

# **Probleme der Evaluation von Maßnahmen aktiver Arbeitsmarktpolitik am Beispiel beruflicher Weiterbildung**

*Dieter Blaschke, Hans-Eberhard Plath*

---

## **1 Evaluation aktiver Arbeitsmarktpolitik**

### **1.1 Notwendigkeit und Anliegen der Evaluation**

Kritik der aktiven Arbeitsmarktpolitik und Forderungen nach ihrem Ausbau wechseln sich ab, nicht nur in der Tagespresse, sondern auch in den Fachveröffentlichungen. Die Forderung nach rational nachvollziehbaren Nachweisen über die Effektivität und die Effizienz der eingesetzten Instrumente und Maßnahmen und damit der Finanzmittel, ist unabweisbar. Einige Evaluationen gibt es bereits, die mittels ökonomischer Rechnerprogramme vorhandene Umfragedaten (SOEP) aufbereiten. Andere Evaluationen, die auf der Basis vorliegender prozessproduzierter Daten der Bundesanstalt für Arbeit (BA) durchgeführt werden, sind angelaufen.

Eine Auseinandersetzung mit solchen Evaluationen ist unter der Fragestellung, welche Erkenntnis daraus, speziell für steuernde Eingriffe in die „Aktive Arbeitsmarktpolitik“ (AAMP), insbesondere in die Förderung der beruflichen Weiterbildung, zu gewinnen sein werde, bereits möglich. Dabei wird zu erörtern sein, mit welchen Restriktionen bei Überblicksanalysen aufgrund bereits vorhandener und auch für andere Zwecke erhobener Verwaltungsdaten zu rechnen ist und was folglich daran ansetzende ökonomisch-statistische Evaluationen für sich genommen zu leisten vermögen.

Seit Anfang der achtziger Jahre des vergangenen Jahrhunderts wurden für die Verbleibsanalysen (und Wirkungsanalysen) nach Qualifizierungs-Maßnahmen („Fortbildung und Umschulung“ - FuU - gem. AFG ab 1969 bis 1997 bzw. „Förderung der beruflichen Weiterbildung“ - FbW - gem. SGB III) Daten herangezogen, die im Verwaltungshandeln anfallen und festgehalten wurden. Nun werden sie auch für ökonomische „Evaluationen“ genutzt. Was wird man erwarten dürfen? Das Ziel besteht darin, etwas über die „Wirksamkeit“ der Instrumente der aktiven Arbeitsmarktpolitik und der einzelnen Maßnahmen in Abhängigkeit von ihrer Gestaltung zu erfahren und damit Ansatzpunkte für die Steuerung auf allen Ebenen kenntlich zu machen. Für die Diskussion ist es wichtig, die Grenzen der Vorgehensweisen zu erkennen und falschen Erwartungen vorzubeugen. So wird sich im Umriss erkennen lassen, in welcher Weise bzw. ob überhaupt solche Analysen für die Praxis oder die Politik hilfreich sein können.

Steuerungsversuche können im Prinzip auf allen Steuerungsebenen ansetzen. Sinnvoll und „gezielt“ steuern kann man jedoch nur, wenn man weiß, wo und wie man eingreifen kann. Dies würden z. B. Ergebnisse ermöglichen, denen die Arbeitsberater Hinweise darauf entnehmen können, wie sie effektiver und effizienter beim Einsatz der Instrumente und Maßnahmen vorgehen können, auf die sich Verwaltung und Selbstverwaltung stützen könnten, wenn es um Schwerpunktsetzungen geht und auf die sich der Gesetzgeber bei Gesetzesänderungen beziehen könnte. Es geht also nicht nur darum, zu wissen, was „erfolgreich war“, sondern, warum das der Fall war oder warum nicht. Werden also notwendige Informationen über *steuerungsrelevante Eingriffsmöglichkeiten* zum Maßstab genommen, so ist Wissen über die tatsächlich wirksamen und beeinflussbaren Bedingungen notwendig. Nur dann kann der Eingriff eine „gezielte Intervention“ werden. Will man steuernd eingreifen, so sind nur jene Nachweise brauchbar, die sich auf gewollte Veränderungen z. B. in der Personenauswahl (Erwachsene unterschiedlichen Alters und Erfahrungshintergrundes), Instrumenten- und Maßnahmenwahl, Abstimmung von Personen und Maßnahmen (z. B. Kurs bestimmter Ausgestaltung bei bekannten Personengruppen), Bedarfsorientierung der Qualifizierung, Ausrichtung an der Reaktion der Arbeitsmärkte, Qualitätskontrolle usw. beziehen. Solchen Fragestellungen wird aber zumeist nicht nachgegangen.

Worin liegen die Probleme? Bei Verbleibsanalysen und Evaluationen im Bereich der beruflichen Weiterbildung für Arbeitslose geht es gleichzeitig um zwei unterschiedliche Themen: Zum einen um die Bedingungen praxisgerechter beruflicher Weiterbildung und zum anderen - da die berufliche Weiterbildung nicht zugleich die notwendigen Arbeitsplätze schafft - um die Bedingungen, unter denen (zusätzliche) Arbeitskräfte mit bestimmter Qualifikation nachgefragt werden, die dann zur Grundlage des erhofften Eingliederungserfolgs werden könnten.

Bei der Evaluierung von Bildungsmaßnahmen im Rahmen der beruflichen Weiterbildung Arbeitsloser steht man vor folgender Situation: Das Ziel der Eingliederung in Arbeit, das im SGB III eindeutig im Vordergrund steht, reicht weit über das hinaus, was Weiterbildung leisten kann. Die Eingliederung in Arbeit ist eine zwar politisch intendierte, aber tatsächlich nur eine „wünschenswerte Nebenwirkung“ von Qualifizierung. Der Nachweis eines stringenten Zusammenhanges von Qualifizierung und Eingliederung als zwingend erwartbarem „outcome“ ist nirgends zu finden.

## **1.2 Zielproblematik der aktiven Arbeitsförderung**

Dass Evaluationen sich auch auf Ziele beziehen sollen, ist unstrittig. Die Frage ist allerdings nicht nur, ob die Ziele erreicht werden, sondern auch und zuvor, ob sie richtig ausgewählt, klar definiert und realistisch sind. Früh wurde erkannt, dass es schwierig war,

die Wirkungsanalysen auf die erklärten Ziele der „Instrumente“ (z. B. FbW, ABM) und der aktiven Arbeitsmarktpolitik bzw. der aktiven Arbeitsförderung zuzuschneiden, da sich darin oft Widersprüche verbergen. Das ist in der Wirkungsforschung an sich nichts ungewöhnliches. Schon mehrmals sind wir mit Nachdruck auch auf das Dilemma eingegangen, das sich aus der doppelten Zielstellung des AFG bzw. des SGB III ergibt. Zum einen handelt es sich um die sozialpolitische Setzung, die Zielgruppen der Arbeitsmarktpolitik besonders zu fördern. Andererseits soll auf Qualität geachtet werden, auf Leistung, auf geringe Abbruchszahlen, auf hohe Eingliederung möglichst in unbefristete und dauerhafte sozialversicherungspflichtige Beschäftigung, also auf Effektivität und Effizienz. Diese beiden Zielvorgaben widersprechen einander - soweit es Qualifizierungsmaßnahmen angeht - fast diametral; sie sind unvereinbar, speziell wegen des Nachdrucks, mit dem die Eingliederung im SGB III gefordert wird.

Dabei ist die „Philosophie“ eigentlich klar: Die Teilnahme an Qualifizierung führt zu einer Aufstockung des Humankapitals der „behandelten“ Person, was möglicherweise durch erworbene Zertifikate belegt werden kann. Dies soll die Aufnahme oder Wiederaufnahme von Arbeit erleichtern, da eine solche Person „Signale“ dafür erworben hat, dass sie mehr an Ertrag bringen könnte als eine andere Person, die sich um die Stelle bewirbt oder die eine solche Stelle innehat (Mismatch-Argument).

Als Orientierungsgrundlagen für die Ableitung der intendierten Ziele kommen die folgenden Quellen in Betracht: Gesetze (AFG, SGB III), Gesetzesbegründungen, politische Strategien im Rahmen der Wirtschafts-, Steuer-, Finanz- und Arbeitsmarktpolitik. Unterschiedliche Ziele können a priori im Konflikt zueinander stehen. Dabei ist das Problem der Vereinbarkeit von Zielen bei Zielkonflikten oder die kompromisshaft Tolerierung bzw. Verknüpfung „konfligierender“ Ziele unvermeidlich. *Mehrzielorientierung* oder - im günstigsten Falle - eine *Mehrzieloptimierung* stellen für die Praxis besonders schwierige Probleme dar.

Unterschwellig geht es stets auch um die Frage, welche der Ziele tatsächlich im *Interventionsbereich* der jeweiligen Maßnahmen liegen. Ist der Interventionsbereich der Qualifizierungsmaßnahmen so beschaffen, dass mit einiger Sicherheit Beschäftigung erreicht werden kann? Oder anders formuliert: Mit welcher Sicherheit des Wirkungseintritts führen welche Maßnahmen unter welchen objektiven und subjektiven Bedingungen zum Erreichen der Interventionsziele? Das System, in dem die „Behandlung“ z. B. durch Qualifizierung erfolgt, unterhält mit dem System, in dem Wirkung erzielt werden soll, nur sehr indirekte Beziehungen. Es wurde deshalb folgerichtig auch immer nur von der *flankierenden Wirkung* aktiver Arbeitsförderung bei der Bekämpfung der Arbeitslosigkeit ausgegangen.

Da jedoch nur selten ein direkter Bezug zu konkreten Arbeitsanforderungen herstellbar ist, dominiert auch hier die Hoffnung, dass es irgendwie gelingen möge, beides in Einklang zu bringen. Wie dies vor sich gehen sollte oder könnte, ist nicht Thema der Evaluation selbst. Darüber ist auch nichts Genaues bekannt. Erschwerend kommt hinzu, dass sowohl auf der Theorie- als auch auf der Phänomenebene weiterhin Klärungsbedarf besteht.

Gleichwohl geht es immer auch um Vorstellungen darüber, was mit aktiver Arbeitsförderung erreicht werden kann. Implizit oder explizit ist stets von Kausalitäten die Rede, z. B. wenn als Ziel von Maßnahmen der beruflichen Weiterbildung formuliert wird, die Teilnehmer sollten dadurch in ein reguläres dauerhaftes Beschäftigungsverhältnis zurückgeführt werden. Die tatsächliche Aufnahme einer Beschäftigung wird dann als der „kausale Effekt der Maßnahmen“ gesehen, der durch „eine ökonomische Analyse des Erfolges von Maßnahmen nach deren Ende“ festgestellt wird (Fitzenberger 1999: 1). Inwieweit derartige Aussagen auch von einem theoretischen Begründungszusammenhang gestützt werden, mag dahingestellt bleiben.

## **2 Probleme der Evaluation aus der Sicht dominierender ökonometrischer Ansätze**

### **2.1 Kontrafaktizität, Vergleichs- bzw. Kontrollgruppen**

Das wichtigste Qualitätsmerkmal der Evaluationen ist die Einhaltung von Standards, insbesondere die eines Vergleichs- oder Kontrollgruppenansatzes. Damit soll das Problem der sog. Kontrafaktizität gelöst werden: Niemand kann zugleich Teilnehmer und Nicht-Teilnehmer an einer Maßnahme sein. Hätte die Person die Aufnahme einer Beschäftigung auch erreicht, ohne teilgenommen zu haben? Der Erfolg der Teilnehmer muss an einer Vergleichsgruppe von Nicht-Teilnehmern, die sich von den Teilnehmern möglichst nur durch die Tatsache der Nicht-Teilnahme unterscheiden, gemessen werden können. Diese Nicht-Teilnehmer können dann praktisch wie Teilnehmer behandelt werden, die - was faktisch nicht möglich ist - nicht teilgenommen haben. Wenn die Teilnehmer in „signifikant“ höherem Maße als die Nicht-Teilnehmer wieder in Arbeit eingegliedert sind, könne man von Erfolg sprechen.

Die *Konstruktion der Vergleichsgruppen* erfolgte auf unterschiedliche Weise. Dabei handle es sich bisher immer nur um grobe Hilfskonstruktionen (Heckman u. a. 1999: 325). Eigentlich müsse man nämlich wissen, wie sich die beiden Personengruppen im Hinblick auf die Teilnahme unterscheiden. Ergebnisse des Projekts „Weiterbildungsabsistenz“, in dem es um die „Nicht-Teilnahme“ an beschäftigungsnaher Weiterbildung ging, belegen die dabei auftretenden Schwierigkeiten.

Nun wird aber in neuere Überlegungen zu ökonometrischen Ansätzen, bei denen eine „state of the art-Evaluation“ angestrebt wird, neben den lange Zeit allein betrachteten Datenauswertungstechniken auch die ex ante erfolgende Gruppenbildung mit einbezogen. Damit kann aber wohl noch nicht der Anspruch eingelöst werden, bereits mit *experimenteller* Stringenz zu arbeiten. Denn experimentelle Ansätze sind auf die Erfassung *bedingungsspezifischer* Wirkungen und mithin auf das Erkennen von „*Wirkungsmechanismen*“ gerichtet, was häufig bereits durch eine fragwürdige Zusammenfassung sehr unterschiedlicher Einflussfaktoren sowie die unklare und ungesicherte Anzeigefunktion vermeintlicher Indikatoren verhindert wird.

Daher prüfen Fachdisziplinen, deren Forschung grundsätzlich (labor- oder feld-)experimentell angelegt ist, nicht nur, ob Gruppen entsprechend den Parallelisierungskriterien homogen sind, sondern ob auch im Hinblick auf die hypothesenrelevanten Hauptkriterien Gruppenhomogenität besteht (hier wären das z. B. Determinanten der Teilnahme; Determinanten der Wiedereingliederung); was zunächst in Voruntersuchungen abgeklärt wird. Dies alles dient dazu, die zwar häufig ganz gut zu kontrollierende, aber nur schwer zu neutralisierende und in ihren Einflüssen auch schwer zu erklärende (Teilnehmer-)Heterogenität und Selektivität so gering wie nur irgend möglich zu halten.

Denn bei Bewährungskontrollen muss möglichst klar erkennbar sein, unter welchen *Bedingungen* welche *Komponenten von Interventionen* (bzw. von *gestaltenden Eingriffen* in Praxisfelder) auf *welche Weise* zu welchen *Effekten* bei den Teilnehmern bzw. behandelten Gruppen führen. Erst dann können angestrebte oder gewünschte Wirkungen aktiv, zielgerichtet und systematisch herbeigeführt werden. Sonst bleibt es bei einer, in bestimmten Situationen bzw. in Einzelfällen mitunter nützlichen und vielleicht auch notwendigen, im Ganzen und auf Dauer aber wenig hilfreichen und „ursachenblinden“ *Symptomtherapie*.

## **2.2 Experiment und Randomisierung**

Eine der schwierigsten Fragen besteht darin, wie mit dem Problem der *Kontrafaktizität* methodisch umgegangen werden sollte. Ein Ausweg ist, über die „Teilnahme“ ein Zufallsverfahren bestimmen zu lassen, wie etwa in den experimentell arbeitenden Wissenschaften. Bei denen geht es allgemein darum, dass Subjekte oder Objekte genau bekannten Bedingungen unterschiedlicher Art ausgesetzt werden. Man spricht dann von der Randomisierung der Treatments bzw. Treatmentstufen. In der Ökonometrie ist das zum Teil übernommen worden, wobei allerdings häufig schon dann von Experimenten gesprochen wird, wenn lediglich eine Randomisierung von Personen bezüglich Teilnahme bzw. Nicht-Teilnahme möglich ist und durchgeführt wurde.

Tatsächlich handelt es sich dabei lediglich um eine nahezu unabdingbare Voraussetzung wirklichen Experimentierens, nicht schon um das Experiment selbst. Mit ihm soll ja nachgewiesen werden, in welcher Weise bestimmte gesetzte und manipulierbare Bedingungen in verschiedener Abstufung auf bestimmte (Gruppen von) Personen im Vergleich untereinander und zu jenen, die keine Treatments erhielten, sondern (lediglich) einem Placebo ausgesetzt waren, wirken. Die Kontrollgruppe ist dort also voll in den Untersuchungsplan eingebunden und kann nicht irgendwelchen anderen (unkontrollierten) Aktivitäten nachgehen. Deshalb ist zu fragen, ob die „Nicht-Teilnahme“ als Merkmal einer Kontrollgruppe überhaupt relevant ist und ob nicht vielmehr eine *andere Behandlung* erforderlich wäre (stattdessen sind Nicht-Teilnehmer mit anderen Einflüssen konfrontiert, deren Wirkungen nach Art, Richtung und Ausmaß nicht bekannt sind). Die Kategorie der *Nicht-Teilnehmer* stellt auch deswegen ein besonderes Problem dar, weil sie nur in einer Bezugsdefinition - eben als Gegensatzgruppe zu den Teilnehmern - gesehen wird. Wichtig ist also das Wissen darüber, wodurch die Nicht-Teilnehmer zu charakterisieren sind. Sie sind, da sie ja nicht deaktiviert sind, nicht nur im Bezug zur Arbeitssuche, sondern auch in Bezug zur Qualifikationsänderung in einer unbekanntem Situation. Mit einem Gegeneffekt z. B. wäre bereits zu rechnen, wenn in der Zeit, in der die Teilnehmer in den Maßnahmen sind, die Nicht-Teilnehmer intensiv beraten und vermittelt werden, und deshalb schneller in eine Beschäftigung einmünden. Der Hauptpunkt, in dem sich Teilnehmer und Nicht-Teilnehmer unterscheiden dürften, ist möglicherweise der Wille bzw. die Bereitschaft zur Teilnahme.

Unter methodologischen Gesichtspunkten ist daher Folgendes zu fragen:

- Ist eine Untersuchung allein schon deshalb als Experiment anzusehen, weil Teilnehmer und Nicht-Teilnehmer vorhanden sind und weil deren Zuweisung durch Randomisierung erfolgte?
- Ist dies nicht um so weniger der Fall, als selbst im Hinblick auf die Treatments lediglich bekannt ist, dass alle Personen einer Vielzahl von Bedingungen unbekannter Beschaffenheit ausgesetzt waren?
- Ist nicht vielmehr die unerlässliche und entscheidende Kernbedingung eines Experiments gerade in der genauen qualitativen und quantitativen Erfassung und Beschreibung der Bedingungen, also der sog. unabhängigen Variablen zu sehen? Dabei kann es sich entweder um gesetzte Bedingungen (wie im Laborexperiment oder im natürlichen Experiment) oder aber um „aufgesuchte“ Bedingungen (wie im umgekehrten Experiment oder im Quasi-Experiment im Feld) handeln. Die Kenntnis der Bedingungen umfasst ihre qualitativen und quantitativen Abstufungen. Aber ist denn bei den hier zur Debatte stehenden Evaluationen hinreichend genau bekannt, was in den Maßnahmen (Treatments) passiert und was die auf ihre Indikatorfunktion zumeist nicht geprüften

Kriterien (die abhängigen Variablen) im Hinblick auf die Fragestellung bzw. Hypothesen eigentlich anzeigen? Sind demzufolge einzelne Wirkbedingungen überhaupt realitätsgerecht erfassbar?

- Werden die Kriterien- und letztlich die Indikatorprobleme durch die Bildung von statistischen Vergleichs- oder experimentellen Kontrollgruppen als automatisch mitgelöst betrachtet?

In aller Regel wird in der einschlägigen ökonomischen Literatur gleichwohl davon ausgegangen, dass ein „Experimentaldesign“, welches lediglich Teilnehmer- und Nicht-Teilnehmer-Gruppen an Maßnahmen vorsieht, kausale Aussagen erlaube, „weil die erforderliche Zeitabhängigkeit der durch die Maßnahme ausgelösten Wirkungen angemessen berücksichtigt“ sei (Fitzenberger/Speckesser 2000: 10). Der gleiche Anspruch wird auch für die ökonometrischen Analysen erhoben, in denen Vergleichsgruppen auf statistischem Wege gebildet werden.

Kann man aber tatsächlich Aussagen über *Kausalität* machen, wenn lediglich eine Kovariation von Ereignissen vorliegt? Aussagen über Kausalität erweisen sich doch wohl erst dann als stichhaltig, wenn mit der Applikation der aufgeklärten Ursachen, also mit deren gerichteter Manipulation, die entsprechenden Wirkungen systematisch herbeigeführt werden können. Um also kausale Aussagen bei Aus- und Weiterbildungsprozessen machen zu können, muss erwiesen sein, *welche Bildungsmaßnahmen* mit welchen Inhalten und Formen (d. h. Lernumgebungen) bei *Teilnehmern* mit welchen Könnens- und sonstigen relevanten *Lernvoraussetzungen* zur Verbesserung welcher *Befähigungen* zur besseren *Bewältigung* welcher *Arbeitsanforderungen* geführt haben. Im Kontext von Arbeitsmarkt und Arbeitsförderung muss zudem noch erwiesen sein, dass genau dieser Zusammenhang der Vermittlung erforderlichen Wissens und Könnens *Grundlage der Eingliederung* in Arbeit war. Anderenfalls muss die Eingliederung nicht unbedingt ein Effekt der Weiterbildung gewesen sein.

### **2.3 Statistisch-ökonometrische Verfahren als alternatives „Untersuchungsdesign“**

Aufgrund der Schwierigkeiten, zu angemessenen Untersuchungskonzepten zu gelangen, aber vor allem wohl durch die Verlockung, die von den „vorhandenen“, d. h. durch die Verwaltung erstellten und in gespeicherter Form verfügbaren Daten ausgeht, kam es zur Entwicklung bemerkenswerter statistisch-ökonometrischer Verfahren. Mit ihrer Hilfe wird versucht, die „Gleichheit“ der Teilnehmer und der Nicht-Teilnehmer im Nachhinein über den Vergleich anhand von verfügbaren Merkmalen sicherzustellen bzw. zu testen. Am beliebtesten, weil in der Regel am besten zugänglich, ist die Konstruktion von Vergleichsgruppen anhand der bekannten *sozialstatistischen Merkmale*, die weithin und seit längerem auch von Behörden standardmäßig erhoben werden.

Es wird davon ausgegangen, dass auf der Basis der bekannten Daten, durch „Matching“ sog. „statistische Zwillinge“ gefunden werden. Das sind jene „Personen“, die in bestimmten Merkmalen oder Merkmalskombinationen (z. B. Geschlecht, Alter bzw. Altersgruppe, Dauer der Arbeitslosigkeit, regionale Zugehörigkeit, Art der Berufsausbildung, letzte Art der Beschäftigung) einander einigermaßen ähnlich sind (die tolerierten Abweichungen sind statistisch wahrscheinlichkeitstheoretisch definiert). Dann unterscheiden sich die Gruppen der Teilnehmer und ihrer statistischen Zwillinge unter den Nicht-Teilnehmern hauptsächlich wiederum nur durch die Teilnahme bzw. Nicht-Teilnahme. Ob die Merkmale, nach denen das „Matching“ häufig erfolgt, für das Anliegen überhaupt von ausreichender Relevanz sind, ist ziemlich fraglich.

Was auch immer herauskommt: Über die genuinen Kontext- und Kontextfaktoren der Weiterbildung ist anhand der bislang verfügbaren Datensätze kaum etwas zu erfahren, weder über die Inhalte und die Form der Vermittlung, noch darüber, was unter dem Begriff „Lernumgebung“ zusammengefasst wird. Man kann in dieser Hinsicht also nichts lernen. Eine so grundlegende Frage wie etwa diejenige, welche Art von Weiterbildung für „nicht formal qualifizierte Personen ohne beruflichen Abschluss“ in welcher Phase ihres Lebenslaufes optimal wäre, ist allein auf dieser Basis nur schwer zu beantworten. Sinnvoll wäre es, die dabei offen bleibenden Fragen durch eigene Forschungsprojekte sorgfältig zu klären und nicht darauf zu hoffen, dass sie als Nebenergebnisse von Evaluationen anfallen. Eine erfolgversprechende Strategie wären deshalb objektkonkrete Untersuchungen, bei denen z. B. einzelne Maßnahmen bzw. Maßnahmetypen klar definiert sind, und wo der Ausgangsstatus der Personen, insbesondere deren Qualifikation, klar umschrieben ist. Beispiele für solches Vorgehen mit pre-post-Messungen gibt es (vgl. z. B. Trube 1995; Luschei/Trube 1999; Richter 2000, zitiert in: Blaschke/Plath 2000).

### **3 Evaluation mit Prozessdaten**

Zu Beginn der achtziger Jahre wurde ein Teil der Statistiken in der Bundesanstalt für Arbeit neu gefasst, insbesondere die Teilnehmerstatistik für Fortbildung und Umschulung. Im Zusammenspiel u. a. mit den Daten der bereits etwa fünf Jahre früher begonnenen Beschäftigtenstatistik lassen sich sowohl Teilaussagen über den Verbleib der Teilnehmer (in sozialversicherungspflichtiger Beschäftigung) gewinnen als auch - in gesonderten Prozeduren - Vergleichsgruppen von Nicht-Teilnehmern konstruieren. In dem Zeitraum von Januar 1983 bis etwa zum 1. Halbjahr 1997 sind Millionen von Personen in Zehntausenden von sehr unterschiedlichen „Maßnahmen“ unter der Verantwortung von zunächst 146, dann, nach der deutschen Einigung, von 181 Arbeitsämtern für regional sehr unterschiedliche Arbeitsmarktgegebenheiten beruflich weitergebildet worden. In den Evaluationen verbirgt sich diese reale Vielfalt hinter den Fachtermini „Heterogenität der Teilnehmer“ und „Heterogenität der Maßnahmen“: Hinzu kommt dann noch die „Heteroge-



nität“ der Situationen auf den regionalen und den berufsfachlichen Arbeitsmärkten bzw. der sog. „Rahmenbedingungen“. Das Denken in Daten und Modellen reduziert die Komplexität der Realität immer, das ist unvermeidlich. Dennoch muss nach unserer Meinung eine Modellbildung immer in relativ enger Beziehung zum Realitäts- bzw. Phänomenbereich erfolgen.

### **3.1 Der individuelle Nutzen der Maßnahme - Erfolgsindikatoren**

Die „Wirkung“ oder der „Erfolg“ beruflicher Qualifizierung wird je nach den verfügbaren Informationen an verschiedenen Größen, „*Indikatoren*“, festgemacht (vgl. Plath/Blaschke 1999). Wenn die (Wieder-)Aufnahme von Arbeit als Datum nicht vorliegt, wird ersatzweise anderes herangezogen, z. B. der Verbleib im Leistungsbezug oder in Arbeitslosigkeit. Seit Einführung der „Eingliederungsbilanzen“ ab Dezember 1998 wird bis auf weiteres ersatzweise der Nachweis über den Nicht-Verbleib in Arbeitslosigkeit (an Stelle der Aufnahme von Arbeit) geführt. Damit wird ein Zusammenhang gedanklich unterstellt und akzeptiert, der nicht zwingend ist. Denn das *direkte* Ergebnis einer Qualifizierung ist Wissens- und Könnenszuwachs, der z. B. an einer mit Erfolg bestandenen Prüfung oder an einem Zertifikat abzulesen wäre. Die - wie auch immer definierte - Eingliederung ist nur ein indirektes Ergebnis.

Es werden sehr viele stillschweigende Annahmen getroffen, damit behauptet werden kann, dass die Beschäftigung im Anschluss an eine Qualifizierungsmaßnahme von dieser abhängt oder dass die Nicht-Beschäftigung (im Vergleich mit Nicht-Teilnehmern) durch diese Qualifizierungsmaßnahme bedingt ist. Ob es im Anschluss daran zur Aufnahme einer Arbeit kommt, hängt, wie immer deutlicher wird, zu - unbekannt - hohen Teilen von der allgemeinen und der konjunkturell bedingten Arbeitsmarktlage ab, zudem auch von der Lage am regionalen und am berufsfachlichen Arbeitsmarkt und schließlich auch von den Möglichkeiten und dem Einsatz der Arbeitsvermittlung.

Die in der BA „vorhandenen Daten“ der Beschäftigtenstatistik enthalten Angaben, aus denen sich die folgenden Sachverhalte mehr oder minder sicher ermitteln lassen:

Der Zeitpunkt der *Aufnahme einer sozialversicherungspflichtigen Beschäftigung* (auch mehrfach im Zeitraum). Beschäftigung in *Voll- oder Teilzeit* (allerdings ohne genaue Stundenzahl). Ggf. deren *Beendigung*. Die daraus errechenbare *Beschäftigungsdauer* für einen bestimmten Zeitraum. Das erzielte *sozialversicherungspflichtige Einkommen* in Form des Brutto-Gehalts (bis zur Kappungsgrenze). Der *Beruf*, den die Person ausübt (als *Berufskennziffer* verschlüsselt, vgl. dazu jedoch Blaschke und Plath 1994).

Es fehlen Angaben darüber, ob es sich um eine *subventionierte Beschäftigung* handelt, also z. B. ABM-Stelle, Eingliederungsbeihilfen, Lohnkostenzuschüsse o.ä. Anhaltspunkte dafür, dass es eine *befristete Beschäftigung* (ab 01.01.1999 „kurzfristig Beschäftigte“

enthalten) ist. Hinweise auf eine sog. *geringfügige Beschäftigung* (erst ab April 1999). Angaben über die Beteiligung der *Arbeitsvermittlung* oder deren maßgebliche Mitwirkung an der Aufnahme einer sozialversicherungspflichtigen Beschäftigung. Angaben darüber, wie weit durch den *Maßnahmeträger* geknüpfte Kontakte an der Aufnahme einer sozialversicherungspflichtigen Beschäftigung beteiligt waren. Angaben über den Verbleib jener Personen, für die es keine Nachweise in der Beschäftigtenstatistik gibt, insbesondere also, ob z. B. ein Übergang in die *Selbständigkeit* vorlag. In der im Aufbau befindlichen Maßnahme-Teilnehmer-Grunddatei und weiteren Bausteinen des in der BA entstehenden „Data Warehouse“ werden einige der hier angeführten Informationen allerdings enthalten sein. Einen Teil der fehlenden Informationen wird man allerdings möglicherweise mittels ergänzender Befragungen, das heißt mit einem gehörigen Zusatzaufwand, gewinnen können.

## **3.2 Erfolgsbedingungen**

### **3.2.1 Teilnehmer**

Schon lange ist bekannt, dass Personen, die eines oder mehrere Risikomerkmale aufweisen, schlechtere Eingliederungsergebnisse erzielen als andere. Könnte es also sein, dass sich Teilnehmer und Nicht-Teilnehmer in gravierender Weise voneinander unterscheiden, sodass der individuelle Erfolg im Anschluss an die Maßnahme sich gerade aus diesen (ex ante-)Unterschieden erklärte? Und hängen solche Unterschiede von Entscheidungen, an Kursen teilzunehmen oder nicht teilzunehmen, davon ab, wie sie von den Personen selbst oder anderen Institutionen getroffen werden? Dann wäre von Selektionsverzerrungen auszugehen.

Verfügt man über eine Reihe von Merkmalen, so kann man zunächst überprüfen, ob sich Teilnehmer und Nicht-Teilnehmer darin unterscheiden, d. h. ob sich hinsichtlich der beobachteten Merkmale Heterogenität nachweisen lässt: In der Regel die üblichen sozialstatistischen Angaben, wie Alter, Geschlecht, schulischer Abschluss, Art der Berufsausbildung, Arbeitslosigkeitsdauer (ggf. Langzeitarbeitslose). Lassen sich danach Unterschiede nachweisen, so wird in der Analyse geprüft, ob dies einen Einfluss auf das Ergebnis hat. Aber welcher theoretische Bezug ließe sich hier zur Teilnahme oder Nicht-Teilnahme herstellen? Gedanklich nahe liegt dann schon das Konstrukt der „Teilnahmeneigung“; über eine wörtlich verstandene Teilnahme“neigung“ weiß man jedoch in der Regel nichts.

Tatsache ist allerdings, dass das SGB III (wie zuvor schon das AFG) von den Fachkräften eine angemessene Berücksichtigung von schwer zu Vermittelnden, also „Zielgruppenangehörigen“, erwartet und dass innerhalb jeder Kategorie unter ihnen eine positive Auswahl der aussichtsreichsten Teilnehmer, also „creaming“, erwartet wird. Wie die Fach-

kräfte bei der Auswahlentscheidung vorgehen, auch wenn es um Zielgruppen geht, ist unbekannt. Auch die Tatsache, dass es Personen gibt, die mit allen Mitteln versuchen, in Maßnahmen bestimmter Art hineinzukommen oder die Teilnahme zu verhindern, ist datenmäßig nicht aufweisbar. Von Selektion und Selbstselektion ist also auszugehen.

In Untersuchungsberichten stößt man zwar immer wieder auf Unterschiede zwischen Personengruppen, die man mit den verfügbaren Daten jedoch nicht erklären kann. Üblicherweise wird dann extemporiert oder „plausibilisiert“. Dabei spielen Begriffe wie „*Motivation*“, „*Fähigkeiten*“, „*Talent*“ von Teilnehmern etc. immer eine ganz prominente Rolle, ohne dass ein adäquater Umgang mit solchen Konzepten oder gar eine empirische Überprüfung erkennbar wäre (z. B. Prey 1999). Versucht man nun, diese Faktoren, denen aufgrund von Hypothesen eine wichtige Funktion zugeschrieben wird, zu neutralisieren, so begibt man sich von vornherein der Möglichkeit zu weiterführenden Einsichten. Die Frage bleibt dann immer noch unbeantwortet, ob Erfolge auf die inhaltliche oder personelle Gestaltung der Maßnahmen, auf die Besonderheiten der Teilnehmer, auf die Teilnehmerrekrutierung oder anderes zurückführbar sind. Vertiefende Untersuchungen hierzu können zur Problemlösung beitragen, das Datenproblem an sich jedoch nicht lösen.

### 3.2.2 Maßnahmen

Mit dem Begriff der „Maßnahme-Heterogenität“ wird auf die der Praxis wohlbekannte Tatsache aufmerksam gemacht, dass Maßnahmen sich voneinander unterscheiden bzw. dass es mehrere Instrumente und eine große Vielfalt von Maßnahmen gibt.

In den vorhandenen Daten der „Statistik über Teilnehmer an Maßnahmen zur beruflichen Fortbildung, Umschulung und Einarbeitung sowie an Deutsch-Sprachlehrgängen“ - St35 - sind Angaben enthalten über: *Umschulung*, *Anpassungsmaßnahme*, *Aufstiegsfortbildung* (bis 1993 auch für Beschäftigte, z. B. Meisterkurse, Technikerkurse; ab 01.01.1994 keine zweckmäßige Förderung mehr, sondern nur noch soweit die Förderung *notwendig* war), *Betriebliche Einarbeitung* (als FuU-Maßnahme gem. § 49 AFG; mit mehreren gesetzlichen Änderungen), *§ 41a-Maßnahmen* zur Verbesserung der Vermittlungsaussichten (bis Ende 1992), jeweils mit der zugehörigen Dauer. Spärliche Hinweise auf *Bildungsinhalte*, z. B. Nachholen eines anerkannten *beruflichen Abschlusses* (d. h. mit Abschluss durch eine Kammerprüfung), Vermittlung von DV-Kenntnissen etc., Beendigung der Maßnahme „*mit Erfolg*“ oder „*vorzeitig beendet*“, d. h. „*Abbrecher*“.

Zu wichtigen Sachverhalten fehlen damit Angaben:

z. B. integrierte *Praktikumsteile*, Art und Qualität von evtl. erworbenen *Zertifikaten*, Informationen über den Kurs (z. B. darüber, ob die „Maßnahme“ von vornherein auf *Abbruch* bzw. *vorzeitiges Ausscheiden* hin angelegt war), *Lernbedingungen im Kurs* (also

Sachverhalte, die in der *Qualitätsdiskussion* um die Qualifizierung von Erwachsenen seit Jahren eine Rolle spielen), Auswahlkriterien für die Teilnehmer, Homogenität bzw. Heterogenität der Teilnehmer in Bezug auf die Vorkenntnisse als Vorbedingung für Erfolg oder Abbruch, Intensität des Lernens (durch Medien unterstützt, frontal, theoretisch oder praktisch, im Schulsystem oder in zusätzlichen Praktika), das zahlenmäßige Verhältnis *Lehrkräfte* zu Lernenden, die Qualifikation der Lehrkräfte, besonders in didaktischer Hinsicht; mit Anteilen allgemeiner Qualifizierung oder ohne (sog. soziales Lernen oder nur Vermittlung von Wissens- und Könnenselementen), Seriosität des *Bildungsträgers*.

Im Laufe der Jahre hat es Änderungen in der „Politik“ der Fachabteilung und der Ämter (auch wegen Gesetzesänderungen) gegeben, die sich jedoch in den Daten bisher nicht wiederfinden. Viele Informationen über Sachverhalte in der Umsetzung der Anordnung FuU oder FbW sind nur Insidern bekannt, d. h. das Wissen darüber ist über die Fachabteilungen aller Arbeitsämter verstreut. Wer sich vornimmt, die einzelnen unter FuU durchgeführten Einzelmaßnahmen oder „Unterprogramme“ zu betrachten, sie bezüglich ihrer Teilnehmerstrukturen und ihrer Wirksamkeit zu analysieren und zu unterscheiden, der muss sich darüber im Klaren sein, dass die dafür notwendigen Informationen bis auf wenige in dem vorhandenen Datenmaterial nicht enthalten sind.

Worauf es ankommt, ist die Aufschließung derjenigen Variablen, die eine definierte Rolle in Wirkungszusammenhängen spielen. Dabei geht es nicht einfach um den Nachweis von „statistischen Wechselwirkungen“ bei vorliegenden Daten, sondern um inhaltlich-funktionale Wirkungsketten. Obwohl bekannt ist, dass viele Entscheidungen im Vorfeld und im Verlauf von Qualifikationsmaßnahmen eine Rolle spielen, kann man keine einzige davon für die Analyse heranziehen, weder bei den Personen, noch bei den Maßnahmen. Wir bleiben damit zwangsläufig an der Oberfläche, denn wir finden die Spezifika hinsichtlich der Teilnehmer, der Kurse mit ihren Lerninhalten und Lernbedingungen in den Daten nicht vor. Damit dürfte zweierlei deutlich geworden sein, nämlich, wie wenig der rein formal-statistische Umgang mit der sog. „Heterogenität der Maßnahmen“ zur Sachaufklärung beizutragen vermag und wie wenig auch zu deren Aufklärung in den vorhandenen Daten verfügbar ist.

Dass schließlich auch die konjunkturelle Situation die Arbeitsnachfrage und das Arbeitsangebot und somit letztlich den „Erfolg“ von Maßnahmen beeinflussen kann, ist bekannt und führte durch die Verpflichtung der Fachkräfte, eine Prüfung der „arbeitsmarktlichen Zweckmäßigkeit der Maßnahmen“ vorzunehmen, zu weiteren Restriktionen in der Auswahl potenzieller Teilnehmer. Neben der Selbstselektion kommt es auch zur Selektion, was in den Daten ebenfalls nicht fassbar ist.

### 3.3 „Trainingsmaßnahmen“ als Demonstrationsbeispiel

Eine der Neuerungen des SGB III war die Einführung der Trainingsmaßnahmen. An ihrem Beispiel lassen sich als Exkurs die von uns diskutierten Probleme verdeutlichen.

Bei den Trainingsmaßnahmen (§ 48 ff. SGB III), die zur Verbesserung der Eingliederungsaussichten beitragen sollen, handelt es sich um variabel kombinierte Maßnahmeteile, die auf den konkreten Einzelfall zugeschnitten sind. Zu diesen gehören:

- „Eignungsfeststellung“ für Tätigkeiten oder für Maßnahmen
- Feststellung von Arbeitsbereitschaft und Arbeitsfähigkeit („Überprüfung der Verfügbarkeit“)
- Bewerbungstraining zur Unterstützung der Selbstsuche
- Vermittlung verschiedener „notwendiger“ Kenntnisse und Fähigkeiten aus einem breiten Bereich (z. B. Arbeitstechniken, EDV-Kenntnisse, Grundkenntnisse in Deutsch, Fremdsprachen oder Mathematik).

Diese variabel kombinierbaren Teile dürfen im konkreten Fall insgesamt nicht mehr als zwölf Wochen dauern, wobei für jedes der Teile eine eigene zeitliche Obergrenze angesetzt ist. Bei einer großen denkbaren Vielfalt von Maßnahmekombinationen, die für sehr unterschiedliche Personengruppen gestaltet werden, lassen sich mögliche Wirkungen schwer oder gar nicht isolieren. Vermutlich wird sich in der Praxis eines Tages eine bestimmte Standardkombination der Maßnahmeteile herausbilden, obwohl gerade dies vom Gesetzgeber nicht gemeint sein dürfte.

Die in einer sog. „Trainingsmaßnahme“ kombinierten konkreten Teile verfolgen offensichtlich sehr unterschiedliche Ziele oder Teilziele, was bereits grundsätzlich jede globale Erfolgsbeobachtung bzw. Evaluation erschwert; die in der Statistik vorgesehenen differenzierenden Merkmale bleiben viel zu grob. Gemeinsam ist der Vielfalt der Ziele der im § 48 ff. SGB III aufgelisteten Teilmaßnahmen oder Maßnahmeteile, dass der Maßnahmeerfolg nicht direkt - wie das etwa bei Bildungsmaßnahmen üblich ist - am Eingliederungserfolg bzw. an Ersatzindikatoren festgemacht werden kann, sondern dass noch längere Wirkungsketten unterstellt werden müssen. In die „Eingliederungsbilanz“ gem. § 11 SGB III sind die Trainingsmaßnahmen möglicherweise deshalb auch nicht aufgenommen worden. Dennoch wird inzwischen nach der Wirkungskontrolle, ähnlich wie bei den Bildungsmaßnahmen, gerufen, d. h. man fragt, wie viele Personen in Arbeit gekommen sind, und welche Erfolge in Hinblick auf andere Zielsetzungen (z. B. bei der Überprüfung der Arbeitsfähigkeit) erzielt werden. Die Heterogenität der Zielsetzungen schließt bei diesem Instrument also Verbleibsanalysen herkömmlicher Art aus.

Der Gesetzgeber hat offensichtlich bei der Anwendung von Trainingsmaßnahmen unterstellt, dass sich für jede Person zunächst zwei Fragen beantworten lassen, bevor man zur Ausgestaltung übergehen kann: 1.) *Was fehlt der Person?* 2.) *Was muss deshalb getan werden?* Die dritte Frage schließt sich später an: 3.) *Was hat sich deshalb geändert?* Die Hoffnung ist, dass die Person sich zum Beispiel für eine Bildungsmaßnahme eignete und dann in sie eingetreten sei, oder dass ihr, nachdem sie sich als arbeitsbereit und arbeitsfähig erwiesen habe, bestimmte Kenntnisse und Fähigkeiten vermittelt worden seien, und sie nach einem Bewerbungstraining dann schließlich auch selbst eine Arbeit gefunden habe, obwohl dies nicht das unmittelbare Ziel ist.

Die ersten beiden Fragen verweisen auf viele Probleme. Grundsatz auch im Umgang mit Hilfesuchenden ist, dass vor jeder „Therapie“ die „sichere Diagnose“ stehen muss: Was also fehlt der Person? Gibt es einigermaßen gesicherte Kenntnisse über Risikofaktoren fachlicher oder überfachlicher Art? Kann man sie zweifelsfrei diagnostizieren? Gibt es dazu Verfahren, die bei unterschiedlichen Personen (z. B. Männer, Frauen; Jüngere, Ältere; Deutsche, Ausländer) richtigerweise zur gleichen Diagnose führen? Sind diese Verfahren unabhängig davon, ob sie von Frauen oder Männern durchgeführt werden (Objektivität und Reliabilität)? Lassen sich die Ergebnisse der Diagnosen weitgehend fehlerfrei in Daten überführen? Hat man diese Fragen in zufriedenstellender Weise beantwortet, dann müssen die Antworten auf die zweite Frage folgen: Was müsste nun mit dieser Person konkret passieren, damit ihre Risikofaktoren minimiert werden? Wie kann das fachliche oder persönliche Defizit verringert werden? Oder könnte man stattdessen auch auf besondere Stärken der Person auf anderem Gebiet setzen? Welche Möglichkeiten bestehen dazu im Rahmen der „Trainingsmaßnahmen“?

Soweit es sich um Befunde handelt, die einigermaßen gesichert sind, wäre es angezeigt, den Institutionen, die solche Maßnahmen durchführen, entsprechende Vorgaben zu machen und ihre Einhaltung auch zu überwachen. Denn ob die Maßnahme für bestimmte Personengruppen auch richtig durchgeführt wird, wird über deren „Erfolg“, also für die Risikominimierung, mitentscheidend sein.

Ist die Maßnahme abgeschlossen worden, so muss festgestellt werden, ob die vorgesehenen Veränderungen herbeigeführt werden konnten. In dem Maße, in dem man Sicherheit in der Beantwortung der drei Fragen gewinnt, zeigt sich, dass man zugleich über „steuerungsrelevante Informationen“ verfügt. Prozessdaten, die solche Informationen in ausreichender Differenzierung enthielten, sind bislang jedoch nicht „vorhanden“ und sie sind auch auf absehbare Zeit nicht zu erwarten. Diese Aussage lässt sich *cum grano salis* auch auf die „berufliche Weiterbildung“ insgesamt übertragen.

Die Hoffnung, dass sich solche Daten von der Verwaltung erzeugen lassen, und damit eines Tages in besserer Qualität vorliegen könnten, ist vermutlich trügerisch. Einerseits deshalb, weil die Menge an Daten, die in jedem Einzelfall zu erfassen wäre, viel zu groß ist, um im normalen Beratungs- und Vermittlungsprozess nebenbei festgehalten zu werden. Außerdem deshalb, weil besondere Schulungen erforderlich wären. Das Verwaltungshandeln ist nicht auf die Erstellung von objektiven, zuverlässigen und validen Daten ausgerichtet. Das ist vielmehr Sache der Wissenschaft. Andererseits ist das Feld der Qualifizierung Arbeitsloser wissenschaftlich längst nicht ausreichend erforscht. Davon aber hängt die Entscheidung ab, welche Daten denn überhaupt zu erheben wären.

#### **4 Vorläufige Schlussfolgerungen**

1. Die kritische Sichtung relevanter Probleme hat gezeigt, dass von Evaluationen auf der Basis vorhandener Daten wenig zu erwarten ist, wenn Gesichtspunkte der Steuerung aktiver Arbeitsmarktpolitik im Vordergrund stehen, d. h. die Frage, welche Steuerungsmöglichkeiten bestehen und wie Effektivität und Effizienz der Instrumente und Maßnahmen der aktiven Arbeitsförderung gesteigert werden können.

2. Auf Problemlagen der Praxis gerichtete Forschung ist nötig. Dazu müssen Transformationen aus der politischen Ebene in sachinhaltliche, fachwissenschaftliche Ebenen vorgenommen werden (z. B. Bildungs-, Lern- und Qualifikationskonzepte, Konzepte der Anforderungsermittlung und -strukturierung, der Aufgaben- und Organisationsgestaltung usw.). Es geht um so grundlegende Fragen wie: In welchem Grade müssen welche Kriterien erfüllt sein, damit *Erfolg* festgestellt werden kann? Wie lange muss oder soll der Zustand erfüllter Kriterien erhalten bleiben, damit Ziele als erreicht gelten können? Reicht es schon aus, dass Personen nicht mehr als „arbeitslos“ registriert sind, oder muss eine „dauerhafte“ Eingliederung in (unbefristete) sozialversicherungspflichtige Arbeit oder in Selbständigkeit vorliegen? Muss die Arbeit „qualifikationsangemessen“ sein?

3. Ein weiteres beträchtliches Problem besteht darin, dass wir es mit zwei Systemen und ihren je spezifischen „Mechanismen“ zu tun haben, zwischen denen eine nicht zu überbrückende Kluft besteht: Das Bildungssystem, in dem das Lernfeld liegt und der Qualifikationserwerb stattfindet, und das Beschäftigungssystem, in dem das jeweilige Funktionsfeld mit seinen Anforderungen liegt. Zwischen beiden gibt es eine Reihe von vermittelnden Instanzen, wie die Arbeitsvermittlung, die Unternehmen mit ihren Rekrutierungsstrategien, Praktika, die Bewerbung und das Vorstellungsgespräch, die bisher kaum in die Betrachtung einbezogen worden sind.

4. Forschungsstrategisch wäre es geraten, auch die inhaltliche Aufklärung wesentlicher Zusammenhänge und Wechselwirkungen der zu evaluierenden Sachverhalte, also die

*sachaufklärende* Forschung, voranzutreiben. Denn die Evaluations“forschung“ kann dies nicht in ausreichendem Umfang zugleich mitleisten, da sie als „Bewährungskontrolle“ von Interventionen streng genommen einen ganz bestimmten Platz und eine ganz bestimmte Funktion in einem Forschungsprozess hat, was sinnvoller Weise auch so bleiben sollte.

**5.** Welche Einsichten können Evaluationen vermitteln, die notwendig sind, um die aktive Arbeitsförderung zu optimieren? Welche Informationen enthalten sie über *steuerungsrelevante* Eingriffsmöglichkeiten, nämlich Personenauswahl, Instrumenten- und Maßnahmenwahl, Abstimmung von Personen und Maßnahmen, Bedarfsorientierung der Qualifizierung, Ausrichtung an der Reaktion der Arbeitsmärkte, Qualitätskontrolle? Nimmt man zum Maßstab, ob hierzu Informationen geliefert werden, so zeigt sich u. E. schnell, dass bislang zu wenig über die tatsächlich wirksamen Bedingungen bekannt ist, die für „gezielte Interventionen“ nötig sind. Das wird bei den Personenmerkmalen besonders deutlich: Bei Personen, deren Könnens- und Wissensdefizite angesichts erwartbarer Arbeitsanforderungen „diagnostiziert“ wären und für die deshalb Kurse, deren Qualität bekannt und bewährt ist, vorgesehen werden, kann man Untersuchungen anstellen. Immer wieder stoßen wir jedoch darauf, dass die Leistungsfähigkeit einer Person abstrakt nicht wirklich beurteilt werden kann. Also ist es auch sehr schwierig, für eine Person Qualifizierungsmaßnahmen anzusetzen, so lange man konkrete Arbeitsanforderungen gar nicht kennt, wie es bei Qualifizierungsmaßnahmen für Arbeitslose die Regel ist.

**6.** Der Bezug zu individuellen Merkmalen der Personen (die über die Sozialstatistik hinausgehen) und zu unterscheidenden Merkmalen der Maßnahmen muss also gesichert werden, weil hier konkrete Eingriffsmöglichkeiten zu finden sind. Evaluationsergebnisse, die das nicht gewährleisten, können nicht als Basis für die Ableitung z. B. nachhaltiger Empfehlungen zur Veränderung, zur Einschränkung oder zur Einstellung aktiver Arbeitsförderung herangezogen werden, obwohl das geschieht.

**7.** Die Bedingungen der Qualifizierung betreffen aber nur eines der beiden genannten Systeme, das Bildungssystem. Forschungen zum Beschäftigungssystem, insbesondere zu Entstehungsbedingungen von Arbeitsplätzen und den Möglichkeiten, dort steuernd einzugreifen, sowie zu Rekrutierungsbedingungen von Arbeitskräften gehören ebenfalls zwingend zu dem Konzept der umfassenden „Evaluation der AAMP“. Dieses umfasst implizit und seinem Anspruch nach beide Systeme, löst diesen Anspruch jedoch nicht ein.

**8.** Die notwendige Forschung muss Vorgaben folgen, die wir bereits vor einiger Zeit im Rahmen eines „Stufenkonzepts“ formuliert haben (vgl. Blaschke u. a. 1992 für FuU, sowie Plath und Blaschke 1999 für Reha). Dazu gehört prinzipiell eine konzeptionsgeleitete



interdisziplinäre Forschungs Kooperation. Insofern erlauben Verbleibsuntersuchungen und ökonomische Evaluationen allenfalls Grobaussagen entsprechend der allgemeinsten Ebene dieses Konzepts. Wenn es jedoch darum geht, Ansatzpunkte für steuernde Eingriffe zu gewinnen, bedarf es gemäß den weiteren Konzeptebenen tiefer greifender Analysen. Dieser Ansatz wurde bei der interventionsorientierten Untersuchung von Abbrüchen der Berufsausbildung bereits nutzbringend umgesetzt (vgl. Blaschke/Plath/Nagel 1997; Faßmann/Funk 1997).

Erforderlich wären demnach durchgearbeitete Untersuchungsansätze, welche neben Datenanalysekonzepten auch die theoretische Ableitungsbasis, spezifizierte Fragestellungen und Hypothesen sowie vor allem sachgerechte Kriterien mit geprüftem Abbildbereich, geprüfter Abbildgüte sowie Angaben zu kritischen Wertebereichen enthalten müssten. Mit solchen Untersuchungsansätzen könnte die Wirkung von Qualifizierung dann ermittelt werden, wenn sie folgende Sachverhalte sowohl im Bildungs- wie im Beschäftigungssystem erfassen würden: Erwerb von Befähigungen nach Art, Umfang und Geschwindigkeit, die bisherigen Lerngelegenheiten, -gewohnheiten, -begrenzungen, die praktizierten Qualifizierungsformen, die zu erwartenden Arbeitsanforderungen und die nötigen Erfüllungsvoraussetzungen, das Rekrutierungsverhalten der Unternehmen u. a. m.

9. Demgegenüber ist durchaus nicht zwingend begründbar, dass Qualifizierungsmaßnahmen der BA „in ihrer Gänze“ oder „gesamtheitlich“ zu evaluieren seien, oder „die gleichzeitige Evaluation heterogener Maßnahmen“ (Lechner/Gerfin 2000: 396) vorzunehmen wäre, auch wenn „Teilnehmerheterogenität“ und „Maßnahmeheterogenität kontrolliert“ würden. Ob in dem ganzen Konglomerat von Einzelmaßnahmen mit unterschiedlichsten Inhalten bei den unterschiedlichsten Personen unter den unterschiedlichsten Vorbedingungen für Bewährungskontrollen unter dem Gesichtspunkt der Steuerung hinreichend differenzierte Ergebnisse erreichbar sind, ist gegenwärtig eher skeptisch zu beurteilen.

## Literatur<sup>1</sup>

Blaschke, D./Plath, H.E. (1994): „Beruf“ und „berufliche Verweisbarkeit“. Kritische Reflexionen zu einer Herausforderung an die Arbeitsmarkt- und Berufsforschung. In: MittAB 4/1994, S. 300-322.

Blaschke, D./Plath, H.E. (2000): Möglichkeiten und Grenzen des Erkenntnisgewinns durch Evaluation aktiver Arbeitsmarktpolitik. In: MittAB 3/2000, S. 462-482.

---

<sup>1</sup> Eine umfangreichere Dokumentation einschlägiger Literatur findet sich bei Blaschke/Plath, MittAB 3/2000, S. 479-482.

- Blaschke, D./Plath, H.E./Nagel, E. (1992): Konzepte und Probleme der Evaluation von Maßnahmen der aktiven Arbeitsmarktpolitik am Beispiel der beruflichen Weiterbildung (FuU). In: MittAB 3/1992, S. 381-405.
- Blaschke, D./Plath, H.E./Nagel, E. (1997): Abbruch der Erstausbildung in der beruflichen Rehabilitation. In: MittAB 2/1997, S. 319-344.
- Faßmann, H./Funk, W. (1997): Früherkennung und Reduzierung von Abbrüchen der Berufsausbildung in Berufsbildungswerken. In: MittAB 2/1997, S. 345-355.
- Fitzenberger, B. (1999): Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland. Ein Überblick. Beitrag zur IAB-Tagung „Evaluation of Active Labour Market Policy in Germany“. Nürnberg, 27. - 28. 10. 1999.
- Fitzenberger, B./Speckesser, St. (2000): Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland. Ein Überblick. ZEW Discussion Paper No. 00-06.
- Heckman, J.J./LaLonde, R.J./Smith, J.A. (1999): The Economics and Econometrics of Active Labour Market Programs, Paper Mannheim Conference June 1999.
- Lechner, M./Gerfin, M. (2000): Wirkung der aktiven Arbeitsmarktpolitik der Schweiz auf die individuellen Beschäftigungschancen von Arbeitslosen. In: MittAB 3/2000, S. 396-404.
- Plath H.-E./Blaschke, D. (1999): Probleme der Erfolgsfeststellung in der beruflichen Rehabilitation. In: MittAB 1/1999, S. 61-69.
- Prey, H. (1999): Wirkungen staatlicher Qualifizierungsmaßnahmen - Eine empirische Untersuchung für die Bundesrepublik Deutschland. Bern - Stuttgart - Wien: Haupt.