

Ergebnisse der beruflichen Rehabilitation und ihre wissenschaftliche Evaluation – ein methodischer Überblick

Zusammenfassung

Ausgehend von zentralen Begriffen der wissenschaftlichen Evaluation und nach Klärung ihrer Beziehungen, werden nicht-experimentelle und experimentelle Methoden der Erfolgsbeobachtung sowie Veränderungsmaße skizziert. Ihre Bewertung kommt zu der Empfehlung, daß die bisherige Nachbefragung der BBW-Absolventen um eine Zufriedenheitserhebung zu ergänzen wäre, wie es auch die „Grundsätze zur Qualitätssicherung und Leistungsbeschreibung in Berufsbildungswerken“ vorgeben. Beide Verfahren erfüllen die Kriterien der Nützlichkeit, Verständlichkeit, Meßbarkeit und Machbarkeit. Weitergehende Fragestellungen obliegen der Bearbeitung in wissenschaftlichen Studien, zu deren Realisierung die Zusammenarbeit mit Hochschul- und Forschungsinstituten verstärkt gesucht werden sollte. Die Vielfalt der Forschungsbemühungen in der medizinischen Rehabilitation ist hier beispielgebend.

Im Januar 2000 wurden die Leiter von Berufsbildungswerken in Bayern von dem Referat „Berufsberatung“ ihres Landesarbeitsamtes gebeten, aussagekräftige Informationen über den Verbleib von Jugendlichen nach ihrer Erstausbildung mit besonderen Hilfen zur Verfügung zu stellen, um beurteilen zu können, inwieweit sie für den Arbeitsmarkt wettbewerbsfähig gemacht werden konnten. Die bereitgestellten Zahlen sollten in eine jährliche gesamtbayrische Erfolgsstatistik der Berufsbildungswerke Eingang finden.

Diese Aufforderung war Anlaß, uns in der Konferenz der Leiter von Berufsbildungswerken in Bayern (KL BiB) und Arbeitsgruppen mit den Themenschwerpunkten „Ergebnisdarstellung“ und „Evaluation“ in der beruflichen Rehabilitation grundlegend vertraut zu machen, deren Erträge wir in diesem Heft dokumentieren und einer breiteren Fachöffentlichkeit zugänglich machen möchten. Vielleicht sind sie ein Anstoß, sich mit der unbehaglichen Bedeutungsvielfalt der verwendeten Begriffe und notwendigen methodischen Klärungen insbesondere im Rahmen der Wirksamkeitsuntersuchungen zu beschäftigen.

1. Effektivität, Effizienz und Evaluation

Die Begriffe Effektivität, Effizienz und Evaluation werden häufig in Zusammenhang mit qualitätsorientierten Präsentationen von Rehabilitationseinrichtungen, ihren Maßnahmen und Leistungen verwendet, allerdings nicht immer mit einer klar umschriebenen und operationalisierbaren Bedeutung.

Im Rahmen der Qualitätssicherung werden diese drei Basisbegriffe vielfach so definiert:

- Die **Effektivität** vergleicht den erreichten Nutzen der erbrachten Leistungen (Ergebnis, Outcome) mit dem angestrebten Nutzen (Ziel).
- Die **Effizienz** setzt die Kosten der eingesetzten Mittel in Relation zur (gewichteten) Menge der erbrachten Leistungen mit definierter (Prozess-)Qualität.

Zu beachten ist, daß üblicherweise nicht nur jene Mittel gesucht sind, die am besten das angestrebte Ziel erreichen (Effektivität), sondern daß es auch die kostengünstigsten sein sollen (Effizienz). Dies ist eine doppelte Optimierung, die sich nach langjährigen Erfahrungen, z.B. im Gesundheitswesen, nur in den seltensten Fällen eindeutig herstellen läßt.

In der folgenden Abbildung sind die bisherigen Festlegungen zusammengefaßt:

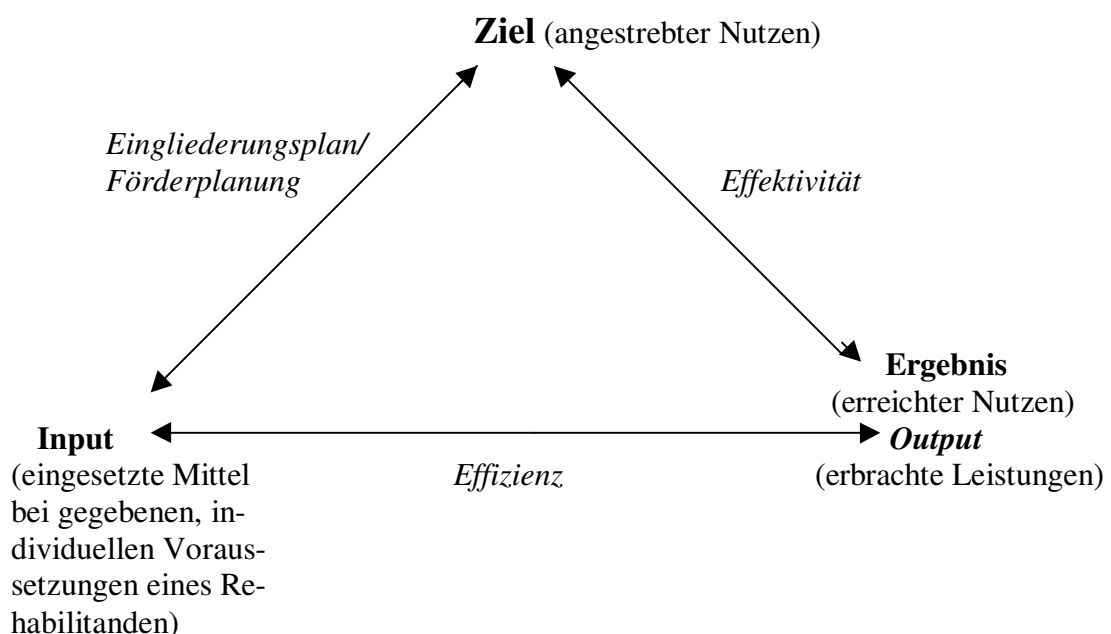


Abb.1: Das Verhältnisdreieck von Eingliederungsplan/Förderplanung, Effektivität und Effizienz

- (Wissenschaftliche) **Evaluation** bezeichnet die systematische, datenbasierte Beschreibung von Programmen, z.B. Rehabilitationsmaßnahmen, zeitlich beschränkten Projekten, z.B. Modellvorhaben, oder Institutionen in Bildung, sozialer Arbeit, Rehabilitation, Gesundheitswesen u.a. mit dem Ziel der expliziten kriteriengestützten Bewertung.

Evaluation beschafft dazu nützliche und abgesicherte Informationen für Auftraggeber, Beteiligte und Betroffene (stakeholders) und unterstützt die schrittweise Verbesserung (Gestaltungsevaluation) oder Bewertung (Bilanzevaluation) der betrachteten Gegenstände.

Sie kann auf alle vier Hauptdimensionen von pädagogischen Prozessen gerichtet werden:

- das *Konzept* und insbesondere seine Zielsetzungen,

- die *Struktur* mit personeller und sächlicher Ausstattung aufgrund gesetzlicher u.a. Bestimmungen etc.,
- den eigentlichen *Hilfeprozess*, z.B. den Ablauf des rehabilitativen Fördergeschehens, Reaktionen der Teilnehmer darauf usw., und
- das *Ergebnis*, z.B. kognitive und emotionale Lerneffekte, Änderungen im Sozialverhalten der Zielgruppe oder ihr Verbleib.

Zur Informationsgewinnung kann sich die Evaluation qualitativer und quantitativer Methoden der Sozialforschung (Beobachtung, Befragung, Inhaltsanalyse etc.) bedienen. Die Ergebnisdarstellung legt der Fach- oder einer weiteren Öffentlichkeit Datenquellen, Instrumente, Resultate und Schlußfolgerungen nachvollziehbar offen.

Die Beziehung der drei fraglichen Termini Effektivität, Effizienz und Evaluation kann im Sinne einer Oberbegriffsrelation geordnet werden. Evaluation als Beschreibungs- und insbesondere Bewertungsprozess ist der weitere Begriff und umfaßt sowohl Effektivität, d.h. Wirksamkeit, als auch Effizienz, d.h. Wirtschaftlichkeit, als (relationale) Parameter der Maßnahmegüte bzw. Ergebnisqualität¹. Der Erfolg einer Maßnahme ist dann eine wertende Aussage auf Grundlage des rehabilitationsinhalten Ziel–Ergebnis- und ggf. des ökonomischen Mittel/Kosten–Leistungs-Vergleiches. Diese bewertungsbegründenden Vergleiche etwa bei Verbleibsanalysen als retrospektiver Erfolgsbeurteilung müssen dabei immer vor dem Hintergrund der jeweiligen Input- und Kontextbedingungen, z.B. anfänglichen Rehabilitandenmerkmalen und der Arbeitsmarktlage, gesehen und interpretiert werden.

2. Maße und Methoden

Die wissenschaftliche Evaluation läßt sich danach untergliedern, ob mit oder ohne Vergleichs- bzw. Kontrollgruppe gearbeitet wird. Üblicherweise stellt die Heranziehung einer Vergleichsgruppe einen methodischen Mindeststandard dar, der nach Hager (2000) für die Hinfälligkeit vieler Evaluationsstudien sorgen würde, weil sie dieses methodische Kriterium nicht erfüllen und damit uninterpretierbar wären. Stellt sich also die Frage nach Bedingungen, unter denen Prüfuntersuchungen ohne Einbeziehung einer Vergleichsgruppe noch interpretierbar sind, und nach deren Grenzen.

2.1 Eingruppen-Prä-Post-Design

Obwohl – oder vielleicht gerade weil – „Eingruppen-Prä-Post-Designs“ nicht auf dem Boden des klassischen Kontrollgruppen-Designs mit seinen für die Evaluation im weiteren Sinne zu strengen Forderungen stehen, liegen sie rehabilitationswissenschaftlichen Forschungsansätzen überwiegend zugrunde. Sie haben routinemäßig durchgeführte Rehabilitationsmaßnahmen, z.B. medizinische Behandlungen oder solche der beruflichen Rehabilitation, im Sinne nicht-experimenteller Studien zum Gegenstand und dienen der Qualitätskontrolle.

Dazu wird eine Gruppe von Rehabilitanden zu zwei (oder mehr) Meßzeitpunkten – z.B. vor und nach einer Intervention – mit dem Ziel untersucht, der Intervention zuschreibbare Verän-

¹ Eine weiterer, allerdings selten gebrauchter und eher theoretischer Parameter wäre die *Zweckmäßigkeit*, die den mit der gewählten Maßnahme angestrebten Nutzen (Ziel) mit dem unter den gegebenen Umständen im optimalen Fall erreichbaren (Maximal-)Nutzen (Potential) vergleicht.

derungen und damit ihre Effektivität (Wirksamkeit) festzustellen. Die Behandlung oder Maßnahme wird häufig als einzige Ursache unterstellt, obwohl weitere denkbare Einflüsse berücksichtigt werden müssen, z.B. Nachreifungsvorgänge bei jungen Menschen mit Behinderungen, die Harmonisierung des familiären und sozialen Umfeldes, Veränderungen am Arbeitsmarkt, Testeffekte, der statistische Effekt der Regression zur Mitte usw.. Eine eindimensionale kausale Interpretation verbietet sich damit von selbst. Allenfalls kann mit diesem Design die durchschnittliche Veränderung zwischen zwei Meßzeitpunkten geschätzt werden.

Untersuchungen nach diesem Muster erlauben also keine Wirkungstrennung, d.h. daß **ausbildungsgebundene** bzw. mit der eigentlichen Ausbildung in Verbindung stehende Wirkungen, **maßnahmegebundene**, wie z.B. Erfahrung der Ausbilder, Lehrer und Erzieher oder das durch das Berufsbildungswerk bereitgestellte Setting (Lernumgebung), und **zeitgebundene** Wirkungen, wie etwa die erwähnten Nachreifungsvorgänge, unentwirrbar verwoben sind. Im Setting eines Berufsbildungswerkes kommt es allerdings und vor allem darauf an, den Rehabilitanden so zu helfen, daß sie nach der Maßnahme gute Voraussetzungen zur Eingliederung in Beruf und Gesellschaft haben. Es ist zweitrangig, ob dies im Anschluss an den einen oder anderen der eben genannten Faktoren erfolgt. Wichtig ist in erster Linie, daß die Eingliederung gelingt. Dabei sind Aussagen über die Wirksamkeit z.B. der Ausbildung allein nicht möglich, wohl aber zur Wirksamkeit der *Institution* Berufsbildungswerk.

Statistische gesehen, reicht hier ein Vortest-Nachtest-Vergleich völlig aus, der mit einfachen Prüfverfahren wie dem t-Test ausgewertet werden kann.

Kontrollierte Interventionsstudien zur Untersuchung der Wirkungsweise von Maßnahmen der beruflichen Rehabilitation auf der Ebene einer „Rahmenkonzeption, eines übergreifenden Untersuchungsansatzes sowie eines komplexen Methodenset (Analyse- und Bewertungsmethoden)“ nach dem Muster des „(Eingruppen)-Prä-Post-Designs“ liegen u.W. bisher nicht vor. Biefang et al. (1997) haben ein solches Vorgehen aus der Perspektive der medizinischen Rehabilitation vorgestellt, dessen **Prinzip** für die Rehabilitationsforschung im Bereich der beruflichen Rehabilitation interessant ist. Es umfaßt:

- eine Taxonomie (Klassifikationssystem) für Prädiktor- und Outcomevariablen (Vorhersage- bzw. Voraussetzungs- und Ergebnisvariablen)
- Empfehlungen für störungsübergreifende und –spezifische methodische Instrumente, die in unserem Falle sicher der Ergänzung bedürften, z.B. um Einschätzungen der Entwicklung lernbedingter Kompensations- und Leistungsmöglichkeiten, d.h. Ressourcen, der Rehabilitanden
- ein Datenerhebungskonzept für die Prä-, Post- und Follow-up-Messungen mit Empfehlungen für den Beobachtungszeitraum
- ein Vorgehen für die Erfassung der direkten und indirekten Kosten.

2.2 Eingruppen-Quasi-Prä-Post-Design

Ein Spezialfall des unter Kapitel 2.1 geschilderten Designs und ebenfalls Beispiel für eine Evaluation im weiteren Sinne sind **Zufriedenheitsuntersuchungen**, d.h. (ex post-facto) Teilnehmerbefragungen nach dem „Eingruppen-()-Post-(follow-up)-Design“, wie sie auch die „Grundsätze zur Qualitätssicherung und Leistungsbeschreibung in Berufsbildungswerken“ fordern. Dabei interessiert die Zufriedenheit von Rehabilitanden z.B. mit den vermittelten

praktischen Ausbildungsinhalten und dem theoretischen Lernstoff, den Ausbildern, Lehrern, Fachdiensten und Erziehern, den Rahmenbedingungen der Ausbildung (Setting) usw. Wenn man bei der Interpretation der Ergebnisse berücksichtigt, daß die Einschätzungen nur etwas über die (subjektive) Einstellung zu diesen Ausbildungsinhalten, zu den Ausbildenden etc. und die einschätzenden Rehabilitanden selbst als Beteiligte an der Ausbildungsmaßnahme aussagen, ist das Vorgehen ohne Vergleichsgruppe sicher zu rechtfertigen. In der „Evaluation der universitären Lehre“ (Lehreevaluation) z.B. wurden bereits eingehende Erfahrungen mit dieser Art von Befragung gesammelt. Um dagegen zu beantworten, ob die Intervention auch effektiv (wirksam) war, wären Leistungsdaten zu vergleichen, z.B. die als Vor-, Nach- und eventuell Follow-up-Test erhobenen Schulnoten oder praktischen Beurteilungen.

Ein anderer Spezialfall ist die **Verbleibsanalyse**, die ebenfalls mit einer Nachbefragungsgruppe und einem (oder mehr) Meßzeitpunkten arbeitet und als „Eingruppen-()-Post-(follow-up)-Design“ umschrieben werden kann. Da der „Prä-Meßzeitpunkt“ bei Erstauszubildenden noch in der Schulzeit oder bestenfalls um die sog. „1. Schwelle“ herum läge und deshalb als Eingliederungsvergleich nicht genutzt werden kann, reduziert sich das Ausgangsdesign auf die „Post“- Messung sowie eventuell auf weitere „Follow-up“- Messzeitpunkte.

Plath und Blaschke (1999) unterscheiden dazu die orientierende Überblicks-Analyse, auch Monitoring² genannt, wie sie als Element der Erfolgsbeobachtung von der BAG BW seit langen Jahren in einer standardisierten Nachbefragung publiziert wird, von der differenzierenden Überblicks-Analyse und der vertieften inhaltlichen Analyse.

Orientierende Überblicks- oder Verbleibsanalysen bewegen sich auf der Dimension „Chancen und Risiken am Arbeitsmarkt“ und erlauben eine fortlaufende diskrete statistische Erfolgskontrolle anhand globaler Kriterien, z.B. Eingliederung in Arbeit, Anschlußarbeitslosigkeit etc.. Sie liefern lediglich grobe Aussagen, da die Kriterien auf hoher Aggregatsebene meist nur formale Sachverhalte betreffen. Grundsätzlich besteht die Möglichkeit einer Ableitung vorwaltender Einflüsse mittels bi- oder multivariater Auswertung der Daten, wenn sie nach sozialstatistischen Kriterien³ sowie nach Situationsmerkmalen⁴ vor und nach der Rehabilitationsmaßnahme geordnet sind. Das zentrale Anliegen der Evaluation, nämlich die Bewertung nachgewiesener Wirkungen als „ausreichend oder unzureichend“, ist damit noch nicht erfüllt und von dieser Art der Verbleibsanalyse auch nicht leistbar.

Differenzierende Überblicks- oder Verbleibsanalysen, die in der Rehabilitationsforschung noch äußerst selten sind, verwerten auf der Dimension „Inhalte von Arbeit und Sozialem“ Befragungsdaten zur beruflichen und sozialen Integration, die die Einbindung von Absolventen in die „Struktur bestimmter Arbeits- und Sozialsysteme“ abbilden (vgl. Plath et al., 1996). Daten zur beruflichen Integration berücksichtigen Eignung – speziell in bisherigen Ausbildungen und Tätigkeiten erworbene Fähigkeiten – und Neigung, Daten zur sozialen Integration Kriterien zur Einbindung in soziale Netze, zur persönlichen Akzeptanz, Anerkennung usw.. Solcherart differenzierende Analysen ermöglichen eine sehr viel genauere Feststellung der Erfolge beruflicher Rehabilitation als orientierende und lassen auch besser Ansatzstellen zur Intervention erkennen. Eine Bewertungsmethodik für die zielorientierte Evaluation ist damit aber auch nicht gegeben.

² Monitoring besteht in der regelmäßigen und wiederholten Erhebung und Analyse von Daten, z.B. über Rehabilitanden, Rehabilitationsverläufe und -ergebnisse.

³ Alter, Geschlecht, Familienstand, Nationalität, Bildungsabschluß etc.

⁴ Vorherige (Erwerbs-) Situation, nachfolgende Dauer der Arbeitssuche, Art der gefundenen Arbeitsstelle, Stellung im Betrieb, Dauer des Verbleibs in Arbeit, Arbeitslosigkeit etc.

Die **vertiefte** inhaltliche Analyse wie sie Plath (1997) beispielhaft für die Dimension „Anforderungs-Fähigkeits-Relationen“ darstellt, prüft z.B. die Gewährleistung einer betrieblich unmittelbar nutzungsrelevanten Ausbildung als wichtiger Zielgröße. Berufliche Lernsituationen sollten aus Gründen der weitestgehenden Vermeidung von Transferbarrieren zwischen Lernfeld und Funktionsfeld einen deutlichen Bezug zur späteren Berufstätigkeit haben. Dazu werden auszuübende Tätigkeiten im Berufseinsatz (Funktionsfeld) und abverlangte Qualifikationen, die in der Berufsausbildung (Lernfeld) erworben wurden, durch (individuelle) Schätzurteile anhand entsprechender Skalen differenziert erfaßt und anschließend auf Entsprechungen überprüft. Voraussetzung ist die nicht unproblematische berufsspezifische Operationalisierung von Arbeitsanforderungen auf der Ebene der Tätigkeiten und von Befähigungen auf der Ebene des Wissens durch Angabe relevanter Merkmale.

2.3 (Systematische) Kasuistik

Im Umfeld der beruflichen Rehabilitation finden häufig Fallberichte als Dokumentationsinstrumente Verwendung. Sie bestehen aus einer mehr oder weniger umfangreichen und detaillierten Darstellung des individuellen Rehabilitationsverlaufes (Eine etwaige Kontrolle durch die vergleichende Darstellung z.B. des Ausbildungsverlaufes nichtbehinderter junger Menschen ist völlig unüblich.). Die Kasuistik als Sammlung von Fallberichten kann wertvolle hypothesengenerierende Hinweise auf mögliche kausale Verkettungen liefern. Auch Unterschiede im diagnostischen und pädagogisch-psychologischen Vorgehen werden in ihrer Wirkung ab einem frühen Stadium offensichtlich. Die Möglichkeit, solche Kausalbeziehungen statistisch zu prüfen oder die Diagnostik und besondere pädagogisch-psychologische Maßnahmen zu evaluieren, besteht allerdings im Rahmen der Kasuistik nicht.

Diese Sammlung möglichst sorgfältiger Beschreibungen von wesentlichen Beobachtungen im Rehabilitationsverlauf ist eine notwendige Vorstufe der Entwicklung eines angemessenen Kategorien- oder Fragenschemas zur systematischen Beobachtung. Durch diese Standardisierung wird es möglich, Fallanalysen i. S. einer **systematischen** Kasuistik zu spezifischen statistischen Aggregaten zusammenzufassen. Voraussetzung ist die Konstruktion eines mehrkriterialen Ansatzes, wie ihn aktuell z.B. der „Diagnosekatalog (DIK)“ von Schopf (1998) bietet, mit Basisdaten z.B. aus Arbeits- und Belastungsproben, der sozialpädagogischen Begleitung und der berufspädagogischen Betreuung auf der Ebene von Leistungen (Noten), Verhalten (Einschätzungen) und Selbst- sowie Fremdurteilen der Betroffenen und kompetenter Dritter, z.B. ihrer Arbeitgeber o.a.. Dabei ergeben sich nicht selten Probleme mit der phänomengerechten Kategorisierung, aber auch mit der Integration von Daten aus unterschiedlichen Quellen. Dieses Vorgehen bietet aber wünschenswerte Chancen der externen Validierung von Ergebnissen zur Erfolgsfeststellung durch die Beobachtung möglicher Koinzidenzen bzw. Kriterienkonvergenzen bei Leistungs-, Verhaltens- und Befragungsdaten.

Deutlich wird mit dem Gesagten, daß die Methodik der systematischen Kasuistik trotz ihrer Bedeutung z.T. noch erhebliche Defizite und Lücken aufweist, die es zu schließen gilt. In Rehabilitationseinrichtungen wie Berufsbildungswerken u.a. sollte sie sogar das Primat haben, da sie *das* methodische Gegenstück der theoretisch geforderten „Individualisierung“ beruflicher Rehabilitation darstellt, indem sie Phänomene am Ursprungsort ihrer Entstehung, nämlich bei den einzelnen Rehabilitanden erfaßt. Nach ihrer methodischen Komplettierung könnte sie ohne weiteres mehrdimensionale Untersuchungen nach dem Muster des „Prä-Post-Designs“ ermöglichen und in kontrollierten Interventionsstudien (s.o.) Eingang finden.

2.4 Experimentelles und quasi-experimentelles Design

Bei Evaluationen pädagogisch-psychologischer oder klinischer Interventionsmaßnahmen stellt das **experimentelle** Operieren mit einer Vergleichs- bzw. Kontrollgruppe eher den Regelfall als die Ausnahme dar. Diese Evaluationen im engeren Sinne zielen explizit und nur auf deren Wirksamkeit (Programmevaluation), unabhängig von bestimmten Institutionen. Vergleichsgruppen können „Wartegruppen“ oder auch Gruppen mit einem anderen, gegenüber der Experimentalgruppe systematisch variierten Interventionsprogramm bzw. einer (Behandlungs-)Maßnahme sein.

Die folgende Abbildung gibt einen Eindruck von der Arbeit mit Experimental- und Vergleichsgruppen bei der Prüfung der Wirksamkeit von Interventionsmaßnahmen:

Evaluationshypothese		Wirksamkeits-hypothes	Wirksamkeits-unterschiedshypothes
Evaluationsparadigma		isolierte Evaluation	vergleichende Evaluation
		Interventionsziele in den Vergleichsgruppen relativ zu denen des Experimentalprogrammes	
Art der Randbedingungen in den Vergleichsgruppen		verschieden	gleich
	verschieden	Wartegruppe	Alternativprogramm
	gut vergleichbar oder gleich	Kontrollprogramm	Quasi-Alternativprogramm

Abb.2: Experimental- und Vergleichsgruppen bei der Wirksamkeitsevaluation⁵ (aus: Hager, 2000)

Es gibt zwei Paradigmen der Wirksamkeitsevaluation mit verschiedenen Vorgehensweisen und Vergleichsgruppen: die isolierte und die vergleichende Evaluation.

Die **isolierte** Evaluation gleicht die Randbedingungen der Kontrollgruppe denen der Experimentalgruppe an und die Vergleichsgruppe erhält als Kontrollgruppe ebenfalls eine spezifische Intervention, allerdings mit anderen Zielen als die Experimental-Intervention. Bei der isolierten Evaluation geht es um die grundlegende Wirksamkeit eines Programmes, z.B. einer Therapie, und um die Wirksamkeitshypothese „Das Programm A wirkt“.

Es gibt aber auch Studien, in denen Programme mit gleichen Zielen untersucht werden, also Alternativprogramme. Die Experimentalgruppen unterscheiden sich hinsichtlich einer Viel-

⁵ „Art der Randbedingungen“ umfaßt das Programm selbst und die Situationen seiner Ausführung im Vergleich zum Experimentalprogramm. Mit „Programm“ wird jede Art von Interventionsmaßnahme einschließlich von Therapien bezeichnet.

zahl von Faktoren, aber die Ziele der Intervention sind in beiden Experimentalgruppen die gleichen. Man spricht dann von **vergleichender** Evaluation, in der zwei grundsätzlich vergleichbare, aber konkurrierende Programme miteinander verglichen werden. Es geht hier um die Prüfung der Wirksamkeitsunterschiedshypothese „Programm A wirkt besser als Programm B“.

Die höchste Interpretationssicherheit („Goldstandard“) innerhalb des experimentellen Designs bietet der randomisierte und kontrollierte Versuchsplan (vgl. Faller et al., 2000). „Randomisiert“ bedeutet, daß die Zuweisung der Teilnehmer zur Experimental- bzw. Kontrollgruppe zufällig erfolgt. „Kontrolliert“ heißt, daß die Experimentalgruppe, in die die experimentelle Bedingung (Interventionsmaßnahme, treatment) eingeführt wird, mit einer Kontrollgruppe verglichen wird, in der diese Bedingung nicht existiert oder eine andere, alternative Bedingung gesetzt wurde.

Bei **quasi-experimentellen** Versuchsplänen kann die Einteilung der Teilnehmer in die experimentellen Gruppen nicht willkürlich von einem Forscher vorgenommen werden, sondern ist schon „natürlicherweise“ vorgegeben. Studien, in denen vorliegende Gruppen in Prä-Post-Messungen miteinander verglichen werden, werden auch als Kohortenstudien bezeichnet. Da man keine Randomisierung durchführen kann, ist die Kontrolle von Störgrößen nicht so einfach zu gewährleisten. Es gibt jedoch Verfahren, die die Kontrolle intervenierender Variablen bewirken, z.B. die Moderatorvariablen-Extraktion, so daß auch quasi-experimentelle Studien Kausalerklärungen mit hoher Gültigkeit ermöglichen.

2.5 Indirekte und direkte Veränderungsmessung

In den zurückliegenden Darstellungen spielen Fragen der Veränderungsmessung eine entscheidende Rolle, wenn man darunter mit Nübling und Schmidt (2000) die Messung von Veränderungen einer oder mehrerer Variablen an einem Rehabilitanden versteht.

Die Bildung von Differenzen zwischen Zustandsbeurteilungen zweier unterschiedlicher Meßzeitpunkte steht für die **indirekte** Veränderungsmessung, wobei die Differenz das Veränderungsmaß darstellt. An der Bildung von einfachen Differenzwerten wurde häufiger Kritik geübt und Alternativen diskutiert, von denen aber unklar ist, ob sie tatsächlich zu präziseren Aussagen führen (vgl. Stieglitz & Baumann, 1994). Der Reliabilitätskoeffizient für Veränderungen sollte dabei über die Meßwiederholung von Differenzwerten bestimmt werden, z.B. durch die Korrelation der Differenzen zweier Paralleltestformen.

Nur einen Meßzeitpunkt braucht die **quasi-indirekte** Veränderungsmessung, wenn der anfängliche Zustand vom Rehabilitanden sozusagen zur Zeit der (nicht erfolgten) Prä-Messung retrospektiv erinnert und mit dem derzeitigen Zustand z.B. am Ende der Maßnahme verglichen wird. Die Urteilsdifferenz kann dann als Veränderungsmaß herangezogen werden. Die Abhängigkeit von zuverlässigen Gedächtnisleistungen ist offensichtlich und könnte z.B. bei jungen Menschen mit Lernbehinderungen zu erheblichen Problemen führen.

Im Gegensatz zu diesen klassischen Verfahren konnte die **direkte** Veränderungsmessung erst in den letzten 15 Jahren Eingang in die Rehabilitationsforschung finden. Sie verlangt als

Selbstbeurteilungsverfahren die direkte Einstufung der von einer Person subjektiv erlebten Veränderung eines Zielindikators zu einem Meßzeitpunkt z.B. anhand einer 5-stufigen Likert-Skala (viel besser – viel schlechter). Auch hier ist die Abhängigkeit von Gedächtnisleistungen zu bedenken. Die Gefahr von Antworttendenzen und subjektiven Unklarheiten über das Ausgangsniveau des Zielindikators sind zu den methodisch unbefriedigenden Momenten dieses Verfahrens zu zählen. Nach Erfahrungen in der medizinischen Rehabilitationsforschung scheint es „bessere“ Veränderungsquoten als das „indirekte“ Differenzmaß zu liefern.

Daneben werden seit einigen Jahren Methoden unter dem Oberbegriff **Zielerreichungsskaliierung** vorwiegend aus der amerikanischen Psychotherapieforschung adaptiert. Dazu zählt das Goal-Attainment-Scaling (GAS), eine einzelfallbezogene und zielorientierte Erfolgsmessung, oder das ZOE-Verfahren („Ziel-orientierte Erfolgsmessung“), das individuelle Rehabilitationsziele aufgrund des Rehabilitandenprofils, gemessen über IRES-Skalen (Indikatoren des Reha-Status), definiert und ihre Veränderungen feststellt. Auch EMEK-Skalen („Einmalig erfaßte multiple Ergebniskriterien“) wurden auf der Basis von 1-Punkt-Erhebungen in letzter Zeit entwickelt und psychometrisch überprüft, um Rehabilitationsmaßnahmen längerfristig bewerten zu können. Dabei werden einzelne Veränderungsinformationen zu multiplen Ergebniskriterien zusammengefaßt und in einem Gesamtindex der „Ergebnisqualität“ abgebildet.

3. Empfehlungen

Anlaß der vorstehenden Darlegungen war der Wunsch nach einer differenzierten bayerischen BBW-Erfolgsstatistik. Im Wissen um die Abhängigkeit der Erfolgskriterien, aber auch der realisierenden Methoden von vorausgehenden Zielbestimmungen müssen sich alle Rechenschaftslegungen darauf richten, den Rehabilitationsträgern und der Öffentlichkeit die erfolgreiche Umsetzung unseres übergreifenden Auftrages und zugleich Zieles zu vermitteln, nämlich „behinderten oder von einer Behinderung bedrohten Menschen die dauerhafte Eingliederung bzw. Wiedereingliederung in das Arbeitsleben und damit die aktive Teilnahme an der Gesellschaft zu ermöglichen“ (vgl. Präambel der „Rahmengrundsätze ... für die Durchführung von Maßnahmen in Einrichtungen der beruflichen Rehabilitation“ vom 21. März 2000).

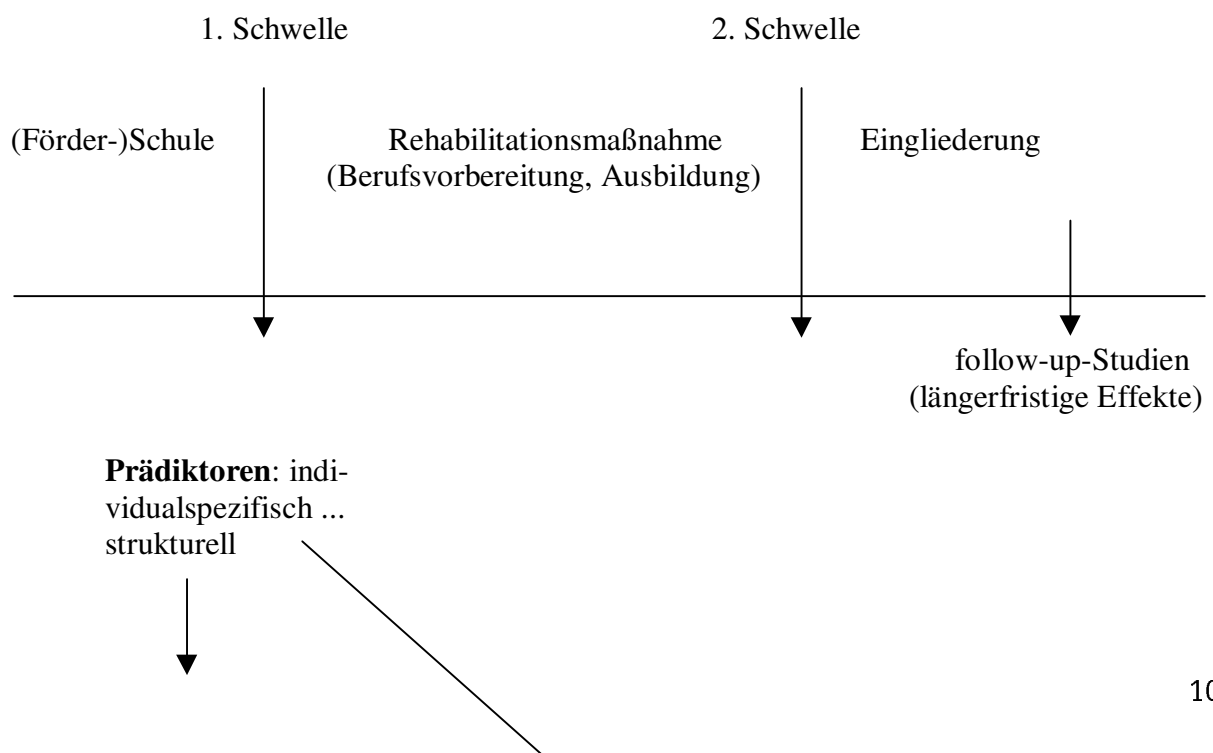
Dem dient nach den „Grundsätzen zur Qualitätssicherung und Leistungsbeschreibung in Berufsbildungswerken“ die Nachbefragung und Statistik der BAG BBW auf Grundlage der von den Berufsbildungswerken erhobenen Daten im Sinne einer Verbleibsanalyse und die geplante regelmäßige Teilnehmerbefragung, die die Akzeptanz der Leistungen der Berufsbildungswerke feststellen soll⁶. Sie gewährleisten als Standarderhebungen (Monitoring) die vereinbarte zielorientierte Erfolgsbeobachtung und müssen natürlich gemäß aktuellen Erfordernissen kontinuierlich weiterentwickelt werden. Ihre jährlichen Ergebnisse sollten dabei – anders als bisher und soweit es die Informationslage zuläßt – anhand paralleler Vergleichszahlen aus der betrieblichen Ausbildung, die sozusagen den „Normalfall“ darstellen, interpretiert werden.

⁶ Die „Qualitätsfördernden Rahmenbedingungen in Berufsförderungswerken (BFW)“ arbeiten mit ähnlichen Elementen der Qualitätssicherung: Leistungshandbuch, Ergebnisdokumentation und Teilnehmerbefragung.

Eine darüber hinausgehende Effektivitätsprüfung von Einzelmaßnahmen vorzunehmen, ist nach Weis (2000) schwierig, wenn nicht gar unmöglich. In der Rehabilitation sind nämlich Einzelmaßnahmen in der Regel eng miteinander verzahnt (Maßnahmekomplex), schwer abzugrenzen und ihre Effekte nur von Fachkräften angemessen zu erheben bzw. zu bewerten, es sei denn, sie sind einem Standardprogramm additiv hinzugefügt. Umfangreichere Fragestellungen sind vor allem als Gegenstände wissenschaftlicher Forschung in Zusammenarbeit mit zu interessierenden Universitäts-, Fachhochschul- oder Forschungsinstituten, z.B. dem Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, vorstellbar, wie es in der medizinischen Rehabilitation schon lange üblich ist und zu erheblichem theoretischen und praktischen Erkenntnisgewinn über Strukturen, Prozesse und Ergebnisse dieser Form von Rehabilitation geführt hat.

Um unsere offiziellen Erhebungsinstrumente – Nachbefragung der BBW-Absolventen und (geplante) Teilnehmerbefragung – zu verorten, wurde versucht, sie in das gegenwärtig vorfindbare Methodenuniversum der Evaluationsforschung einzuordnen. Als Spezialfall des „Eingruppen-Prä-Post-(follow-up)-Designs“ geben sie einen breiten, grob orientierenden Eingliederungs- und Zufriedenheitsüberblick und leisten das, was von einer Standarderhebung zu erwarten ist, indem sie die Kriterien der Nützlichkeit, Verständlichkeit, Meßbarkeit und Machbarkeit erfüllen. Auch einzelne bedeutsame Problemlagen, wie z.B. hinsichtlich der Abbrecherproblematik (s. Faßmann & Wasilewski, 1998), können hierdurch vororientierend erfaßt und danach einer gezielteren Analyse mit differenzierteren Methoden zugeführt werden.

Das Netz der Basisvariablen in der rehabilitativen Erfolgsforschung (vgl. Bullinger & Ravens-Sieberer, 2000), wie es in der folgenden Abbildung aufgespannt ist, ermöglicht die grundsätzliche Ableitung von einfachen und komplexen, inhaltlichen und methodischen Fragestellungen im Zusammenspiel von Ursachen-, Wirkungs- und Moderatorvariablen. Wissenschaftliche Bemühungen könnten sich daran orientieren, um sukzessive die „blinden Flecken“ auf der Landkarte wichtiger Wirkungszusammenhänge in der beruflichen Rehabilitation zu erkunden.



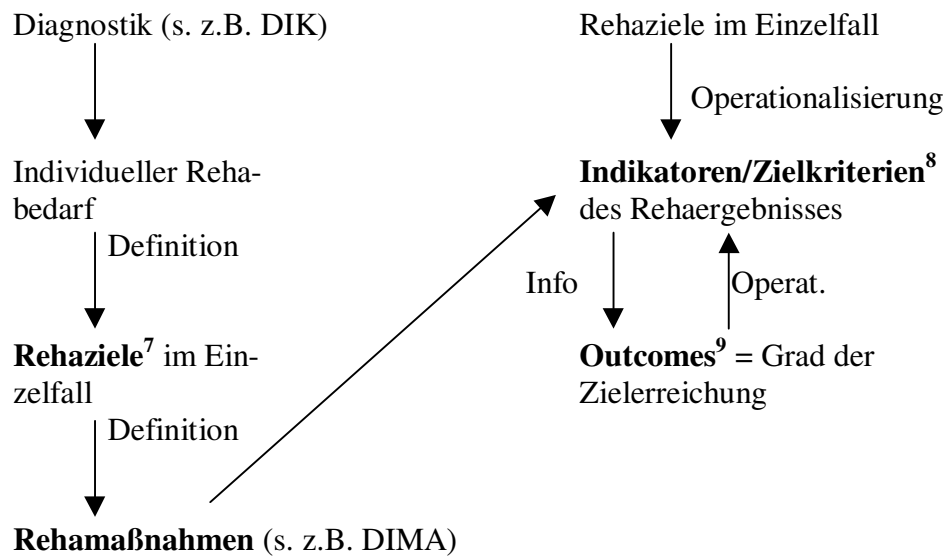


Abb.3: Basisvariablen in der rehabilitativen Erfolgsforschung

4. Literatur

BIEFANG, S., BIRKNER, B., THIEN, U., HÄRTEL, U., & BULLINGER, M. (1997). Harmonisierung der Messung von Outcomes, Prädiktoren und Kosten sowie Prüfung geschlechtspezifischer Unterschiede in der rehabilitationswissenschaftlichen Forschung. *Rehabilitation*, 36, 213-223.

BULLINGER, M. & RAVENS-SIEBERER, U. (2000). Indikatoren des Rehabilitationsergebnisses. In J. Bengel & U. Koch (Hrsg.), *Grundlagen der Rehabilitationswissenschaften. Themen, Strategien und Methoden der Rehabilitationsforschung* (S. 305-322). Berlin: Springer.

FALLER, H., HAAF, H.-G., LÖSCHMANN, C., MAURISCHAT, C. & SCHULZ, H. (2000). Experimentelle und nicht-experimentelle Studienpläne in der Rehabilitationsforschung. In J. Bengel & U. Koch (Hrsg.), *Grundlagen der Rehabilitationswissenschaften. Themen, Strategien und Methoden der Rehabilitationsforschung* (S. 271-286). Berlin: Springer.

FASSMANN, H. & WASILEWSKI, R. (1998). Wissenschaftliche Erprobung des Instruments zur Früherkennung und Reduzierung von Ausbildungsabbrüchen in Berufsbildungswerken und BüEs. Abschlußbericht zum Projekt Nr. 69/97 des Inst. f. empirische Soziologie Nürnberg (IfeS). Veröffentl. i. Vorber.

⁷ Rehaziele im Einzelfall bestehen in der Regel aus übergreifenden, z.B. Steigerung der Selbständigkeit, behinderungsspezifischen, z.B. besseres Verstehen von (logischen) Zusammenhängen, und individualspezifischen Zielanteilen, z.B. größere Sicherheit in den Grundrechenarten.

⁸ Indikatoren müssen durch Meßinstrumente hinterlegt sein.

⁹ Die Outcomes sind zugleich Prädiktoren der Eingliederung zusammen mit weiteren individualbezogenen, z.B. soziale Unterstützung, und strukturellen Prädiktoren, z.B. der Arbeitsmarktlage.

HAGER, W. (2000). Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien (S. 180-201). Bern: Huber.

NÜBLING, R. & SCHMIDT, J. (2000). Methodische Grundlagen der Ergebnisevaluation. In J. Bengel & U. Koch (Hrsg.), Grundlagen der Rehabilitationswissenschaften. Themen, Strategien und Methoden der Rehabilitationsforschung (S. 323-346). Berlin: Springer.

PLATH, H.-E. (1997). Behinderung und das Problem mit der Einfacharbeit. MittAB, 2, 424-431.

PLATH, H.-E., KÖNIG, P. & JUNGKUNST, M. (1996). Verbleib sowie berufliche und soziale Integration jugendlicher Rehabilitanden nach der beruflichen Erstausbildung. MittAB, 2, 247-278.

PLATH, H.-E. & BLASCHKE, D. (1999). Probleme der Erfolgsfeststellung in der beruflichen Rehabilitation. In M. Niehaus (Hrsg.), Erfolg von Maßnahmen zur beruflichen Rehabilitation (S. 9-28). Freiburg: Lambertus.

SCHOPF, P. (1998). Lernschwäche, Lernbeeinträchtigung, Lernbehinderung – Aspekte zur Klärung. ibv - Zeitschrift für berufskundliche Information und Dokumentation, 6, 341-449.

STIEGLITZ, H.-D. & BAUMANN, U. (1994). Veränderungsmessung. In H.-D. Stieglitz & U. Baumann (Hrsg.), Psychodiagnostik psychischer Störungen (S. 21-36). Stuttgart: Enke.

Weis, J. (2000). Interventionsmethoden in der Rehabilitation. In J. Bengel & U. Koch (Hrsg.), Grundlagen der Rehabilitationswissenschaften. Themen, Strategien und Methoden der Rehabilitationsforschung (S. 121-138). Berlin: Springer.

Karl-Heinz Eser

Geb. 1948, Dipl.-Psychologe, seit 1997 Gesamtleiter des Förderungswerkes St. Nikolaus und Leiter des Berufsbildungswerkes Dürrlauingen.

Karl-Heinz Eser

Results of vocational rehabilitation and their scientific evaluation – a methodical summary

Starting with central ideas of scientific evaluation and their relations non-experimental and experimental methods of watching effective results and measures of change will outlined. The evaluation of these objects will lead to the recommendation to complete the previous interrogation of BBW-absolvents through an inquiry of contentment as the „principles of quality securing and description of results in establishments for vocational education“ pre-tend it, too.

Both methods are fulfilling the criteria of utility, intelligibility, measurability, variability and practicability. More extensive questions need to be handled by scientific studies in cooperation with university- or research-institutes. The variety of investigations in medical rehabilitation set an example of best practice.