

Gebel, Michael

Research Report

Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen

ZEW-Dokumentation, No. 06-01

Provided in Cooperation with:

ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Gebel, Michael (2006) : Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen, ZEW-Dokumentation, No. 06-01, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, <https://nbn-resolving.de/urn:nbn:de:bsz:180-madoc-12573>

This Version is available at:

<https://hdl.handle.net/10419/39168>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen

Michael Gebel

Dokumentation Nr. 06-01

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen

Michael Gebel

Dokumentation Nr. 06-01

Laden Sie diese ZEW Dokumentation von unserem ftp-Server:

<ftp://ftp.zew.de/pub/zew-docs/docus/dokumentation0601.pdf>

Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen

Michael Gebel

Februar 2006

Zentrum für Europäische Wirtschaftsforschung (ZEW)

ISSN 1611-681X

Kontakt und weitere Informationen:

Michael Gebel
Zentrum für Europäische
Wirtschaftsforschung (ZEW)
L 7, 1 · 68161 Mannheim
Telefon: +49-621-1235-367
Telefax: +49-621-1235-225
E-Mail: gebel@zew.de

© ZEW 2006

Das Wichtigste in Kürze

Die Notwendigkeit einer wissenschaftlichen Begleitforschung zur Messung der Effizienz und Effektivität von arbeitsmarktpolitischen Maßnahmen hat in den letzten Jahren, insbesondere in Hinblick auf die finanziellen Restriktionen durch die angespannte Haushaltslage, zunehmend an Akzeptanz gewonnen. Während die Evaluation als Instrument zur Identifikation der kausalen Programmwirkungen auf die Teilnehmer und Nichtteilnehmer der Maßnahmen fundamentaler Bestandteil der Begleitforschung ist, gibt es erst wenige Arbeiten, die sich mit Monitoring und Benchmarking beschäftigen. Diese beiden Konzepte sind ebenfalls Instrumente der wissenschaftlichen Analyse arbeitsmarktpolitischer Maßnahmen. Monitoring als Zielerreichungskontrolle und Benchmarking als Leistungsvergleich zwischen Arbeitsagenturen bzw. Arbeitsmärkten sind jedoch speziell im Rahmen der Dezentralisierung und Zielsteuerung im Bereich der Arbeitsmarktpolitik unerlässlich. Ziel dieser Dokumentation ist es, einen Überblick über den Forschungsstand zum Monitoring und Benchmarking im Bereich der Arbeitsmarktpolitik zu geben. Der Fokus richtet sich dabei auf allgemeine Definitionen, generelle Vorgehensweisen, ausgewählte Fallbeispiele, sowie bestehende Probleme und die Kritik an den Konzepten.

Monitoring in Form des Arbeitsmarktpolitik-Monitoring (AMPM) meint die Dokumentation, jedoch nicht die Erklärung der Erreichung gesetzter Programmziele. Die Dokumentation erfolgt durch kontinuierlich wiederholte Beobachtungen statistischer Indikatoren, mittels derer der Zielerreichungsgrad gemessen wird. Dieses AMPM ist vom reinen Arbeitsmarkt-Monitoring (AMM) zu differenzieren, das speziell auf die Ermittlung und Prognose des Qualifikationsbedarfs zielt, der sich aus dem Mismatch zwischen Arbeitsangebot und Arbeitsnachfrage ergibt. Auf Basis des AMPM lassen sich nur deskriptive Aussagen gewinnen. Um Aussagen zu den kausalen Effekten der Maßnahmeteilnahme treffen zu können, ist eine ergänzende mikroökonomische und makroökonomische Evaluationsstudie notwendig.

Benchmarking meint im Bereich der Arbeitsmarktforschung den Leistungsvergleich zwischen Arbeitsmärkten bzw. den Vergleich regionaler Arbeitsagenturen bezüglich der Effizienz der Maßnahmenimplementierung. Analog zur Differenzierung zwischen Monitoring und Evaluation lässt sich das Benchmarking in einen deskriptiven und einen kausalanalytischen Part dividieren. Beim deskriptiven Benchmarking stehen verschiedene Methoden zur Verfügung, um Effizienzunterschiede zu messen, wie beispielsweise Radar-Charts mit SMOP-Analyse und Analyseverfahren der Effizienzgrenze, die durch spezifische Vor- und Nachteile gekennzeichnet sind. Zur kausalen Erklärung der Leistungsunterschiede zwischen Arbeitsagenturen bzw. Arbeitsmärkten müssen zusätzlich die Kontextfaktoren berücksichtigt werden in Form der strukturellen Komposition der Bevölkerung, der wirtschaftlichen Situation und den institutionellen Rahmenbedingungen. Erst nach der Korrek-

tur um diese Umweltfaktoren können Effizienzunterschiede zwischen Benchmarking-Einheiten einer Untersuchung organisationsinterner Faktoren unterzogen werden.

Um die beiden Konzepte Monitoring und Benchmarking im System der Begleitforschung zu verorten, wird in dieser Dokumentation ein Versuch unternommen, die einzelnen Elemente der Begleitforschung voneinander abzugrenzen, die Interdependenzen aufzudecken, sowie die Rückkoppelungsschleifen hin zur Definition und Implementierung arbeitsmarktpolitischer Maßnahmen zu skizzieren. Die Interdependenzen zeigen, dass eine isolierte Betrachtung einzelner Elemente einschränkend ist und folglich die Begleitforschung in einem integrativen Rahmen aller Teilelemente aufgestellt werden sollte.

Inhaltsverzeichnis

1. Einleitung	1
2. Monitoring	2
2.1. Arbeitsmarkt-Monitoring (AMM).....	3
2.2. Arbeitsmarktpolitik-Monitoring (AMPM)	4
2.3. Vorgehensweise beim AMPM	5
2.4. Abgrenzung des Monitorings gegenüber existierenden Konzepten.....	8
2.5. Probleme und Grenzen des AMPM.....	9
2.6. Zwischenfazit	13
3. Benchmarking	14
3.1. Definition des Benchmarkings	14
3.2. Vorgehensweise beim Benchmarking	15
3.3. Generelle Probleme beim Benchmarking.....	16
3.4. Radar-Chart und SMOP-Analyse	17
3.4.1. Vorgehensweise.....	17
3.4.2. Kritik	20
3.5. Technische Effizienzanalyse	21
3.5.1. Vorgehensweise.....	21
3.5.2. Kritik	23
3.6. Kausalanalytisches Benchmarking.....	24
3.7. Zwischenfazit	25
4. Einordnung von Monitoring und Benchmarking in die Begleitforschung.....	26
4.1. Generelle Vorgehensweise in der Begleitforschung	26
4.2. AMM versus AMPM.....	27
4.3. Monitoring versus Evaluation	28
4.4. Benchmarking in der Begleitforschung.....	30
4.5. Rückkoppelungsschleifen.....	31
5. Zusammenfassung	32
6. Literaturverzeichnis.....	35

1. Einleitung

Die Begleitforschung bei arbeitsmarktpolitischen Maßnahmen ist innerhalb der letzten Jahre fester Bestandteil in der Arbeitsmarktpolitik geworden. Während früher oftmals ausschließlich der finanzielle Mittelzufluss oder die Anzahl der Teilnehmer als Indiz für die Wirksamkeit einer Maßnahme angesehen wurde, wird heute ein breites Instrumentarium eingesetzt, um die Effektivität und Effizienz arbeitsmarktpolitischer Programme zu beurteilen. Ziel ist es, eine höhere Transparenz in der Förderungslandschaft im Bereich der Arbeitsmarktpolitik zu kreieren, um damit Fehlentwicklungen zu identifizieren und bei Bedarf durch Aktionen gegenzusteuern. Diese Neuausrichtung der Arbeitsmarktpolitik spiegelt sich in neueren gesetzlichen Regelungen wie z.B. dem Sozialgesetzbuch (SGB) III wider. Beispielsweise fordert § 11 SGB III die Erstellung von Eingliederungsbilanzen, um „(...)Aufschluss über den Mitteleinsatz, die geförderten Personengruppen und die Wirksamkeit der Förderung zu geben(...)“. Oder § 280 SGB III verlangt, „(...)Wirkungen der aktiven Arbeitsförderung zu beobachten, zu untersuchen und auszuwerten(...)“.

Als Ursache für diese Entwicklung wird die angespannte öffentliche Haushaltslage genannt, die angesichts der hohen Fördermittel für die Arbeitsmarktpolitik eine Überprüfung deren Wirksamkeit unumgänglich macht (Erhel, 2003: 421f). So sind innerhalb der Begleitforschung in der Arbeitsmarktpolitik Evaluationsstudien fest etablierter Bestandteil zur Leistungsbeurteilung von Programmen. Dabei hat sich in dieser Forschungsrichtung größtenteils die Einsicht durchgesetzt, dass nicht die Bruttowirkungen als Erfolgsmaßstab einer Maßnahme angesehen werden können, sondern dass mittels ökonomischer Verfahren Nettoeffekte in Form kausaler Ursache-Wirkungs-Zusammenhänge zwischen Programmmaßnahme und Erfolg identifiziert werden müssen (Brinkmann und Wießner, 2002: 373).

Ergänzend zur Evaluation gewinnen Monitoring und Benchmarking als weitere Elemente der Begleitforschung in jüngerer Zeit zunehmend an Bedeutung. Monitoring als Zielerreichungskontrolle und Benchmarking als Leistungsvergleich zwischen Arbeitsagenturen bzw. Arbeitsmärkten sind speziell im Rahmen der Dezentralisierung und Zielsteuerung im Bereich der Arbeitsmarktpolitik unerlässlich. Als ursächlich für diese Entwicklungstendenzen wird das New Public Management genannt als Konglomerat von Managementinstrumenten, die zur Leistungssteigerung in öffentlichen Verwaltungen beitragen sollen (Auer und Kruppe, 1996: 903f). Diese Verwaltungsmodernisierung führt zu einem Wechsel von der Input- zur Outputsteuerung und einer Übertragung von Entscheidungsspielräumen auf untere Verwaltungsebenen (Brinkmann 1999: 7f). Zur Kontrolle müssen einerseits Monitoring-Systeme eingerichtet werden, die kontrollieren und darüber informieren, ob die Ziele arbeitsmarktpolitischer Programme erreicht wurden. Andererseits besteht ein weiterer Kontrollmechanismus im Benchmarking der Leistungen der dezentralen Ar-

beitsagenturen. Als weitere Ursachen der wachsenden Bedeutung von Monitoring und Benchmarking wird in der Literatur der technologische Fortschritt in der Datenverarbeitung diskutiert, der die zeitnahe und umfangreiche Datensammlung auch auf dezentraler Ebene erlaubt (Auer, 1998: 166). Außerdem bestehen seitens der Europäischen Union (EU) Vorgaben zum Monitoring arbeitsmarktpolitischer Maßnahmen, die mit Geldern der EU unterstützt werden (Erhel, 2003: 422), und Benchmarking wird von der Europäischen Kommission als Instrument zur Verbesserung der Arbeitsmarkt-Performance gefordert (Tronti, 1998: 489).

In der Literatur finden sich allerdings erst wenige Versuche, diese beiden Instrumente theoretisch und empirisch näher zu beleuchten bzw. im Rahmen der Begleitforschung trennscharf einzuordnen. Diese Dokumentation soll daher einen Überblick über den Forschungsstand zum Monitoring (Kapitel 2) und Benchmarking (Kapitel 3) im Bereich der Arbeitsmarktpolitik geben. Der Fokus richtet sich dabei auf allgemeine Definitionen, generelle Vorgehensweisen, ausgewählte Fallbeispiele, sowie bestehende Probleme und die Kritik an den Konzepten. In Kapitel 4 findet ein Versuch statt die beiden Elemente Monitoring und Benchmarking im System der Begleitforschung zu verorten. Die Konzepte werden gegenüber dem Konzept der Evaluation abgegrenzt, die wechselseitigen Interdependenzen werden offen gelegt und Rückkoppelungsschleifen zur Programmdefinition und Programmimplementierung skizziert. Kapitel 5 fasst die Ergebnisse zusammen.

2. Monitoring

Im Bereich der Arbeitsmarktpolitik lässt sich Monitoring in die beiden großen Teilgebiete Arbeitsmarkt-Monitoring (AMM) und Arbeitsmarktpolitik-Monitoring (AMPM) differenzieren (Auer und Kruppe, 1996). Demnach fokussiert sich AMM auf das Monitoring der Vorgänge und Trends auf dem Arbeitsmarkt, während sich AMPM der Beobachtung einzelner arbeitsmarktpolitischer Maßnahmen widmet. Diese Unterscheidung liegt auch im Gesetzestext des SGB III vor, wonach „Arbeitsmarktmonitoring (...) ein System wiederholter Beobachtungen, Bilanzierungen, Trendbeschreibungen und Bewertungen der Vorgänge auf dem Arbeitsmarkt einschließlich der den Arbeitsmarktausgleich unterstützenden Maßnahmen (ist)“ (§ 9, Abs. 2 SGB III). Zwar soll der Fokus dieser Dokumentation auf dem Monitoring arbeitsmarktpolitischer Maßnahmen, d.h. dem AMPM liegen, aber Analysen zum Monitoring haben gezeigt, dass beide Teilgebiete wegen bestehender Interdependenzen nicht isoliert voneinander betrachtet werden können (u.a. LASA, 2000). Daher erfolgt zunächst in Kapitel 2.1 ein kurzer Abriss des theoretischen und empirischen Forschungsstands zum AMM. Dem schließt sich eine detailliertere Darstellung des AMPM an: eine Definition (Kapitel 2.2), die Beschreibung der Vorgehensweise beim AMPM (Ka-

pitel 2.3), eine Abgrenzung gegenüber ähnlichen Konzepten (Kapitel 2.4) und schließlich Probleme und Kritikpunkte in Kapitel 2.5. Die Interdependenzen zwischen beiden Monitoring-Systemen werden im größeren Kontext zusammen mit weiteren Elementen der Begleitforschung in Kapitel 4 skizziert.

2.1. Arbeitsmarkt-Monitoring (AMM)

Beim Arbeitsmarkt-Monitoring (AMM) steht generell die Analyse der Arbeitsmarktsituation im Vordergrund und speziell die Ermittlung des Qualifikationsbedarfs, der sich aus dem Mismatch zwischen Arbeitsangebot und Arbeitsnachfrage ergibt (LASA, 2000: 11). Der Radius des untersuchten Arbeitsmarktes kann sich auf eine Region bzw. eine Nation beziehen. In einer an informations- und kommunikationstheoretischen Überlegungen orientierten Definition subsumieren Larsen et al. (2005: 9) unter AMM die Beobachtung des Arbeitsmarktes, Sammlung, Aufbereitung, Vergleich und Kommunikation von Informationen, sowie Prognosen von Entwicklungen und Diagnosen der Arbeitsmarktsituation.

Ziel solcher systematischen Beobachtungen und Prognosen des Arbeitskräfteangebots und der Arbeitskräftenachfrage ist eine Verbesserung der Matchingeffizienz auf dem Arbeitsmarkt. Nur bei Beachtung der zukünftigen Arbeitsmarktsituation können qualifikations-spezifische Fehlentwicklungen wie z.B. „Akademikerschwemmen“ und „Fachkräftemängel“ gedämpft werden, die kostenintensive Kapazitätsengpässe in Unternehmen und qualifikationsbedingte Arbeitslosigkeit induzieren (Hilbert und Mytzek, 2002: 1f). Folglich betonen Larsen et al. (2005: 13f) das konstitutive Merkmal des AMM, den Zielerreichungsgrad in Form der Anpassung des Arbeitsangebots an die Arbeitsnachfrage durch kontinuierliche Beobachtungen zu dokumentieren und zu kommunizieren. Diese Informationen können Grundlage für notwendige Gegensteuerungen sein, indem insbesondere arbeitsmarktpolitische Maßnahmen anhand der Erkenntnisse des AMM ausgerichtet werden, um die Matching-Effizienz auf dem Arbeitsmarkt und die Effizienz der Programme zu erhöhen.

Als methodisches Instrumentarium wird ein Mix aus qualitativen Analysen (z.B. Fallstudien) und quantitativen Methoden (z.B. ökonometrische Prognosemodelle) empfohlen, die jeweils mit spezifischen Vor- und Nachteilen behaftet sind (vgl. dazu Hilbert und Mytzek, 2002: Tabelle 2). Insbesondere besteht ein Trade-off zwischen Detaillierungsgrad und Repräsentativität bei der Wahl zwischen qualitativen und quantitativen Verfahren.

In Deutschland¹ kann man AMM-Systeme mit nationaler und regionaler Perspektive differenzieren (Hilbert und Mytzek, 2002: 7-13). So bietet z.B. das IAB/Prognos-Modell Qua-

¹ Einen Überblick zu Erfahrungen und Ansätzen in internationaler Perspektive geben Hilbert und Mytzek (2002).

lifikationsprognosen auf nationaler Ebene bis zum Jahr 2010, die wegen des hohen Aggregationsgrads nicht für regionale Prognosen verwendet werden können (Schüssler et al., 1999). Die meisten regionalen AMM gehen auf zwei Projekte zurück: die Prospect-Methode und das Konzept „Regionales Monitoring-System Qualifikationsentwicklung“ (RMQ), das im Rahmen des Projekts „Entwicklungsplanung Qualifikation im Land Bremen“ (Equib) entstanden ist (Larsen et al., 2005: 14). Fallstudien, Experteninterviews und regionale Betriebspanel als Grundlagen des RMQ sollen der Früherkennung der kurz- bis mittelfristig zu erwartende Qualifikationsbedarfe vor dem Hintergrund technologischer und wirtschaftsstruktureller Veränderungen auf regionaler Ebene dienen (Bendix und Hammer, 2005). Die Prospect-Methode orientiert sich hingegen an einem Matching-Ansatz, indem es sowohl Bedarfsermittlung als auch Arbeitskräfteangebotsermittlung integriert und einander abgleicht (GIB, 2002). Damit lassen sich Bereiche identifizieren, in welchen kurz- und mittelfristig spezifische Qualifikationen gebraucht werden. Gleichzeitig kann aber auch geprüft werden, ob das Angebot an Qualifikationen diese Nachfrage passgenau bedienen kann oder ob in kurz- oder mittelfristiger Perspektive mit einem Mismatch in einzelnen Qualifikationssegmenten zu rechnen ist.

2.2. Arbeitsmarktpolitik-Monitoring (AMPM)

Arbeitsmarktpolitik-Monitoring (AMPM) ist gemäß der Leitdefinition von Auer und Kruppe (1996), die sich in der Literatur durchgesetzt hat, durch drei Merkmale charakterisiert. Erstens handelt es um eine Dokumentation, aber keine Erklärung der Erreichung bzw. Verfehlung vorher gesetzter Ziele. Aus diesem Merkmal leitet sich eindeutig die Notwendigkeit einer ergänzenden kausalanalytischen Wirkungsforschung ab. Die Dokumentation erfolgt zweitens durch die kontinuierlich wiederholte Beobachtung statistischer Indikatoren, mittels derer der Zielerreichungsgrad gemessen wird. Drittens ist das AMPM charakterisiert durch systematische Rückkoppelungsschleifen zu den unterschiedlichen Steuerungsebenen mit dem Ziel die Programmimplementierung und das Programmdesign der arbeitsmarktpolitischen Maßnahmen zu verbessern. Demnach soll das Monitoring der Früherkennung von Performance-Problemen und der Aufdeckung möglicher Performance-Verbesserungen dienen (Affholter, 1994). Die Idee eines solchen „Feedback-Loop“ koinzidiert mit der Definition des Begriffs durch Schmid (1996), der Monitoring als einen Teil der Prozessanalyse sieht. Der Prozesscharakter des Monitorings offenbart sich dabei insbesondere im zweiten Charakteristikum der kontinuierlichen Erhebung der Indikatoren, d.h. beim Monitoring handelt es sich um eine begleitende Programmebeobachtung.

Affholter (1994) betont zudem noch die Funktion des Monitorings, ein regelmäßiges *Reporting der Programmergebnisse* zu gewährleisten.² Demnach sollen beim Reporting die notwendigen Daten und Informationen den beteiligten Akteuren zugespielt werden, so dass eine kontinuierliche Kommunikation und Abstimmung der involvierten Einheiten garantiert wird.

2.3. Vorgehensweise beim AMPM

In Anlehnung an Auer und Kruppe (1996) lässt sich der idealtypische Ablauf des AMPM in vier verschiedene Stufen unterteilen³. Abbildung 1 skizziert schematisch die einzelnen Prozessschritte beim Monitoring arbeitsmarktpolitischer Maßnahmen:

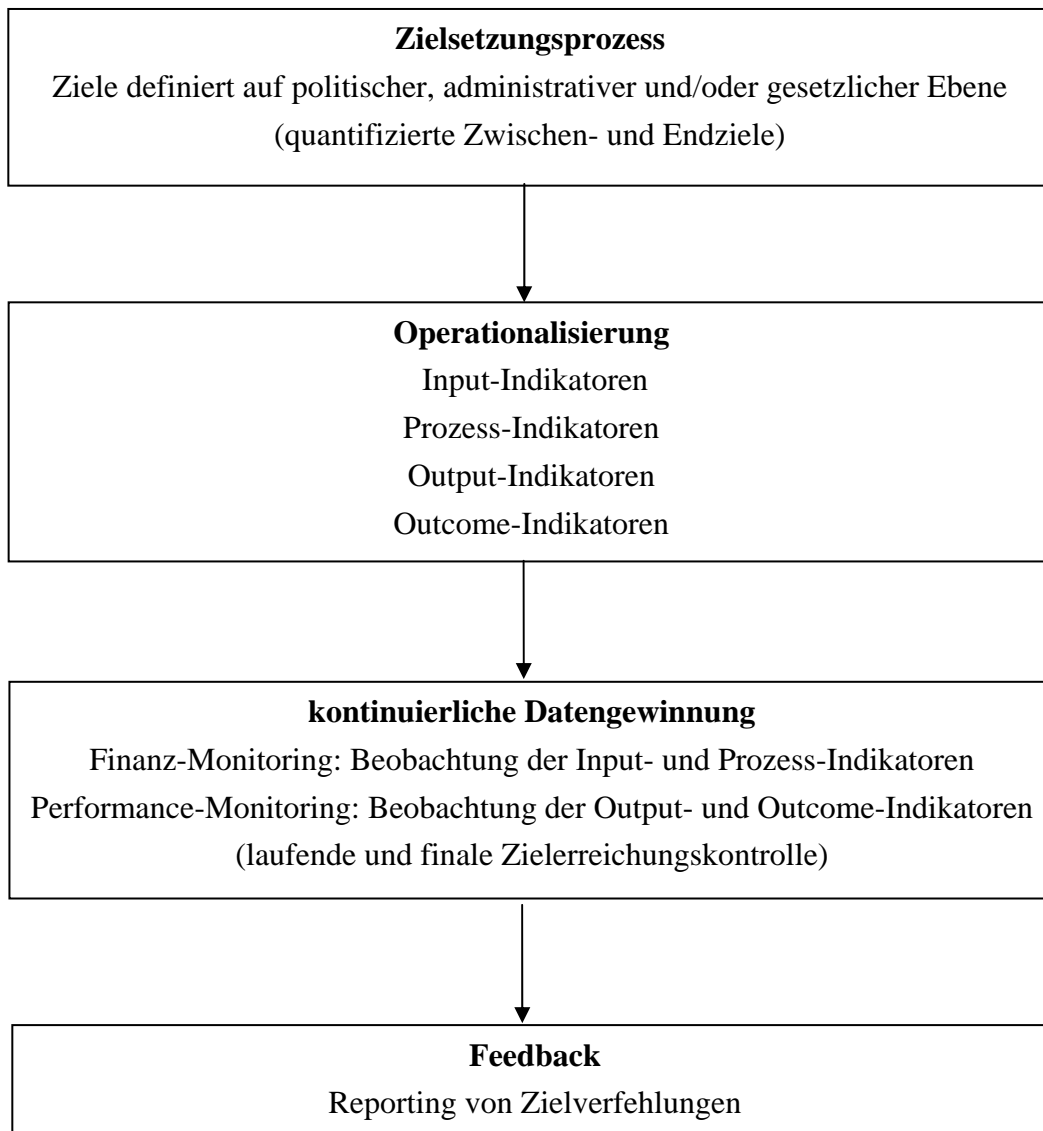
(1) In einem vorgelagerten Prozess werden die Programmziele auf der gesetzlichen, politischen und/oder administrativen Ebene definiert. Als Grundlage für die Zielauswahl können u.a. die Ziele des SGB III oder das übergeordnete Zielsystem der Bundesregierung bzw. der EU dienen (Brinkmann, 2000: 485f). Beispielsweise leiten Fitzenberger und Hujer (2002: 1f) als wichtigste Ziele der aktiven Arbeitsmarktpolitik den effizienten Instrumenteneinsatz, die Förderung bestimmter Problemgruppen wie z.B. Langzeitarbeitslose und die Eingliederung von Arbeitslosen in reguläre Beschäftigung aus § 7 SGB III ab. Solche Zieldefinitionen können mit expliziten Zielsetzungen in Form von Zwischen- und Endzielen kombiniert werden. Die Zielsetzungen können sich an Ergebnissen eines AMM orientieren, d.h. sie sollten auf Basis der Ausgangslage unter Beachtung ökonomischer Trends, struktureller und institutioneller Veränderungen erfolgen (Brinkmann, 1999). Solche Zielsteuerungen (MbO-Management by Objectives) dienen der Klärung von Prioritäten ex ante, woran sich die Programmaktivitäten ausrichten können und das Verwaltungshandeln beurteilt werden kann (Schütz, 2001).

(2) Auf Basis der spezifizierten Programmziele werden im nächsten Schritt Indikatoren zur Operationalisierung der einzelnen Zieldimensionen konstruiert. Im Idealfall sollten die Ziele eindeutig und vollständig formuliert werden, so dass bei der Indikatorenauswahl die Operationalisierung erleichtert wird. Affholter (1994: 97) differenziert zwischen Input-, Prozess-, Output- und Outcome-Indikatoren. Die Indikatoren sollten die generellen Anforderungen der Quantifizierbarkeit, intersubjektiven Nachprüfbarkeit, Validität und Reliabilität genügen und auf einen klaren Zeitraum oder –punkt bezogen sein.

² Ob das Reporting als integraler Bestandteil des Monitorings gesehen wird oder ob es ein eigenständiges Element der Begleitforschung ist, verbleibt Definitionssache.

³ Im Bereich des New Public Management gibt es ein ähnliches Konzept in Form des Performance-Management, das auf die effiziente Leistungserstellung ausgerichtet ist. Dabei unterscheidet man die Instrumente Performance-Contracting (Zielvereinbarungen), Performance-Monitoring (regelmäßige Messung von Performance-Indikatoren) und das Performance-Reporting (regelmäßige Berichterstattung).

Abbildung 1: Stufen eines idealtypischen AMPM



Quelle: Darstellung angelehnt an Schütz (2001: Abb. 1) und Auer und Kruppe (1996: 908).

Input- und Prozess-Indikatoren umfassen Angaben zu den Teilnehmern und den eingesetzten Finanzmitteln. Daraus lassen sich Aussagen zur Teilnehmerstruktur, den administrativen Abläufen und den verwendeten Finanzmitteln gewinnen (Schütz, 2003: 1). Während *Output-Indikatoren* die direkten Ergebnisse des Programms messen, wie z.B. den Anteil der abgeschlossenen arbeitsmarktpolitischen Maßnahmen, erfassen die *Outcome-Indikatoren* die Wirkungen der Maßnahmen auf die Programmteilnehmer. Untersucht man beispielsweise eine Maßnahme, die auf die Eingliederung in den ersten Arbeitsmarkt zielt, dann lässt sich die Wirkung u.a. durch den Anteil der Übergänge in eine ungeforderte Beschäftigung nach dem Ende der Maßnahme bestimmen. Dies entspricht aus deskriptiver Sicht einer positiven Programmwirkung. Ob diese Wirkung jedoch kausal auf die Pro-

grammteilnahme zurückzuführen ist, lässt sich nur in einer anschließenden Evaluation feststellen. Die Outcome-Indikatoren sollten zu mehreren Zeitpunkten nach der Maßnahmeteilnahme im Rahmen des Monitorings erfasst werden. Wegen des zeitnahen Charakters des Monitorings ist es jedoch schwierig langfristige Wirkungen zu messen. Handelt es sich allerdings um ein explizites Ziel, dass z.B. die Maßnahmeteilnehmer nicht nur in eine ungeforderte Beschäftigung wechseln, sondern darin auch möglichst lange verweilen, ist eine Ausdehnung des zeitlichen Horizonts des Monitorings notwendig. Generell überwiegen jedoch kurz- und mittelfristige Informationen über den Zielerreichungsgrad beim Monitoring (Brinkmann und Wiesner, 2002).

Orientiert an dieser Differenzierung der Indikatoren dividieren Auer und Kruppe (1996) das AMPM in ein *Finanz-Monitoring*, bei dem Input- und Prozess-Indikatoren erhoben werden, und ein *Performance-Monitoring*, das entsprechend die Output- und Outcome-Indikatoren umfasst. Alle Input-, Prozess-, Output- und Outcome-Indikatoren können zudem disaggregiert nach spezifischen Problemgruppen, wie z.B. Langzeitarbeitslosen oder Geringqualifizierten erhoben werden, um detaillierte Beobachtungen für diese Personen zuzulassen. Die konkrete Auswahl der Indikatoren zur Bewertung der Ergebnisse der Arbeitsförderung hat sich situationsadäquat nach den Förderinstrumenten, nach den Zielen und den regionalen Kontextbedingungen zu richten (LASA, 2000: 57). Liegen mehrere Indikatoren für die Messung einer Zieldimension vor, so existieren verschiedene methodische Instrumente, um stark korrelierende Indikatoren zu eliminieren, wie z.B. die Hauptkomponentenanalyse (u.a. Jolliffe, 2002).

(3) Der Indikatorendefinition schließt sich ein Prozess der Datengewinnung und –aufbereitung durch eine kontinuierliche Beobachtung der Indikatoren an. Die regelmäßige Beobachtung der Indikatoren über die Zeit hinweg erlaubt es Entwicklungstrends in der Programmimplementierung und bezüglich des Programmerfolgs abzulesen. Voraussetzung dafür ist allerdings die Reliabilität der Indikatoren. Als Zeitintervall für die Erhebung der Indikatoren gibt Auer (1998: 173) ein Quartal an. Bei kurzen Programmen ist jedoch eine höhere Frequenz der Datenerhebung empfehlenswert. Wurden ex ante quantitative Zwischen- und Endziele fixiert, so empfiehlt sich beim Prozess der Datengewinnung die jeweiligen Zielerreichungsgrade zu ermitteln.

(4) Auer und Kruppe (1996) betonen die Wichtigkeit der Rückkoppelungsschleifen vom Monitoring hin zur Programmimplementierung. So sollen Zielverfehlungen bei Indikatoren zu Anpassungen des Programms führen. Denn speziell Monitoring als zeitnahes Instrument der Begleitforschung ermöglicht direkte Korrekturen bei Fehlentwicklungen. Allerdings sollten keine solchen Korrekturen auf Basis der deskriptiven Ergebnisse des Monitorings durchgeführt werden, da gesicherte Ursache-Wirkungszusammenhänge erst in einer Kausalanalyse identifiziert werden können. Stattdessen sollte sich der Rückkoppe-

lungsmechanismus auf das Reporting von Zielverfehlungen und die Identifikation möglicher Ansatzpunkte für Kausalanalysen beschränken.

2.4. Abgrenzung des Monitorings gegenüber existierenden Konzepten

In der Literatur finden sich bereits die beiden Konzepte des Controllings und der Eingliederungsbilanzen, die eng mit dem Monitoring verwandt sind. Allerdings unterscheidet sich das Monitoring in einigen Punkten von diesen ähnlichen Ansätzen, weshalb eine Abgrenzung notwendig ist.

Schütz (2003:1) definiert *Controlling i.w.S.* als Zielsteuerung im Rahmen eines Kreislaufs, der übergeordnete Zieldefinitionen, Festlegung operativer Indikatoren und Zielhöhen, Implementierung, Monitoring und Evaluation erfasst. Diese weite Definition des Controllings entspricht fast der gesamten Begleitforschung, weshalb eine engere Definition angebracht scheint. Entsprechend hat *Controlling i.e.S.* gemäß Caliendo und Jahn (2004) die Aufgabe, zeitnah Informationen zur Bestimmung von Zielerreichung bzw. -verfehlung zu sammeln, aufzubereiten und auszuwerten, was die Grundlage für steuernde Eingriffe zur Korrektur von Fehlentwicklungen liefert. Diese enge Definition liegt sehr nah am allgemeinen Verständnis von Monitoring. Brinkmann (2000: Übersicht 3) sieht hingegen Monitoring nur als ein Teilelement des Controllings. So entspricht es dem Controlling-Element „Ziel-Controlling“. Analog zur Monitoring-Definition werden im Ziel-Controlling quantitative, auf den Arbeitsmarkt bzw. auf Programmmaßnahmen bezogene Wirkungsziele vereinbart, deren Erreichung durch die Überwachung von Indikatoren kontrolliert wird. Zielabweichungen sollen über Rückkoppelungsmechanismen zu Anpassungen führen. In Anlehnung an die Controlling-Definition von Postlep (1994) betont Brinkmann (1999: 22f), dass sich Controlling vom Monitoring abhebt, indem es zusätzlich die langfristigen, strategischen Arbeitsmarktziele erfasst. Hingegen dominieren beim Monitoring wegen seines zeitnahen Charakters eher die kurzfristigen Ziele.

Allerdings ist diese Abgrenzung problematisch, soweit im AMM langfristige Prognosen unternommen werden bzw. im AMPM explizit auch langfristige Programmziele definiert werden, deren Einhaltung dann durch die Kontrolle von Indikatoren überprüft wird. Folglich empfiehlt es sich, Controlling als ein Konzept zu definieren, das über das Monitoring hinausgeht, wobei die Abgrenzung vom spezifischen Monitoring-Projekt abhängt. So bestimmen die Zeitperspektive und Zieldefinitionen des jeweiligen AMM bzw. AMPM, ob das Monitoring neben kurzfristigen auch langfristige Ziele wie das Controlling misst.

Gemäß § 11 SBG III existiert bereits ein gesetzlich fixiertes Instrument in Form der *Eingliederungsbilanz (EGB)*, das dem AMPM sehr nahe kommt. So muss eine EGB u.a.

- Umfang und Verteilung der Geldmittel,
- Durchschnittsausgaben je Gefördertem,
- Beteiligungszahlen insbesondere für förderungsbedürftige Personen,
- Vermittlungsquote,
- Verbleibsquote und
- Entwicklungen über den Zeitverlauf enthalten.

Solche Eingliederungsbilanzen können jedoch nicht das AMPM substituieren. Beispielsweise liegen die Ergebnisse der EGB nur stark zeitlich verzögert zum Ende des fiskalischen Jahres vor, was eine wie vom Monitoring geforderte zeitnahe Dokumentation der Zielerreichung behindert. Zudem dienen EGB speziell der Erstellung von Geschäftsstatistiken unter Ausblendung der Belange der Begleitforschung (Brinkmann und Wießner, 2002: 382). Fundamentaler Kritikpunkt ist, dass die EGB zwar detaillierte standardisierte Input- und Output-Indikatoren, aber keine adäquaten Outcome-Indikatoren ausweisen. Zwar wird der Verbleib der Arbeitnehmer „in angemessener Zeit im Anschluss an die Maßnahme“ erfasst, aber es erfolgt keine Differenzierung, ob der Maßnahmeteilnehmer in ein gefördertes, reguläres befristetes oder reguläres unbefristetes Beschäftigungsverhältnis eintritt (Caliendo und Jahn, 2004). Folglich sind für das Monitoring zeitnahe, detaillierte und komplexere Datenerhebungen notwendig, die über die EGB hinausgehen.

2.5. Probleme und Grenzen des AMPM

Beim AMPM können zahlreiche Probleme auftreten. Zudem wurde die begrenzte Aussagekraft des Monitorings in der Literatur kritisiert. Die *Probleme*, die im Monitoring-Prozess selbst auftreten können, lassen sich folgendermaßen systematisieren:

(1) Insbesondere bei Heterogenität im Zielsystem können Zielkonflikte bestehen. Potenzielle Konflikte existieren vor allem zwischen kurzfristigen und langfristigen Zielen, Mikro- und Makrozielen, sowie quantitativen und qualitativen Zielen (Brinkmann und Wießner, 2002: 384). Ein unpräzises Monitoring kann zudem aus unklar bzw. nicht definierten Programmzielen resultieren (Erhel, 2003: 423). Um dieses Problem zumindest teilweise zu bekämpfen, empfiehlt sich eine konsistente, klare Zielstruktur.

(2) Der quantitative Kennzahlenvergleich beim AMPM verleitet dazu, dass qualitative Zieldimensionen, die sich nur schwer operationalisieren lassen, wie z.B. die Verbesserung der Vermittlungs- bzw. Beschäftigungsfähigkeit, ausgeblendet werden (Brinkmann, 1999: 14). Stattdessen werden einfache Indikatoren für quantitative Ziele, wie beispielsweise die Anzahl der Übergänge in reguläre Beschäftigungsverhältnisse, wegen der leichten Opera-

tionalisierbarkeit präferiert. Dem kann durch ein verstärktes Bemühen auch qualitative Aspekte in Indikatoren zu erfassen begegnet werden.⁴

(3) Ein inkonsistentes Monitoring resultiert aus mangelhaften Indikatoren. Wenn die Indikatoren nicht mit den entsprechenden Zieldimensionen des Programms korrespondieren, kann das AMPM nicht die wahre Zielerreichung des Programms feststellen. Zudem droht bei nicht-reliablen Messkonzepten, dass die Vergleichbarkeit der Indikatoren über die Zeit nicht mehr gewährleistet ist. Ist das Monitoring mit konkreten Zielsetzungen verbunden, kann ein Moral-Hazard-Problem entstehen, wenn die datenerhebenden Personen einen Anreiz haben Zielerreichungsgrade zu manipulieren (Schütz, 2001: 208).

(4) Auer (1998: 172f) mahnt die Gefahr einer schlechten Datenqualität bei der Erhebung der Indikatoren an. Dies kann sich in einer mangelnden Aktualität der Daten sowie dem Fehlen bestimmter Daten äußern. Ursache sind oftmals personelle und finanzielle Engpässe bei den mit den Monitoring-Aktivitäten betrauten Personen.

(5) Durch den zeitnahen Charakter des Monitoring-Prozesses werden tendenziell langfristige Effekte der Programmteilnahme nicht erfasst. Insbesondere bei Zielkonflikten zwischen kurz- und langfristigen Zielen kann die Erhebung der langfristigen Effekte notwendig sein. So transferieren zwar z.B. Maßnahmen der aktiven Arbeitsmarktpolitik Personen in ein (gefördertes) Beschäftigungsverhältnis, was zu einem kurzfristigen Eingliederungserfolg führt. Aber eventuell wird gerade dadurch die Reintegration in den ersten Arbeitsmarkt behindert (Brinkmann und Wießner, 2002: 379). Gehören solche langfristige Ziele zum Zielsystem, dann ist der Monitoring-Prozess entsprechend über die Zeit nach dem Maßnahmenende auszudehnen.

(6) Der im vorigen Kapitel dargestellte stilisierte Ablauf eines idealtypischen Monitoring wird in der Realität oftmals nicht eingehalten (Deeke und Kruppe, 2002: 404). Insbesondere können sich Problemdefinition und Zielformulierung im Rahmen der Programmimplementierung ändern. Dies kann zu Fehlern im Monitoring-Prozess führen, wenn die Indikatoren nicht mehr mit den gesetzten Zielen übereinstimmen.

Während sich die bisher genannten Probleme auf potenzielle Defizite im Monitoring-Prozess beziehen, lässt sich das Monitoring des Weiteren hinsichtlich der *Grenzen* seiner Aussagekraft kritisieren. Um weitergehende Aussagen treffen zu können, ist eine Ergänzung des AMPM durch weitere Elemente der Begleitforschung, insbesondere Evaluation und auch Benchmarking, notwendig:

⁴ Im betriebswirtschaftlichen Controlling wurden z.B. quantitative Kennzahlensystem erfolgreich um qualitative Dimensionen erweitert mit dem Konzept der Balanced Score Card (BSC) (Kaplan und Norton, 1992). BSC umfassen die Schwerpunktfelder Finanzperspektive, interne Geschäftsprozesse, Innovationsperspektive und Kundenperspektive als einzelne Leistungsdimensionen mit spezifischen Zielen und Indikatoren zur Performance-Messung. Während die klassische Finanzperspektive eine quantitative Dimension ist, erfassen die drei anderen Schwerpunktfelder v.a. qualitative Aspekte mittels quantitativer Indikatoren.

(1) Entscheidender Kritikpunkt am AMPM ist, dass zwar die Zielerreichungsgrade deskriptiv erfasst werden, aber keine kausalen Schlüsse aus den Monitoring-Ergebnissen, auch nicht auf Basis der Outcome-Indikatoren, gezogen werden dürfen. Aus den Monitoring-Daten lassen sich direkt nur *Bruttoeffekte* der Maßnahme ablesen. Ein solcher Bruttoeffekt misst beispielsweise im Rahmen des AMPM den Anteil der Maßnahmeteilnehmer, die nach Beendigung einer Maßnahme in ein reguläres Beschäftigungsverhältnis wechseln. Dieser Bruttoeffekt ist bei der Erfolgsbeurteilung der Maßnahme nicht mit dem *Nettoeffekt*, der den kausalen Effekt der Maßnahmeteilnahme auf die Ergebnisvariablen misst, gleichzusetzen (Brinkmann und Wießner, 2002: 378). Denn der Beschäftigungsstatus nach Ende der Maßnahme kann nicht ausschließlich auf die spezifische Maßnahmeteilnahme attribuiert werden. Ein reguläres Beschäftigungsverhältnis hätte sich für den Teilnehmer auch ohne die Teilnahme ergeben können. Nur eine sich dem Monitoring anschließende *Evaluation* kann kausale Effekte der Maßnahmeteilnahme identifizieren.

Um den Kausaleffekt der Teilnahme für ein Individuum zu bestimmen, müsste man die Differenz bilden zwischen dem Wert der Ergebnisvariable bei Teilnahme und dem Wert, der sich für dasselbe Individuum bei Nichtteilnahme ergeben würde (Fitzenberger und Speckesser, 2000: 361). Das *fundamentale Evaluationsproblem* besteht nun darin, dass sich die Ergebnisse der Teilnehmer nicht im hypothetischen Fall der Nichtteilnahme beobachten lassen (Michaelis und Spermann, 2004: 224). Um diese hypothetischen Ergebnisse zu bestimmen, muss eine geeignete Kontrollgruppe gefunden werden. Dazu verwendet man entweder vergleichbare Nichtteilnehmer oder die Teilnehmer vor ihrer Teilnahme als Kontrollgruppe. Im Fall eines idealisierten Experiments, d.h. wenn der Teilnahmezustand zufällig zugewiesen wird, reicht der Vergleich der Ergebnisse für die Teilnehmer und die Nichtteilnehmer. Da in Deutschland solche Experimente in der Vergangenheit nicht durchgeführt wurden, konzentriert sich die mikroökonomische Evaluationsforschung auf eine nichtexperimentelle Vorgehensweise (Fitzenberger und Hujer, 2004: 3f).

Im nichtexperimentellen Untersuchungsdesign besteht jedoch das Problem einer möglichen Selektionsverzerrung, falls sich Teilnehmer und Kontrollgruppe in beobachteten oder unbeobachteten Merkmalen unterscheiden, die einen Einfluss auf die Ergebnisvariable haben. Beispielsweise kann eine *positive Selektion* auftreten, wenn der Anreiz existiert, ausschließlich erfolgsversprechende Personen in der Maßnahme teilnehmen zu lassen. Insbesondere bei quantitativen Zielsetzungen, wie z.B. bezüglich dem Anteil erfolgreicher Übergänge in ungeforderte Beschäftigungen, existiert der Anreiz eines solchen „Teilnehmercreaming“ unter Ausklammerung der Problemgruppen. Analog besteht eine *negative Selektion*, wenn insbesondere Problemgruppen an der Maßnahme teilnehmen. Aus den Monitoring-Daten sind diese Selektionseffekte nicht direkt ersichtlich und die im Monitoring ausgewiesenen Bruttoeffekte können entsprechend zu falschen Schlüssen zur Wirksamkeit der Maßnahme verleiten. Hingegen stellt die mikroökonomische Evaluationsli-

teratur ein breites Instrumentarium bereit, um solche Selektionsverzerrungen zu korrigieren, soweit die notwendigen, detaillierten und validen Mikrodaten zur Verfügung stehen (u.a. Heckman et al., 1999, Hagen und Fitzenberger, 2004). Folglich kann in Abhängigkeit von der zu evaluierenden arbeitsmarktpolitischen Maßnahme und den verfügbaren Daten eine entsprechend speziell zugeschnittene mikroökometrische Evaluationsmethode ausgewählt werden, um den kausalen Effekt der Maßnahmeteilnahme zu identifizieren. Erst mit diesen in der Evaluation geschätzten Nettoeffekten und nicht bereits auf den im Monitoring erfassten Bruttoeffekten, lassen sich letztlich Aussagen zur Effizienz und Effektivität des Programms treffen.

(2) Im einfachen Monitoring der Teilnehmer an Programmmaßnahmen können keine *indirekten Programmeffekte* auf Nichtteilnehmer erfasst werden. Solche Effekte auf Nichtteilnehmer bzw. allgemeine Gleichgewichtseffekte der Maßnahme auf die gesamte Volkswirtschaft sind vor allem bei umfangreichen arbeitsmarktpolitischen Programmen zu erwarten (Fitzenberger und Hujer, 2002: 7). Eine mikroökonomische Evaluation kann nur die direkten Maßnahmeeffekte messen. Ergänzend können jedoch die indirekten Effekte in einer makroökonomischen Evaluation ermittelt werden. Gemäß der Klassifikation von Calmfors (1994) unterscheidet man zwischen folgenden indirekten bzw. allgemeinen Gleichgewichtseffekten:

- *Mitnahmeeffekte*: Der Erfolg wäre auch ohne die Maßnahme eingetreten.
- *Substitutionseffekte*: Wegen Veränderung der relativen Lohnkosten werden bestimmte Typen von Nichtteilnehmern weniger nachgefragt.
- *Steuereffekte*: Die Finanzierung der Maßnahme führt zu einer höheren Abgabenbelastung bei Nichtteilnehmern bzw. der Gesamtwirtschaft.
- *Verdrängungseffekte*: Maßnahmeteilnehmer verdrängen Nichtteilnehmer aus regulären bzw. ungeforderten Beschäftigungsverhältnissen in anderen Unternehmen.

Zur Erfassung solcher makroökonomischen Effekte können allgemeine Gleichgewichtsmodelle und spezifische makroökonomische Verfahren eingesetzt werden (u.a. Caliendo et al. 2004). Solche Verfahren müssen in der Lage sein potentielle Simultanitäts- oder Endogenitätsproblem zu lösen, falls der Einsatz der arbeitsmarktpolitischen Maßnahme nicht exogen ist. Nur mit solchen in makroökonomischen Evaluationsstudien geschätzten indirekten Effekten lässt sich die Frage beantworten, ob die Maßnahme einen Nettogewinn für die Volkswirtschaft darstellt (Fitzenberger und Hujer, 2002: 7). Das Monitoring stößt hier an seine Grenzen, da es nicht solche indirekten Effekte messen kann.

(3) Monitoring-Daten erlauben keinen direkten, deskriptiven *Leistungsvergleich zwischen regionalen Arbeitsagenturen*. Ein deskriptiver Leistungsvergleich verlangt die klare Definition der Vergleichsobjekte, der Vergleichspartner und die Vergleichbarkeit der erhobenen Daten. Die darauf aufbauend gemessenen Unterschiede zwischen Arbeitsagenturen können nicht direkt auf unterschiedliche Programmeffizienzen zurückgeführt werden.

Denn die regionalen Agenturen agieren mit unterschiedlich dosierten Maßnahmeneinsätze und unter verschiedenen Kontextbedingungen (Brinkmann, 2000: 488f). Folglich muss sich zusätzlich noch eine Kausalanalyse der Leistungsunterschiede anschließen. Im Rahmen des Benchmarkings als weiterem Instrument der Begleitforschung existieren solche Methoden für einen deskriptiven und kausalanalytischen Leistungsvergleich zwischen Arbeitsagenturen bzw. Arbeitsmärkten, die in Kapitel 3 ausführlich beschrieben werden.

2.6. Zwischenfazit

Zusammenfassend lässt sich festhalten, dass Monitoring in Form des Arbeitsmarktpolitik-Monitoring (AMPM) die Dokumentation, jedoch nicht die Erklärung der Erreichung gesetzter Programmziele meint. Die Dokumentation erfolgt durch kontinuierlich wiederholte Beobachtungen statistischer Indikatoren, mittels derer der Zielerreichungsgrad gemessen wird. Wegen des zeitnahen Charakters des Monitorings können aus den beobachteten Daten erste Hinweise auf Fehlsteuerungen bei der Programmimplementierung gewonnen werden. Das AMPM ist vom reinen Arbeitsmarkt-Monitoring (AMM) zu differenzieren, das speziell auf die Ermittlung und Prognose des Qualifikationsbedarfs zielt, der sich aus dem Mismatch zwischen Arbeitsangebot und Arbeitsnachfrage ergibt. Eng verwandt ist das AMPM mit dem Zielcontrolling, das ein Teilelement des Controllings darstellt. AMPM geht über die im SGB III gesetzlich vorgeschriebenen Eingliederungsbilanzen hinaus, da es die Daten zeitnaher erfasst und zudem Outcome-Indikatoren einschließt.

AMPM ist in seiner Aussagekraft auf die Dokumentation begrenzt. Folglich lassen sich nur deskriptive Aussagen z.B. zu den Bruttoeffekten einer Maßnahme gewinnen. Um Aussagen zu den Nettoeffekten, also den kausalen Effekten der Maßnahmeteilnahme treffen zu können, ist eine ergänzende mikroökonomische Evaluationsstudie notwendig. Indirekte bzw. allgemeine Gleichgewichtseffekte der Maßnahme können nur in einer anschließenden makroökonomischen Evaluationsstudie gemessen werden. Für deskriptive und kausalanalytische Leistungsvergleiche zwischen regionalen Arbeitsagenturen ist eine Benchmarking-Studie das passende weitere Element der Begleitforschung zu arbeitsmarktpolitischen Maßnahmen.

3. Benchmarking

3.1. Definition des Benchmarkings

Benchmarking ist ein in der Unternehmenspraxis entwickeltes Management-Instrument, bei dem möglichst ähnliche Organisationseinheiten miteinander verglichen werden (Camp, 1989).⁵ Dieser Vergleich soll Lernprozesse und die Übertragung erfolgreicher Modelle („best practise“) ermöglichen, mit dem Ziel die eigene Performance im Wettbewerb zu verbessern. In einer engeren Definition konzentriert sich das Benchmarking ausschließlich auf den Vergleich mit dem „Klassenbesten“, da es das explizite Ziel ist, zum „Besten der Besten“ zu avancieren (Horváth und Herter, 1992: 5). Insbesondere im Public- und Non-Profit-Management wird dieses Steuerungsinstrument als Surrogat für den oftmals fehlenden Preiswettbewerb eingesetzt, um die erbrachten Leistungen bezüglich Effizienz und Effektivität bewerten zu können (Löffler, 1996). Entsprechend wird dieses Verfahren auch im Bereich der Arbeitsmarktpolitik beim Leistungsvergleich zwischen nationalen Arbeitsmärkten (u.a. Schmid et al., 1999) bzw. zwischen einzelnen regionalen Arbeitsagenturen (u.a. Mosley et al., 2003) angewandt.

In der Literatur differenziert man zwischen verschiedenen Arten des Benchmarkings. Eine Systematisierung ist erstens bezüglich des *Benchmarking-Objekts* möglich. Für den öffentlichen Sektor unterscheiden Schütz et al. (1998) das Benchmarking von Ergebnissen, Prozessen und Standards. Letzteres bezieht sich auf den Vergleich der Erreichung gesetzter Standards. Beispielhaft kann bezüglich einer spezifischen arbeitsmarktpolitischen Maßnahme ein Benchmarking der Ergebnisse, Prozesse und/oder erreichter Standards zwischen regionalen Arbeitsmarktagenturen durchgeführt werden. Der Vergleich von Prozessen gestaltet sich dabei jedoch umso schwieriger, je unstandardisierter der Prozess ist (Schütz et al., 1998). Wird die Performance regionaler Arbeitsagenturen generell verglichen, so ist es schwierig zwischen der Effizienz einzelner arbeitsmarktpolitischer Maßnahmen zu diskriminieren.

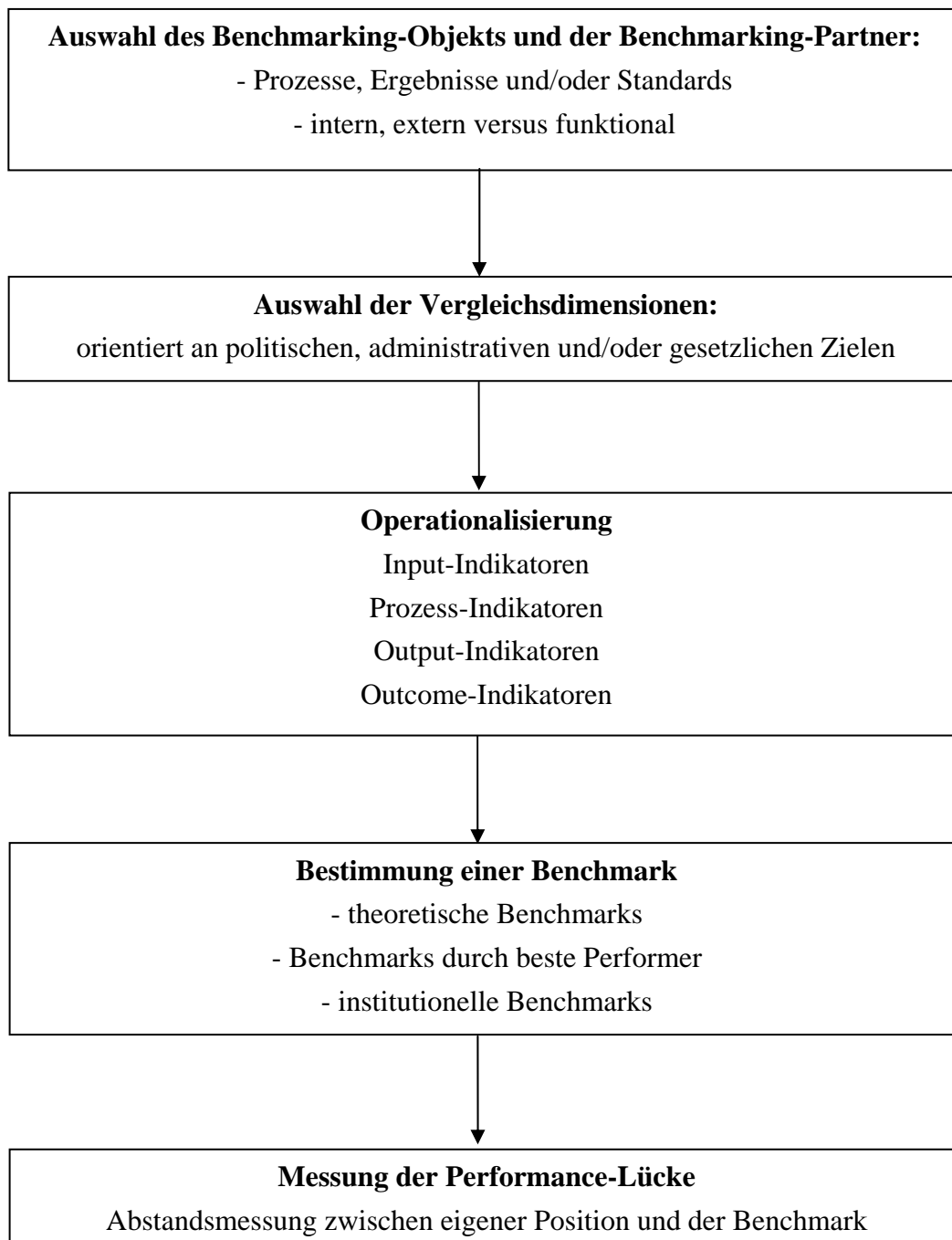
Als zweites Systematisierungskriterium lässt sich das Auswahlkriterium für die *Benchmarking-Partner* anführen. Man differenziert i.d.R. zwischen internem, externem und funktionalem Benchmarking (Camp, 1989: 60ff). Beim *internen* Benchmarking erfolgt ein Vergleich innerhalb einer Organisation. So können Ergebnisse, Prozesse oder die Erreichung von Standards der Subeinheiten einer Organisation einander gegenüber gestellt werden. Beim *externen* Benchmarking werden ähnliche externe Organisationseinheiten herangezogen. Werden diese Einheiten unabhängig von ihrem Operationsbereich ausgewählt, weil sich das Benchmarking ausschließlich auf spezifische Funktionen und Prozesse beschränkt, so spricht man vom *funktionalen* Benchmarking.

⁵ Ähnliche Ideen finden sich bereits in der älteren betriebswirtschaftlichen Literatur unter dem Begriff „Betriebsvergleich“ (Wöhe, 1973).

3.2. Vorgehensweise beim Benchmarking

Der Benchmarking-Prozess lässt sich grob in einen deskriptiven und einen kausalanalytischen Teil dividieren (Jones, 2004: 496). Der deskriptive Part dient dazu, die eigene Position im Vergleich zur Benchmark zu bestimmen. In der anschließenden Kausalanalyse versucht man die Ursachen für die Abweichung von der Benchmark mittels theoretischer Überlegungen und ökonometrischer Analysen zu identifizieren (Tronti, 1998: 503).

Abbildung 2: Stufen des deskriptiven Benchmarkings in der Arbeitsmarktpolitik



Quelle: eigene Darstellung

Der *deskriptive Teil* des Benchmarking-Prozesses lässt sich weiter in mehrere Stufen strukturieren (vgl. Abbildung 2). Erstens werden Vergleichsobjekt und die zu vergleichenden Einheiten festgelegt. Im zweiten Prozessschritt werden Dimensionen definiert, anhand derer die Effizienz, Effektivität bzw. Performance gemessen werden. Die Dimensionen entsprechen den politischen, administrativen und gesetzlich definierten Zielen. Drittens werden diese Dimensionen mittels Indikatoren operationalisiert, womit sich analog zum Monitoring ein mehrdimensionales Kennzahlensystem ergibt, so dass die Daten auch vom AMPM geliefert werden können. So kommen Input-, Output- und Outcome-Indikatoren für die Leistungsmessung in Frage. Die Indikatoren können differenziert nach Maßnahmenarten und Zielgruppen erhoben werden. Im vierten Schritt wird in jeder Dimension eine Benchmark bestimmt, anhand derer die eigene Leistung gemessen werden kann. Die Differenz zwischen der eigenen Position und der Benchmark misst letztlich im fünften Schritt die Performance-Lücke, die das Ausmaß der erforderlichen Leistungsverbesserung repräsentiert.

Für den deskriptiven Teil des Benchmarking-Prozesses stehen verschiedene Methoden zur Bestimmung und Visualisierung der Performance-Lücke bereit (u.a. Bogan und English, 1994). Beim Benchmarking der Performance von Arbeitsmärkten und Arbeitsagenturen fanden v.a. die Methoden des Radar-Charts verbunden mit der Surface Measure of Overall Performance (SMOP)-Analyse, die in Kapitel 3.4, und der Effizienzgrenze, die in Kapitel 3.5 näher erläutert werden, Anwendung.

Die Ursachen der Performance-Lücke können jedoch ausschließlich in einem anschließenden *kausalanalytischen Part* des Benchmarkings erklärt werden. Benchmarking geht also über einen bloßen Kennzahlenvergleich hinaus, indem kausale Analysen über die Performance-Lücke angestellt werden, mit dem Ziel von Organisationseinheiten mit besserer Performance in einem Know-How-Transfer zu lernen. Die ermittelten Ursache-Wirkungs-Zusammenhänge sind Grundlage für konkrete Verbesserungsvorschläge für die Leistungserstellung.

3.3. Generelle Probleme beim Benchmarking

(1) Bei der Wahl der Benchmarking-Partner wird man tendenziell mit einem Trade-off zwischen Vergleichbarkeit und Innovationspotenzial konfrontiert. Vorteil des internen Benchmarking gegenüber dem externen Benchmarking ist eine höhere Vergleichbarkeit der Benchmarking-Objekte und damit auch eine erleichterte Transferierbarkeit der „best practise“. Diese Problematik zeigt sich beispielsweise beim Benchmarking der Arbeitsmarkt-Performance. Vergleicht man die Performance international, so ist man mit einem hohen Maß an institutioneller und struktureller Heterogenität konfrontiert, was die Identi-

fikation und Diffusion von erfolgsversprechenden Faktoren erschwert (Tronti, 1998: 491f). Hingegen ist der Vergleich von regionalen Einheiten innerhalb eines Landes wegen der Ähnlichkeit der Strukturen einfacher. Nachteil ist jedoch das mangelnde Innovationspotenzial beim internen Benchmarking, wobei beim externen Vergleich speziell wegen der Heterogenität innovative Lösungskonzepte eher zu erwarten sind.

(2) Probleme beim Benchmarking können ähnlich wie beim Monitoring durch eine mangelhafte Datenqualität und inadäquate Messinstrumente auftreten. Um die Vergleichbarkeit und Akzeptanz der Daten zu garantieren, ist zum einen die Validität der Indikatoren zu sichern, d.h. die Indikatoren sollen die operationalisierten Ziele möglichst exakt messen. Zum anderen sollen die Indikatoren verlässlich sein, damit die Wiederholbarkeit der Analysen in Zeitreihen gewährleistet ist. Wegen Datenverfügbarkeitsproblemen wird tendenziell auf existierende bzw. einfach zu erhebende Indikatoren für leicht quantifizierbare Ziele zurückgegriffen, während die Operationalisierung qualitativer Ziele gar nicht bzw. oft nur mangelhaft erfolgt. Es gibt jedoch bereits Ansätze das Benchmarking auf Balanced Score Cards anzuwenden, die auch qualitative Dimensionen erfassen, wie z.B. der Arbeitsmarktservice (AMS) in Österreich (Schütz, 2001: 215f).

(3) Beim Benchmarking besteht die Gefahr des einfachen „Kopierens“ der Strategie des Besten (Schütz et al., 1998). Dieser Fehler lässt sich durch eine detaillierte Kausalanalyse der Ursachen der Performance-Lücke vermeiden. So kann eine unterschiedliche Performance auch das Ergebnis unterschiedlicher Kontextbedingungen sein. Daher ist eine Kontrolle struktureller Kontextvariablen im Rahmen der Kausalanalyse notwendig.

(4) Leistungsvergleiche auf Basis von Output-Indikatoren sind problematisch, soweit nicht Input-Indikatoren wie z.B. die Höhe der eingesetzten Mittel oder die Zahl der Mitarbeiter pro Arbeitssuchendem berücksichtigt werden (Caliendo und Jahn, 2004). Demnach ist bei der Ermittlung der Performance-Lücke bzw. spätestens bei deren Erklärung, für die unterschiedlichen Inputs der Benchmarking-Partner zu korrigieren.

3.4. Radar-Chart und SMOP-Analyse

Radar-Charts dienen der Illustration des Vergleichs verschiedener Performance-Indikatoren mit den jeweiligen Benchmarks und sind Grundlage für die Konstruktion eines zusammengefassten Indexes für die Performance-Lücke mit der SMOP-Methode.

3.4.1. Vorgehensweise

Analog zur generellen Vorgehensweise beim Benchmarking werden Performance-Dimensionen und –Indikatoren basierend auf politischen, gesetzlichen oder administrativen Zielerklärungen ausgewählt. Zwar existiert keine theoretische Grenze für die Dimen-

sionszahl, aber es empfiehlt sich aus Gründen der Übersichtlichkeit und potenzieller Korrelationsstrukturen die Begrenzung auf wenige, voneinander unabhängige Dimensionen (Mayer und Mosley, 1998: 24). Denn eine hohe Korrelation führt dazu, dass es bei gutem Abschneiden in einer Dimension automatisch auch zu einer guten Performance in der stark korrelierenden Dimension kommt. Methodisch stehen verschiedene Verfahren zur Aufdeckung der Korrelationsstruktur zwischen einzelnen Dimensionen zur Verfügung, wie z.B. Korrelationsmatrizen. Zusätzlich kann eine Hauptkomponentenanalyse eine Reduktion auf möglichst unabhängige Dimensionen erreichen. Auf die Elimination von Dimensionen sollte man jedoch verzichten, wenn die eng korrelierenden Dimensionen aus theoretischer Sicht unabhängige Aspekte der Performance repräsentieren.

Die ausgewählten Dimensionen bilden im nächsten Schritt der Visualisierung im Radar-Chart die radialen Achsen, auf denen jeweils der Zielerreichungsgrad in Bezug auf die jeweils gesetzte Benchmark markiert wird (vgl. Abbildung 3). Mayer und Mosley (1998: 11) differenzieren zwischen drei verschiedenen Typen von Benchmarks:

- *theoretische Benchmark*: Die Benchmark wird auf Basis theoretischer Überlegungen fixiert. Folglich ist z.B. die Benchmark bezüglich der Arbeitslosenquote Null, da aus mathematischer Perspektive kein geringerer Wert erreicht werden kann, auch wenn dieser aus praktischer Sicht nicht erreichbar scheint.
- *Benchmark durch beste Performer*: Die Benchmark wird gemäß der besten Performance der Vergleichspartner definiert.
- *institutionelle Benchmark*: Hierbei werden explizite Zielwerte seitens Politik und Verwaltung gesetzt.

Um verschieden skalierte Performance-Indikatoren in einem Radar-Chart sinnvoll abbilden zu können, empfiehlt sich eine Standardisierung der Performance-Werte in Bezug auf die Benchmark auf das Intervall [0,1] (Schütz et al., 1998: 39f). Das Radar-Chart wird letztlich komplettiert, indem gerade Verbindungslinien von den erreichten Performance-Niveaus in jeder Dimension zu den erreichten Niveaus auf den jeweils benachbarten Achsen gezogen werden.

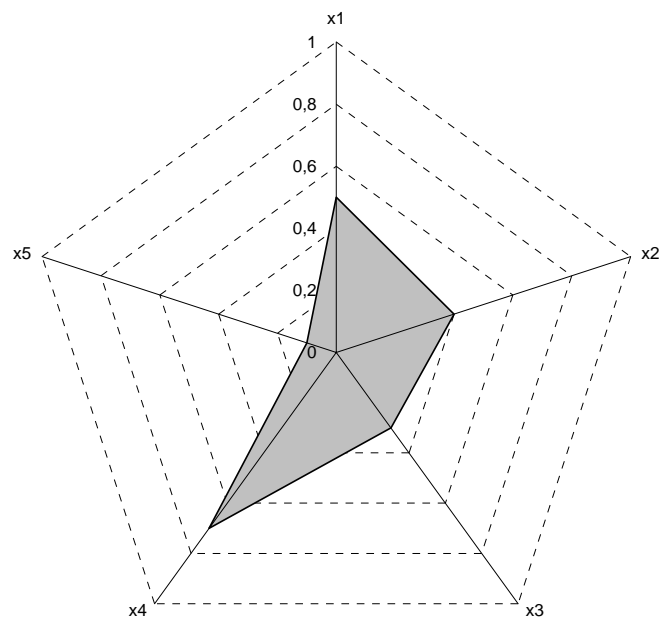
Abbildung 3 repräsentiert ein Beispiel für ein Radar-Chart mit fünf Dimensionen (x1 bis x5). Bei der anschließenden SMOP-Analyse wird ein zusammengesetzter Performance-Index über alle Performance-Dimension gemessen, indem mittels einer mathematischen Formel die Oberfläche des von den Verbindungslinien zwischen den Zielerreichungspunkten aufgespannten Polygons berechnet wird.⁶ In Abbildung 3 ist diese Fläche, die den SMOP-Wert bestimmt, grau unterlegt. Folglich vereint die Radar-Chart-Methode mit SMOP-Analyse den Vorteil einer kompakten Visualisierung der Performance im Vergleich zu den Benchmarks in jeder Zieldimension und den Vorteil einer Messung der Ge-

⁶ Für n Dimensionen lautet die SMOP-Formel:

$$SMOP = ((x_1 * x_2) + (x_2 * x_3) + (x_3 * x_4) + \dots + (x_n * x_1)) * \sin(360/n) / 2$$

samt-Performance in einem zusammengefassten Index, dem SMOP. Dieses Instrumentarium kann nicht nur dem Benchmarking zwischen Vergleichseinheiten, wie z.B. Arbeitsagenturen dienen, sondern es lassen sich bei gegebener Datenbasis auch zeitliche Veränderungen darstellen (Schütz et al., 1998: 39). So kann einerseits die zeitliche Veränderung der eigenen Performance im Vergleich zur Benchmark im Radar-Chart in Form von Polygonen für jeden Untersuchungszeitpunkt visualisiert werden. Andererseits dient ein Vergleich der SMOP-Werte zu unterschiedlichen Zeitpunkten der Quantifizierung der im Radar-Chart dargestellten Veränderungen der Zielerreichungsgrade.

Abbildung 3: Radar-Chart



Quelle: eigene Darstellung

Ein *Fallbeispiel der Radar-Chart/SMOP-Analyse für das Benchmarking von nationalen Arbeitsmärkten* ist u.a. Schmid et al. (1999), die mehrere Indikatoren entlang der Dimensionen soziale Integration, Integration Jugendlicher, Chancengleichheit zwischen den Geschlechtern und Arbeitslosigkeitsdauer mittels mehrerer Indikatoren auswählen. Ein *Fallbeispiel der Radar-Chart/SMOP-Analyse für das Benchmarking von internationalen Arbeitsmärkten* ist u.a. Mosley und Mayer (1999), die Zieldimensionen und Indikatoren orientiert an den arbeitsmarktpolitischen Zielen der EU ableiten. Für die Analyse der Performance in Hinblick auf die Beschäftigung wählen die Autoren die Dimensionen Beschäftigungswachstum, Beschäftigungsniveau, Chancengleichheit und Integration älterer Arbeitnehmer. Als zweite Analyse untersuchen sie die Performance hinsichtlich der Arbeitslo-

sigkeit basierend auf den Dimensionen Bekämpfung der Arbeitslosigkeit, Chancengleichheit, Arbeitsmarktintegration Jugendlicher und Arbeitsintegration von Problemgruppen.

3.4.2. Kritik

Bei der Radar-Chart/SMOP-Methode existieren allerdings einige nicht zu vernachlässigende theoretische und methodologische Probleme:

(1) SMOP-Werte sind sensitiv bezüglich der Auswahl der Performance-Dimensionen, der Indikatoren und der Benchmarks (Mayer und Mosley, 1998: 7f). Es handelt sich dabei um normative Setzungen, die theoretisch und empirisch gerechtfertigt werden müssen. Die Gefahr besteht insbesondere darin, dass aus Gründen der Datenverfügbarkeit auf eine Quantifizierung qualitativer Zieldimensionen verzichtet wird bzw. bei quantitativen Zielen auf inadäquate Indikatoren zurückgegriffen wird. Diesem Datenproblem kann jedoch mit einem gezielten Monitoring in Form zusätzlicher Datenerhebungen begegnet werden.

(2) Problematisch erweist sich die Abhängigkeit der SMOP-Werte von der Gewichtung der einzelnen Dimensionen (Tronti, 1998: 502). Auch bei Gewichtungen handelt es sich um normative Setzungen. Schütz et al. (1998: 43) verweisen jedoch als Lösung auf das methodische Instrumentarium der Hauptkomponentenanalyse hin, die nicht nur für die Dimensionsreduktion verwendet werden kann, sondern auch ein Gewichtungsschemata aus den Daten ermittelt.

(3) Es existiert eine Abhängigkeit der SMOP-Werte von der Anordnung der Radialachsen im Radar-Chart. Mayer und Mosley (1998: Appendix B) schlagen als Lösung vor, den Durchschnitt aller SMOP-Werte über alle theoretisch möglichen Anordnungskombinationen der Dimensionen zu bilden.

(4) Außerdem sieht sich die SMOP-Analyse der Kritik der Substituierbarkeit zwischen den Dimensionen ausgesetzt, die an allen additiven Indizes geübt wird (Mayer und Mosley, 1998: 8). So ist bei der Ermittlung der Gesamt-Performance eine Kompensation zwischen einzelnen Dimensionen möglich, d.h. eine niedrige Performance in einer Dimension kann durch eine hohe Performance in einer anderen Dimension ausgeglichen werden.

(5) Fundamentaler Kritikpunkt ist, dass zwar mittels der Radar-Chart Methode und Berechnung der SMOP-Werte eine grafische Darstellung und Quantifizierung der Performance-Lücke möglich ist, aber nicht eine kausalanalytische Erklärung dieser Unterschiede mittels dieses Instrumentariums. Dem kann man entgegenhalten, dass diese Methode einen rein deskriptiven Charakter hat und gar keinen Anspruch auf Erklärungen erhebt. Zur Erklärung der Performance sind folglich weitere methodische Werkzeuge notwendig (Mosley und Mayer, 1998: 15). Tronti (1998: 502f) verweist auf die Notwendigkeit neben theoretischen Überlegungen alle potenziellen ökonometrischen Evaluationstechniken einzusetzen. Beispielsweise schlagen Schmid et al. (1999 897) vor, die SMOP-Werte als ab-

hängige Variable in einer Regression zu verwenden und damit die Erklärungskraft potenzieller Einflussfaktoren auf die Performance zu bestimmen.

(6) Ein weiterer zentraler Kritikpunkt am SMOP-Ansatz lautet, dass die Performance-Messung unter Ausblendung des Produktionsprozesses geschieht, weil die Inputfaktoren nicht in die Analyse einfließen. Folglich wird die Effizienz als Relation zwischen Input- und Outputfaktoren nicht erfasst. Tronti (1998) schlägt daher vor, Benchmarking als Effizienzvergleiche mit dem Instrumentarium der technischen Effizienzanalyse zu implementieren.

3.5. Technische Effizienzanalyse

Die Methode der Effizienzgrenze (EFA – efficiency frontier approach) hat ihren Ursprung in der Produktionstheorie (Farrell, 1957). Im Gegensatz zu anderen Benchmarking-Methoden finden bei der EFA auch die Inputfaktoren explizit Beachtung beim Performance-Vergleich.

3.5.1. Vorgehensweise

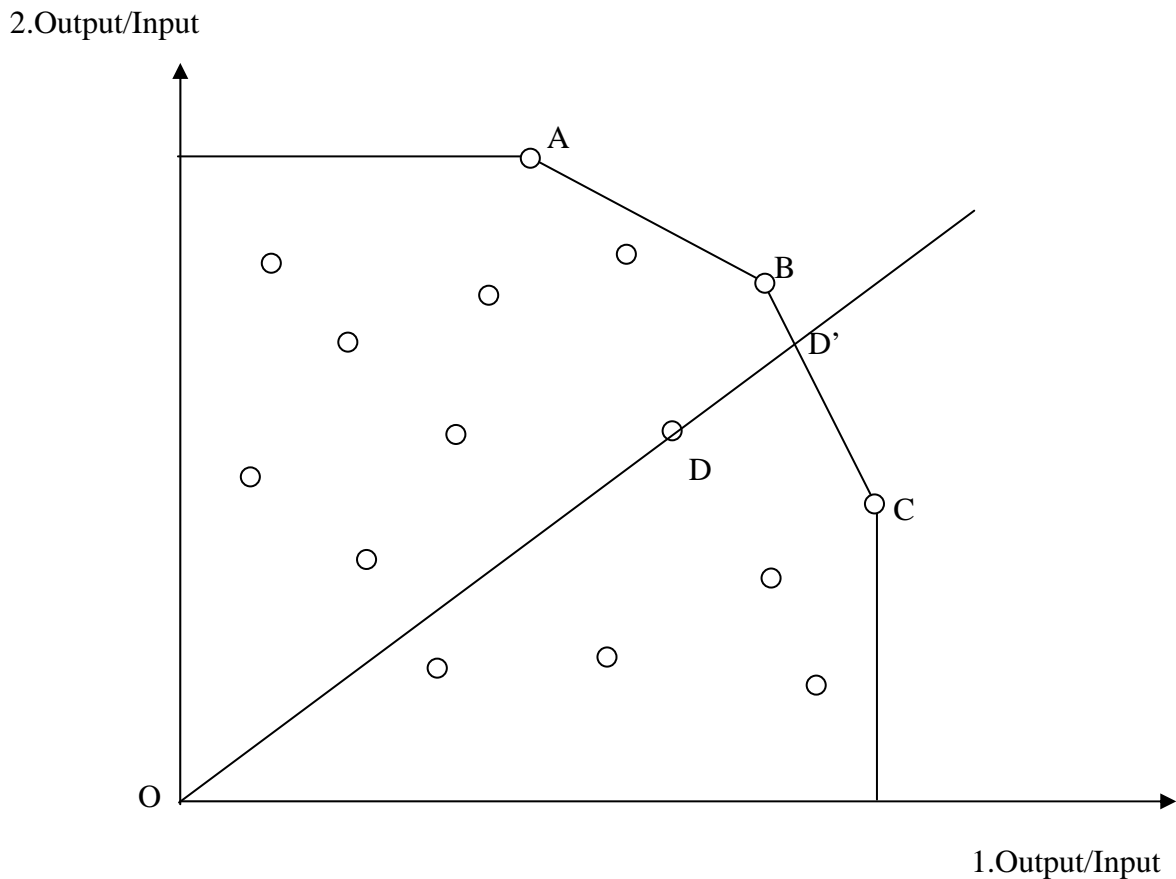
Aus der Produktionstheorie ist bekannt, dass die Produktionsfunktion den Rand der Produktionsmöglichkeitenmenge als Effizienzgrenze bildet, auf der die Inputs und Outputs effizient kombiniert werden. Zur Schätzung dieser Effizienzgrenze existieren zwei Methoden: ein stochastischer und ein mathematischer Ansatz. Beim stochastischen Ansatz werden spezifische funktionale Formen bezüglich der Technologie angenommen und zwischen Ineffizienz und Zufallseinflüssen differenziert, wobei spezifische Verteilungsannahmen bezüglich der Zufallskomponente unterstellt werden müssen. Einige Benchmarking-Studien zu Arbeitsmärkten (u.a. Storrie und Bjurek, 2000) und von Arbeitsagenturen (u.a. Mosley et al., 2003) verwenden den mathematischen Ansatz der Effizienzgrenze in Form der Data Envelopment Analysis (DEA). Im Gegensatz zur stochastischen Version ist dies ein nicht-parametrischer Ansatz, beim dem die Effizienzgrenze mit Hilfe der Linearen Programmierung konstruiert wird. Zudem wird keine Trennung zwischen Ineffizienz und Zufallseinflüssen modelliert. Beispielhaft für die Effizienzanalyse soll hier die generelle Vorgehensweise der DEA skizziert werden.

Die Effizienzgrenze repräsentiert die Benchmark und wird bei der DEA aus den Benchmarking-Partnern gebildet, die in ihrer Input-Output-Kombination nicht dominiert werden. Eine Einheit wird genau dann dominiert, wenn eine andere Einheit bei gegebenem Input einen höheren Output erreicht bzw. einen gegebenen Output mit einem geringeren Input erreicht. Zwischen diesen nicht-dominierten Einheiten auf der Produktionsgrenze besteht jedoch keine weitere Möglichkeit zu differenzieren. Um einen stetigen Verlauf der Pro-

duktionsgrenze zu garantieren, trifft man die Konvexitäts-Annahme, dass Linearkombinationen nicht-dominierter Einheiten auch auf der Produktionsgrenze liegen. D.h. auf Basis beobachteter Benchmarks werden virtuelle Benchmarks durch Kombination der beobachteten Benchmarks konstruiert. Die Effizienzlücke wird anschließend aus der Distanz der Untersuchungseinheiten zur Effizienzgrenze gemessen (Mosley et al., 2003: 33ff). Im Fall von einem Input- und zwei Outputfaktoren lässt sich die Vorgehensweise noch grafisch illustrieren.

In Abbildung 4 sind verschiedene Benchmarking-Einheiten gemäß ihren Input-Output-Kombinationen positioniert. Der effiziente Rand wird von den Einheiten A, B und C, sowie den Verbindungslinien gebildet. Beobachtungseinheit D ist ineffizient, da sie von der Einheit B dominiert wird, weil B für beide Outputindikatoren ein besseres Output-Input-Verhältnis erreicht als D. Zeichnet man einen Fahrstrahl aus dem Ursprung O durch D, der die Effizienzgrenze in D' schneidet, so bemisst sich die Effizienzlücke als Quotient DD'/OD' .

Abbildung 4: Effizienzgrenze bei einem Input- und zwei Outputfaktoren



Quelle: eigene Darstellung

Ein *Fallbeispiel der DEA für das Benchmarking von nationalen Arbeitsmärkten* ist Storrie und Bjurek (2000). Es handelt sich dabei jedoch eher um eine Index-Theorie, da die Effizienzgrenze unter Ausblendung von Inputfaktoren nur auf Basis von Performance-Indikatoren konstruiert wird. Gegenüber dem Radar-Chart/SMOP-Analyse hat diese Vorgehensweise jedoch immer noch den Vorteil, dass die Gewichtung der verschiedenen Performance-Indikatoren zu einem zusammengesetzten Index nicht explizit per Annahmen erfolgt, sondern aus den Daten selbst generiert wird (Storrie und Bjurek, 2000: 2). Zudem wird auch die Effizienzgrenze aus beobachteten Einheiten gebildet und die Performance-Lücke nur zu Benchmarks mit ähnlichen Ausprägungen der Performance-Indikatoren gemessen.

Ein *Fallbeispiel der DEA für das Benchmarking von Arbeitsverwaltungen* ist Mosley et al. (2003). Sie verwenden den Ansatz einer outputorientierten DEA, d.h. das Ziel der linearen Programmierung ist, bei gegebenem Input den maximalen Output zu erzeugen. Die Autoren rechtfertigen diese Wahl damit, dass die Personalbesetzung und die Programmausgaben als Inputvariablen außerhalb der Kontrolle der Arbeitsämter liegen.⁷ Für die Untersuchung werden die beiden Konzepte der Aktivitäts-Effizienz und der Wirkungs-Effizienz eingeführt. Aktivitäts-Effizienz misst das Verhältnis von Output-Indikatoren (z.B. die Zahl der beendeten Maßnahmen) zu den Input-Indikatoren (z.B. Personaleinsatz im Arbeitsamt). Im Gegensatz dazu bezieht die Wirkungs-Effizienz, die auch als Effektivität bezeichnet wird, Outcome-Indikatoren (z.B. die Verbleibsquote) auf Output-Indikatoren (z.B. die Zahl der beendeten Maßnahmen). Folglich repräsentieren die Output-Indikatoren bei der Messung der Effektivität die DEA-Inputs, während sie bei der Messung der Effizienz die DEA-Outputs darstellen.

3.5.2. Kritik

Klarer Vorteil der DEA im Gegensatz zu anderen deskriptiven Benchmarking-Methoden wie z.B. der Radar-Chart/SMOP-Analyse ist, dass die Inputfaktoren in die Performance-Analyse einfließen. Die DEA zeichnet sich zudem dadurch aus, dass zum einen die Benchmarks nicht theoretisch fixiert werden, sondern durch die beobachteten Einheiten mit den besten Input-Output-Relationen definiert werden. Zum anderen ist bei der Aggregation der unterschiedlichen Dimensionen zu einer Maßzahl in der Form der Effizienzlücke kein explizites Gewichtungsschema notwendig. Die Gewichte entstehen induktiv aus den Daten. Folglich wird keine monetäre Bewertung der Input- und Outputgrößen benötigt wird, weshalb die DEA insbesondere für die Effizienzmessung im öffentlichen Sektor

⁷ Im Gegensatz dazu wird bei einer inputorientierten DEA angenommen, dass das Niveau des Outputs gegeben ist und Effizienzsteigerungen durch Reduktionen des Inputs erreicht werden können. Beispielsweise verwenden Althin und Behrenz (2004) ein inputorientiertes Modell für das Benchmarking schwedischer Arbeitsämter.

attraktiv ist, da eine Gewichtung des Outputs im öffentlichen Sektor oftmals am Mangel an Preisen scheitert. Ein weiterer Vorteil der DEA ist, dass die Effizienzlücke nur im Vergleich zu ähnlichen Einheiten (sog. „peers“) berechnet wird, die auf der Effizienzgrenze liegen (Mosley et al., 2003: 35).

Die DEA birgt jedoch die Problematik, dass ohne weitere Annahmen keine Diskriminierung zwischen nicht-dominierten Untersuchungseinheiten auf der Effizienzgrenze möglich ist. Insbesondere bei einer hohen Anzahl an Dimensionen steigt die Wahrscheinlichkeit, nicht-dominierte Einheiten zu finden zwischen denen keine Diskriminierung mehr möglich ist, womit die Aussagekraft der DEA schrumpft.

Der Nachteil der DEA besteht außerdem darin, dass mit den ermittelten Effizienz- bzw. Performance-Lücken keine kausalen Aussagen getroffen werden können. Auch hier muss sich ein kausalanalytischer Forschungsschritt anschließen, der die Ursachen der Performance-Unterschiede identifiziert, analog zur Arbeitsteilung zwischen dem deskriptiven Monitoring und der kausalanalytischen Evaluation.

3.6. Kausalanalytisches Benchmarking

Ein Ansatz zum kausalanalytischen Benchmarking findet sich in der oben beschriebenen Arbeit von Mosley et al. (2003), die nach Implementierung der DEA eine Regressionsanalyse des Einflusses von Umweltfaktoren auf die in der DEA ermittelten Effizienz- und Effektivitätswerte durchführen.⁸ So erfolgt eine Kontrolle für die Stärke der regionalen Problemlagen, z.B. gemessen über die durchschnittliche Dauer und Höhe der Arbeitslosigkeit in der Region, und für weitere Strukturmerkmale, wie z.B. der Unterschied zwischen Stadt und Land, die Wirtschaftsstruktur, sowie die sozioökonomische Struktur. Als Methodik verwenden sie ein Tobit-Modell, da die Effizienz- und Effektivitätswerte aus der DEA auf das Intervall [0; 1] beschränkt sind. Die so ermittelten, um Umwelteinflüsse korrigierten Effizienz- und Effektivitätswerte versuchen Mosley et al. (2003) mit Organisationscharakteristika in den Arbeitsämtern auf Basis von Fallstudien in Form systematischer Paarvergleiche zu erklären. Diese Vorgehensweise ist jedoch dann problematisch, wenn nicht sämtliche relevanten Kontextvariablen bei der Korrektur der Effizienz- und Effektivitätswerte kontrolliert werden. In diesem Fall wären die Werte nicht vollkommen von Umwelteinflüssen bereinigt, weshalb sie nicht als das ausschließliche Ergebnis der Organisationseffizienz der Arbeitsämter interpretierbar wären. Insbesondere bei nicht erfassten räumlichen Spill-over Effekten zwischen benachbarten Arbeitsämtern würden verzerrte Ergebnisse entstehen.

⁸ Folglich wird davon ausgegangen, dass die Umweltfaktoren keine Inputfaktoren sind. Alternativ könnten die Umwelteinflüsse als exogene Inputs direkt in die DEA einbezogen werden.

Folglich stellt sich für die weitere empirische Forschung zum Benchmarking die Aufgabe diese Kontextvariablen und räumlichen Spill-over Effekte besser zu erfassen und zu kontrollieren. Zudem sollte der Einfluss der Organisationscharakteristika der regionalen Arbeitsagenturen auf die Performance-Lücke nicht nur in Fallstudien wie bei Mosley et al. (2003), sondern auch mittels ökonometrischer Verfahren bestimmt werden. Weiterhin könnte die Sensitivität der DEA-Ergebnisse beim Benchmarking von Arbeitsagenturen bezüglich unterschiedlicher DEA-Modellwahlen näher analysiert werden. Alternativ können auch stochastische Modelle der Effizienzgrenze zum Benchmarking von regionalen Arbeitsagenturen analog zu Analysen der regionalen Matching-Effizienz (u.a. Fahr und Sunde, 2005) verwendet werden. Auch bei dieser Methode existieren verschiedene Modellspezifikationen, die eingesetzt werden können (u.a. Greene, 2005).

3.7. Zwischenfazit

Benchmarking meint bezogen auf die Arbeitsmarktpolitik den Leistungsvergleich zwischen Arbeitsmärkten bzw. den Vergleich regionaler Arbeitsagenturen bezüglich der Effizienz der Maßnahmenimplementierung. Der Leistungsvergleich erfolgt analog zum Monitoring entlang spezifischer Dimensionen, die sich aus übergeordneten Zielen ableiten und mittels Indikatoren operationalisiert werden. Als Vergleichsmaßstab werden Benchmarks theoretisch oder durch die Position der besten Benchmarking-Partner gesetzt. Für diesen deskriptiven Part des Benchmarkings existieren unterschiedliche Methoden, wie beispielsweise Radar-Charts mit SMOP-Analyse und Analyseverfahren der Effizienzgrenze, die durch spezifische Vor- und Nachteile gekennzeichnet sind. Zur kausalanalytischen Erklärung der Leistungsunterschiede zwischen Arbeitsagenturen bzw. Arbeitsmärkten müssen die Kontextfaktoren berücksichtigt werden in Form der strukturellen Komposition der Bevölkerung, der wirtschaftlichen Situation und den institutionellen Rahmenbedingungen. Erst nach der Korrektur um diese Umweltfaktoren können Effizienzunterschiede zwischen Benchmarking-Einheiten einer Untersuchung organisationsinterner Faktoren unterzogen werden. Für die weitere Forschung stellt sich hierbei die Aufgabe, weitere adäquate Methoden in empirischen Analysen anzuwenden.

4. Einordnung von Monitoring und Benchmarking in die Begleitforschung

Es besteht eine Abgrenzungsproblematik zwischen verschiedenen Elementen der Begleitforschung, insbesondere AMM, AMPM, Benchmarking und Evaluation. Die Ursache liegt darin, dass die Begriffe eng miteinander verflochten sind und oftmals unterschiedlich je nach Autor definiert werden, was eine klare Begriffstrennung erschwert. In der Literatur finden sich verschiedene Ansätze zwischen einzelnen Teilelementen zu differenzieren und Interdependenzen offen zu legen (u.a. Auer und Kruppe 1996, Brinkmann 1999, 2000, LASA 2000). Unter Rückgriff auf diese existierenden Konzepte wird im Folgenden der Zusammenhang zwischen diesen Elementen illustriert und erörtert. Während sich die bisherigen Konzepte hauptsächlich auf die Abgrenzung von Monitoring und Evaluation konzentrieren, soll hier auch das Benchmarking explizit in die Übersicht integriert werden.

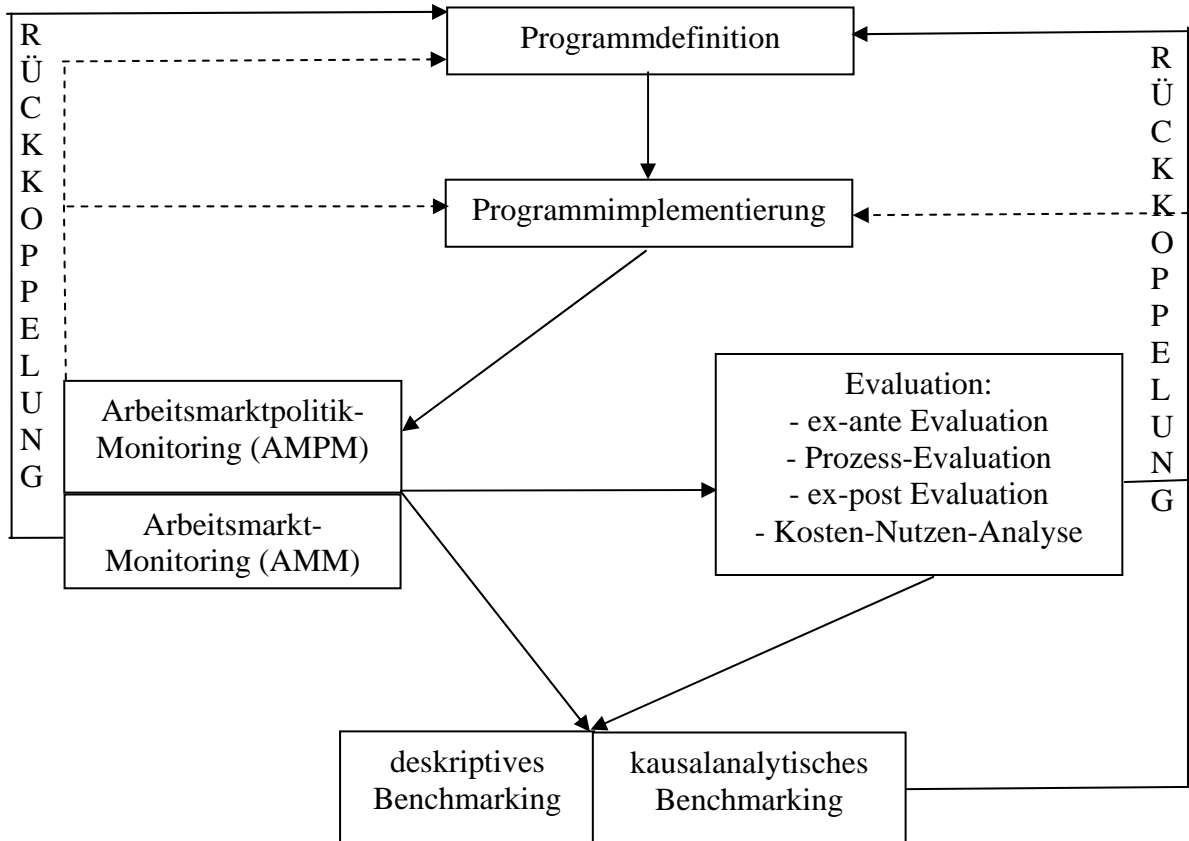
4.1. Generelle Vorgehensweise in der Begleitforschung

Abbildung 5 versucht die genannten Elemente der Begleitforschung in einen Rahmen zu integrieren. Die Begleitforschung zielt generell mit all ihren Elementen darauf, Politikinterventionen hinsichtlich ihrer Effizienz und Effektivität zu analysieren und darauf aufbauend Handlungsempfehlungen für die Steuerung auszusprechen.

Auf Basis von Zielen, die auf politischer, administrativer und/oder gesetzlicher Ebene formuliert werden, orientiert sich die Definition der arbeitsmarktpolitischen Programmmaßnahmen. Im nächsten Schritt werden diese Programme in der Praxis implementiert. Zur Bewertung des Erfolgs der eingeleiteten Maßnahmen stehen die Instrumente Monitoring, Benchmarking und Evaluation bereit. Diese Unterteilung spiegelt die klassische dreiphasige Struktur eines Programms mit Entwicklung, Umsetzung und Ergebnisfeststellung wider (Deeke und Kruppe, 2002: 400). Die Elemente der Begleitforschung stehen in engen Austauschbeziehungen zueinander, die durch Pfeile in Abbildung 5 symbolisiert werden. Die Pfeilrichtung repräsentiert die Richtung des Informationsflusses und damit auch die zeitliche Abfolge der einzelnen Elemente. Die Ergebnisse der Begleitforschung werden zudem über Rückkoppelungsschleifen zum einen zur Programmimplementierung und zum anderen insbesondere zur Programmdefinition übertragen. Durchgezogene Linien stehen für stärkere Rückkoppelungen als unterbrochene Linien. Die rückgekoppelten Informationen können zu Änderungen des Programmablaufs genutzt werden, z.B. wenn fehlgesteuerte Wirkungen in der Evaluation festgestellt werden. Zeigt die Begleitforschung fundamentale Defizite eines laufenden Programms auf, so können diese Informationen zu generellen Änderungen der Programmdefinition auf der politischen und/oder ad-

ministrativen Ebene führen. Ist die arbeitsmarktpolitische Maßnahme bereits abgeschlossen, können die Forschungsergebnisse auch bei der Definition zukünftiger arbeitsmarktpolitischer Maßnahmen berücksichtigt werden.

Abbildung 5: Einordnung von Monitoring und Benchmarking in die Begleitforschung



Quelle: eigene Darstellung (angelehnt an Auer und Kruppe, 1996: Abb. 30.1)

4.2. AMM versus AMPM

Das Verhältnis zwischen dem Monitoring von Arbeitsmärkten (AMM) und dem Monitoring arbeitsmarktpolitischer Maßnahmen (AMPM) ist durch eine wechselseitige Abhängigkeit gekennzeichnet, wobei die Suche nach Schnittstellen erst am Anfang der theoretischen und praktischen Entwicklung steht (LASA 2000: 12). Bereits Auer und Kruppe (1996: 910) erkannten die Interaktion, indem sie postulierten, dass die Arbeitsmarktsituation, die durch das AMM erfasst wird, Kriterium für die Mittelvergabe bei arbeitsmarktpolitischen Maßnahmen ist, was sich letztlich auf das AMPM auswirkt. Wer-

den beispielsweise im AMM spezifische Qualifikationsdefizite von Arbeitnehmern aufgedeckt, die zu einem Mismatch zwischen Arbeitsangebot und Arbeitsnachfrage führen, so kann eine Weiterbildungsmaßnahme zur Problembehebung definiert werden. Ob diese Maßnahme die erhofften Wirkungen zeigt, wird in der Begleitforschung zur Maßnahme analysiert. Die Auswahl der Zieldimensionen und die konkrete Programmausgestaltung kann auch von Resultaten des AMM beeinflusst werden. Speziell die Zielhöhen beim AMPM orientieren sich an den Ergebnissen des AMM in Form von Trendberechnungen unter Beachtung der wirtschaftlichen und institutionellen Kontextbedingungen (Schütz, 2001: 213).

Zum anderen können umfangreiche arbeitsmarktpolitische Programme makroökonomische Wirkungen entfalten, die sich im AMM widerspiegeln. Solche makroökonomischen Effekte sind insbesondere bei umfangreichen arbeitsmarktpolitischen Maßnahmen zu erwarten. Demnach empfiehlt es sich für das AMM, die Effekte aus dem AMPM in die Prognosen einzuarbeiten.

Eine weitere Interdependenz besteht darin, dass die Effekte der arbeitsmarktpolitischen Maßnahmen, die mit dem AMPM gemessen werden, abhängig von der Arbeitsmarktsituation sind, die im AMM erfasst wird (Auer, 1998: 175). In der anschließenden Evaluation sind diese Kontextbedingungen zu berücksichtigen. Insbesondere die makroökonomische Evaluation benötigt Informationen des AMM. Dieser Zusammenhang gilt auch für das sich dem Monitoring anschließende kausalanalytische Benchmarking, bei dem die Kontextbedingungen Berücksichtigung finden, wenn man die relative Performance einzelner Arbeitsagenturen bestimmt. Folglich ist ein Informationsfluss sowohl vom AMPM als auch vom AMM hin zur Evaluation und dem Benchmarking notwendig, wie es die Pfeile in Abbildung 5 illustrieren.

4.3. Monitoring versus Evaluation

In der Literatur finden sich zahlreiche Versuche, um zwischen Monitoring und Evaluation zu differenzieren. Bereits die Standarddefinition des Monitorings nach Auer und Kruppe (1996), die in Kapitel 2 vorgestellt wurde, betont den Unterschied zwischen Monitoring und Evaluation, wonach Monitoring ausschließlich der Deskription der Performance dient und die Evaluation ergänzend versucht, die Zielerreichung zu erklären. Entsprechend charakterisieren Brinkmann und Wießner (2002: 380) das Verhältnis zwischen der primär kausalanalytisch orientierten Wirkungsforschung und den vorwiegend routinisierten und quantitativen Monitoring-Aktivitäten der Verwaltung als nicht konkurrierend, sondern notwendigerweise *komplementär*. Das Monitoring liefert in diesem Prozess detaillierte Daten zu den einzelnen Indikatoren, aus denen sich Bruttoeffekte bestimmen lassen. Sol-

che Bruttoeffekte sind jedoch wie in Kapitel 2.5 erörtert wurde, nicht als kausale Programmwirkungen zu interpretieren. Stattdessen bedarf es nichtexperimenteller Evaluationsmethoden, um die kausalen Nettoeffekte zu identifizieren. Als Methoden dieser *ex-post Evaluation* stehen zum einen mikroökonomische Verfahren zur Messung der direkten Programmeffekte und zum anderen makroökonomische Verfahren zur Quantifizierung auch der indirekten Effekte zur Verfügung.

Neben der *ex-post Evaluation* ist auch die *ex-ante Evaluation* integraler Bestandteil der Begleitforschung. Solche Studien dienen zur Programmvorbereitung, indem sie potenzielle direkte und indirekte Programmwirkungen unter Beachtung der existierenden Arbeitsmarktsituation unter Verwendung von Daten des AMM auf Mikro- und/oder Makroebene simulieren. Im Fokus steht dabei insbesondere die Frage, in welchem Maße die erhofften Effekte bei den Zielgruppen eintreten werden. Ergebnisse von *ex-ante Evaluationen* können daher genutzt werden, um Vorschläge zur Programmgestaltung zu machen (Deeke und Kruppe, 2002: 400).

Die Erhebung von Input- und Prozessindikatoren im Rahmen des AMPM ist Voraussetzung für eine *begleitende Evaluation* (Implementierungsanalyse) als weiteres Evaluationsinstrument, das auf die Erfassung der Effizienz der Programmsteuerung zielt. Mittels detaillierter Analysen lassen sich Probleme bei der organisationsinternen Programmumsetzung identifizieren und bei Bedarf Programmkorrekturen durchführen.

Unter den Begriff Evaluation werden i.d.R. auch *Kosten-Nutzen-Analysen* subsumiert. Während die Mikroevaluationen mit der Messung kausaler Effekte und die Makroevaluation mit der zusätzlichen Messung indirekter Effekte nur den Nutzen der Maßnahme analysieren, versuchen Kosten-Nutzen-Analysen zusätzlich auch die individuellen und volkswirtschaftlichen Kosten der Maßnahme zu erfassen. In einem Vergleich untersucht dieses Verfahren, ob die identifizierten Nettoeffekte die Kosten der Maßnahme rechtfertigen. Idealerweise werden nicht nur die individuellen Nettoeffekte, sondern auch indirekte Effekte auf die Nichtteilnehmer berücksichtigt. Auf der Kostenseite fließen neben den direkten Kosten auch die indirekten Kosten in Form entgangener Löhne, Sozialbeiträge und Steuerzahlungen in die Kalkulation ein. Solche Kosten-Nutzen-Analysen stellen damit einen zusätzlichen Informationsgewinn zur Beurteilung der Maßnahmeneffizienz dar, da sie den Nutzen in Relation zu den Kosten setzen.

Monitoring und Evaluation ergänzen sich zudem dadurch, dass Monitoring zeitnah erste Ergebnisse mit gesetzten Zielen vergleicht und damit erste Hinweise gibt, wo spätere Evaluationsstudien ansetzen können. Der prozessuale Charakter des Monitorings erlaubt begleitende Implementierungsanalysen, die auf die Identifikation von Vollzugsdefiziten zielt (Deeke und Kruppe, 2002: 400). Komplexere Evaluationsstudien können meist erst nach Programmende durchgeführt werden, so dass diese Ergebnisse erst zeitverzögert im Ge-

gensatz zum Monitoring vorliegen, womit die Evaluation eher zur mittel- bis langfristigen Steuerung geeignet sind.

In Abbildung 5 ist die Beziehung zwischen AMPM und Evaluation illustriert. Die zeitliche Abfolge ist klar definiert, wonach zunächst das AMPM die Daten zu den einzelnen Indikatoren dokumentiert und dann im Rahmen der Evaluation diese Daten zur Identifikation kausaler Effekte der Programmpartizipation verwendet werden (Affholter, 1994: 98). Das AMM liefert komplementär Daten für die makroökonomische Evaluation der indirekten Effekte. Demnach kommt dem Monitoring generell die Funktion zu, Daten für ex-ante, begleitende und ex-post Evaluationsstudien zu sammeln und aufzubereiten. Folglich erleichtert Monitoring die Evaluation, da das Vorhandensein von Daten die Neuerhebung von Daten für die Evaluation teilweise erspart (Erhel, 2003: 421). Detaillierte Informationen zu Inputs, Prozessen, Outputs und Outcomes wie sie im Rahmen eines idealtypischen AMPM erhoben werden, verbessern insbesondere mit den differenzierten vermittlungs- und personenbezogenen Informationen die Konstruktion statistischer Kontrollgruppen im Rahmen der mikroökonomischen Evaluation. Je besser die vom Monitoring bereitgestellten Daten für die Evaluation sind, desto leichter sind die identifizierenden Annahmen bei der Schätzung kausaler Effekte erfüllt. Die detaillierten Daten des Finanz-Monitoring eignen sich zudem, um die direkten Kosten arbeitsmarktpolitischer Maßnahmen im Rahmen einer Kosten-Nutzen-Analyse zu messen.

4.4. Benchmarking in der Begleitforschung

Analog zur Differenzierung zwischen Monitoring und Evaluation kann man beim Benchmarking zwischen einem deskriptiven Part und einem kausalanalytischen Part unterscheiden (Tronti, 1998: 503). Daher ist eine eindeutige Zuordnung des Benchmarkings entweder zum Monitoring oder zur Evaluation nicht angebracht. Die Zuweisung des deskriptiven Benchmarkings zum Monitoring und die des kausalanalytischen Benchmarkings zur Evaluation würden den Charakter des Benchmarkings als eigenständiges und zusätzliches Element der Begleitforschung schmälern. Daher ist es in Abbildung 5 als eigenständiger, entsprechend zweigeteilter Aspekt aufgeführt wird.⁹

Der *deskriptive Teil* zur Visualisierung und Quantifizierung der Performance-Lücke basiert auf Daten, die vom AMPM und AMM bereitgestellt werden. Führt man beispielsweise ein Benchmarking von Arbeitsagenturen bezüglich des Erfolgs arbeitsmarktpolitischer Maßnahmen durch, so benötigt man Daten des AMPM aus den einzelnen Agenturen. Möchte man hingegen einen Leistungsvergleich zwischen regionalen oder nationalen Ar-

⁹ Schütz et al. (1998: 12) ordnen Benchmarking komplett der Evaluation zu, da Benchmarking über das Monitoring hinausgeht.

beitsmärkten durchführen, dann sind Daten des AMM angebracht. Es empfiehlt sich außerdem, die in der Evaluation ermittelten kausalen Nettoeffekte für ein Benchmarking zu verwenden, um einen Leistungsvergleich hinsichtlich der interessierenden Nettoeffekte anstellen zu können. Folglich verlaufen sowohl vom Monitoring als auch von der Evaluation Pfeile in Richtung des deskriptiven Benchmarkings in Abbildung 5.

Der *kausalanalytische Teil* des Benchmarkings baut auf den Ergebnissen des deskriptiven Benchmarking auf, da die identifizierten Performance-Lücken in diesem Teil ursächlich erforscht werden sollen. Um die Ursachen identifizieren zu können, sind auch im kausalanalytischen Benchmarking Informationen des AMPM und AMM notwendig. Das AMPM liefert speziell Daten zu der Höhe und Struktur der Inputfaktoren, die eine potenzielle Erklärung für die unterschiedliche Performance der Benchmarking-Einheiten darstellen. Man benötigt vor allem Daten zum personellen und finanziellen Input, die in den Input- und Prozess-Indikatoren des AMPM erfasst sind. Das AMM liefert zusätzlich Auskunft über die Unterschiede der Arbeitsmarktbedingungen, unter denen die einzelnen Benchmarking-Einheiten agieren müssen (Auer, 1998: 175). Generell gilt, dass je detaillierter das AMPM und das AMM Daten für das kausalanalytische Benchmarking liefern können, desto leichter lassen sich die ursächlichen Faktoren für die Performance-Unterschiede identifizieren.

4.5. Rückkoppelungsschleifen

Die Hauptfunktion des Monitorings ist die Bereitstellung zeitnaher, detaillierter und exakter Daten zu den aus den Zieldimensionen abgeleiteten Finanz- und Performance-Indikatoren. Zudem kann die Aufdeckung von Zielverfehlungen erste Hinweise für Forschungsfragen einer sich anschließenden Evaluations- bzw. kausalanalytischen Benchmarking-Studie liefern. Direkte Rückkoppelungsschleifen vom Monitoring hin zur Programmimplementierung und zur Programmdefinition sollten sich auf das Reporting von Zielverfehlungen beschränken. Korrekturmaßnahmen auf Basis der deskriptiven Monitoring-Maßnahmen, wie sie in der Literatur gefordert werden (u.a. Auer, 1998: 173), sind zu hinterfragen, auch wenn der zeitnahe, prozessuale Charakter des Monitorings zu Adjustierungen des laufenden Programms verleitet. Denn Korrekturmaßnahmen sollten eigentlich nur auf Basis von fundierten Ursache-Wirkungszusammenhängen, die man in Evaluationsstudien und kausalanalytischen Benchmarking-Studien identifiziert, eingeleitet werden. Zudem werden Evaluationsstudien zu Ursache-Wirkungszusammenhänge der Programmteilnahme erheblich in ihrer Aussagekraft geschmälert, wenn sich die Ursache (=Programm) und die Komposition der Teilnehmer und Nichtteilnehmer im Zeitablauf ändert. Auch aus dieser Perspektive sind die direkten Rückkoppelungseffekte im Rahmen

des AMPM hin zur Programmimplementierung zu hinterfragen, soweit sie zu Korrekturmaßnahmen führen. Starke Rückkoppelungsmechanismen zur Programmdefinition bestehen hingegen beim AMM. Beispielsweise kann sich die Zielsetzung und Ausgestaltung arbeitsmarktpolitischer Maßnahmen an den vom AMM aufbereiteten Informationen orientieren (Hilbert und Mytzek 2002: 20).

Ergebnisse der Evaluation können im Gegensatz zum Monitoring für Änderungen der Programmumsetzung und Programmdefinition ausgenutzt werden, soweit die identifizierten Annahmen für die kausalen Effekte erfüllt sind. Direkte Korrekturen bei der Programmimplementierung sind nur möglich, wenn das Programm bei Vorliegen der Evaluationsergebnisse noch nicht beendet ist. Da die identifizierten Nettoeffekte im Rahmen von ex-post Evaluationen oftmals erst nach Ende des Programms vorliegen, sind Eingriffe in den Ablauf des untersuchten Programms nicht mehr möglich, aber Empfehlungen für die Gestaltung zukünftiger Programme können abgeleitet werden. Das Feedback von ex-ante Evaluationen richtet sich vor allem auf die Programmdefinition, da auf Basis der simulierten Wirkungen die Programme im Voraus auf die Ziele und insbesondere Zielgruppen ausgerichtet werden können.

Die Ergebnisse des kausalanalytischen Benchmarking können sowohl genutzt werden, um Änderungen der Programmdefinitionen als auch der Implementierung durchzusetzen, soweit das Programm noch nicht beendet ist. So fließen die Erfahrungen der besten Performer ein und es kann zu einem Lernprozess kommen.

5. Zusammenfassung

Die Notwendigkeit einer wissenschaftlichen Begleitforschung zur Messung der Effizienz und Effektivität von arbeitsmarktpolitischen Maßnahmen hat in den letzten Jahren zunehmend an Akzeptanz gewonnen, was sich teilweise auch in den neuen Regelungen des SGB III zeigt. Gerade im Hinblick auf die finanziellen Restriktionen durch die angespannte Haushaltslage in Deutschland, ist die Antwort auf die Frage notwendig, welche arbeitsmarktpolitischen Programme effizient bezüglich ihres Ablaufs und mikro- bzw. makroökonomisch auch wirksam sind. Die Evaluation als Instrument zur Identifikation der kausalen Programmwirkungen auf die Teilnehmer und Nichtteilnehmer der Maßnahmen mittels nicht-experimenteller Methoden hat sich dabei als fundamentaler Bestandteil der Begleitforschung etabliert, was sich in den zahlreichen theoretischen und empirischen wissenschaftlichen Beiträgen widerspiegelt. Hingegen gibt es erst wenige theoretische und empirische Arbeiten, die sich mit Monitoring und Benchmarking beschäftigen, die ebenfalls Instrumente der wissenschaftlichen Analyse arbeitsmarktpolitischer Maßnahmen sind.

Die unterschiedlichen Definitionen der einzelnen Elemente Evaluation, Monitoring und Benchmarking in der Literatur machen eine trennscharfe Verortung im System der Begleitforschung schwierig. In dieser Dokumentation wurde jedoch ein weiterer Versuch unternommen, die Elemente voneinander abzugrenzen und die Interdependenzen aufzudecken, sowie die Rückkoppelungseffekte hin zur Definition und Implementierung der arbeitsmarktpolitischen Programme zu skizzieren. Die Interdependenzen zeigen, dass eine isolierte Betrachtung einzelner Elemente einschränkend ist und folglich die Begleitforschung in einem integrativen Rahmen aller Teilelemente aufgestellt werden sollte.

Speziell meint Arbeitsmarktpolitik-Monitoring (AMPM) die Dokumentation, jedoch nicht Erklärung der Erreichung gesetzter Programmziele. Die Dokumentation erfolgt durch kontinuierlich wiederholte Beobachtungen statistischer Finanz- und Performance-Indikatoren, mittels derer der Zielerreichungsgrad gemessen wird. AMPM ist vom reinen Arbeitsmarkt-Monitoring (AMM) zu differenzieren, das auf die Ermittlung und Prognose des Qualifikationsbedarfs zielt, der sich aus dem Mismatch zwischen Arbeitsangebot und Arbeitsnachfrage ergibt. Verwandt ist das AMPM mit dem Zielcontrolling, das ein Teilelement des Controllings darstellt. AMPM geht zudem über die im SGB III gesetzlich vorgeschriebenen Eingliederungsbilanzen hinaus, da es die Daten zeitnaher erfasst und zudem Outcome-Indikatoren einschließt.

Aus den beobachteten Daten des AMPM können erste Hinweise auf Fehlsteuerungen bei der Programmimplementierung gewonnen werden. Allerdings sollten daraus keine Programmkorrekturen abgeleitet werden, weil Monitoring einen rein deskriptiven Charakter hat. Solche Korrekturen empfehlen sich ausschließlich auf Basis fundierter Ursache-Wirkungszusammenhänge, die in der Evaluation oder dem kausalanalytischen Benchmarking als weitere Elemente der Begleitforschung gewonnen werden. Beide Monitoring-Methoden sollten sich darauf beschränken, den weiteren Instrumenten der Begleitforschung detaillierte Daten zur Verfügung zu stellen, insbesondere für die Evaluation, die auf Basis dieser Daten die Nettoeffekte der Maßnahme identifiziert und für das Benchmarking, das Leistungsunterschiede deskriptiv erfasst und kausalanalytisch erklärt.

Im Bereich der Arbeitsmarktforschung meint Benchmarking den Leistungsvergleich zwischen Arbeitsmärkten bzw. den Vergleich regionaler Arbeitsagenturen bezüglich der Effizienz der Maßnahmenimplementierung. Analog zur Differenzierung zwischen Monitoring und Evaluation lässt sich das Benchmarking in einen deskriptiven und einen kausalanalytischen Part dividieren. Beim deskriptiven Benchmarking stehen verschiedene Methoden zur Verfügung, um Effizienzunterschiede zu identifizieren, wie beispielsweise Radar-Charts mit SMOP-Analyse und Analyseverfahren der Effizienzgrenze, die durch spezifische Vor- und Nachteile gekennzeichnet sind. Zur kausalen Erklärung der Leistungsunterschiede zwischen Arbeitsagenturen bzw. Arbeitsmärkten müssen die Kontextfaktoren berücksichtigt werden in Form der strukturellen Komposition der Bevölkerung, der wirt-

schaftlichen Situation und den institutionellen Rahmenbedingungen. Erst nach der Korrektur um diese Umweltfaktoren können Effizienzunterschiede zwischen Benchmarking-Einheiten einer Untersuchung organisationsinterner Faktoren unterzogen werden. Hier besteht allerdings noch Bedarf, weitere adäquate Forschungsmethoden in empirischen Analysen anzuwenden.

6. Literaturverzeichnis

Affholter, D. (1994), Outcome Monitoring, in Wholey, J., H. Hatry und K. Newcomer (Hrsg.), *Handbook of Practical Program Evaluation*, San Francisco, Jossey-Bass, 96-118.

Althin, R. und L. Behrenz (2004), An Efficiency Analysis of Swedish Employment Offices, *International Review of Applied Economics* 18 (4), 471-482.

Auer, P. (1998), Monitoring of Labour Market Policy in EU Member States, in Lange, T. (Hrsg.), *Unemployment in Theory and Practise*, Cheltenham, Elgar, 164-178.

Auer, P. und T. Kruppe (1996), Monitoring of Labour Market Policy in EU Member States, in Schmid, G., J. O'Reilly und K. Schömann (Hrsg.), *International Handbook of Labour Market Policy and Evaluation*, Cheltenham, Elgar, 899-922.

Bendix, U. und G. Hammer (2005), EQUIB: Regionales Monitoring Qualifikationsentwicklung in Bremen, in Larsen, C., S. Dera, C. Knobel und A. Schmid (Hrsg.), *Regionales Arbeitsmarktmonitoring – Ansätze, Konzepte und Entwicklungen in Deutschland*, München, Rainer Hampp Verlag, 43-52.

Bogan, C. und M. English (1994), *Benchmarking for Best-Practices. Winning through Innovative Adaption*, New York.

Brinkmann, C. (1999), *Controlling and Evaluation of Employment Promotion and the Employment Services in Germany*, IAB Labour Market Research Topics, Bd. 36, Nürnberg.

Brinkmann, C. (2000), *Zur Evaluation der aktiven Arbeitsförderung nach dem SGB III – Monitoring und Wirkungsforschung im Umbruch*, MittAB, Bd. 3, Nürnberg, 483-499.

Brinkmann, C. und F. Wießner (2002), *Zur Wirkungsforschung aktiver Arbeitsmarktpolitik – neue Herausforderungen für Monitoring und Evaluation*, BeitrAB, Bd. 250, 373-386.

Caliendo, M. und E. Jahn (2004), Verbleibsquote ein Controlling-Indikator für den Eingliederungserfolg von ABM!?, *Zeitschrift für Evaluation* 3 (1), 51-69.

Caliendo, M., T. Hagen und R. Hujer (2004), Makroökonomische Methoden auf Basis von Regionaldaten, in Hagen, T. und A. Spermann (Hrsg.), *Hartz-Gesetze: Methodische Ansätze zu einer Evaluierung*, ZEW Wirtschaftsanalysen, Bd. 74, Baden-Baden, 73-97.

Calmfors, L. (1994), Active Labour Market Policy and Unemployment – A Framework for the Analysis of Crucial Design Features, *OECD Economic Studies*, OECD, Bd. 22, 7-47.

Camp, R. (1989), *Benchmarking – The Search for Industry Best Practices that Lead to Superior Performance*, Milwaukee, Quality Press.

Deeke, A. und T. Kruppe (2002), *Prozess- und kontextorientierte Begleitforschung zu arbeitsmarktpolitischen Programmen*, BeitrAB, Bd. 250, 373-386.

Erhel, C. (2003), Evaluation and Monitoring of Active Labor Market Policies, in Marshall, K. (Hrsg.), *New Social Policy Agendas for Europe and Asia: Challenges, Experience and Lessons*, Washington D.C., World Bank, 419-437.

Fahr, R. und U. Sunde (2005), *Regional Dependencies in Job Creation: An Efficiency Analysis for Western Germany*, IZA Discussion Paper, Nr. 1660, Bonn.

Farrel, M. (1957), The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society* 120, 253-281.

Fitzenberger, B. und R. Hujer (2002), *Stand und Perspektiven der Evaluation der aktiven Arbeitsmarktpolitik in Deutschland*, ZEW Discussion Paper, Nr. 02-13, Mannheim.

Fitzenberger, B. und S. Speckesser (2004), *Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland: Ein Überblick*, MittAB, Bd. 3, 357- 370.

Gesellschaft für innovative Beschäftigungsförderung (G.I.B.) (2002), *Modellvorhaben Prospect – dialogorientiertes regionales Arbeitsmarktmonitoring, Abschlussbericht*, Bottrop.

Greene, W. (2005), Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model, *Journal of Econometrics* 126, 269-303.

Hagen, T. und B. Fitzenberger (2004), Mikroökonometrische Methoden zur Ex-post-Evaluation, in Hagen, T. und A. Spermann (Hrsg.), *Hartz-Gesetze: Methodische Ansätze zu einer Evaluierung*, ZEW Wirtschaftsanalysen, Bd. 74, Baden-Baden, 45-72.

Heckman, J., R. Lalonde und J. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, in Ashenfelter, O. und D. Card (Hrsg.), *Handbook of Labor Economics* 3, 1865-2097.

Hilbert, C. und R. Mytze (2002), *Strategische und methodische Ansatzpunkte zur Ermittlung des regionalen Qualifikationsbedarfs*, Discussion Paper FS | 02 – 211, Wissenschaftszentrum Berlin für Sozialforschung.

Horváth, P. und R. Herter (1992), Benchmarking – Vergleich mit den Besten der Besten, *Controlling* 4, 4-11.

Hujer, R., L. Bellmann und C. Brinkmann (2000), *Evaluation aktiver Arbeitsmarktpolitik – Probleme und Perspektiven*, MittAB, Bd. 3, 341- 344.

Jones, M. (2004), The Dynamic Benchmarking of Labour Markets, *Regional Studies* 38 (5), 495-506.

Jolliffe, J. (2002), *Principal Component Analysis*, 2. Auflage, New York.

Kaplan, R. und D. Norton (1992), The Balanced Scorecard – Measures that drive Performance, *Harvard Business Review* 70 (1), 71-79.

Landesagentur für Struktur und Arbeit (LASA) Brandenburg GmbH (2000). *Arbeitsförder-Monitoring – Regulierung zwischen „top down“ und „bottom up“*. Machbarkeitsstudie zum Aufbau eines Monitoring-Systems als Begleitinstrument regionalisierter Landes-Arbeitsförderung, Potsdam.

Larsen, C., S. Dera, C. Knobel und A. Schmid (2005), *Regionales Arbeitsmarktmonitoring – Ansätze, Konzepte und Entwicklungen in Deutschland*, München, Rainer Hampp Verlag.

Löffler, E. (1996), A Survey on Public Sector Benchmarking Concepts, in Hill, H., H. Klages und E. Löffler (Hrsg.), *Quality, Innovation and Measurement in the Public Sector*, Frankfurt a.M., Lang, 137-159.

Michaelis, J. und A. Spermann (2004), Evaluation von Mini- und Midijobs sowie Arbeitslosengeld II: Methodische Grundprobleme und Lösungsansätze, *Zeitschrift für Evaluation* 3 (2), 223-241.

Mosley, H., H. Schütz und G. Schmid (2003), *Effizienz der Arbeitsämter. Leistungsvergleich und Reformpraxis*, Berlin.

Mosley, H. und A. Mayer (1999), *Benchmarking National Labour Market Performance: A Radar Chart Approach*, Discussion Paper FS I 99-202, Wissenschaftszentrum Berlin.

Schmid, G., J. O'Reilly und K. Schömann (1996), Theory and Methodology of Labour Market Policy and Evaluation: An Introduction, in Schmid, G., J. O'Reilly und K. Schömann (Hrsg.), *International Handbook of Labour Market Policy and Evaluation*, Cheltenham, Elgar, 1-33.

Schüssler, R., K. Spiess, D. Wendland und M. Kukuk (1999), *Quantitative Projektion des Qualifikationsbedarfs bis 2010*, BeitrAB, Bd. 221, Nürnberg.

Schütz, H., S. Speckesser und G. Schmid (1998), *Benchmarking Labour Market Performance and Labour Market Policies: Theoretical Foundations and Applications*, Discussion Paper, FS | 98-205, Wissenschaftszentrum Berlin für Sozialforschung.

Schmid, G., H. Schütz und S. Speckesser (1999), Broadening the Scope of Benchmarking: Radar Charts and Employment Systems, *Labour* 13 (4), 879-899.

Schütz, H. (2001), *Zielsteuerung in europäischen Arbeitsverwaltungen*, MittAB, Bd. 34 (2), 207-225.

Schütz, H. (2003), *Controlling von Arbeitsverwaltungen im internationalen Vergleich*, Discussion Paper SP I 2003-103, Wissenschaftszentrum Berlin für Sozialforschung.

Storrie, D. und H. Bjurek (2000), *Benchmarking European Labour Market Performance with Efficiency Frontier Techniques*, Discussion Paper FS | 00-211, Wissenschaftszentrum Berlin für Sozialforschung.

Tronti, L. (1998), Benchmarking Labour Market Performances and Policies, *Labour* 12 (3), 489-513.

Wöhe, G. (1973), *Einführung in die allgemeine Betriebswirtschaftslehre*, 11. Auflage, München.

Das Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW) ist ein Wirtschaftsforschungsinstitut mit Sitz in Mannheim, das 1990 auf Initiative der Landesregierung Baden-Württemberg, der Landeskreditbank Baden-Württemberg und der Universität Mannheim gegründet wurde und im April 1991 seine Arbeit aufnahm. Der Arbeit des ZEW liegen verschiedene Aufgabenstellungen zugrunde:

- ▷ interdisziplinäre Forschung in praxisrelevanten Bereichen,
- ▷ Informationsvermittlung,
- ▷ Wissenstransfer und Weiterbildung.

Im Rahmen der Projektforschung werden weltwirtschaftliche Entwicklungen und insbesondere die mit der europäischen Integration einhergehenden Veränderungsprozesse erfaßt und in ihren Wirkungen auf die deutsche Wirtschaft analysiert. Priorität besitzen Forschungsvorhaben, die für Wirtschaft und Wirtschaftspolitik praktische Relevanz aufweisen. Die Forschungsergebnisse werden sowohl im Wissenschaftsbereich vermittelt als auch über Publikationsreihen, moderne Medien und Weiterbildungsveranstaltungen an Unternehmen, Verbände und die Wirtschaftspolitik weitergegeben.

Recherchen, Expertisen und Untersuchungen können am ZEW in Auftrag gegeben werden. Der Wissenstransfer an die Praxis wird in Form spezieller Seminare für Fach- und Führungskräfte aus der Wirtschaft gefördert. Zudem können sich Führungskräfte auch durch zeitweise Mitarbeit an Forschungsprojekten und Fallstudien mit den neuen Entwicklungen in der empirischen Wirtschaftswissenschaften vertraut machen.

Die Aufgabenstellung des ZEW in der Forschung und der praktischen Umsetzung der Ergebnisse setzt Interdisziplinarität voraus. Die Internationalisierung der Wirtschaft, vor allem aber der europäische Integrationspro-

zeß werfen zahlreiche Probleme auf, in denen betriebs- und volkswirtschaftliche Aspekte zusammentreffen. Im ZEW arbeiten daher Volkswirte und Betriebswirte von vornherein zusammen. Je nach Fragestellung werden auch Juristen, Sozial- und Politikwissenschaftler hinzugezogen.

Forschungsprojekte des ZEW sollen Probleme behandeln, die für Wirtschaft und Wirtschaftspolitik praktische Relevanz aufweisen. Deshalb erhalten Forschungsprojekte, die von der Praxis als besonders wichtig eingestuft werden und für die gleichzeitig Forschungsdefizite aufgezeigt werden können, eine hohe Priorität. Die Begutachtung von Projektanträgen erfolgt durch den wissenschaftlichen Beirat des ZEW. Forschungsprojekte des ZEW behandeln vorrangig Problemstellungen aus den folgenden Forschungsbereichen:

- ▷ Internationale Finanzmärkte und Finanzmanagement,
- ▷ Arbeitsmärkte, Personalmanagement und Soziale Sicherung,
- ▷ Industrieökonomik und Internationale Unternehmensführung,
- ▷ Unternehmensbesteuerung und Öffentliche Finanzwirtschaft,
- ▷ Umwelt- und Ressourcenökonomik, Umweltmanagement sowie der Forschungsgruppe
- ▷ Informations- und Kommunikationstechnologien und der Querschnittsgruppe
- ▷ Wachstums- und Konjunkturanalysen.

Zentrum für Europäische
Wirtschaftsforschung GmbH (ZEW)

L 7, 1 · D-68161 Mannheim

Postfach 10 34 43 · D-68034 Mannheim

Telefon: 06 21 / 12 35-01, Fax - 224

Internet: www.zew.de

In der Reihe ZEW-Dokumentation sind bisher erschienen:

Nr.	Autor(en)	Titel
93-01	Johannes Velling Malte Woydt	Migrationspolitiken in ausgewählten Industriestaaten. Ein synoptischer Vergleich Deutschland - Frankreich - Italien - Spanien - Kanada.
94-01	Johannes Felder, Dietmar Harhoff, Georg Licht, Eric Nerlinger, Harald Stahl	Innovationsverhalten der deutschen Wirtschaft. Ergebnisse der Innovationserhebung 1993
94-02	Dietmar Harhoff	Zur steuerlichen Behandlung von Forschungs- und Entwicklungsaufwendungen. Eine internationale Bestandsaufnahme.
94-03	Anne Grubb Suhita Osório-Peters (Hrsg.)	Abfallwirtschaft und Stoffstrommanagement. Ökonomische Instrumente der Bundesrepublik Deutschland und der EU.
94-04	Jens Hemmelskamp (Hrsg.)	Verpackungsmaterial und Schmierstoffe aus nachwachsenden Rohstoffen.
94-05	Anke Saebetzki	Die ZEW-Umfrage bei Dienstleistungsunternehmen: Panellaufbau und erste Ergebnisse.
94-06	Johannes Felder, Dietmar Harhoff, Georg Licht, Eric Nerlinger, Harald Stahl	Innovationsverhalten der deutschen Wirtschaft. Methodenbericht zur Innovationserhebung 1993.
95-01	Hermann Buslei	Vergleich langfristiger Bevölkerungsvorausberechnungen für Deutschland.
95-02	Klaus Rennings	Neue Wege in der Energiepolitik unter Berücksichtigung der Situation in Baden-Württemberg.
95-03	Johannes Felder, Dietmar Harhoff, Georg Licht, Eric Nerlinger, Harald Stahl	Innovationsverhalten der deutschen Wirtschaft. Ein Vergleich zwischen Ost- und Westdeutschland.
95-04	Ulrich Anders	G-Mind – German Market Indicator: Konstruktion eines Stimmungsbarometers für den deutschen Finanzmarkt.
95-05	Friedrich Heinemann Martin Kukuk Peter Westerheide	Das Innovationsverhalten der baden-württembergischen Unternehmen – Eine Auswertung der ZEW/infas-Innovationserhebung 1993
95-06	Klaus Rennings Henrike Koschel	Externe Kosten der Energieversorgung und ihre Bedeutung im Konzept einer dauerhaft-umweltgerechten Entwicklung.
95-07	Heinz König Alfred Spielkamp	Die Innovationskraft kleiner und mittlerer Unternehmen – Situation und Perspektiven in Ost und West
96-01	Fabian Steil	Unternehmensgründungen in Ostdeutschland.
96-02	Norbert Ammon	Financial Reporting of Derivatives in Banks: Disclosure Conventions in Germany, Great Britain and the USA.
96-03	Suhita Osório-Peters Karl Ludwig Brockmann	Nord-Süd Agrarhandel unter veränderten Rahmenbedingungen.
96-04	Heidi Bergmann	Normsetzung im Umweltbereich. Dargestellt am Beispiel des Stromeinspeisungsgesetzes.
96-05	Georg Licht, Wolfgang Schnell, Harald Stahl	Ergebnisse der Innovationserhebung 1995.
96-06	Helmut Seitz	Der Arbeitsmarkt in Brandenburg: Aktuelle Entwicklungen und zukünftige Herausforderungen.
96-07	Jürgen Egel, Manfred Erbsland, Annette Hügel, Peter Schmidt	Der Wirtschaftsstandort Vorderpfalz im Rhein-Neckar-Dreieck: Standortfaktoren, Neugründungen, Beschäftigungsentwicklung.
96-08	Michael Schröder, Friedrich Heinemann, Kathrin Kölbl, Sebastian Rasch, Max Steiger, Peter Westernheide	Möglichkeiten und Maßnahmen zur Wahrung und Steigerung der Wettbewerbsfähigkeit der Baden-Württembergischen Wertpapierbörse zu Stuttgart.
96-09	Olaf Korn, Michael Schröder, Andrea Szczesny, Viktor Winschel	Risikomessung mit Shortfall-Maßen. Das Programm MAMBA – Metzler Asset Management Benchmark Analyzer.
96-10	Manfred Erbsland	Die Entwicklung der Steuern und Sozialabgaben – ein internationaler Vergleich.
97-01	Henrike Koschel Tobias F. N. Schmidt	Technologischer Wandel in AGE-Modellen: Stand der Forschung, Entwicklungsstand und -potential des GEM-E3-Modells.
97-02	Johannes Velling Friedhelm Pfeiffer	Arbeitslosigkeit, inadäquate Beschäftigung, Berufswechsel und Erwerbsbeteiligung.
97-03	Roland Rösch Wolfgang Bräuer	Möglichkeiten und Grenzen von Joint Implementation im Bereich fossiler Kraftwerke am Beispiel der VR China.
97-04	Ulrich Anders, Robert Dornau, Andrea Szczesny	G-Mind – German Market Indicator. Analyse des Stimmungsindikators und seiner Subkomponenten.
97-05	Katinka Barysch Friedrich Heinemann Max Steiger	Bond Markets in Advanced Transition: A Synopsis of the Visegrád Bond Markets.
97-06	Suhita Osório-Peters, Nicole Knopf, Hatice Aslan	Der internationale Handel mit Agrarprodukten – Umweltökonomische Aspekte des Bananenhandels.
97-07	Georg Licht, Harald Stahl	Ergebnisse der Innovationserhebung 1996.
98-01	Horst Entorf, Hannes Spengler	Kriminalität, ihr Ursachen und ihre Bekämpfung: Warum auch Ökonomen gefragt sind.
98-02	Doris Blechinger, Alfred Kleinknecht, Georg Licht, Friedhelm Pfeiffer	The Impact of Innovation on Employment in Europe – An Analysis using CIS Data.

98-03	Liliane von Schuttenbach Krzysztof B. Matusiak	Gründer- und Technologiezentren in Polen 1997.
98-04	Ulrich Kaiser Herbert S. Buscher	Der Service Sentiment Indicator – Ein Konjunkturklimaindikator für den Wirtschaftszweig unternehmensnahe Dienstleistungen.
98-05	Max Steiger	Institutionelle Investoren und Corporate Governance – eine empirische Analyse.
98-06	Oliver Kopp, Wolfgang Bräuer	Entwicklungschancen und Umweltschutz durch Joint Implementation mit Indien.
98-07	Suhita Osório-Peters	Die Reform der EU-Marktordnung für Bananen – Lösungsansätze eines fairen Handels unter Berücksichtigung der Interessen von Kleinproduzenten .
98-08	Christian Geßner Sigurd Weinreich	Externe Kosten des Straßen- und Schienenverkehrslärms am Beispiel der Strecke Frankfurt – Basel.
98-09	Marian Beise, Birgit Gehrke, u. a.	Zur regionalen Konzentration von Innovationspotentialen in Deutschland
98-10	Otto H. Jacobs, Dietmar Harhoff, Christoph Spengel, Tobias H. Eckerle, Claudia Jaeger, Katja Müller, Fred Ramb, Alexander Wünsche	Stellungnahme zur Steuerreform 1999/2000/2002.
99-01	Friedhelm Pfeiffer	Lohnflexibilisierung aus volkswirtschaftlicher Sicht.
99-02	Elke Wolf	Arbeitszeiten im Wandel. Welche Rolle spielt die Veränderung der Wirtschaftsstruktur?
99-03	Stefan Vögele Dagmar Nelissen	Möglichkeiten und Grenzen der Erstellung regionaler Emittentenstrukturen in Deutschland – Das Beispiel Baden-Württemberg.
99-04	Walter A. Oechsler Gabriel Wiskemann	Flexibilisierung von Entgeltsystemen – Voraussetzung für ein systematisches Beschäftigungsmanagement.
99-05	Elke Wolf	Ingenieure und Facharbeiter im Maschinen- und Anlagenbau und sonstigen Branchen – Analyse der soziodemographischen Struktur und der Tätigkeitsfelder.
99-06	Tobias H. Eckerle, Thomas Eckert, Jürgen Egel, Margit Himmel, Annette Hügel, Thomas Kübler, Vera Lessat, Stephan Vaterlaus, Stefan Weil	Struktur und Entwicklung des Oberrheingraben als europäischer Wirtschaftsstandort (Kurzfassung).
00-01	Alfred Spielkamp, Herbert Berteit, Dirk Czarnitzki, Siegfried Ransch, Reinhard Schüssler	Forschung, Entwicklung und Innovation in produktionsnahen Dienstleistungsbereichen. Impulse für die ostdeutsche Industrie und Perspektiven.
00-02	Matthias Almus, Dirk Engel, Susanne Prantl	The „Mannheim Foundation Panels“ of the Centre for European Economic Research (ZEW).
00-03	Bernhard Boockmann	Decision-Making on ILO Conventions and Recommendations: Legal Framework and Application.
00-04	Otto H. Jacobs, Christoph Spengel, Gerd Gutekunst, Rico A. Hermann, Claudia Jaeger, Katja Müller, Michaela Seybold, Thorsten Stetter, Michael Vituschek	Stellungnahme zum Steuersenkungsgesetz.
00-05	Horst Entorf, Hannes Spengler	Development and Validation of Scientific Indicators of the Relationship Between Criminality, Social Cohesion and Economic Performance.
00-06	Matthias Almus, Jürgen Egel, Dirk Engel, Helmut Gassler	Unternehmensgründungsgeschehen in Österreich bis 1998. ENDBERICHT zum Projekt Nr. 1.62.00046 im Auftrag des Bundesministeriums für Wissenschaft und Verkehr (BMWV) der Republik Österreich.
00-07	Herbert S. Buscher, Claudia Stirböck, Tereza Tykvová, Peter Westerheide	Unterschiede im Transmissionsweg geldpolitischer Impulse. Eine Analyse für wichtige Exportländer Baden-Württembergs in der Europäischen Währungsunion.
00-08	Helmut Schröder Thomas Zwick	Identifizierung neuer oder zu modernisierender, dienstleistungsbezogener Ausbildungsberufe und deren Qualifikationsanforderungen Band 1: Gesundheitswesen; Botanische/Zoologische Gärten/Naturparks; Sport Band 2: Werbung; Neue Medien; Fernmeldedienste; Datenverarbeitung und Datenbanken Band 3: Technische Untersuchung und Beratung; Architektur- und Ingenieurbüros; Unternehmens- und Public-Relations-Beratung Band 4: Verwaltung von Grundstücken, Gebäuden und Wohnungen; Mit dem Kredit- und Versicherungsgewerbe verbundene Tätigkeiten; Wirtschaftsprüfung und Steuerberatung; Messewirtschaft Band 5: Vermietung beweglicher Sachen ohne Bedienungspersonal; Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften; Personen- und Objektschutzdienste; Verkehrsvermittlung; Reiseveranstalter und Fremdenführer
00-09	Wolfgang Franz, Martin Gutzeit, Jan Lessner, Walter A. Oechsler, Friedhelm Pfeiffer, Lars Reichmann, Volker Rieble, Jochen Roll	Flexibilisierung der Arbeitsentgelte und Beschäftigungseffekte. Ergebnisse einer Unternehmensbefragung.
00-10	Norbert Janz	Quellen für Innovationen: Analyse der ZEW-Innovationserhebungen 1999 im Verarbeitenden Gewerbe und im Dienstleistungssektor.

00-11	Matthias Krey, Sigurd Weinreich	Internalisierung externer Klimakosten im Pkw-Verkehr in Deutschland.
00-12	Karl Ludwig Brockmann Christoph Böhringer Marcus Stronzik	Flexible Instrumente in der deutschen Klimapolitik – Chancen und Risiken.
00-13	Marcus Stronzik, Birgit Dette, Anke Herold	„Early Crediting“ als klimapolitisches Instrument. Eine ökonomische und rechtliche Analyse.
00-14	Dirk Czarnitzki, Christian Rammer Alfred Spielkamp	Interaktion zwischen Wissenschaft und Wirtschaft in Deutschland. Ergebnisse einer Umfrage bei Hochschulen und öffentlichen Forschungseinrichtungen.
00-15	Dirk Czarnitzki, Jürgen Egel Thomas Eckert, Christina Elschner	Internetangebote zum Wissens- und Technologietransfer in Deutschland. Bestandsaufnahme, Funktionalität und Alternativen.
01-01	Matthias Almus, Susanne Prantl, Josef Brüderl, Konrad Stahl, Michael Woywode	Die ZEW-Gründerstudie – Konzeption und Erhebung.
01-02	Charlotte Lauer	Educational Attainment: A French-German Comparison.
01-03	Martin Gutzeit Hermann Reichold Volker Rieble	Entgeltflexibilisierung aus juristischer Sicht. Juristische Beiträge des interdisziplinären Symposiums „Flexibilisierung des Arbeitsentgelts aus ökonomischer und juristischer Sicht“ am 25. und 26. Januar 2001 in Mannheim.
02-01	Dirk Engel, Helmut Fryges	Aufbereitung und Angebot der ZEW Gründungsindikatoren.
02-02	Marian Beise, Thomas Cleff, Oliver Heneric, Christian Rammer	Lead Markt Deutschland. Zur Position Deutschlands als führender Absatzmarkt für Innovationen. Thematische Schwerpunktstudie im Rahmen der Berichterstattung zur Technologischen Leistungsfähigkeit im Auftrag des bmb+f (Endbericht).
02-03	Sandra Gottschalk, Norbert Janz, Bettina Peters, Christian Rammer, Tobias Schmidt	Innovationsverhalten der deutschen Wirtschaft: Hintergrundbericht zur Innovationserhebung 2001.
03-01	Otto H. Jacobs, Ulrich Schreiber, Christoph Spengel, Gerd Gutekunst, Lothar Lammersen	Stellungnahme zum Steuervergünstigungsabbaugesetz und zu weiteren steuerlichen Maßnahmen.
03-02	Jürgen Egel, Sandra Gottschalk, Christian Rammer, Alfred Spielkamp	Spinoff-Gründungen aus der öffentlichen Forschung in Deutschland.
03-03	Jürgen Egel, Thomas Eckert Heinz Griesbach, Christoph Heine Ulrich Heublein, Christian Kerst, Michael Leszczensky, Elke Middendorf, Karl-Heinz Minks, Brigitta Weitz	Indikatoren zur Ausbildung im Hochschulbereich. Studie zum Innovationssystem Deutschlands.
03-04	Jürgen Egel, Sandra Gottschalk, Christian Rammer, Alfred Spielkamp	Public Research Spin-offs in Germany.
03-05	Denis Beninger	Emploi et social en France: Description et évaluation.
03-06	Peter Jacobebbinghaus, Viktor Steiner	Dokumentation des Steuer-Transfer-Mikrosimulationsmodells STSM.
03-07	Andreas Ammermüller, Bernhard Boockmann, Alfred Garloff, Anja Kuckulenz, Alexander Spermann	Die ZEW-Erhebung bei Zeitarbeitsbetrieben. Dokumentation der Umfrage und Ergebnisse von Analysen.
03-08	David Lahl Peter Westerheide	Auswirkungen der Besteuerung von Kapitaleinkünften und Veräußerungsgewinnen auf Vermögensbildung und Finanzmärkte – Status quo und Reformoptionen.
03-09	Margit A. Vanberg	Die ZEW/Creditreform Konjunkturumfrage bei Dienstleistern der Informationsgesellschaft. Dokumentation der Umfrage und Einführung des ZEW-Indikators der Dienstleister der Informationsgesellschaft.
04-01	Katrin Schleife	Dokumentation der Ruhestandsregelungen in verschiedenen Ländern.
04-02	Jürgen Egel, Thomas Eckert, Christoph Heine, Christian Kerst, Birgitta Weitz	Indikatoren zur Ausbildung im Hochschulbereich.
05-01	Jürgen Egel Christoph Heine	Indikatoren zur Ausbildung im Hochschulbereich.
05-02	Margit Kraus Dan Stegarescu	Non-Profit-Organisationen in Deutschland. Ansatzpunkte für eine Reform des Wohlfahrtsstaats.
06-01	Michael Gebel	Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen.