

Evaluation von GM-Implementationsprozessen

Angela Wroblewski / Lorenz Lassnigg / Elisa Kahlhammer

1. Einleitung	82
2. Funktion und Zweck von Evaluationen	83
3. Rolle der Evaluation im GM-Implementationsprozess	86
4. Evaluationsansätze und ihre Eignung für die Evaluation von GM-Implementationsprozessen	88
4.1 Analyse von Wirkungen: Experimentelle und quasi-experimentelle Evaluationsdesigns	88
4.2 Responsive Evaluation	90
4.3 Empowerment-Evaluation	91
5. Herausforderungen bei der Evaluation von GM-Implementationsprozessen	92
5.1 Konkret formulierte Zielsetzung	92
5.2 Adäquates Set an Indikatoren	93
5.3 Hohe Anforderungen an Know-how von EvaluatorInnen	94
6. Schlussfolgerungen	95
7. Literatur	96

Evaluation von GM-Implementationsprozessen

Angela Wroblewski / Lorenz Lassnigg / Elisa Kahlhammer

1. Einleitung

Die Implementierung von GM in Institutionen oder Organisationen zielt im Allgemeinen darauf ab, bestehende geschlechtsspezifische Ungleichheiten abzubauen. Ausgehend von einer umfassenden Analyse der Ausgangssituation wird eine Strategie entwickelt, um Strukturen und Rahmenbedingungen so zu verändern, dass sich für Frauen und Männer gleiche Partizipationschancen ergeben. Es geht also um die Veränderung von Strukturen und Prozessen innerhalb der Institution, die geschlechtsspezifische Benachteiligungen bedingen oder die Reproduktion tradiert Rollenbilder begünstigen. Ein zentrales Element des Implementationsprozesses ist dessen Evaluation, die – so sie entsprechend konzipiert und durchgeführt wird – einen wesentlichen Beitrag zur Qualität des Implementationsprozesses leisten kann.

Die Evaluation kann im Zuge eines GM-Implementationsprozesses unterschiedliche Aufgaben oder Funktionen übernehmen: Diese kann darin bestehen, die Wirksamkeit der gesetzten Maßnahmen im Nachhinein zu überprüfen und Informationen für die weitere Entwicklung des GM-Prozesses zu liefern. Die Evaluation kann in einem begleitenden Setting als integraler Bestandteil des Implementationsprozesses verankert werden, um sicherzustellen, dass die Implementation entsprechend der Konzeption erfolgt. Evaluation kann aber auch ex ante, d.h. bereits im Planungsprozess, durchgeführt werden, um bei gegebenen Rahmenbe-

dingungen zu erwartende Wirkungen der geplanten Intervention abzuschätzen.

Je nachdem welche Aufgabe der Evaluation zukommt, sind jeweils andere Evaluationszugänge und -modelle geeignet. Die spezifischen Charakteristika eines GM-Implementationsprozesses stellen die Evaluation vor nicht unbeträchtliche Herausforderungen. So stellt sich neben der nach wie vor häufig unzureichenden Datengrundlage für Gender-Analysen auch die Frage, welcher methodische Zugang am besten geeignet ist, um die für GM zentralen Umsetzungs- und Wirkungsdimensionen zu erfassen, wie z.B. die von Ursula Rosenbichler und Karl Schörhuber angesprochenen Veränderungen auf der Tiefenstruktur von Organisationen (vgl. Rosenbichler & Schörhuber 2007, S. 16).

Im vorliegenden Beitrag liegt der Schwerpunkt auf methodischen Aspekten der Evaluation von GM-Implementationsprozessen, ohne auf deren Ausgestaltung inhaltlich näher einzugehen. Konkret wird zunächst die Funktion der Evaluation in einem idealtypischen GM-Implementationsprozess präzisiert und ein Überblick über unterschiedliche Evaluationsansätze gegeben. Im Anschluss daran werden ausgewählte Evaluationsansätze dahingehend diskutiert, inwieweit sie für die Evaluation eines GM-Implementationsprozesses geeignet erscheinen bzw. welche Anforderungen nicht gelöst werden können. Dabei erfolgt aus pragmatischen Gründen eine Konzentration auf Ansätze, die entweder häufig verwendet werden oder aber für die Evaluation von GM besonders geeignet erscheinen.

¹ Wolfgang Beywl (1988) entwickelt eine Typologie mit acht Evaluationsmodellen, die Offenheit versus Geschlossenheit sowie Tatsachen- versus Wertorientierung als Kriterien heranzieht.

Stufflebeam (2001) unterscheidet 22 Modelle, die er in drei Kategorien einteilt: Pseudo- und Quasi-Evaluationen, verbesserungs- und rechenschaftslegungsorientierte sowie sozial engagierte bzw. an-

2. Funktion und Zweck von Evaluationen

Unter Evaluation wird die „systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes“ verstanden (DeGEval-Standards 2002, S. 11). Nach dem diesem Beitrag zugrunde liegenden Verständnis handelt es sich bei Evaluation um eine wissenschaftsbasierte Dienstleistung, die insbesondere immaterielle Gegenstände unabhängig, systematisch und datengestützt beschreibt und bewertet, sodass Beteiligte und Betroffene die erzeugten Ergebnisse für ausgewiesene Zwecke nutzen können (vgl. Beywl 1999). Dieses Verständnis von Evaluation als wissenschaftsbasierte Dienstleistung verdeutlicht auch die Abgrenzung zu Wissenschaft und Forschung. Evaluation ist vielmehr ein anwendungsbezogenes Derivat, das Wissen und Methoden aus unterschiedlichen Disziplinen verknüpft. Der Evaluation geht es primär um eine durch wissenschaftliche Untersuchungsmethodik vorbereitete Praxiseinwirkung, während empirisch-wissenschaftliche Forschung allgemeingültige Erkenntnisse gewinnen will (Bortz & Döring 1995, S. 29). Evaluation verfolgt damit andere Zwecke als Forschung. Durch Evaluation bereitgestelltes Wissen stellt eine Ressource für die Praxis dar und soll instrumentell genutzt werden (vgl. Patton 1997 und Weiss 1975). Beywl (1999, S. 35) bringt den Unterschied zwischen Evaluation und Forschung folgendermaßen auf den Punkt: *„Forschung und Evaluation sind wissenschaftliche Praktiken mit identischen Anforderungen an die Genauigkeit der Informationsgewinnung, aber je eigenem Geltungsanspruch*

an die Reichweite der bereitgestellten Erkenntnisse und die hierauf abzielenden Untersuchungsanwendungen.“

In den letzten Jahrzehnten hat sich ein breites Spektrum unterschiedlicher Evaluationsansätze und ein umfassendes Set an methodischen Zugängen herausgebildet. Es gibt mittlerweile mehrere Versuche, diese Vielfalt zu klassifizieren und fassbar zu machen.¹ In den klassischen Evaluationslehrbüchern wird das Spektrum der Evaluationsansätze üblicherweise nach folgenden Kriterien eingeteilt (vgl. z.B. Stockmann 2006; Rossi, Freeman & Lipsey 1999; Weiss 1998; Beywl 1988):

- ▶ Nach dem Zeitpunkt im „Lebenszyklus eines Programms oder einer Maßnahme“ zu dem die Evaluation einsetzt (ex-ante, begleitend, ex-post).
- ▶ Nach der Funktion, die der Evaluation zukommt (proaktiv, klärend, interaktiv, dokumentierend, wirkungsfeststellend).
- ▶ Nach dem gewählten methodischen Zugang (qualitativ, quantitativ).
- ▶ Nach der Positionierung der Evaluation zur Institution oder Organisation (intern, extern).

Einteilung nach dem Zeitpunkt im Lebenszyklus

Eine ex-ante Evaluation wird vor Einführung einer Maßnahme durchgeführt. Dabei werden die erwarteten Wirkungen sowie die dem zugrunde liegenden Wirkungsmechanismen expliziert und einer Prüfung unterzogen. Es geht darum festzustellen, ob die getroffenen Annahmen realistisch sind und ob im Falle

wissenschaftliche Evaluationen. Wolfgang Beywl (2006) klassifiziert Evaluationsmodelle nach ihrem Umgang mit Werten und unterscheidet vier Gruppen von Evaluationsmodellen: Wertedistanzierte,

werterealistische, wertepriorisierende und wertepositionierte Modelle.

der zutreffenden Annahmen die angestrebten Wirkungen bei gegebenen Rahmenbedingungen tatsächlich realisiert werden können. Es wird also nicht nur das Wirkungsmodell einer Prüfung unterzogen, sondern dieses in den Kontext der gegebenen Rahmenbedingungen gestellt. Zielsetzung einer ex-ante Evaluation ist es, unrealistische Annahmen möglichst frühzeitig aufzuzeigen und ggf. eine Veränderung der zentralen Parameter der Maßnahme vorzunehmen oder eine Umformulierung des Wirkungsmodells durchzuführen. Im Gegensatz dazu wird bei einer ex-post Evaluation (häufig auch als „summative Evaluation“ bezeichnet) im Nachhinein nach den durch die Maßnahme erzielten Wirkungen gefragt. Es wird versucht, möglichst alle durch die Maßnahme hervorgerufenen Wirkungen zu erfassen, d.h. nicht nur intendierte sondern auch nicht-intendierte Wirkungen in die Analyse einzubeziehen. Zielsetzung ist zumeist eine abschließende Bewertung, die nach Ablauf einer bestimmten Umsetzungsfrist durchgeführt wird, um fundiert Entscheidungen über die Fortführung, den Ausbau oder die Adaptierung eines Programms treffen zu können.

Eine begleitende Evaluation (häufig auch als „formative Evaluation“ bezeichnet) erfolgt dagegen in der Einführungs- oder Durchführungsphase einer Maßnahme. Aufgabe der Evaluation ist es, die Umsetzung dahingehend zu analysieren, ob sie plangemäß erfolgt bzw. wo sich in der Umsetzung der Maßnahme Abweichungen vom Plan ergeben. Häufig sind Abweichungen von der Planung im Zuge der Umsetzung auf nicht ausreichend konkret oder widersprüchlich formulierte Zielvorgaben zurückzuführen oder aber auf Veränderungen der Rahmenbedingungen. So kann es beispielsweise in der Umsetzung dazu kommen, dass aufgrund eines schwierigen Zugangs zur angestrebten Zielgruppe diese nicht durch die Maßnahme angesprochen wird oder aber bestimmte Gruppen innerhalb der Zielgruppe unterrepräsentiert bleiben. Zielsetzung ist es, bei Abweichungen von der Planung möglichst frühzeitig Rückmeldung an die verantwortlichen Stellen zu geben, um eine Neuausrichtung der Umsetzung zu erreichen und damit die Zielerreichung zu unterstützen.

Einteilung nach Funktion der Evaluation

Im Hinblick auf die Funktionen einer Evaluation unterscheiden Owen und Rogers (1999) fünf unterschiedliche Evaluationsarten. Die proaktive Evaluation führt vor dem Start einer Maßnahme eine Bedarfsanalyse

und eine Machbarkeitsstudie durch und versucht durch Simulationen die Resultate der Maßnahme abzuschätzen. Eine klärende Evaluation überprüft in der Konzeptphase eines Programms und vor Beginn der Umsetzung die Stimmigkeit und Prägnanz der Programmziele und gegebenenfalls deren Anpassung an bestehende Bedarfs- und Problemlagen. Es erfolgt eine Konkretisierung und Operationalisierung der Zielsetzungen und damit eine wesentliche Grundlage für die Wirkungsanalyse. Diese beiden Evaluationsarten entsprechen der oben beschriebenen ex-ante Evaluation. Die interaktive Evaluation nach Owen und Rogers entspricht im Wesentlichen der begleitenden Evaluation. Diese läuft parallel zur Programmumsetzung und liefert Ergebnisse zur Prozessqualität und unterstützt die Feinabstimmung von Zielen, gegebenenfalls notwendige Anpassungen des Programms an unterschiedliche Rahmenbedingungen (z.B. bei verschiedenen Standorten) und macht Vorschläge zur Neuausrichtung des Programms. Es sollen kurzfristig Rückmeldungen an die Programmverantwortlichen gegeben werden, um gegebenenfalls möglichst frühzeitig notwendige Adaptierungen vornehmen zu können. Unter dokumentierender Evaluation verstehen Owens und Rogers den Aufbau eines Monitoringsystems, d.h. die Formulierung eines Indikatorensets, das über die Umsetzung des Programms in regelmäßigen Abständen informiert. Und last but not least wird die wirkungsfeststellende Evaluation angeführt, die nach der Zielerreichung fragt und inwieweit das erzielte Ergebnis auf die Maßnahme zurückzuführen ist.

Einteilung nach der methodischen Vorgangsweise

In den meisten Fällen können den beschriebenen Ansätzen aufgrund der zentralen Fragestellung relativ eindeutig qualitative oder quantitative Methoden zugeordnet werden. So basiert eine ex-post Evaluation mit Schwerpunkt auf Wirkungsanalyse zumeist auf einer Analyse quantitativer Daten (z.B. Analyse von Daten eines Monitoringsystems), während bei einer begleitenden Evaluation typischerweise qualitative Verfahren angewendet werden (z.B. ExpertInneninterviews, Beobachtung, Dokumentenanalyse). Bei einer ex-ante Evaluation steht die Abschätzung realistischerweise zu erwartender Wirkungen einer Intervention im Vordergrund. Dabei werden zumeist auf Basis vorliegender Strukturdaten sowie ausformulierter und begründeter Annahmen über Wirkungsmechanismen, die der In-

tervention zugrunde liegen, Prognosen über die zu erzielenden Wirkungen erstellt. Die Formulierung der Annahmen basiert häufig auf qualitativen Verfahren, wie beispielsweise ExpertInneninterviews, Fokusgruppendifkussionen, Dokumenten- und Literaturanalyse.

Einteilung nach der Positionierung der Evaluation in der Organisation

Die Evaluation kann intern (d.h. durch Angehörige der Institution selbst) oder aber durch externe ExpertInnen durchgeführt werden. Beide Vorgangsweisen haben Vor- und Nachteile, die es im jeweiligen Fall gegeneinander abzuwägen gilt. Auf der einen Seite verfügen Angehörige der Institution über umfassende Kenntnisse über die Struktur, Prozessabläufe etc.; sie sind in die spezifischen Themenbereiche stärker eingearbeitet als externe EvaluatorInnen und können die Umsetzungsrelevanz von Empfehlungen besser einschätzen. Auf der anderen Seite kann ein Blick von außen wertvolle neue Perspektiven einbringen und häufig weisen Empfehlungen externer ExpertInnen eine höhere Akzeptanz auf. So verfügen externe EvaluatorInnen häufig über Umsetzungserfahrungen in ähnlichen Kontexten oder Institutionen und können vor diesem Hintergrund fundierte Empfehlungen ableiten.

In diesem Kontext stellt sich die Frage, wie das Verhältnis von Evaluations-Know-how und fachlich-inhaltlichem Know-how optimal gestaltet sein sollte. Bei einer vor mehr als 25 Jahren in den USA durchgeführten Untersuchung zur Professionalisierung von Evaluation bzw. EvaluatordInnen stellte sich als ein gravierendes Problem heraus, dass Evaluationen häufig ohne ausreichende Kenntnisse und Erfahrungen im Feld durchgeführt werden. Bereits damals wurde festgestellt, dass Grundkenntnisse über den Evaluations-

gegenstand neben den methodischen Kenntnissen eine zentrale Rolle spielen (Newman & Brown 1992). Gerade im GM-Kontext scheint es schwer vorstellbar, dass eine Evaluation nur mit fachlich-methodischem Evaluations-Know-how ohne einschlägiges Fachwissen durchführbar ist. Externe EvaluatordInnen sollten neben dem methodischen Rüstzeug auch umfassende Gender-Kompetenz aufweisen und zumindest Grundkenntnisse über die jeweilige Organisation (ihre Struktur, Logik, Vision, Kultur etc.) mitbringen. Da sich nur selten all diese Kompetenzen in einer Person vereint finden, ist es oft notwendig, die Evaluation in einem Team durchzuführen, dessen Mitglieder im Kollektiv über diese Qualifikationen verfügen (vgl. dazu auch Sanders 2006).

Weiters ist Evaluation von Beratung im Zuge von Organisations- oder Strukturveränderungen strikt zu trennen. Eine Vermischung dieser beiden Funktionen würde bedeuten, dass die Evaluation Teil des Evaluandums (des zu evaluierenden Gegenstandes) würde und sich selbst evaluieren müsste. Aufgabe der Evaluation ist es, die Ergebnisse anderer begleitender bzw. reflektierender Maßnahmen zu sichten, durch eigene Erhebungen zu ergänzen und diese zueinander in Beziehung zu setzen. Weiters ist darauf zu achten, dass es nicht zu Doppelgleisigkeiten kommt und die verschiedenen Interventionen möglichst klar abgegrenzte Bereiche betreffen (vgl. zur Abgrenzung zwischen Beratung und Evaluation auch Zepke 2005). Die Kombinationsmöglichkeiten der bisher beschriebenen Elemente sind vielfältig und schlagen sich in dem angesprochenen breiten Spektrum unterschiedlicher Evaluationsansätze nieder, wobei sich auch Überschneidungen bzw. Abgrenzungsprobleme ergeben, wie die folgende Übersicht verdeutlicht.

Kriterium	Ausprägungen		
Zeitpunkt	Ex-ante	Begleitend (formativ)	Ex-post (summativ)
Funktion	Proaktiv Klärend	Interaktiv Dokumentierend	Wirkungsfeststellend
Methode	Qualitativ	Eher qualitativ	Eher quantitativ
Position der Evaluation	Eher intern	Intern / extern	Eher extern

Übersicht 1: Spektrum der Evaluationsansätze

Im Folgenden wird die mögliche Rolle der Evaluation im GM-Implementationsprozess diskutiert und werden Evaluationsfragestellungen für die einzelnen Implementationschritte formuliert. Daran anschließend werden unterschiedliche Evaluationsmodelle vorgestellt, die zur Beantwortung typischer Fragen für den GM-Implementationsprozess geeignet erscheinen. Weiters wird gezeigt, warum die am häufigsten verwendeten Evaluationsmodelle für die Evaluation von GM-Implementationsprozessen den AutorInnen nicht geeignet erscheinen. In einem dritten Schritt werden ausgewählte Aspekte der Evaluation angesprochen, die nicht nur im GM-Kontext – aber hier besonders – Herausforderungen darstellen.

3. Rolle der Evaluation im GM-Implementationsprozess

GM-Implementationsprozesse zielen im Wesentlichen auf eine Veränderung von bestehenden Strukturen ab, die geschlechtsspezifische Benachteiligungen bedingen. Gender Mainstreaming ist also eine Strategie, die zur Gleichstellung der Geschlechter in allen Lebensbereichen führen soll. Diese Strategie impliziert eine weitreichende Veränderung von Strukturen und Prozessen. *„Gender mainstreaming is the (re)organisation, improvement, development and evaluation of policy processes, so that a gender equality perspective is incorporated in all policies at all levels and at all stages, by the actors normally involved in policy-making.“* (Europarat 1998; zitiert nach Directorate General of Human Rights 2004, S. 48).

Das bedeutet, dass der GM-Implementationsprozess strukturelle Veränderungen auslösen soll und damit Gender Mainstreaming-Aktivitäten und Organisationsentwicklung eng miteinander verzahnt sind. Dieses extrem komplexe Setting stellt für die Evaluation insbesondere dann eine Herausforderung dar, wenn kausale Effekte ermittelt werden sollen, d.h. wenn nachgewiesen werden soll, dass die festgestellten Veränderungen auf die GM-Aktivitäten zurückzuführen sind.

In den meisten Handbüchern oder Leitfäden zur GM-Implementation (vgl. z.B. Stiegler 2001 oder auch

Bergmann & Pimminger 2004, S. 27) besteht der GM-Prozess aus vier, aufeinander aufbauenden Schritten: Gender-Analyse, Festlegung von Gleichstellungszielen, Umsetzung und Evaluation. Zunächst wird eine Analyse der Ausgangssituation durchgeführt, d.h. eine Bestandsaufnahme gemacht, in der bestehende strukturelle Benachteiligungen erfasst werden. Auf dieser Basis werden konkrete Zielsetzungen formuliert, die durch die GM-Implementierung erreicht werden sollen. Die angestrebten Ergebnisse sollten möglichst konkret ausformuliert sein, wobei es hier die Gratwanderung zu schaffen gilt, keine zu allgemein formulierten Zielsetzungen zu definieren, gleichzeitig aber auch Zielsetzungen nicht ausschließlich zu quantifizieren. Im Extremfall kann ersteres dazu führen, dass bestehendes Verhalten unter den GM-Vorgaben subsumiert werden kann, d.h. es zu keiner Veränderung des Verhaltens kommt und bestehende Rollenbilder verstärkt werden. Im zweiten Fall kann die starke Konzentration auf die Erfüllung von Kennzahlen dazu führen, dass die inhaltliche Dimension von GM unter den Tisch fällt (vgl. zu diesen Problematiken Leitner 2005).

Auf dieser Basis werden konkrete Umsetzungsschritte definiert, für die im Idealfall ein Monitoringsystem aufgebaut wird, das laufend einen Überblick über den aktuellen Stand der Umsetzung der Maßnahmen gibt. Als abschließender Schritt wird schließlich die Evaluation der durchgeführten Maßnahmen gesehen. Die durchgeführten Maßnahmen werden dahingehend analysiert, ob sie die intendierten Zielsetzungen erreicht haben bzw. welche Veränderungen gegenüber der Ausgangslage verzeichnet werden können. Gleichzeitig stellt dies eine neuerliche Analyse der Ausgangslage dar und der Prozess beginnt von Neuem („GM-Spirale“ nach Bergmann & Pimminger 2004, S. 28).

Die Evaluation ist als einer der vier Schritte also integraler Bestandteil des GM-Prozesses, wodurch ihr auch eine zentrale Rolle für dessen Weiterentwicklung zukommt. Die Aufgaben bzw. Funktionen von Evaluation bleiben damit jedoch tendenziell eher auf eine ex-post Bewertung des jeweiligen Zyklus ausgerichtet. Eine ernsthafte Evaluation sollte sich jedoch nicht auf diese Rolle beschränken, sondern vielmehr eine Querschnittsfunktion im gesamten GM-Prozess einnehmen, d.h. systematisch in jedem Implementa-

tionsschritt einbezogen werden. In Übersicht 2 sind typische Fragestellungen angeführt, die dabei von der Evaluation behandelt werden können.

Einen Schritt in diese Richtung gehen Rosenbichler und Schörghuber (2003), die die 4-Schritte-Spirale durch die Anwendung eines Organisationsentwicklungsansatzes auf 8 Schritte erweitern. Als Zwischenschritt zwischen der Festlegung von Zielen und der Entwicklung von Maßnahmen wird die Entwicklung von Hypothesen über Diskriminierungsmechanismen verankert, die in die Ausarbeitung von Maßnahmen einfließen sollen. Der Ausarbeitung von Maßnahmen folgt eine ex-ante Evaluation, indem die voraussichtlichen Wirkungen unterschiedlicher Maßnahmen analysiert werden sowie die Entscheidung für eine bestimmte Intervention getroffen wird. Durch diese Weiterentwicklung wurden - neben der Überprüfung der Wirkungen als vorläufig letzten Schritt des Prozesses - weitere Evaluationsschritte in den Prozess eingebaut (für eine detailliertere Darstellung siehe den Beitrag von Ursula Rosenbichler und Karl Schörghuber in Band 2 dieser Schriftenreihe, „Grundlagen“).

Die geforderte systematische Integration von Evaluationsfragestellungen in alle Schritte des GM-Implementationsprozesses könnte zu einer Qualitätssicherung und -verbesserung des Prozesses beitragen, da nicht intendierte Entwicklungen frühzeitig erkannt und rückgemeldet werden. Gleichzeitig stellt die Evaluation durch die systematische Begleitung des Imple-

mentationsprozesses Informationen bereit, die möglichst viele unterschiedliche Perspektiven berücksichtigen und so eine breite und fundierte Informationsbasis für EntscheidungsträgerInnen darstellt. So trägt die Evaluation beispielsweise zu einer umfassenden Gender-Analyse durch die Mitarbeit am Aufbau entsprechender Datengrundlagen, durch die systematische Analyse verfügbarer Informationen sowie durch die Entwicklung aussagekräftiger Indikatoren bei.

In die Formulierung der Gleichstellungsziele kann die Evaluation in moderierender Weise eingebunden werden. Zum einen kann die Evaluation einen Rahmen schaffen, indem alle relevanten Beteiligten (z.B. Leitung, Umsetzende, Frauen- bzw. GM-Beauftragte) gemeinsam Gleichstellungszielsetzungen entwickeln, die von allen mitgetragen werden. Die Evaluation kann aber auch Rückmeldung zur Zielformulierung im Hinblick darauf geben, ob GM-Zielsetzungen in sich widersprüchlich sind oder im Widerspruch zu anderen Zielsetzungen der Institution stehen. In einer erweiterten Variante könnte die Evaluation auch prüfen, inwieweit GM-Zielsetzungen innerhalb einer Institution mitgetragen bzw. allgemein akzeptiert sind.

Das Vorliegen einer aussagekräftigen Gender-Analyse und eine auf dieser Basis entwickelte Zielsetzung sind Voraussetzung für die Entwicklung konkreter Umsetzungsschritte. Wurde im Zuge der Gender-Analyse eine aussagekräftige Datenbasis bzw. ein entsprechendes Indikatorensystem entwickelt, so

GM-Implementationschritte	Evaluationsfragestellungen
Gender Analyse (Analyse der Ausgangslage)	Sind alle genderrelevanten Dimensionen berücksichtigt? Wird auf Gender-Theorie Bezug genommen? Sind die Schlussfolgerungen schlüssig? Sind die relevanten Kontextbedingungen und deren Gender-Relevanz berücksichtigt?
Festlegung von Gleichstellungszielen	Sind die Zielsetzungen gleichstellungsorientiert formuliert? Sind zugrunde liegende Wirkungsmechanismen schlüssig?
Umsetzung	Erfolgt die Umsetzung entsprechend der Planung? Zeigen sich nicht intendierte geschlechtsspezifische Selektionseffekte (z.B. im Auswahlverfahren für die Teilnahme)? Ist die Zielerreichung möglich? Verändern sich Rahmenbedingungen während der Umsetzungsphase? Wenn ja, sind davon beide Geschlechter gleichermaßen betroffen?
Evaluation	Evaluation im engeren Sinn: Analyse der Zielerreichung.

Übersicht 2: GM-Implementations-Schritte und typische Evaluationsfragestellungen

stellt dies die Grundlage für die Entwicklung eines internen Monitoring-Systems dar. Auf Basis eines solchen Monitoring-Systems ist es laufend und mit wenig Aufwand möglich, einen Überblick über die aktuelle Umsetzung zu bekommen. Dadurch können unvorhergesehene Abweichungen von der ursprünglichen Konzeption erkannt und entsprechende Anpassungsschritte eingeleitet werden. Ebenso werden dadurch nicht intendierte Effekte (wie z.B. die Verstärkung traditioneller Rollenbilder in nicht unmittelbar durch die Maßnahme adressierten Bereichen) frühzeitig erkennbar. Diese beschriebenen Schritte sind schließlich die zentrale Voraussetzung für eine ex-post Bewertung der Zielerreichung nach Ablauf einer bestimmten Zeit.

Wird die Einbeziehung der Evaluation in alle GM-Implementierungsschritte konsequent durchgeführt, so ergibt sich ein komplexes Evaluationsdesign, das Elemente der ex-ante, begleitenden und ex-post Evaluation beinhaltet. Der zentrale Bereich liegt unseres Erachtens jedoch auf den begleitenden Elementen, die eine regelmäßige Feedbackschleife für den Gesamtprozess ermöglichen und damit die Voraussetzungen für eine ständige kritische Reflexion des Prozesses durch die Angehörigen der Institution selbst schaffen.

Der Anspruch, die Evaluation begleitend zur Konzepterstellung und Umsetzung von GM-Maßnahmen zu gestalten, stellt eine Reihe von Anforderungen an die EvaluatorenInnen aber auch an die Umsetzenden der Maßnahme. So müssen die EvaluatorenInnen in alle Entwicklungs- und Umsetzungsschritte einbezogen sein oder zumindest detailliert darüber informiert werden. Dies erfordert fachliches Grundwissen der EvaluatorenInnen in den jeweiligen Bereichen, die gegenseitige Akzeptanz von EvaluatorenInnen und Umsetzenden und last but not least Genderkompetenz auf beiden Seiten. Insgesamt erhöht sich dadurch der Entwicklungs- und Konzeptionierungsaufwand für die Evaluation, der sich allerdings dann relativiert, wenn dies als ein Beitrag zur Qualitätssicherung des Prozesses verstanden wird.

4. Evaluationsansätze und ihre Eignung für die Evaluation von GM-Implementationsprozessen

Es stellt sich nun die Frage, welcher Evaluationsansatz am besten geeignet ist, um die im vorherigen Kapitel beschriebenen Fragestellungen im Kontext der GM-Implementierung zu beantworten.

Im Folgenden wird zunächst auf jenes Design eingegangen, das am häufigsten mit Evaluation assoziiert wird, nämlich die Analyse der Wirkungen von Interventionen. Diskutiert wird, welche Voraussetzungen für eine Wirkungsanalyse im Kontext von GM-Implementationsprozessen gegeben sein müssen bzw. welche Herausforderungen sich in der Praxis stellen. Im Anschluss daran wird das Modell der responsiven Evaluation beschrieben, das prozessbegleitend angelegt ist und auf die Weiterentwicklung der bestehenden Praxis abzielt. Als dritter Ansatz wird die Empowerment-Evaluation vorgestellt, die zwar hohe Anforderungen an die Institutionen stellt, aber gerade im GM-Kontext großes Potential beinhaltet.

4.1 Analyse von Wirkungen: Experimentelle und quasi-experimentelle Evaluationsdesigns

Lange Zeit hindurch wurde die Feststellungen von Wirkungen einer Maßnahme oder eines Programms im Nachhinein (ex post) als primäres Anwendungsgebiet der Evaluation gesehen. Im Bereich der Analyse der Wirkungen von Interventionen hat sich auch ein relativ einheitlicher und anerkannter Mindeststandard im Hinblick auf die methodische Vorgangsweise herausgebildet. Um die Frage nach den Wirkungen, die durch eine Intervention entstanden sind, beantworten zu können, gelten experimentelle Designs als einzig wirklich adäquater Zugang, da nur sie es erlauben, kausale Effekte zu ermitteln. Die Kausalität zwischen Intervention und Ergebnis wird in vielen Fällen zwar unterstellt, aber nur selten tatsächlich nachgewiesen. Dieser Nachweis ist v.a. dann wichtig, wenn die durch eine Intervention erzielten Ergebnisse als Grundlage dafür verwendet werden, die Interventionen weiterzu-

² Bei quasi-experimentellen Designs wird von „Vergleichsgruppen“ gesprochen, beim Experiment spricht man von „Kontrollgruppen“.

³ Es handelt sich dabei um ein fiktives Beispiel.

führen oder weiter zu entwickeln. Diese Entscheidungen gehen implizit davon aus, dass die Ergebnisse durch die Intervention hervorgerufen wurden und nicht durch andere Faktoren (wie z.B. geänderte Rahmenbedingungen oder Interventionen in anderen Bereichen). Allerdings sind experimentelle Evaluationsdesigns insbesondere im Bereich der Sozialwissenschaften aufgrund der verfügbaren Datenlage, praktischen und ethischen Vorbehalten sowie des komplexen Settings an Einflussfaktoren nur selten anzutreffen (vgl. Cook 2002). Als die am besten geeignete Alternative gelten quasi-experimentelle Forschungsdesigns, die auch in den sozialwissenschaftlichen Anwendungsgebieten der Evaluation mittlerweile zum anerkannten Mindeststandard zählen (vgl. Shadish, Cook & Campell 2002).

Bei experimentellen Evaluationsdesigns wird nach einem Zufallsprinzip eine TeilnehmerInnen- und eine Vergleichsgruppe gebildet (z.B. wenn BewerberInnen für die Teilnahme an einer Maßnahme nach einem Zufallsprinzip – z.B. Münzwurf – einer TeilnehmerInnen- und einer Kontrollgruppe zugeordnet werden). Ein solches Design ist außerhalb der Medizin nur selten anzutreffen. In den Sozialwissenschaften kommen eher quasi-experimentelle Designs vor. Der zentrale Unterschied zum klassischen Experiment liegt darin, dass eine Vergleichsgruppe² nicht nach einem Zufallsprinzip gebildet, sondern nach unterschiedlichen Verfahren konstruiert wird, wie z.B. durch ein Matching-Verfahren. Zielsetzung dieses Verfahrens ist es, eine Vergleichsgruppe zu bilden, die der TeilnehmerInnen-Gruppe in den zentralen Charakteristika möglichst ähnlich ist. Sowohl ein experimentelles wie auch ein quasi-experimentelles Evaluationsdesigns sollte möglichst von Beginn an im Rahmen der Maßnahmenkonzeption mitgeplant werden. In der Praxis wird jedoch nur in den seltensten Fällen die Evaluation bereits in einem derart frühen Stadium mitgedacht.

Wie bereits erwähnt sind experimentelle Forschungsdesigns in den Sozialwissenschaften nur selten anzutreffen, wobei potentielle Anwendungsfelder aufgrund

des hohen Aufwandes, der mit Experimenten verbunden ist, nicht genutzt werden. Ein möglicher Anwendungsfall für ein experimentelles Design wäre z.B. die Frage nach den Wirkungen der Einführung von Gender-Trainings für BeraterInnen des Arbeitsmarktservice (AMS).³ Diese Trainings könnten als Maßnahme in einem GM-Implementationsprozess vorgesehen sein. Nachdem es sich um eine umfassende Maßnahme handelt, ist es wahrscheinlich nicht möglich, diese auf einen Schlag flächendeckend in allen regionalen Geschäftsstellen (RGSn) gleichzeitig einzuführen. Vielmehr ist anzunehmen, dass die Trainings in den einzelnen Geschäftsstellen nacheinander durchgeführt werden, die Intervention also stufenweise eingeführt wird. Ein experimentelles Design bietet sich immer bei der stufenweisen Einführung einer Intervention an, wobei die Reihenfolge, in der die Trainings durchgeführt werden, auf Basis eines Zufallsmechanismus erstellt wird. Es werden nach einem Zufallsprinzip zwei Gruppen gebildet: Eine Gruppe von RGSn, in denen die Trainings zuerst durchgeführt werden (TeilnehmerInnengruppe) und eine Gruppe von RGSn, in denen dies erst später der Fall ist (Kontrollgruppe).

Der Effekt der Trainings wird durch den Vergleich von TeilnehmerInnen- und Kontrollgruppe in der interessierenden Variable ermittelt (z.B. gender-gerechte Beratungstätigkeit). Es handelt sich dabei um den Vergleich der Situation nach der Maßnahme mit einer Situation, die sich ergeben hätte, wenn keine Maßnahme stattgefunden hätte, das sogenannte „Counterfactual“. Nur wenn das Counterfactual gebildet werden kann, ist es möglich kausale Effekte einer Intervention zu erfassen. Eine weitere Voraussetzung dafür ist, dass quantitative bzw. quantifizierbare Zielgrößen vorliegen, im Idealfall für mehrere Zeitpunkte vor und nach der Intervention. In dem oben angeführten Beispiel würde dies erfordern, ein Mess- oder Beobachtungsinstrument zu entwickeln, das Gender-Gerechtigkeit in der Beratung „misst“ und das bereits vor der Einführung der Trainings herangezogen wird, um die „Ausgangslage“ zu erfassen, die dann als zentraler Bewertungsmaßstab für Bewertung der Wirkungen der Trainings herangezogen wird.⁴

⁴ Ein mögliches Evaluationskriterium wäre ein Vorher-Nachher-Vergleich wie häufig in Beratungsgesprächen Situationen auftreten, in denen Rollenstereotype reproduziert bzw. gestärkt werden. Bei

einer erfolgreichen Intervention müsste nach dem Training die Häufigkeit derartiger Gesprächssituationen zurückgehen.

Ist kein experimentelles Design möglich, wird versucht, im Rahmen eines quasi-experimentellen Designs ein adäquates Counterfactual zu konstruieren (z.B. über ein Matchingverfahren). Die Herausforderung dabei ist, die Vergleichbarkeit zwischen jenen RGSn, die die TeilnehmerInnen-Gruppe stellen, und jenen in der Vergleichsgruppe sicherzustellen. Es gilt auszuschließen, dass sich diese beiden Gruppen bei zentralen Einflussfaktoren (wie z.B. Engagement, Motivation, Vorwissen) systematisch unterscheiden und so eine Vergleichbarkeit nicht gegeben ist.

Neben diesen methodischen Herausforderungen setzt eine experimentelle oder quasi-experimentelle Evaluation voraus, dass es eindeutig formulierte Zielsetzungen des Programms gibt, die von den EvaluatorenInnen übernommen werden können. Im Gegensatz zu den im folgenden beschriebenen Evaluationsdesigns sind die EvaluatorenInnen kaum in die Umsetzung und Weiterentwicklung der Intervention eingebunden. Die Zielerreichung wird anhand von quantitativen Kennzahlen gemessen, wodurch der Eindruck des „objektiven“ Charakters der Evaluation hervorgerufen wird. Fehlt eine eindeutig formulierte Zielsetzung für das Programm, so ist im Grunde eine Wirkungsanalyse nicht möglich, da nicht klar ist, woran das Ergebnis gemessen werden soll bzw. wie ein erzielt Ergebnis zu interpretieren ist. Voraussetzung für eine Wirkungsanalyse ist daher eine von vornherein von allen relevanten Stakeholdern geteilte und explizit formulierte Zielsetzung. Gerade im Zusammenhang mit der Evaluation von GM ist diese Voraussetzung häufig jedoch nicht gegeben. Vielmehr wird ganz bewusst mit einer offenen Definition von GM gearbeitet und die Zielformulierung als Teil des GM-Implementationsprozesses verstanden.

Zusammenfassend kann festgehalten werden, dass dieser Evaluationsansatz eher für die Evaluation von einzelnen Interventionen innerhalb eines GM-Implementationsprozesses geeignet ist, als für die Evaluation eines gesamten Prozesses. Letzterer ist aufgrund des komplexen Settings an Einflussfaktoren und relevanter Rahmenbedingungen kaum im Zuge eines Experiments zu evaluieren. Dies deshalb, da durch ein

experimentelles Design versucht wird, alle relevanten Rahmenbedingungen konstant zu halten und nur einige wenige Faktoren bewusst zu verändern. Es ist folglich kaum vorstellbar, für einen Gesamtprozess ein adäquates Counterfactual zu definieren. Weiters geht dieser Ansatz von einer abgeschlossenen Intervention aus und widerspricht im Grunde dem Prozesscharakter der GM-Implementierung. Als Teil des GM-Implementationsprozesses würde eine derartige Wirkungsanalyse nur dann Sinn machen, wenn die Implementationsphase beendet ist, bereits die Konsolidierungsphase begonnen hat und die Ergebnisse der Wirkungsanalyse genutzt werden, um eine neuerliche Gender-Analyse durchzuführen und den bisherigen Prozessverlauf zu reflektieren. Für die Implementationsphase ist dieser Evaluationsansatz jedoch nicht geeignet; vielmehr sollte hier auf begleitende Evaluationsansätze, wie z.B. das Modell der responsiven Evaluation, zurückgegriffen werden.

4.2 Responsive Evaluation

Das Modell der responsiven Evaluation wurde von Stake (1975) entwickelt und im deutschsprachigen Raum durch Beywl (1988) bekannt gemacht. Eine radikale Weiterentwicklung des responsiven Modells erfolgte durch Guba und Lincoln (1989).⁵

Responsive Evaluation ist in erster Linie auf die Nützlichkeit in der Praxis ausgerichtet, wobei die Weiterentwicklung und Verbesserung der bestehenden Praxis (des Programms oder der Maßnahme) im Vordergrund steht. Die Interessen und Anliegen aller am Programm und seiner Umsetzung beteiligten Personengruppen („stakeholder“) sind daher in die Evaluation explizit einzubeziehen – auch bzw. gerade wenn diese einander widersprechen.

Eine weitere Besonderheit der responsiven Evaluation ist deren konsequente Offenheit im Prozess. Die zentralen Aspekte einer Evaluation – wie z.B. die Formulierung der Fragestellung, das Evaluationsdesign und die Methodenentscheidung werden erst im Laufe des Prozesses festgelegt und sind wiederum Ergebnis eines Aushandlungsprozesses zwischen allen Beteiligten. Es gibt also keinen typischen Ablauf von

⁵ Guba und Lincoln (1989) entwickeln ihr Konzept der „Fourth Generation Evaluation“ aus dem Modell der responsiven Evaluie-

rung heraus. Sie führen jedoch zwei neue Aspekte ein: Ein konstruktivistisches Paradigma und Verhandlung („negotiation“) als

Evaluationsstudien dieses Typs. Grundprinzip ist, dass „die verschiedenen Schritte der Beschreibung und Bewertung, der Erstellung von Schlussfolgerungen und Empfehlungen flexibel an das jeweilige Evaluationsobjekt, seine lokale Umwelt und die Interessenslagen der Beteiligengruppen angepasst werden.“ (Beywl 1999, S. 247)

Aufgrund der Offenheit im Prozess kommt den EvaluatorInnen ein spezifisches Aufgaben- und Rollenset zu: Sie sind nicht nur für eine systematische Datenerhebung und Analyse zuständig, sondern sind auch in beratender und moderierender Funktion in den Entwicklungsprozess involviert. Dies primär über die Bereitstellung von Feedback und Reflexionsmöglichkeiten über den Prozess, wodurch sie jedoch in den Prozess eingreifen und diesen mitgestalten.

Durch die Evaluation soll es zu einer Verbesserung des Programms kommen, d.h. Evaluation ist eine Serviceleistung für das Programm. Aus diesem Grund stellt die Nützlichkeit der Evaluationsergebnisse das zentrale Qualitätskriterium für die Evaluation dar. Durch die starke Anwendungsorientierung und Ausrichtung an der Verbesserung eines bestehenden Programms in Verbindung mit dem Prinzip der Offenheit im Prozess erscheint dieser Ansatz eher für die Evaluation von GM geeignet, als wirkungsorientierte Ansätze. Voraussetzung dafür, dass die Nützlichkeit der Ergebnisse tatsächlich erreicht wird, ist die Einbeziehung der AdressatInnen der Evaluation (z.B. Leitung, Umsetzende, Betroffene) in den Evaluationsprozess. Diese müssen ihren Informationsbedarf formulieren sowie eine Vorstellung über die Verwendung der Ergebnisse bzw. über den Verwendungskontext haben und diese auch den EvaluatorInnen kommunizieren.

Damit kommt den EvaluatorInnen eine wesentlich stärker gestaltende Rolle zu und ihre Sichtweise auf GM fließt deutlich stärker ein, als dies im wirkungsfeststellenden Design der Fall ist. Um die Akzeptanz der Evaluation bei den AdressatInnen herzustellen, ist eine Abstimmung bzw. Annäherung der Position der EvaluatorInnen und AdressatInnen notwendig, wenn

es z.B. darum geht, ein gemeinsames Verständnis von GM zu finden oder die Zielsetzungen des Prozesses zu konkretisieren. In jenen Fällen, in denen die EvaluatorInnen auch die Funktion haben, Gender-Know-how in der jeweiligen Organisation aufzubauen, kommt diesem Abstimmungsprozess eine noch größere Bedeutung zu.

In Summe bedeutet dies, dass das Modell der responsiven Evaluation insbesondere dann geeignet ist, wenn es eine explizit formulierte GM-Strategie sowie ein klares Bekenntnis zur Weiterentwicklung des Prozesses gibt.

4.3 Empowerment-Evaluation

Der Ansatz der Empowerment-Evaluation wurde in den 1990er Jahren von David M. Fetterman entwickelt (vgl. Fetterman 2001). Es handelt sich dabei um einen Ansatz der Selbstevaluation mit externer Begleitung (durch EvaluationsexpertInnen), deren Aufgabe es ist, gemeinsam mit den Beteiligten die Evaluation zu konzipieren, das dafür notwendige Evaluations-Know-how zu vermitteln und sich im Laufe der Zeit immer mehr aus dem Evaluationsprozess zurückzuziehen. Ziel der Evaluation ist, dass die Beteiligten durch die Umsetzungserfahrungen lernen und auf dieser Basis das Programm weiterentwickeln.

Der erste Schritt einer Empowerment-Evaluation ist die Formulierung eines „Leitbildes“ („mission, vision or unifying purpose“, siehe Fetterman 2001, S. 23). Dieser Idealvorstellung wird der aktuelle Status-Quo gegenüber gestellt, wobei auf Stärken und Schwächen eingegangen wird. In einem dritten Schritt werden von allen Beteiligten gemeinsam jene Schritte entwickelt bzw. geplant, die zur Zielerreichung notwendig sind.

Die Formulierung einer gemeinsamen Zielsetzung des Programms ist Voraussetzung für die Bestandsaufnahme wie auch für die Weiterentwicklung des Programms. Liegt diesbezüglich kein Konsens vor, sind die nächsten Schritte der Evaluation sinnvollerweise nicht zu setzen. In diesem Zusammenhang nehmen externe EvaluatorInnen die Rolle eines Coaches ein.

zentralen Bestandteil von Evaluierungen.

Ihre Aufgabe ist es, den Prozess der Konsensfindung zu moderieren und zu begleiten.

Im Zuge der Bestandsaufnahme werden die einzelnen Elemente des Leitbildes von allen Beteiligten in Hinblick auf den Grad ihrer Erfüllung bewertet (z.B. von 1 bis 10). In jenen Bereichen, die im Schnitt über alle Bewertungen einen niedrigeren Erfüllungsgrad aufweisen, besteht offensichtlicher Handlungsbedarf. Auf Basis dieser Bestandsaufnahme werden nun Strategien entwickelt, deren Umsetzung in regelmäßigen Abständen überprüft wird.

Die Rolle des/der Evaluators/in in diesem Evaluationssetting ist nach Fettermann (2001, S. 30) „to serve as a coach, facilitator and critical evaluative friend“. Zu Beginn des Evaluationsprozesses kommt dem/der Evaluator/in auch noch stärker eine Qualifizierungsfunktion zu, d.h. er/sie hat den Aufbau des für die Selbstevaluation notwendigen Know-hows zu organisieren und zu moderieren. Mit der Zeit kommt es zu einem Wandel der Aufgabe, wobei dann jene des Coaches dominanter wird.

Dieser Evaluationsansatz korrespondiert am stärksten mit der Idee von GM-Implementierung als Prozess- und Strukturveränderung. Er erfordert auch die aktive Einbeziehung aller relevanten AkteurInnen und bietet einen Rahmen für die Reflexion des eigenen Alltagshandelns, das bei GM-Implementationsprozessen häufig zu kurz kommt (vgl. Leitner 2005, S. 195). Gleichzeitig stellt dieser Ansatz die höchsten Anforderungen an die Beteiligten, sich auf den Prozess einzulassen, ihre Alltagspraxis kritisch zu reflektieren und auch zu verändern. Fettermann (2001, 144f) hält als eine der Beschränkungen von Empowerment-Evaluation fest: „*Empowerment evaluation depends on people. (...) Even though empowerment evaluation is a group activity, it is dependent on motivated and conscientious people.*“

Eine weitere zentrale Einschränkung ist, dass dieser Ansatz nur für kleine Organisationseinheiten durchführbar ist. Für die Implementierung in größeren Institutionen ist darüber hinaus eine Gesamtstrategie für die Organisation zu entwickeln, die als Rahmen für die Initiativen auf Ebene von kleineren Subeinheiten geeignet ist und gleichzeitig die Möglichkeit der Zusammenführung der einzelnen Prozesse bietet.

5. Herausforderungen bei der Evaluation von GM-Implementationsprozessen

Bei der Evaluation von GM-Implementationsprozessen zeigen sich in der Praxis derzeit die größten Herausforderungen im Zusammenhang mit der Zielformulierung, der unzureichenden Datengrundlage und dem Aufbau von relevantem Gender- bzw. Evaluations-Know-how. Es handelt sich dabei um zentrale Punkte, die generell bei der Wahl des Forschungsdesigns zu berücksichtigen sind. Da gerade diesen Fragen aber im GM-Implementationsprozess eine besondere Bedeutung zukommt, werden sie kurz im GM-Kontext diskutiert.

5.1 Konkret formulierte Zielsetzung

Eine zentrale Herausforderung für die Evaluation von GM-Implementationsprozessen ist die häufig fehlende oder sehr unkonkret formulierte Zielsetzung für den Implementationsprozess. Mieke Verloo (2002, S. 4) hat dies als den Stretch-Faktor in der GM-Definition bezeichnet, mit dem sowohl Vor- wie auch Nachteile einhergehen. „*As all successful political and policy concepts, gender mainstreaming has a high “stretch” factor. It is stretched to mean gender equality, or equal opportunities, or just gender impact assessment, or attention for diversity issue, or more women in higher positions, and so on. Its ability to be stretched is part of its success. One concept fits all. (...) Still, the stretch factor is also a problem if we want to improve, or even monitor its results.*“

Ein Vorteil ist die leichtere Implementierung des Ansatzes, da Gender-Mainstreaming ein eher neutraler Begriff ist, der gut an die Managementsprache anzudocken ist. Der Nachteil liegt in dem damit verbundenen Interpretationsspielraum, der u.a. dazu führen kann, dass unterschiedliche AkteurInnen jeweils andere Vorstellungen von der Zielsetzung und auch der Umsetzungspraxis von Gender Mainstreaming haben (vgl. dazu auch den Beitrag von Andrea Leitner und Christa Walenta in diesem Band). Diese unterschiedliche Zieldefinition ist häufig nicht auf den ersten Blick ersichtlich und wird erst im Zusammenhang mit der Definition von Erfolgskriterien deutlich. Werden unterschiedliche Zielsetzungen verfolgt, so kann ein erzielttes Ergebnis je nach Sichtweise als Zielerreichung

oder Misserfolg interpretiert werden. Der damit einhergehende Eindruck der Beliebigkeit führt nicht nur dazu, dass der Verpflichtungscharakter von GM sinkt, sondern auch dazu, dass die Zielerreichung seriöserweise nicht mehr evaluiert werden kann. Eine konkrete Zielformulierung erhöht nicht nur die Selbstverpflichtung innerhalb der Organisation, sie stellt gleichzeitig auch die Voraussetzung dafür dar, dass der Erfolg des Prozesses evaluierbar ist.

Ist eine Zielformulierung für Gender Mainstreaming gefunden und auch innerhalb der Organisation kommuniziert und mitgetragen, so stellt sich die Frage nach dem Verhältnis der GM-Zielsetzung zu anderen Zielsetzungen der Organisation. Der Anspruch Chancengleichheit zu erhöhen, kann auch dazu führen, dass kurzfristig andere Zielsetzungen der Organisation nicht erreicht werden. Dies ist beispielsweise der Fall, wenn Frauen, die nicht als arbeitslos bzw. arbeitssuchend beim AMS registriert waren, durch arbeitsmarktpolitische Maßnahmen angesprochen werden sollen, um ihre Reintegration zu unterstützen. Dadurch wird zunächst die Zahl der registrierten Arbeitslosen erhöht, v.a. jene der arbeitslos gemeldeten Frauen. Ähnlich verhält es sich auch, wenn durch eine Beschäftigungsausweitung im Gesundheits- und Pflegebereich zusätzliche Arbeitsplätze für Frauen geschaffen werden. Die Frage, ob damit auch ein Schritt in Richtung Chancengleichheit erfolgt, ist nicht leicht zu beantworten. Wie sind die damit einhergehende zunehmende Segregation, der steigende Gender pay gap oder der Anteil atypischer Beschäftigungsverhältnisse zu bewerten? Die Tatsache, dass dieselbe Entwicklung auf der einen Seite als Erfolg verbucht, auf der anderen Seite aber kritisiert werden kann, verdeutlicht die Notwendigkeit einer Zielhierarchie bzw. einer eindeutigen Prioritätensetzung. Dies ist die Voraussetzung, um für ein bestimmtes Ergebnis einen eindeutigen Bewertungsmaßstab anlegen zu können.

Im Rahmen eines begleitenden Evaluationssettings bzw. einer Empowerment-Evaluation kann die Evaluation den Prozess der Zielfindung bzw. -konkretisierung in moderierender Weise unterstützen bzw. vortreiben. Die Evaluation sollte daher möglichst früh parallel zur GM-Implementierung geplant und durchgeführt werden. Ein wirkungsfeststellendes Design ist jedoch bei fehlender oder unkonkreter Zielformulierung seriöserweise nicht durchführbar. Eine Wir-

kungsanalyse für einen GM-Prozess ist folglich nur dann möglich, wenn bereits ein etablierter und über einen längeren Zeitraum laufender Prozess vorliegt.

5.2 Adäquates Set an Indikatoren

Derzeit besteht im Zusammenhang mit Geschlechterunterschieden massiver Weiterentwicklungsbedarf an bestehenden Indikatoren und Datengrundlagen. Auf der einen Seite ist es noch nicht in allen Bereichen möglich, Frauen und Männer getrennt auszuweisen. Auf der anderen Seite bedeutet aber die Aufspaltung in Frauen und Männer noch lange nicht, dass die Lebensrealitäten beider Geschlechter gleichermaßen abgebildet werden können (vgl. z.B. Wroblewski, Leitner & Steiner 2005 sowie den Beitrag von Andrea Leitner und Christa Walenta diesem Band).

Das Fehlen aussagekräftiger Indikatoren für das „Messen“ von Geschlechterunterschieden erschwert den GM-Implementationsprozess insofern, als die Analyse der Ausgangslage häufig auf unzureichende Daten abstellen muss und das „gendern“ von Statistiken selbst, d.h. die Weiterentwicklung von Datengrundlagen dahingehend, dass sie die spezifischen Lebenssituationen unterschiedlicher Gruppen adäquat abbilden, ein Teil des Prozesses wird. Für die Evaluation stellt dies insofern ein Problem dar, da insbesondere bei Wirkungsanalysen der Bewertungsmaßstab für die erzielten Ergebnisse unklar ist. In vielen Fällen ist – wie bereits dargestellt – ein quasi-experimentelles oder experimentelles Design zur Analyse der Wirkungen nur schwer durchführbar. Insbesondere bei Forschungsdesigns, die Fallstudien analysieren, ist es schwierig ein „Counterfactual“ zu bilden, d.h. einen vergleichbaren „Fall“ ohne Intervention zu finden. Soll jedoch eine Wirkungsanalyse erfolgen, stellt eine aussagekräftige Analyse der Ausgangssituation, die dann als Bewertungsmaßstab für die Ergebnisse herangezogen wird, die Minimalvoraussetzung dar.

Eine weitere Herausforderung für die Entwicklung aussagekräftiger Indikatoren für die Evaluation von GM-Implementationsprozessen ist, dass die angestrebte Veränderung von Prozessen und Strukturen nicht direkt beobachtbar ist. Häufig wird eine Annäherung über den jeweils erzielten Output versucht, die jedoch zumeist unbefriedigend bleibt. Dies deshalb, da nicht alle relevanten Aspekte berücksich-

tigt werden können und daher eine Konzentration auf quantifizierbare Dimensionen erfolgt. So wird z.B. die Veränderung des Frauenanteils unter Geförderten als ein Indikator für die Veränderung des Prozesses herangezogen, ohne dass dies aber etwas über die veränderten Rollenbilder in den „Köpfen“ der Angehörigen der Institution aussagen muss.

5.3 Hohe Anforderungen an Know-how von EvaluatorInnen

Und last but not least stellt die Evaluation von GM-Implementationsprozessen hohe Anforderungen an das Know-how der EvaluatorsInnen. Diese müssen zum einen die Strukturen, Prozesse und Abläufe in der Organisation kennen und gleichermaßen über Gender- und Evaluationskompetenz verfügen. Diese Kombination ist nur extrem selten anzutreffen und bedeutet, dass innerhalb eines Evaluationsteams alle diese Kompetenzen vertreten sein müssen. Damit stellt sich auch die Frage, wer die Evaluation durchführen soll: organisations-interne ExpertInnen oder externe EvaluatorsInnen? Wird die Evaluation von Personen aus der Organisation selbst durchgeführt, so kann davon ausgegangen werden, dass Informationen über den Aufbau und die Abläufe in der Organisation bekannt sind, dass Prozesse nachvollzogen werden können und vor allem, dass auch das nicht codifizierte (implizite) Wissen über die Organisation selbst vorliegt. Internen EvaluatorsInnen werden häufig auch weniger Vorbehalte entgegen gebracht, d.h. sie finden leichter Zugang zum Forschungsfeld und relevanten Informationen. Demgegenüber ist häufig wenig Wissen über Evaluationstheorie bzw. über adäquate methodische Verfahren vorhanden. Weiters setzt eine interne Evaluation eine hohe Reflexionsbereitschaft und -fähigkeit all jener Personen voraus, die an der Durchführung der Evaluation und der Nutzung der Ergebnisse beteiligt sind. Ist diese gegeben, so bestehen auch die Voraussetzungen für eine interne Evaluation.

Externe EvaluatorsInnen verfügen darüber hinaus häufig auch über Erfahrungen aus ähnlichen Kontexten, die bei der Formulierung von Empfehlungen zur Weiterentwicklung der Maßnahmen hilfreich sind und können auch einen kritischen Blick von außen einbringen, wobei sie weniger durch organisationsinterne Sichtweisen oder Traditionen geprägt bzw. eingeengt sind.

In vielen Fällen wird es daher zu einer Kombination von Elementen externer und interner Evaluation kommen (z.B. externe Evaluation mit intensiver Einbindung organisationsinterner ExpertInnen oder interne Evaluation unter Einbeziehung externer Evaluations-expertInnen). Im Rahmen einer solchen – häufig auch interdisziplinär angelegten – Kooperation treffen unterschiedliche institutionelle und fachliche Hintergründe sowie methodische Zugänge aufeinander. Für das Gelingen der Evaluation ist es dann zentral, dass ausreichend Zeit und Ressourcen für den Prozess der Teambildung und des Findens einer gemeinsamen Sprache zur Verfügung gestellt werden. Rolfsen und Torvatn (2005, S. 301ff) sehen eine gute Kommunikationsbasis als Voraussetzung für eine erfolgreiche Einspeisung von Evaluationsergebnissen in den Prozess und betonen dabei die Bedeutung einer „gemeinsamen Sprache“: *„By trying to speak their language, however, we had accepted their worldview as valid and because we communicated using their vocabulary, we started to build trust between us. (...) One important aspect of good communication is creating a shared language. This language has references to shared experiences interpreted similarly by both parties.“*

Inwieweit Evaluationsergebnisse tatsächlich genutzt werden und beispielsweise in eine Neukonzeptionierung von Maßnahmen einfließen, hängt stark vom Engagement der obersten Leitungsebene und damit auch vom erwarteten Nutzen für die Organisation ab. Je nachdem, welche Bedeutung Gleichstellung für die Organisation hat und wie groß das Bewusstsein für Gleichstellungsfragen bzw. Diskriminierungen ist, umso wahrscheinlicher ist es, dass umfassende Gender Mainstreaming Maßnahmen gesetzt werden.

6. Schlussfolgerungen

Im vorliegenden Beitrag wird davon ausgegangen, dass eine systematisch und frühzeitig geplante Evaluation den GM-Implementationsprozess in Institutionen wesentlich unterstützen und vorantreiben kann. Dies erfordert jedoch, dass die Evaluation von Beginn an entweder integriert in den Prozess oder parallel dazu konzipiert und durchgeführt wird. Es wird daher vorgeschlagen, die Evaluation systematisch in alle Schritte des GM-Implementationsprozesses zu integrieren, sie also als Querschnittsmaterie zu verstehen. Die Evaluation kann damit einen Beitrag zur Qualitätssicherung des Prozesses leisten, indem Reflexionsschleifen für den Prozess ermöglicht werden und ein systematisches Monitoring des Prozesses erfolgt.

Je nach konkreter Ausgestaltung bzw. aktuellem Stand des GM-Prozesses variieren die Fragestellungen für die Evaluation und kommen andere Evaluationsansätze in Frage. In der Startphase des Implementationsprozesses erscheinen primär begleitende Evaluationsdesigns sinnvoll, wobei die Evaluation sowohl bei der Analyse der Ausgangslage wie auch bei der Zielfindung oder –konkretisierung einen Beitrag leisten kann. Dies ist insofern zentral, als in der Praxis derzeit unkonkrete Zielformulierung und unzureichende Datengrundlagen die zentralen Herausforderungen für die Evaluation von GM-Implementationsprozessen darstellen.

Werden – wie eingangs dargelegt – GM-Implementationsprozesse mit Organisations- und Strukturveränderungen assoziiert, so stellt dies hohe Ansprüche an alle Beteiligte. Aus diesem Grund sollten auch alle relevanten AkteurInnen in diesen Prozess eingebunden werden und die mit der GM-Implementierung verbundenen Zielsetzungen mittragen. Ein derartiger partizipativer Ansatz in der Umsetzung sollte sich auch in der Evaluation widerspiegeln. Darüber hinaus sollte Evaluation möglichst frühzeitig als Teil des Prozesses mitkonzipiert werden, um den Beitrag, den die Evaluation zur Qualitätssicherung des Prozesses leisten kann, auch nutzen zu können. Aus diesen Gründen wird stark für begleitende und partizipative Evaluationsansätze plädiert, die je nach konkretem Setting als begleitende Evaluation durch externe ExpertInnen oder aber auch als spezifische Form der Selbstevaluation konzipiert sein können.

Unabhängig von der Entscheidung für ein konkretes Evaluationsmodell setzt eine erfolgreiche Evaluation jedoch eine konkrete Zielformulierung für die GM-Implementierung voraus sowie eine prinzipielle Bereitschaft, den Erfolg der gesetzten Interventionen zu überprüfen und aus den Ergebnissen zu lernen. Durch die Formulierung von konkreten Zielen, die auch explizit im Evaluationsauftrag enthalten sind, wird der Stellenwert von GM sowohl innerhalb als auch außerhalb der Institution betont und der Selbstverpflichtungscharakter erhöht. Dies verlangt jedoch eine Reflexions- und Entwicklungsbereitschaft der Institution, die durch entsprechende Begleitmaßnahmen bewusst zu fördern ist.

7. Literatur

Bergmann, Nadja; Pimminger, Irene (2004). *Praxis-handbuch Gender Mainstreaming. Konzept. Umsetzung. Erfahrung*. Wien: GeM-Koordinationsstelle für Gender Mainstreaming im ESF.

Beywl, Wolfgang (1988). *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegungen, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt am Main: Peter Lang.

Beywl, Wolfgang (1999). *Programmevaluation in pädagogischen Praxisfeldern – begriffliche und konzeptionelle Grundlagen*. In Klaus Künzel (Hrsg.), *Internationales Jahrbuch der Erwachsenenbildung* (S. 29-48). Köln: Böhlau Verlag.

Beywl, Wolfgang (2006). *Evaluationsmodelle und qualitative Methoden*. In Uwe Flick (Hrsg.), *Qualitative Evaluationsforschung. Konzepte. Methoden. Umsetzung* (S. 92-116). Hamburg: Rowohlt Verlag.

Bortz, Jürgen; Döring, Nicola (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. Auflage). Berlin: Springer.

Bustelo, Maria (2003). *Evaluation of Gender Mainstreaming. Ideas from a Meta-evaluation Study*. *Evaluation*, 9 (4) (S. 383-403).

Cook, Tom D. (2002). *Randomized Experiments in Educational Policy Research: A critical Examination of the Reasons the Educational Evaluation community has Offered for not Doing Them*. *Educational Evaluation and Policy Analysis*, 24 (3) (S. 175-199).

Deutsche Gesellschaft für Evaluation (DeGEval) (2004). *Standards für Evaluation* (2. Aufl.). Köln.

Fetterman, David M. (2001). *Foundations of Empowerment Evaluation*. London: Sage Publications.

Forss, Kim; Rebien, Claus C; Carlsson, Jerker (2002). *Process Use of Evaluations. Types of Use that Precede Lessons Learned and Feedback*. *Evaluation*, 8 (1) (S. 29-45).

Lassnigg, Lorenz; Leitner, Andrea; Wroblewski, Angela; Steiner, Mario; Steiner, Peter; Mayer, Kurt; Schmid, Günther; Schömann, Klaus (2000). *Evaluationsschema für Maßnahmen der aktiven Arbeitsmarktpolitik in Wien*. Wien: IHS-Projektbericht.

Leitner, Andrea (2005). *Von Frauenförderung zu Gender Mainstreaming. Gleichstellung in der Arbeitsmarktpolitik*. Dissertation, Universität Wien.

Bendl, Regine; Leitner, Andrea; Rosenbichler, Ursula; Walenta, Christa (2007). *Geschlechtertheoretische Perspektiven und Gender Mainstreaming*. In: *Equal-Entwicklungspartnerschaft QE GM (Hrsg.), Qualitätsentwicklung Gender Mainstreaming, Band 2: Grundlagen* (S. 31-61).

Newman, Diana L.; Brown, Robert D. (1992). *Violations of Evaluation Standards*. *Evaluation Review*, 16 (3) (S. 219-234).

Owen, John M.; Rogers, Patricia J. (1999). *Program Evaluation. Forms and Approaches*. London: Sage Publications.

Patton, Michael Quinn (1997). *Utilization-Focused Evaluation (3rd edition)*. Thousand Oaks: Sage Publications.

Piccotto, Robert (2002). *The Logic of Mainstreaming. A Development Evaluation Perspective*. *Evaluation*, 8 (3) (S. 322-339).

Rolfen, Monica; Torvatn, Hans (2005). *How to ‚Get Through‘. Communication Challenges in Formative Evaluation*. *Evaluation*, 11 (3) (S. 297-309).

Rosenbichler, Ursula; Schörghuber, Karl (2003). *Frauen, Männer und gleiche Chancen. Warum Gender Mainstreaming Unternehmen nutzt*. (Booklet zum Film). Wien: waff.

Rosenbichler, Ursula; Schörghuber, Karl (2007). *Integratives Konzept zu: Gender-Mainstreaming als Systementwicklung*. In: *Equal-Entwicklungspartnerschaft QE GM (Hrsg.), Qualitätsentwicklung Gender Mainstreaming, Band 2: Grundlagen* (S. 12-29).

Rossi, Peter H.; **Freeman**, Howard E; **Lipsey**, Mark W. (1999). *Evaluation. A systematic Approach (6th edition)*. Thousand Oaks: Sage Publications.

Sanders, James R. (ed) (2006). *Handbuch der Evaluationsstandards. Die Standards des Joint Committee on Standards for Educational Evaluation*“, 3. Auflage, Wiesbaden: VS-Verlag.

Sanderson, Ian (2000). *Evaluation in Complex Policy Systems*. Evaluation, 6 (4), 433-454.

Shadish, William R.; **Cook**, Tom D.; **Campbell**, Donald T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Stiegler, Barbara (2001). *Wie Gender in den Mainstream kommt. Konzepte, Argumente und Praxisbeispiele zur EU-Strategie des Gender Mainstreaming*. Bonn: Friedrich Ebert Stiftung.

Stockmann, Reinhard (2006). *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement (Sozialwissenschaftliche Evaluationsforschung Band 5)*. Münster: Waxmann.

Stufflebeam, Daniel L. (2001). *Evaluation Models. New Directions for Evaluation*, Volume 2001, Issue 89, 7-98.

Vedung, Evert (2004). *Evaluation Research and Fundamental Research*. In Reinhard Stockmann (Hrsg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (2. Auflage)* (S. 397-416). Opladen: Leske+Budrich.

Weiss, Carol H. (1998). *Evaluation (2nd edition)*. Upper Saddle River: Prentice Hall.

Weiss, Carol H. (1975). *Evaluation Research in the Political Context*. In Elmer Struening, Marcia Guttentag (eds.): *Handbook of Evaluation Research (Volume 1)* (S. 13-26). London, Beverly Hills: Sage Publications.

Wroblewski, Angela; **Leitner**, Andrea; **Steiner**, Peter (2005). *Gendersensible Statistik - Vom Sex-Counting*

zur Genderanalyse. Statistische Mitteilungen der Stadt Wien. 2005 (Heft 1,2), 7-45.

Zepke, Georg (2005). *Reflexionsarchitekturen. Evaluierung als Beitrag zum Organisationslernen*. Heidelberg: Verlag für Systemische Forschung.

Elektronische Quellen:

<http://www.degeval.de/calimero/tools/proxy.php?id=70>
Deutsche Gesellschaft für Evaluation (DeGEval) (2005, 22. März) Standards der Evaluation der Deutschen Gesellschaft für Evaluation (DeGEval-Standards). Zugriff am 21. März 2007

http://www.coe.int/T/E/Human_Rights/Equality/PDF_EG-S-MS_98_2rev_E.pdf
Directorate General of Human Rights (2004). Gender mainstreaming. Conceptual framework, methodology and presentation of good practices. Final report of Activities of the Group of Specialists on Mainstreaming (EG-S-MS). Strasbourg. Zugriff am 21. März 2007

<http://www.bpb.de/files/XUWTOH.doc>
Verloo, Mieke (2002). The Development of Gender Mainstreaming as a Political Concept for Europe, paper presented at the Conference „Gender Learning“, Leipzig, 6-8. September 2002. Zugriff am 21. März 2007

<http://www.equi.at/pdf/genderstatistik.pdf>
Wroblewski, Angela; Leitner, Andrea; Steiner Peter (2005). Gendersensible Statistik - Vom Sex-Counting zur Genderanalyse. Statistische Mitteilungen der Stadt Wien. 2005 (Heft 1,2), 7-45. Zugriff am 21. März 2007