisch
				PvB	
Morger M., M. Rudin, H. Stutz, J. Jäggi und V. Legler	2017	-	AG	Diverse	Qual. / Desk.
Morlok M. und D. Liechti	2018a	2017	AG, SZ, ZH	Diverse	Qual. / Desk.
Morlok M. und D. Liechti	2018b	2015 - 2017	SO	Basisprogramm	Qual. / Desk.
Mühlebach C.	2016	2015 - 2016	CH	PvB	Qual. / Desk.
Neuenschwander M.P.	2014	2010 - 2012	CH	Coaching	Bereits abgedeckt
Notter D.	2011	-	AG	Diverse	Qual. / Desk.
Oegerli T.	2010	2010	SG	PvB	Qual. / Desk.
Ragni T.	2007	-	CH	Diverse	Bereits abgedeckt
Schaffhausen, Kantonales Arbeitsamt	2014	2012 - 2014	SH	Andere	Qual. / Desk.
Schaffhausen, Kantonales Arbeitsamt	2013	2012 - 2013	SH	Andere	Qual. / Desk.
Schallberger P. und B. Wyer	2009	2007 - 2009	TG	PvB	Qual. / Desk.
Schläfli K.	2010	2009 - 2010	BE	PvB	Qual. / Desk.
Schmugge S., K. Künzi und J. Guggisberg	2007	-	SG	Diverse	Qual. / Desk.
Stastny Hoffet M. und B. Zimmermann	2014	2013	BS	PvB	Qual. / Desk.
Stremlow J. und B. Trommsdorff	2013	2012 - 2013	ZH	Andere	Qual. / Desk.
Thurnherr G. und W. Fuchs	2012	2011 - 2012	LU, AG	Fachkurs	Qual. / Desk.

5.3. Workshop Liste der Teilnehmenden

Name	Vorname	Institution	Funktion / Rolle
Allemann	Rolf	AWA Solothurn	Leiter RAV Olten
Ceppi	Mario	SECO	Mitglied Begleitgruppe
d'Alessandro	Prisca	Arbeitsintegration Schweiz	Geschäftsleiterin
Degen	Katharina	SECO	Mitglied Begleitgruppe
Denoth	Sascha	KIGA Baselland	Stv. Leiter RAV Oberwil
Fischer	Gabriel	Travail.Suisse	Mitglied AK ALV / Begleitgruppe
Flückiger	Stephan	AMA Freiburg	Leiter ORP Nord
Gasser	Martin	SECO	Begleitgruppe (Stv.)
Huwiler	Seraina	VSAA	Stv. Direktorin
Käch	Annamarie	wira Luzern	Leiterin RAV Pilatus
Karrer	Martin	AWA Zürich	Leiter LAM, Mitglied Begleitgruppe
Leist	Stefan	SECO	Mitglied Begleitgruppe
Liechti	David	B,S,S.	Projektteam
Lützelshwab	Daniella	SAV	Mitglied Begleitgruppe
Morlok	Michael	B,S,S.	Projektteam
Moser	Nathanael	B,S,S.	Projektteam
Pagelli	Maja	AWA St. Gallen	Leiterin LAM
Simonet	Arnaud	AWA Basel-Stadt	Leiter LAM
Soncini	Lorenza	SDL Tessin	LAM Produkteleiterin
Zumofen	Alain	DIHA Wallis	Sektionschef LAM

Staatssekretariat für Wirtschaft (SECO)

Holikofenweg 36, CH-3003 Bern

Tel 031 322 42 27, Fax 031 323 50 01

www.seco.admin.ch, seco@seco.admin.ch

Eidgenössisches Departement für Wirtschaft, Bildung und Forschung WBF