SSOAR

Open Access Repository
www.ssoar.info

Indikatorenentwicklung: eine praxisorientierte Einführung

Meyer, Wolfgang

Preprint / Preprint
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:
Meyer, W. (2004). *Indikatorenentwicklung: eine praxisorientierte Einführung.* (CEval-Arbeitspapier, 10). Saarbrücken: Universität des Saarlandes, Fak. 05 Empirische Humanwissenschaften, CEval - Centrum für Evaluation. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-111245

Nutzungsbedingungen:
Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Centrum für Evaluation

CEval

Center for Evaluation

ARBEITSPAPIERE 10

CENTRUM FÜR EVALUATION – UNIVERSITÄT DES SAARLANDES

Indikatorenentwicklung:

Eine praxisorientierte Einführung

(2. Auflage)

Dr. Wolfgang Meyer

2004

Meyer, Wolfgang: Indikatorenentwicklung. Eine praxisorientierte Einführung (2. Auflage)
Saarbrücken: Centrum für Evaluation, 2004.
(CEval-Arbeitspapiere; 10

NICHT IM BUCHHANDEL ERHÄLTLICH

SCHUTZGEBÜHR: 5 €

BEZUG: Centrum für Evaluation (CEval)
Universität des Saarlandes
Postfach 15 11 50
D-66041 Saarbrücken
info@ceval.de

oder kostenfrei zum Download:
http://www.ceval.de

LAYOUT + SATZ: Alexandra Caspari

INHALT

1 EINFÜHRUNG 2
2 ERGÄNZENDE BEMERKUNGEN ZUR ZWEITEN AUFLAGE 3

TEIL I: NUTZEN VON INDIKATOREN

3 DEFINITIONEN, EIGENSCHAFTEN UND FUNKTIONEN VON INDIKATOREN 5
4 ARTEN VON INDIKATOREN 7
5 ... UND FÜR WAS SIE GEBRAUCHT WERDEN 8
6 INDIKATOREN – VON EINER NEUEN AURA UMGEBEN 10
7 INDIKATORENTYPEN? TYPISCH INDIKATOR ! 11
8 WIE MAN ÄPFEL MIT BIRNEN VERGLEICHT 12
9 ... UND WARUM APFEL AUCH NICHT IMMER GLEICH APFEL IST 13
10 QUANTITATIVE ODER QUALITATIVE INDIKATOREN? ALLES EINE FRAGE DES NIVEAUS! 14
11 KEINE ANGST VOR ZAHLEN 16
12 ... ABER AUCH KEIN BLINDES VERTRAUEN ! 17
13 EINFACHE INDIKATOREN 18
14 ... RELATIVE MAßZAHLEN 19
15 ... UND KOMPLEXE INDIZES 20
16 DER NUTZEN VON INDIKATOREN: EINE ERSTE ZUSAMMENFASSUNG 21

TEIL II: ANFORDERUNGEN AN INDIKATOREN

17 WANN IST EIN INDIKATOR EIN GUTER INDIKATOR? 24
18 ES GIBT NICHTS PRAKTISCHERES ALS EINE GUTE THEORIE 25
19 ... UND NICHTS THEORETISCHERES ALS EINE GUTE PRAXIS! 26
20 DAS UNMESSBARE MESSEN: OPERATIONALISIERUNG UND INDIKATORENBILDUNG 27
21 HAUPTSACHE GUT GEMESSEN 28
22 ... BIS ZU DEN GRENZEN DER MACHBARKEIT 29
23 ANFORDERUNGEN AN INDIKATOREN: EINE ZWEITE ZUSAMMENFASSUNG 30

TEIL III: ENTWICKLUNG VON INDIKATOREN

24 DIE UNENDLICHE GESCHICHTE: INDIKATORENENTWICKLUNG ALS SOZIALER PROZESS 33
25 AUF DEN RICHTIGEN RAHMEN KOMMT ES AN: DAS KONZEPT DER LOGICAL FRAMEWORKS ... 34
26 VORGEHENSWEISE BEI DER INDIKATORENENTWICKLUNG – 10 SCHRITTE ZUM INDIKATOR 36
27 MESSEN UM DES MESSENSWILLEN? KRITERIEN DER INDIKATORENAUSWAHL 39
28 WENIGER IST HÄUFIG MEHR... 45
29 ... MANCHMAL ABER AUCH NICHT 46
30 DIE UNENDLICHE GESCHICHTE TEIL II: MIT INDIKATOREN LEBEN 47
31 DIE ENTWICKLUNG VON INDIKATOREN: EINE DRITTE ZUSAMMENFASSUNG 48
32 LESSONS LEARNT 49
LITERATUR: 53

1 Einführung

Die folgende Einführung in den praktischen Umgang mit Indikatoren ist im Rahmen des Fortbildungsprogramms Evaluation für entwicklungspolitische Gutachter entstanden. Bei der Vorbereitung auf das entsprechende Lernmodul fiel dem Autor auf, dass es zwar eine Vielzahl von Einführungen zu Indikatoren gibt, diese sich aber einseitig entweder auf die theoretisch-methodischen Grundlagen oder auf den praktischen Einsatz und die inhaltliche Umsetzung konzentrieren. Dabei werden die z.T. sehr gut ausformulierten und begründeten Ausführungen gegenseitig kaum wahrgenommen. Dies beginnt bereits bei dem Begriff „Indikator", der speziell in der entwicklungspolitischen Praxis teilweise haarsträubende Entfremdungen erfährt. So ist z.B. in einer praxisorientierten Einführung des Verbands Entwicklungspolitik deutscher Nicht-Regierungsorganisationen folgende Definition von Indikator zu finden (VENRO 2000: 6ff.):

> „Die beispielhafte, konkrete Beschreibung eines Merkmales (Indikator = Anzeiger) eines Zustandes nennen wir Indikator. Diese Beschreibung, also der Indikator, sollte möglichst objektiv nachprüfbar sein, d.h. wenn er von verschiedenen Akteuren beobachtet wird, sollten diese nicht zu unterschiedlichen, sondern zu gleichen Beobachtungsergebnissen kommen. Es wird nicht immer gelingen, objektiv nachprüfbare Indikatoren zu finden, weil viele Merkmale der Problematik in der Entwicklungszusammenarbeit im Bereich der weichen (nicht eindeutig erfassbaren) Faktoren liegen und außerdem individuell (subjektiv) nur sehr unterschiedlich wahrgenommen werden können."

Folgt man dieser Definition, die sämtliche wissenschaftlichen Ausführungen zum Thema Indikatoren ignoriert, so ist bereits jede persönliche Meinung und jede Beschreibung eines Gegenstandes ein „Indikator". Durch diese allen Konventionen widersprechende begriffliche Aufweichung verlieren Indikatoren allerdings ihre spezifischen Nutzwert: Indikatoren sollen uns helfen, Beobachtungen realer Gegenstände zu klassifizieren, diese nicht beobachtbaren Phänomenen zuordnen und dadurch schließlich eine seriöse Grundlage für eine Bewertung dieser Phänomene bilden. Dies spricht den subjektiven Eindrücken und Beschreibungen nicht ihren Sinn ab – sie sind eben nur keine Indikatoren. Und: wenn es nicht gelingt „objektiv nachprüfbare" Indikatoren zu finden, heißt dies schlicht, dass es nicht gelungen ist überhaupt einen Indikator (zumindest einen, der mit einer vertretbaren Validität und Reliabilität misst) zu finden. Selbst „weiche" Indikatoren müssen in der Lage sein, eindeutige Zuordnungen vor zunehmen, damit sie insbesondere für eine Evaluation von Nutzen sind. Das „Gefühl", die „Meinung", der „Verdacht" ein Projekt ist gut oder schlecht mag als Ausgangspunkt einer Evaluation nützlich sein – am Ende zählen nur Fakten, die mit geeigneten Indikatoren gewonnen werden konnten. Diese Fakten mögen manchmal nur eine Sammlung von mehr oder weniger beweiskräftigen „Indizien" (also Ergebnisse schlecht konstruierter und schlecht messender Indikatoren) sein, aber sie sind eben mehr als nur „individuelle Wahrnehmung".

Das folgende Papier versucht in kurzer und prägnanter Form die wesentlichen Aufgaben von Indikatoren systematisch herauszuarbeiten und dabei sowohl die wissenschaftlichen Ansprüche an Indikatoren als auch die praktische Umsetzung sowie die dabei gewonnenen Erfahrungen in Einklang zu bringen. Die Ausführungen zu den einzelnen Kapiteln sind bewusst knapp gehalten, nähere Erläuterungen und Beweisführungen sind der angegebenen Literatur zu entnehmen.

Inhaltlich gliedert sich das Papier in drei Teile, die sich zunächst mit dem Nutzen von Indikatoren, dann mit den Anforderungen an Indikatoren und schließlich mit der Entwicklung von Indikatoren beschäftigen. Die zentralen Aussagen der einzelnen Abschnitte sind schließlich unter der Rubrik „Lessons learnt" nochmals zusammengestellt. Das Papier erhebt nicht den

Anspruch, in dieser Form bereits „fertig“ zu sein. Der Autor freut sich über jede Kritik und jeden Hinweis zur Ergänzung sowie über Hinweise auf neue Literatur zu Indikatoren. Eine Überarbeitung und Weiterentwicklung der Ausführungen ist geplant.

2 ERGÄNZENDE BEMERKUNGEN ZUR ZWEITEN AUFLAGE

Die vorliegende zweite Auflage der Einführung zur Indikatorenentwicklung entspricht weitgehend der im letzten Jahr erstellten ersten Auflage. Modifikationen und Ergänzungen wurden an den Stellen vorgenommen, die einen Bezug zum neuen Auftragsrahmen (AURA) der Deutschen Gesellschaft für Technische Zusammenarbeit (GTZ) aufweisen. AURA stellt einen umfassenden Perspektivenwechsel der gesamten Auftragsvergabe im Bereich der Technischen Zusammenarbeit dar: anstelle der erbrachten Leistungen während eines Projektes stehen nun die erzielten Wirkungen (insbesondere in Relation zu den angestrebten Wirkungszielen) im Fordergrund der Vorgehensweise.

Dies hat auch tiefgreifende Auswirkungen auf die zu formulierenden Indikatoren und die an sie gerichteten Anforderungen. Folgende Aspekte sollen bei der Formulierung von Indikatoren für das Gesamtziel (und analog für die Phasenziele) beachtet werden:

- ✓ *Nutzenorientierung*: „Die Indikatoren müssen sich auf die im Gesamtziel angestrebte logische Ebene beziehen, also normalerweise den direkten Nutzen, nicht aber auf die Leistungen des Vorhabens.“
- ✓ *Prüfbarkeit*: „Sie müssen zwar noch nicht zu Beginn eines Vorhabens, aber doch im Laufe der Durchführung wertbestückt, d.h. entweder quantifiziert oder mit einem objektiv nachprüfbaren Qualitätsmerkmal beschrieben werden. Indikatoren, bei denen das voraussichtlich nicht möglich sein wird, sollten auch nicht benannt werden.“
- ✓ *Absolute Größe und Relative Veränderungen*: „Wenn immer möglich, sollten nicht nur relative Veränderungen angegeben werden, sondern auch die absolute Größenordnung“.
- ✓ *Abgrenzung zwischen Wirkungsindikatoren und Zielformulierungen*: „Der Indikator muss Auskunft darüber geben, woran etwas beobachtet werden kann. Er darf nicht selbst eine solche Frage generieren.“ (alle Zitate aus: BMZ-GTZ 2003).

Durch diese im AURA festgelegten Qualitätsstandards hoffen BMZ und GTZ auf eine Verbesserung hinsichtlich der Formulierung von Indikatoren, die zusätzlich durch die Orientierung auf Wirkungen eine größere Bedeutung für die Projektsteuerung erhalten. Dies stellt auf der anderen Seite auch an die Evaluatorinnen und Evaluatoren neue Anforderungen sowohl hinsichtlich der Bewertung existierender Indikatorensysteme als auch bezüglich der Unterstützung bei der Entwicklung von Indikatoren zum Aufbau geeigneter M&E-Systeme. Damit erhält der Aspekt der Indikatorenentwicklung eine Aufwertung, der im Rahmen von FEEZ Rechnung getragen werden musste.

Für die vorliegenden Arbeitsunterlagen bedeutet diese keine grundlegende Revision, da die meisten Sachverhalte auf allgemeiner Ebene vorgestellt werden und die Beispiele sich bewusst nicht nur auf die TZ und die Arbeit der GTZ beschränken. Die mit AURA verbundenen begrifflichen Änderungen bei Projektplanung und -management sowie die durch die Wirkungsorientierung bedingten Implikationen für die Indikatorenbildung sind aber an den entsprechenden Stellen ergänzend aufgenommen worden.

Teil I: Nutzen von Indikatoren

3 DEFINITIONEN, EIGENSCHAFTEN UND FUNKTIONEN VON INDIKATOREN

Wie bei jedem wissenschaftlichen Fachbegriff lassen sich auch für den Begriff „Indikator“ eine Reihe, z.T. konkurrierende, zumindest aber unterschiedliche Aspekte betonende, Definitionen finden. Ein Blick in das „Duden Fremdwörterbuch“ z.B. ergibt bereits vier verschiedene Verwendungsformen, deren allgemeinstes Verständnis Indikator als „*Umstand oder Merkmal, das als [beweiskräftiges] Anzeichen oder als Hinweis auf etwas anderes dient*“ fasst. Die stärker inhaltlichen Begriffserklärungen beziehen sich auf bibliothekarische, technische und chemische Anwendungsbereiche (die aktuelle Ausgabe des sehr populären Computerlexikon „Encarta“ führt sogar ausschließlich die Begriffsverwendung in der Chemie an), der sozialwissenschaftliche Gebrauch wird dagegen nicht einmal erwähnt – obwohl sicherlich einige sozialwissenschaftliche Indikatoren wie z.B. das Bruttosozialprodukt weit über die engen Wissenschaftszirkel hinausgehend bekannt geworden sind.

Selbstverständlich finden sich aber auch in gängigen Fachlexika und –wörterbüchern Definitionen des Begriffs „Indikator“. So wird z.B. in dem weitverbreiteten „Wörterbuch der Soziologie“ unter einem Indikator „*eine empirisch direkt (z.B. durch Beobachtung oder Befragung) ermittelbare Größe, die Auskunft gibt über etwas, das selbst nicht direkt ermittelbar ist*“ verstanden (Hartmann 2002). Im Unterschied zu der sehr allgemeinen Duden-Definition wird also der *empirische Bezug* und die *Nichtmessbarkeit des hinter dem Indikator stehenden theoretischen Konstrukts* hervorgehoben. Zur Bestimmung „sozialer Indikatoren“ grenzt in demselben Wörterbuch Wolfgang Glatzer dieses Verständnis weiter ein und bezieht es ausschließlich auf „*statistische Maßzahlen, die eine quantitative Abbildung gesellschaftlich bzw. gesellschaftspolitisch relevanter Sachverhalte darstellen*“ (Glatzer 2002).

Diese Auffassung wird sicherlich nicht allgemein geteilt. Insbesondere im Evaluationskontext wird darauf verwiesen, dass auch „qualitative“ Indikatoren möglich und sinnvoll einsetzbar sind. Dafür beschränken Autoren, die sich im Rahmen der konkreten Projektarbeit in der Entwicklungszusammenarbeit mit Indikatorenentwicklung beschäftigen, diese häufig auf die Projektziele. Indikatoren „sollen konkret beschreiben, was mit den Zielsetzungen des Projektes gemeint ist“ (Bolay 1989; zit. nach Werner 2000) und stellen „one of a variety of mechanisms that can answer the question of how much (or whether) progress is being made towards a certain objective“ (USAID 1998; zit. ebd.) dar. Ein ähnliches Verständnis zeigt auch die EU im Rahmen ihrer Programmevaluationen, wenn sie einen Indikator als „a characteristic or attribute which can be measured to assess a programme in terms of outputs or impacts“ (Nagarajan & Vanheukelen 1997) bezeichnet. *Indikatoren sind dementsprechend empirisch (quantitativ oder qualitativ) zu erfassende Kenngrößen, die einen Soll-Ist-Vergleich bezüglich der Zielsetzungen von Projekten oder Programmen ermöglichen sollen*.

Problematisch an einer solchen Beschränkung der Indikatorenkonstruktion auf Maßzahlen, die sich ausschließlich an den Zielsetzungen von Projekten oder Programmen orientieren ist die Tatsache, dass dadurch *nicht-intendierte Handlungsfolgen oder Nebeneffekte* aus dem Blick geraten. Welche katastrophalen Folgen die fehlerhafte Berücksichtigung solcher Wirkungen haben kann führte z.B. der bekannte Fall des Beruhigungsmittels Contagan der Firma Chemie Grünenthal drastisch vor Augen. Dieses Beispiel einer fehlerhaften Einschätzung von Nebenwirkungen zeigt allerdings auch, dass ein hundertprozentiger Schutz trotz hohem Aufwand für die Wirkungskontrolle und strengen Richtlinien letztendlich nicht möglich ist. Nichtsdestotrotz sollte es die Aufgabe einer Evaluation sein, auch nicht-intendierte Hand-

lungsfolgen und Nebeneffekte mit geeigneten Indikatoren zu erfassen und bei der Bewertung mit zu berücksichtigen.

Ein weiteres Problem der Beschränkung auf Soll-Ist-Vergleiche mittels Indikatoren, die sich ausschließlich an den Zielsetzungen der Projekte orientieren, ergibt sich durch die festzulegenden *Zielgrößen*. Ein bekanntes Beispiel für die daraus resultierenden Schwierigkeiten sind die endlosen Expertendiskussionen über Grenzwerte im Bereich des Umweltschutzes. Ein Verfahren zur Vermeidung dieser Effekte ist die Orientierung an *„best practices“* und die daraus folgende Ableitung von *„benchmarks“*, die an Stelle der im Einzelprojekt definierten Ziele für Soll-Ist-Vergleiche von Indikatoren verwendet werden.

Entsprechend dieser Einwände macht es Sinn, definitorische zwischen „**Indikatoren**“ und „**Indikatorenwerten**“ zu unterscheiden wie dies z.B. die GTZ in ihrer „Begriffswelt“ tut. Demzufolge sind Indikatoren „*Kenngrößen, die zur Abbildung eines bestimmten, nicht messbaren und oftmals komplexen Sachverhalts herangezogen werden. Sie beschreiben, woran das Eintreten einer geplanten Veränderung beobachtet oder gemessen werden kann. Sie können quantitativ oder qualitativ sein. Um die Erreichung eines Ziels mithilfe eines Indikators bewerten zu können, benötigt man Indikatorwerte. Diese definieren den Erwartungshorizont, bei dessen Eintreten das Ziel als erreicht gelten soll. Ein Indikatorwert kann quantitativ oder qualitativ sein; in letzterem Fall besteht er aus einem objektiv nachprüfbaren Qualitätsmerkmal*“ (GTZ 2003: 119).

Aber auch ein solches Vorgehen kann nicht die notwendige *Interpretationsleistung bei der Bewertung von Indikatoren* ersetzen. Letztendlich gibt es keine „sich-selbst-erklärende“ Indikatoren und eine bürokratische Bewertung von Projekten oder Programmen anhand von Indikatorenwerten birgt die Gefahr, die Gründe, warum Zielwerte erreicht oder nicht erreicht wurden, aus dem Blick zu verlieren. In diesem Sinne stellen Indikatorensysteme für eine an der *Verbesserung der Praxis* orientierter Evaluation lediglich den Ausgangspunkt, nicht den Endpunkt der Arbeiten dar.

Wolfgang Zapf (1977) hat in einem Aufsatz über soziale Indikatoren vier verschieden Aufgaben, die Indikatoren erfüllen können, unterschieden.

- ✓ Indikatoren beschreiben einen Ist-Zustand und beantworten dabei die Frage: **„Wie ist es?“**
- ✓ Werden die Messungen in regelmäßigen Abständen wiederholt, kann auch die Frage beantwortet werden: **„Wie verändert es sich?“**
- ✓ Für Prognosen und Erklärungen müssen allerdings Theorien entworfen und empirisch in Kausalanalysen überprüft werden, um die Frage zu beantworten: **„Wie funktioniert es?“**
- ✓ Zur Steuerung schließlich muss noch ein weiterer Schritt gegangen werden und anhand gezielter Experimente die Beeinflussbarkeit der zentralen Merkmale erprobt werden. Nur durch Indikatoren ist die Frage **„Wie kann man es beeinflussen?“** nicht zu beantworten.

Indikatoren sind somit Hilfsmittel zur frühzeitigen Identifikation von Problemen (**„Früherkennung“**), zur Verfolgung von Veränderung im Zeitverlauf (**„Dauerbeobachtung“**) und zur Einordnung des Erreichten im Hinblick auf die eigenen Ziele (**„Fortschrittskontrolle“**) sowie im Vergleich zu alternativen Lösungswegen (**„Lösungsoptimierung“**). Damit sie diese Funktionen erfüllen können, müssen sie zu Vergleichswerten in Bezug gesetzt werden. Solche Vergleichswerte können *normativ begründet* (durch Festlegung kritischer Grenzwerte

oder anzustrebender Zielwerte) oder *empirisch produziert* (durch wiederholte oder vergleichbare Messungen) sein.

Indikatoren sind Kenngrößen, die über einen festgelegten, nicht oder nur sehr schwer messbaren Tatbestand Auskunft geben sollen. Dank der ermittelten quantitativen oder qualitativen Informationen sind über Vergleiche mit kritischen Schwellenwerten (Grenzwertperspektive), früheren Messwerten (Entwicklungsperspektive), ex-ante bestimmten Zielwerten (Zielereichungsperspektive) oder den Ergebnissen anderer Beobachtungseinheiten (Bilanzierungsperspektive) Bewertungen möglich. Diese lenken die Aufmerksamkeit auf Stärken und Schwächen der Beobachtungseinheiten und stellen bei Evaluationen den Ausgangspunkt für die Ursachenforschung dar.

4 ARTEN VON INDIKATOREN ...

Entsprechend der unterschiedlichen Anforderungen an und Aufgaben von Indikatoren lassen sich diese anhand von Klassifizierungen in verschiedene Typen einordnen. Im Kontext der Programm- und Projektevaluation wird häufig zwischen

- ✓ Inputindikatoren
- ✓ Outputindikatoren
- ✓ Outcomeindikatoren und
- ✓ Impactindikatoren

unterschieden (vgl. z.B. EC 2002). Mit Hilfe von *Inputindikatoren* werden die zur Zielerreichung eingesetzten Ressourcen (Finanzmittel, Personal, Technik) abgebildet. *Outputindikatoren* stellen die unmittelbaren und konkreten Ergebnisse der durchgeführten Maßnahmen dar. *Outcomeindikatoren* dagegen beziehen sich auf den Nutzen bzw. die Resultate, welche durch den Output im Hinblick auf die Zielsetzungen erreicht werden konnten. *Impactindikatoren* schließlich messen die längerfristigen Konsequenzen und dauerhaften Veränderungen, die durch den Projekt- oder Programmoutcome erzielt werden konnten. Outcome- und Impactindikatoren sind somit beides Wirkungsindikatoren, die sich im Wesentlichen durch den zeitlichen Horizont der beobachteten Wirkungen unterscheiden. Sie beziehen sich ferner auf unterschiedliche Ebenen des Zielsystems (und der dabei mittels *Zielindikatoren* festgelegten Zielgrößen): Outputindikatoren verstehen sich als Überprüfung der konkreten Projektergebnisse, Outcomeindikatoren verweisen auf die unmittelbar angestrebten Projektwirkungen während Impactindikatoren die Nachhaltigkeit der erreichten Veränderungen (und damit die Oberziele von Projekten und Programmen) in das Blickfeld rücken. Schließlich lassen sich die erzielten Ergebnisse und Wirkungen nochmals mittels *Bedarfsindikatoren* mit den (sich ebenfalls verändernden) Zielen, Wünschen und Bedürfnissen der Zielgruppen (oder anderer stakeholder) in Beziehung setzen.

Ein Beispiel:

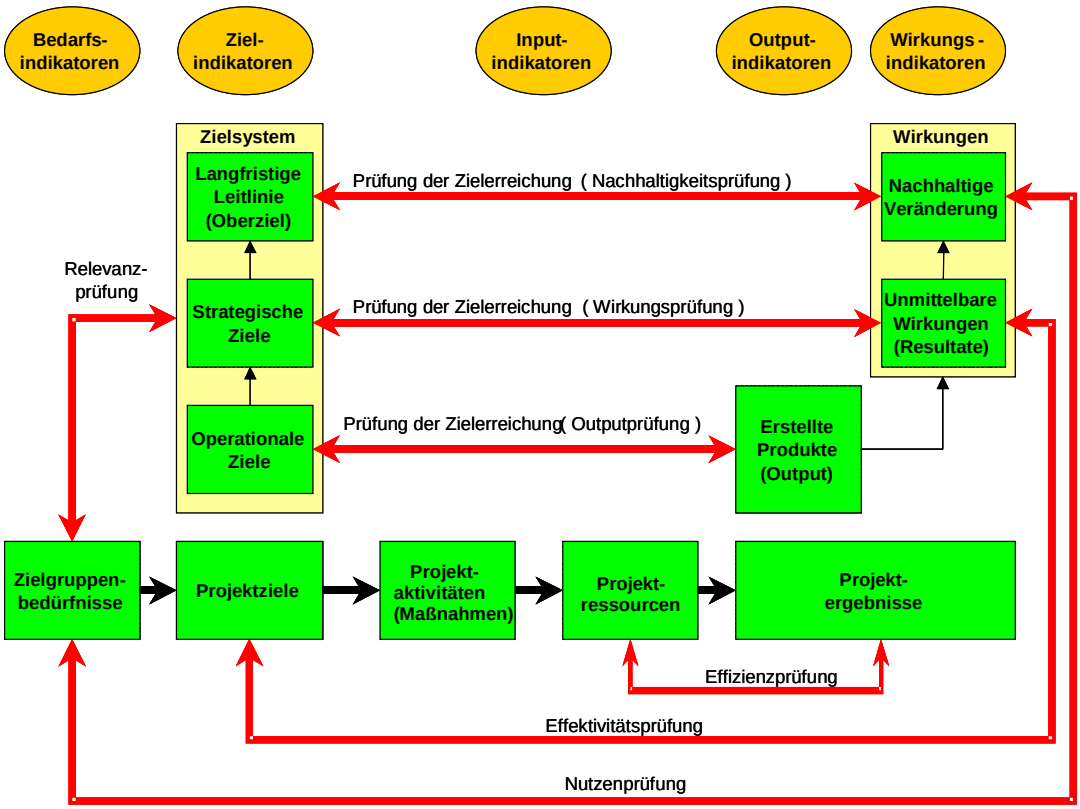
Langfristiges (Ober-)ziel eines Projektes ist es, die Analphabetenrate in einem Land zu senken. Dementsprechend stellt die Veränderung der landesweiten Analphabetenraten im Vergleich zum Ausgangspunkt einen *Impactindikator* dar. Zur Erreichung dieses Ziels werden

Schulen in bilateraler Zusammenarbeit gefördert, in denen Schreiben und Lesen gelehrt wird. Die Anzahl der Schulabsolventen kann als *Outputindikator* dienen. Die Veränderungen ihrer Schreib- und Lesefähigkeiten im Vergleich zu gleichaltrigen Personen ohne Schulbesuch stellen einen *Outcomeindikator* dar. Mit Hilfe von *Inputindikatoren* kann z.B. bestimmt werden, wie viel Lehrer zur Ausbildung dieser Schüler notwendig gewesen sind. Mittels *Bedarfsindikatoren* wird schließlich überprüft, ob die gewünschte Veränderung der landesweiten Analphabetenrate überhaupt noch angestrebt wird oder sich die Ziele z.B. bei der politischen Elite des Landes während der Projektlaufzeit gewandelt haben.

Wirkungsindikatoren lassen sich entsprechend der unterschiedlichen Ebenen des Zielsystems in Output-, Outcome- und Impactindikatoren einteilen. Diese Wirkungsindikatoren können mit den Zielwerten der Zielindikatoren zur Beurteilung der Zielerreichung in Beziehung gesetzt werden (Effektivitätsprüfungen). Sie können ferner auch mit den eingesetzten Ressourcen und dem theoretisch zur Zielerreichung notwendigen Ressourcenbedarf verglichen werden (Effizienzprüfungen).

5 … UND FÜR WAS SIE GEBRAUCHT WERDEN

Häufig ist weniger der einzelne Wert eines Indikators von Interesse sondern der Vergleich mit bereits erhobenen bzw. noch zu erhebenden anderen Werten. Prinzipiell lassen sich eine Reihe verschiedener Arten solcher Vergleiche unterscheiden (vgl. Abb. 1).



Quelle: Meyer (2000)

- *Relevanzprüfung (Vergleich zwischen Zielgruppenbedürfnissen und Zielsetzungen)*

  Insbesondere im Rahmen von Machbarkeitsstudien kann mittels Vergleich geeigneter Indikatoren überprüft werden, ob die entwickelten Ziele tatsächlich den Bedürfnissen der angestrebten Zielgruppen entsprechen oder nicht.

- *Effizienzprüfung (Input-Output- oder Kosten-Nutzen-Vergleiche)*

  Durch den Indikatorenvergleich zwischen Aufwendungen für die Leistungserstellungen und den tatsächlich realisierten Projektergebnissen wird die Effizienz einer Maßnahme geprüft. Diese Vergleiche sind vor allem im Projektverlauf interessant, wenn durch geeignete Steuerungsentscheidungen eventuell Effizienzsteigerungen erzielt werden können. Entsprechende Indikatoren stellen deshalb häufig ein wesentliches Element von Monitoringsystemen dar.

- *Effektivitätsprüfungen (Ziel-Wirkungs-Vergleiche)*

  Der Vergleich zwischen den erreichten Wirkungen und den angestrebten Zielwerten gibt dagegen Auskunft über die Effektivität der Maßnahmen. Bei einem solchen Vergleich steht weniger die Bewertung der Zielerreichung als die Quantität der Wirkungen im Vordergrund des Interesses. Prüfungen dieser Art werden häufig in ex-post-Evaluationen (auch Projektvergleichend) durchgeführt.

- *Zielerreichungsprüfungen (Ziel-Ergebnis-Vergleiche)*

  Zielerreichungsprüfungen unterscheiden sich von Effektivitätsprüfungen durch die stärkere Gewichtung der Zielsetzungen im Vergleich mit den Ergebnissen. Überprüfungen der Zielerreichung können gemäß des Zielsystems auf drei verschiedenen Ebene durchgeführt werden. Bei einer einfachen Outputprüfung wird festegestellt, ob die ursprünglich im Rahmen der operationalen Ziele geplanten Produkte und Ergebnisse tatsächlich in der gewünschten Form erreicht wurden oder nicht. Wirkungsprüfungen dagegen bewegen sich auf der Ebene der strategischen Ziele und sind mit den erzielten Resultaten des Projektes zu vergleichen. Bei Nachhaltigkeitsprüfungen geht es dagegen um die Dauerhaftigkeit der Wirkungen, die in Beziehung zu den formulierten Oberzielen der Projekte stehen. Während Outputprüfungen zu den am Förderende durchgeführten Standardverfahren zur Überprüfung der Zielerreichung in fast allen Projektbereichen und bei der überwiegenden Zahl der Mittelgeber gehören, gilt dies für Wirkungsprüfungen, die im Rahmen einer Abschlussevaluierung (manchmal auch bereits in Zwischenevaluierungen) durchgeführt werden sollten, leider noch nicht im selben Umfang. Vergleichsweise selten werden bisher Nachhaltigkeitsprüfungen im Rahmen umfassender ex-post-Evaluierungen vorgenommen, weshalb die Erreichung der Oberziele häufig nicht kritisch hinterfragt wird.

- *Nutzenprüfungen (Bedarf-Wirkungs-Vergleiche)*

  Ebenfalls sehr selten werden die erzielten dauerhaften Projektwirkungen mit den Bedürfnissen der stakeholder (oder auch nur der angestrebten Zielgruppe) ins Verhältnis gesetzt. Die Durchführung solcher Nutzenprüfungen ist insbesondere im Rahmen langfristiger bilateraler Zusammenarbeit und zur Weiterentwicklung des gemeinsamen strategischen Vorgehens in umfangreichen Programmen unabdingbar notwendig. Entsprechende Aktivitäten sind dementsprechend am ehesten auf der Systemebene und weniger im Rahmen der Projektarbeit vorzufinden.

Die Zuordnung von Maßzahlen zu den hier vorgestellten Indikatorenklassen bedeutet nicht, dass sich z.B. Wirkungsindikatoren von Zielindikatoren prinzipiell unterscheiden. Es gilt im Gegenteil, dass beide Maßzahlen möglichst gleich erhoben werden sollten, damit die Vergleichbarkeit überhaupt zu gewährleisten ist. Dies bedeutet, dass bei der Zieldefinition bereits konkret mit Hilfe von Wirkungsindikatoren messbare Werte auf einer gleichen Skala vorgegeben werden. Aus Gründen der Vergleichbarkeit werden beispielsweise in Kosten-Nutzen Rechungen immaterielle Werte ausschließlich monetär berechnet. Zur Beurteilung von Erkrankungen werden die zur Heilung im Durchschnitt notwendigen finanziellen Aufwendungen berechnet, wobei auch die menschliche Zuwendung des Pflegepersonals ausschließlich in Geldeinheiten definiert wird. Diese abstrakte (manchmal sogar makabere) Art der *Transformation* sozialer Sachverhalte auf eine gemeinsame Skala (Geldeinheiten) erleichtert den Vergleich und hat den Vorteil, leicht zu interpretierende Ergebnisse zu liefern.

Zur Bewertung von Sachverhalten werden Indikatoren bzw. die mit ihnen gemessenen Werte miteinander verglichen. Dies erfordert eine möglichst einheitliche Formulierung der Indikatoren mit einer gemeinsamen Vergleichsdimension. So muss z.B. bereits bei der Zielentwicklung an die Möglichkeiten der Prüfung der Zielerreichung und die hierfür notwendigen Indikatoren gedacht werden.

6 INDIKATOREN – VON EINER NEUEN AURA UMGEBEN

Durch die Einführung des neuen Auftragsrahmens „AURA“ wurden innerhalb der GTZ zur Kennzeichnung der veränderten Vorgehensweise die Begrifflichkeiten der Zielsysteme und Wirkungen erheblich verändert. So werden z.B. keine Oberziele für Programme und Projekte mehr definiert und auch die oben beschriebene Unterscheidung zwischen strategischen und operationale Zielsetzungen wird zugunsten einer stärker zeitlich orientierten Trennung zwischen Gesamt- und Komponentenzielen aufgegeben. Damit hat sich die GTZ nicht nur begrifflich von der Formulierung hierarchischer Zielsysteme, wie sie von der Projektmanagementliteratur empfohlen und von vielen Organisationen (auch anderen Akteuren innerhalb der EZ) verwendet wird, abgesetzt.

Im Zentrum von AURA stehen nun „Wirkungsketten“ mit den folgenden Komponenten:

- ✓ **Leistungen** des Vorhabens (z.B. die erstellten Produkte, die durchgeführte Beratung)
- ✓ **Nutzung** dieser Leistungen (z.B. der Produkteinsatz, die Inanspruchnahme der Beratung)
- ✓ **Direkter Nutzen** (Gesamtziel) (z.B. messbare Produkt- oder Beratungswirkung)
- ✓ **Indirekter Nutzen** (z.B. von Produkten oder Beratungen mitbeeinflusste Veränderungen)
- ✓ **Hoch aggregierter Nutzen** (z.B. Wandel innerhalb des Sektors, in dem die Produkte oder Maßnahmen Wirkungen entfalten konnten).

Bei der Projektbeantragung sollen nun anstelle von Projekt- oder Programmzielen ein Gesamtziel formuliert werden. Dieses **Gesamtziel** „*soll den direkten Nutzen beschreiben, d.h. eine intendierte, positive, entwicklungspolitische Wirkung, die dem Vorhaben noch eindeutig*

zugeordnet werden kann. An der Erreichung des Gesamtziels wird der Erfolg des Vorhabens gemessen... Wichtig ist, dass das Ziel nicht lediglich eine Leistung beschreibt, die im Rahmen des Vorhabens erbracht wird, oder die Nutzung dieser Leistung.“ (BMZ-GTZ 2003). Von einer Anbindung des direkten Nutzens an den hoch aggregierten Nutzen (entspricht in etwa den heutigen Oberzielen der Projekte) wird abgesehen, da nach Auffassung der GTZ eine „Zuordnungslücke“ die kausale Verbindung zwischen Projektwirkungen und aggregierten Veränderungen verhindert.

Da das Gesamtziel erst am Ende der Projektlaufzeit erreicht werden soll, ist eine Konkretisierung für die einzelnen Projektphasen durch Formulierung von **Phasenzielen** vorzunehmen. Diese Phasenziele sollen sich genauso wie die Gesamtziele auf den direkten Nutzen (und damit die messbaren Wirkungen eines Projekts oder Programms beziehen), müssen sich aber nicht unbedingt auf der gleichen Ebene bewegen. So ist z.B. bei einem Berufsbildungsprojekt denkbar, dass in einer ersten Phase durch Maßnahmen wie die Veränderung von Curricula eine höhere praktische Leistungsfähigkeit der Auszubildenden während der Ausbildung erreicht werden soll und erst in der nächsten Phase Wirkungen auf der Betriebsebene angestrebt werden (und hierfür wiederum eine Reihe anderer, stärker betriebsbezogener Maßnahmen implementiert werden). Im Unterschied zum Gesamtziel können sich Phasenziele auch auf die Nutzung von Leistungen beziehen (nicht jedoch auf die Leistung selbst) und sie müssen zum Ende einer definierten Projektphase (und nicht erst zum Auslaufen des Projektes) realisiert werden.

Für die Indikatoren bedeutet diese Veränderung, dass ausschließlich auf den direkt messbaren Nutzen bezogene **Wirkungsindikatoren** (differenziert nach verschiedenen Projektphasen) formuliert werden sollen und die **Zielsetzung als Definition von zu erreichenden Indikatorenwerte** zu erfolgen hat. Die während eines Projekts erstellten Leistungen sind nun nur noch Mittel zum Zweck und sollen nicht zur Projektbewertung herangezogen werden.

7 INDIKATORENTYPEN? TYPISCH INDIKATOR !

Indikatoren lassen sich selbstverständlich nicht nur wie in der bereits beschriebenen Form klassifizieren. Es sind im Gegenteil unendlich viele Typologien anhand unterschiedlicher Kriterien denkbar. Welche Kriterien zur Klassifikation verwendet und wie viele Typen gebildet werden, bleibt letztlich ausschließlich dem Interesse des Analytikers vorbehalten.

Damit ist zugleich auch ein zentrales Problem der Indikatorenbildung angesprochen: Indikatoren sind Hilfsmittel zur Ordnung meist nicht direkt beobachtbarer Tatbestände und sie ordnen die Realität in die vom Analytiker vorgegebenen „Schubladen“ anhand ebenfalls künstlich bestimmter Kriterien ein. Über den Sinn dieser Zuordnung geben Indikatoren keine Auskunft und es gibt auch keine Eigenschaften, die eine allgemeingültige Entscheidung über eine für alle Anwendungsfälle „gute“ oder „schlechte“ Klassifizierung zulassen. Im Gegenteil geraten häufig unterschiedliche Ordnungsprinzipien in Konflikt miteinander (z.B. mag es bei der Ordnung von Kleidung im Kleiderschrank nach dem Waschen effizienter sein, diese nach Waschtemperatur zu sortieren – beim Ankleiden ist dieses Sortierungsschema allerdings eher hinderlich). Dementsprechend gibt es natürlich auch nicht „die“ Indikatorentypologie, sondern verschiedene Formen mit unterschiedlichen Zielsetzungen.

Eine weitere Form der Ordnung von Indikatoren bezieht sich beispielsweise auf die *Inhalte* der verwendeten Indikatoren. So spricht man von naturwissenschaftlichen, technischen, ökonomischen oder sozialen Indikatoren. Eine solche Einteilung sagt selbstverständlich nichts über die messtechnische Güte eines Indikators aus (ein sozialer Indikator ist keineswegs prinzipiell schlechter messbar oder weniger aussagekräftig als ein technischer Indikator) noch bedeutet dies eine normative Wertung (ein sozialer Indikator ist sicherlich nicht in diesem Sinne „sozial", ein ökonomischer nicht „ökonomisch"). Ähnlich wie im obigen Beispiel bedeutet diese Klassifizierung auch nicht, dass eine bestimmte Maßzahl nicht unterschiedlichen Klassen zugeordnet werden kann. Die Messung der Raumtemperatur mit einem Thermometer muss nicht unbedingt ein „naturwissenschaftlicher" Indikator sein, sondern kann je nach Einsatzgebiet und *Verwendungszusammenhang* „technisch" (z.B. zur Überprüfung der Funktion einer Klimaanlage), „ökonomisch" (z.B. zur Bewertung der Heizkosten) oder „sozial" (z.B. zur Einschätzung der Wohnqualität) sein.

Bei der Konstruktion von Indikatoren sind zumeist die inhaltlichen Ausrichtungen (und eine entsprechende Klassifizierung der Indikatoren) weniger relevant als messtechnische Kriterien, die zu einer völlig anderen Unterscheidung von Indikatorenarten führt. Auf zwei dieser Ordnungsprinzipien soll später noch näher eingegangen werden. So kann hinsichtlich dem *Grad der Komplexheit der verwendeten Maßzahlen* zwischen einfachen, relationalen und komplexen Indikatoren unterschieden werden. Die Einteilung nach der *Art der Skala* in nominale, ordinale und metrische Indikatoren betrifft den bereits erwähnten Unterschied zwischen quantitativen und qualitativen Messungen.

Indikatoren sind Instrumente zur Klassifizierung von Merkmalen anhand eines Ordnungsprinzips. Sowohl dieses Ordnungsprinzip als auch die Art und Zahl der verwendeten Klassen sind Ergebnis menschlicher Entscheidungen. Entsprechende des Verwendungszusammenhangs können unterschiedliche Klassifizierungsstrategien sinnvoll sein. In der Praxis können allerdings die mit diesen Ordnungsprinzipien verbundenen Ansprüchen auch in Konflikt geraten.

8 WIE MAN ÄPFEL MIT BIRNEN VERGLEICHT …

Der Spruch „man kann Äpfel nicht mit Birnen vergleichen" gehört nicht nur im Alltag, sondern leider immer noch auch unter Evaluatoren zu den häufig verwendeten Aussagen. Damit soll zumeist auf die großen vorhandenen *Unterschiede zwischen Evaluationsobjekten* hingewiesen und die Unmöglichkeit eines Vergleichs zwischen ihnen behauptet werden. Dabei wird übersehen, dass diese dem Volksmund entnommene Aussage schlicht falsch ist – selbstverständlich vergleichen wir ständig Äpfel mit Birnen und fahren sehr gut damit!

Wenn wir auf den Markt gehen, um Äpfel zu kaufen, so verwenden wir ein bestimmtes, über die Zeit erlerntes Vorwissen, wie Äpfel *im Unterschied* zu Birnen aussehen, damit wir diese überhaupt finden können. Es gelingt uns – auch ohne Auszeichnung der Ware und in Ländern, deren Sprache uns fremd ist – aufgrund dieses *Konzepts* problemlos, Äpfel als solche zu identifizieren und nicht mit Birnen zu verwechseln.

Ohne Vergleich wüssten wir dagegen nicht, dass sich Äpfel überhaupt von Birnen unterscheiden! Wir *differenzieren* Äpfel von Birnen anhand unterschiedlicher *Ausprägungen* eines *Indikators* auf einer gemeinsamen *Vergleichsdimension*. Wir ordnen z.B. die Form der beiden Gegenstände (Indikator) einer von verschiedenen Formenart (Vergleichsdimension), die u.a. auch die Klassen „rund“ sowie „birnenförmig“ (Ausprägungen) umfassen, eindeutig zu, wobei wir gleichzeitig wissen, das auf dieser Dimension Äpfel immer zur Klasse „rund“ und Birnen immer zur Klasse „birnenförmig“ gehören (theoretische Annahme).

Wenn wir Äpfel mit Birnen vergleichen wollen brauchen wir also:

- ✓ einen Indikator, der als *Variable* verschiedene Ausprägungen besitzt
- ✓ einen Vergleichsmaßstab mit eindeutig zu unterscheidenden Ausprägungen (*Skala*)
- ✓ mindestens zwei unterschiedliche Ausprägungen, denen beide Objekte eindeutig zuordenbar sind (*Skalenwerte*)

Wollen wir den Unterschied beider Obstsorten herausarbeiten, so müssen wir mindestens eine Variable benennen, auf deren Skala aufgrund unserer *theoretischen Vorüberlegungen* Äpfel und Birnen immer unterschiedliche Skalenwerte besitzen.

Prinzipiell lassen sich eine Vielzahl von Indikatoren (ein *Indikatoren-Set oder –Tableau*) definieren, anhand derer Äpfel und Birnen miteinander verglichen und sowohl die Gemeinsamkeiten als auch die Unterschiede bestimmt werden können. Auf diese Weise gelingt es z.B. nicht nur Biologen, Äpfel und Birnen der gemeinsamen Klasse „Obst“ – die sich von „Gemüse“ unterscheidet – zuzuordnen, obwohl es durchaus bei einzelnen Indikatoren größere Gemeinsamkeiten zwischen bestimmten Obst- oder Gemüsesorten gibt. So unterscheidet sich z.B. die Form des Kürbis weniger stark von der des Apfels als von der der Birne. Dementsprechend gelingt eine korrekte (d.h. der Zielsetzung entsprechende) Klassifikation normalerweise nicht bei der Verwendung eines einzigen, sondern nur bei der Verwendung eines Satzes von Indikatoren, die verschiedene Dimensionen abdecken können.

Indikatoren sind Vergleichsinstrumente, die Unterschiede oder Gemeinsamkeiten verschiedener Objekte auf einem gemeinsamen Vergleichsmaßstab (Skala) mit mindestens zwei verschiedenen Ausprägungen (Skalenwerte) anzeigen. Notwendige Voraussetzung ist die eindeutige Zuordenbarkeit eines Objektmerkmals zu einem bestimmten Skalenwert.

9 … UND WARUM APFEL AUCH NICHT IMMER GLEICH APFEL IST

Die Tatsache, dass wir anhand von Indikatoren mühelos Äpfel von Birnen durch die Einordnung in verschiedene Klassen auf einer gemeinsamen Vergleichsdimension unterscheiden können, bedeutet nun wiederum nicht, dass alle einer Klasse zugeordneten Objekte gleich sind. Es lassen sich problemlos weitere Dimensionen finden, die zu einer immer tieferen Differenzierung führen können: so gibt es z.B. unterschiedliche Sorten von Äpfeln, wobei Äpfel der gleichen Sorte wiederum aus verschiedenen Anbaugebieten stammen können usw. Letztendlich kommen wir am Ende der Vergleiche zu dem Schluss, dass jeder Apfel einmalig und „mit keinem anderen Apfel vergleichbar“ ist. Je feiner das von uns gewählte Raster oder

je größer die Zahl der von uns gewählten Indikatoren, umso wahrscheinlicher werden wir zu dieser Schlussfolgerung der Einmaligkeit gelangen!

Aber es gilt auch die umgekehrte Schlussfolgerung: wird das Klassifikationsschema nur grob genug gewählt, gelingt es, auch die unterschiedlichsten Sachverhalte in eine Klasse einzubinden (letztlich sind alle Lebewesen und Pflanzen Teil der Erde – auch Äpfel und Birnen – und sind dementsprechend von der Vegetation anderer Planeten zu trennen – auch wenn wir diese gar nicht kennen ...). Grundsätzlich gibt es keine „objektiven", unveränderlichen und allgemein gültigen Klassifikationskriterien oder Ordnungsprinzipien – sie sind immer das Ergebnis menschlicher Entscheidungen, die auf der Basis von Theorien getroffen werden – auch wenn diese Theorien nicht unbedingt in einem wissenschaftlichen Sinne formuliert worden sind!

Jede Klassifikation ist im Gegenteil das Ergebnis der gewählten Vergleichsmaßstäbe und der definierten Ausprägungen auf dieser Skala. Sie sind mit anderen Worten Artefakte und ausschließlich Ergebnisse der verwendeten Indikatoren! Je nach *Aggregatniveau* (also der Klassifizierungsebene) werden die Gemeinsamkeiten oder Unterschiede einzelner Elemente hervorgehoben. Dementsprechend gibt es auch keine „richtige" oder „falsche" Zuordnung in einem absoluten Sinn des Wortes, sondern nur in Relation zu den verwendeten Indikatoren und deren Beurteilungsmaßstab.

Für das Beispiel der Äpfel und Birnen bedeutet dies, dass „Äpfel" eine gemeinsame Kategorie, die sich von „Birnen" unterscheidet, darstellen kann, wenn wir das entsprechende Aggregatniveau wählen. Sie können aber auch Elemente der gemeinsamen Kategorie „Obst" sein, die sich von der Kategorie „Gemüse" unterscheidet – wenn wir ein höheres, allgemeineres Aggregatniveau und entsprechende Indikatoren wählen. Sie können aber auch irrelevante Unterscheidungskriterien darstellen, wenn wir ein niedrigeres Aggregatniveau wählen, bei dem z.B. die Sorten „Boskop" und „Granny Smith" voneinander anhand geeigneter Indikatoren unterschieden werden. Eventuell können auch bestimmte Birnensorten den verschiedenen Kategorien gleichrangig zugeordnet werden (wird z.B. der Geschmack als Indikator mit den Ausprägungen „süß" und „sauer" gewählt). Die Unterscheidung zwischen „Äpfel" und „Birnen" (sowie deren Vergleich) ist letztlich nur das Ergebnis einer Entscheidung über Indikatoren, den verwendeten Vergleichsdimensionen und der Aggregationsebene, welche schließlich diese Aussagen ermöglichen oder nicht.

Gemeinsamkeiten oder Unterschiede von Objekten auf einer Vergleichsdimension werden von Indikatoren in Abhängigkeit von dem gewählten Aggregationsniveau sowie dem Differenzierungsgrad des Vergleichsmaßstabs künstlich hergestellt. Es gibt kein allgemeingültiges „natürliches" Kriterium zur Klassifikation.

10 QUANTITATIVE ODER QUALITATIVE INDIKATOREN? ALLES EINE FRAGE DES NIVEAUS!

Insbesondere Anfang der achtziger Jahre beherrschte der Streit zwischen den Anhängern quantitativer und qualitativer Verfahren die sozialwissenschaftliche Methodendiskussion. Auch wenn dieser Streit mittlerweile längst überwunden ist, so hält sich doch vielfach noch das (Vor-)Urteil, dass mit Hilfe von quantitativen Indikatoren keine qualitativen Tatbestände

abgebildet werden können und dementsprechend qualitative Indikatoren nicht quantifizierbar seien. Der Unterschied zwischen quantitativen und qualitativen Indikatoren ist in der Tat eine Frage des Niveaus – allerdings nicht unbedingt des wissenschaftlichen, sondern des *Skalenniveaus*. Grundsätzlich lassen sich drei Typen von Skalen mit spezifischen Eigenschaften unterscheiden:

- ✓ *Nominalskalen*

  Eine Zuordnung der beobachteten Objekte zu einer Klasse erfolgt ausschließlich nach dem Kriterium der Gleichheit. Voraussetzung ist allein die Möglichkeit, dass jedes Objekt bezüglich der interessierenden Dimension *genau einer Klasse zugeordnet werden kann* und sowohl Zuordnung zu mehr als einer Klasse als auch zu keiner der angebotenen Klassen nicht vorkommen dürfen. Die einzelnen Klassen entscheiden sich lediglich hinsichtlich ihrer *Qualität*, eine irgendwie bewertete Abstufung oder Festlegung der Abstände zwischen den Klassen erfolgt nicht. Wenn ein Indikator dieses Kriterium nicht erfüllen kann, darf er auch nicht den Anspruch erheben, ein Indikator zu sein!

- ✓ *Ordinalskalen*

  Im Unterschied zu Nominalskalen erfordern Ordinalskalen zusätzlich die Möglichkeit, dass eine eindeutige *Rangordnung zwischen den einzelnen Klassen* auf der interessierenden Dimension herzustellen ist. Dementsprechend stellt die Zuordnung eines Objekts zu einer Klasse auf einer Ordinalskala bereits eine Bewertung dar, da zumindest eine relative Positionierung im Vergleich zu anderen Objekten auf diese Dimension vorgenommen wird. In der Regel ist eine solche Zuordnung für die Zielsetzungen von Evaluationen vollkommen ausreichend (natürlich nur unter der Voraussetzung der Exaktheit einer solchen Zuordnung durch den Indikator).

- ✓ *Metrische Skalen*

  Die höchsten messtechnischen Anforderungen werden an metrische Skalen gestellt, die zusätzlich die Annahme gleicher Abstände zwischen den einzelnen Klassen in Form von *quantifizierbaren Intervallen* erfordern. Von einer Ratioskala spricht man dann, wenn zusätzlich auch noch ein „natürlicher“ Nullpunkt auf der Skala angegeben werden kann. Während Nominalskalen keine Aussagen über „besser“ oder „schlechter“ zulassen, gehen Metrische Skalen über diese Möglichkeit der Ordinalskalen hinaus und lassen sogar die Quantifizierung der Größe der Abstände zu.

Mit dem Skalenniveau steigt nicht nur der Informationsgehalt des Indikators, sondern auch die messtechnische Anforderung an die Datenerhebung. Die überzeugende Interpretierbarkeit quantitativer Indikatoren kaschiert oft dieses Problem der Datengewinnung. Umgekehrt gilt aber auch, dass mit sinkendem Skalenniveau mehr Aufwand zur Begründung der Bewertung des Indikators erforderlich wird. Die messtechnischen Probleme werden bei qualitativen Indikatoren nicht gelöst, sondern lediglich auf die Interpretation verschoben.

Quantitative und qualitative Indikatoren unterscheiden sich durch das Messniveau ihrer Skalen. Methodische Probleme treten eher bei der Erhebung (quantitativ) oder eher bei der Auswertung (qualitativ) auf.

11 KEINE ANGST VOR ZAHLEN ...

Insbesondere bei Bewertungen haben viele Menschen eine gewisse Scheu vor der Verwendung von Zahlen. Eventuell ist dies ein Ergebnis traumatischer Kindheitserlebnisse, denn bereits in jungen Jahren werden wir in der Schule mit Zahlenwerten benotet. Diese Noten werden gesellschaftlich als gute Indikatoren zur Beurteilung der Leistungsfähigkeit eines Schülers in einem bestimmten Schulfach angesehen und z.B. direkt zur Selektion beim Zugang zu den verschiedenen Schulebenen, zu beruflichen Bildungseinrichtungen oder – über den Numerus Clausus – zu einzelnen Hochschulfächern verwendet. Die Unzulänglichkeiten des Benotungsprozesses und die vielfältigen Einflussmöglichkeiten auf die Notengebung (z.B. bedingt durch die Präferenzen der Lehrer, die Manipulationen der Schüler durch Spicken oder die individuellen Rahmenbedingungen des Lernens) sowie die harten Konsequenzen, die aus einer solchen Bewertung folgt, haben wir am eigenen Leib erfahren.

Allerdings ist es schlicht falsch, die negativen Erfahrungen mit dieser Form der Bewertung der Verwendung von Zahlenwerten bei der Benotung zu zuschreiben (die Ursachen der Probleme sind eher in der Messung des Indikators „Schulnote“ zu sehen). Eigentlich sind nämlich Noten etwas sehr praktisches: sie ermöglichen sowohl den Vergleich zwischen den Schülern und Schülerinnen in einem bestimmten Schulfach als auch den Vergleich zwischen verschiedenen Schulfächern eines Schülers bzw. einer Schülerin. Unter der Voraussetzung gleicher Curricula können sogar die Schülerinnen und Schüler unterschiedlicher Schulen miteinander verglichen werden. Eine rein qualitative Beschreibung der Leistungen – wie sie z.B. von einer Reihe Privatschulen vorgenommen werden – erschwert solche Vergleiche, auch wenn sie der Realität näher kommen mögen als ein einziger Zahlenwert auf einer wenig differenzierenden Notenskala. Letztendlich ist eine Zahl immer eine *Vereinfachung der Realität* mit dem Ziel, Vergleiche zu ermöglichen oder zumindest zu beschleunigen.

Es ist allerdings möglich, über einen *Codierungsprozess* qualitative Aussagen zu quantifizieren. Wenn z.B. bei Umfragen offene Antwortmöglichkeiten vorgegeben werden (insbesondere auch bei qualitativen Interviews), so werden häufig im Nachhinein Typen und Kategorien gebildet, denen die vielfältigen und differenzierten Antworten der Personen zugeordnet werden (vgl. Kluge 1999). Diesen Kategorien können dann wiederum Zahlenwerte zugeordnet und eventuell Rangordnungen oder sogar Abstände zwischen den Klassen festgelegt werden. Damit entspricht dieser Prozess der Vorgehensweise während der Erstellung standardisierter Fragebögen, bei denen allerdings die Antwortvorgaben im vorhinein den Befragten vorgegeben werden. Bezüglich der Codierung hat diese Vorgehensweise allerdings nicht nur den Vorteil, schneller zu sein, sondern zusätzlich auch den Vorteil, dass den Befragten die später dann verwendeten Zahlenwerte vorgelegt werden können und die Auswahl der Kategorie (und des Zahlenwertes) die Entscheidung der Befragten darstellt.

Die Codierung qualitativer Informationen veranschaulicht auch sehr gut, mit welchen Problemen ein Indikator „bei seiner Arbeit“ konfrontiert wird. Für die Zuordnung bestimmter Aussagen zu den einzelnen Kategorien benötigt man eine *allgemeingültige Regel*, der diese Entscheidung zu folgen hat. Ist diese Regel unzureichend, gelingt die Zuweisung aller Fälle zu den vorhandenen Klassen nicht oder es entstehen Zweifelsfälle.

Zahlen stellen Orientierungshilfen dar und vereinfachen den Vergleich von Beobachtungen. Darüber hinaus wird die Verwendung mathematischer Rechenregeln und Transformationen (z.B. zur Erstellung von Indizes) ermöglicht. Soweit ein Indikator nicht direkt zählbare Ergebnisse produziert, können über geeignete Codierungsregeln auch qualitative Befunde quantifiziert werden.

12 … ABER AUCH KEIN BLINDES VERTRAUEN !

Zahlen sind allerdings nicht unproblematisch, da wir mit ihnen eine Präzision verbinden, die nicht immer gegeben ist. Dies kann im Alltag Vorteile haben, wie das folgende Beispiel zeigt:

„Kürzlich saß ich an meiner Steuererklärung. Man kennt das ja. So und so viele Ausgaben für das und das, aber wo ist der Beleg? Vergessen und verschlampt. Ärgerlich, wenn man deshalb zuviel Steuern zahlen muss. Aber halt! Akzeptiert nicht das Finanzamt bestimmte Kosten auch ohne Beleg? Nach einem Urteil des Bundesfinanzgerichts reicht es aus, wenn ‚nicht nur ein geringes Maß an Wahrscheinlichkeit' für deren Echtheit spricht, und diese Wahrscheinlichkeit, obwohl am einfachsten durch Belege zu untermauern, ist grundsätzlich auch anders herstellbar. Die Frage ist nur, wie. Ich habe es so gemacht: statt ‚60 Mark', was ‚etwa 60 Mark, aber genau weiß ich das nicht mehr' verräterisch ähnlich klingt, habe ich geschrieben ‚Ausgaben für Büromaterial: DM 58,24'. Das akzeptiert der Sachbearbeiter vermutlich eher als ‚etwa 60 Mark'. Die ‚DM 58,24' klingen präzise und seriös. Die Gefahr, dass der Finanzbeamte weitere Beweise sehen will, ist hier nur halb so groß. Mit anderen Worten, ich habe mit Daten manipuliert."

Aus: Walter Krämer (1991), So lügt man mit Statistik, München: Piper, S. 13 (Kap. 1: Die Illusion der Präzision)

Ob Zahlenwerte verwendet werden oder nicht (und erst recht nicht die Zahl der Kommastellen), sagt nichts über die Güte der Messung aus und nur diese entscheidet über die Qualität der Ergebnisse. Manchmal sagt der Umgang mit Zahlen allerdings etwas über die Güte des Auswertenden, wenn z.B. bei weniger als 100 Fällen Prozentangaben mit Kommastellen (und damit eine gar nicht messbare Präzision) angegeben werden…

Natürlich gilt auch umgekehrt, dass qualitative Aussagen nicht „qualitativ besser" sind als quantitative, nur weil Zahlenwerte vermieden werden. Die Tatsache, dass Zahlenwerte die Realität „vergröbern" bedeutet nicht, dass eine verbale Beschreibung den Sachverhalt richtiger wiedergibt. Ausführliche Erläuterungen und umfangreiche Zitate (die „Kommastellen" der qualitativen Sozialforschung) müssen ebenfalls nicht unbedingt auf eine höhere Präzision der Messung hindeuten.

Für die inhaltliche Bewertung von Indikatoren ist es wichtig, die *Grenzen der Präzision* von Messungen deutlich zu machen. In manchen Fällen lassen sich statistische Maßzahlen zur Beurteilung der Präzision (z.B. Varianz und Standardabweichung) berechnen und dadurch Schwankungsbereiche für die Interpretation festlegen. Sollten sich diese Schwankungsbereiche aber zu groß für die praktische Nutzung der Indikatoren herausstellen, so ist dringend eine Überarbeitung des Indikators oder sein Austausch angeraten.

Die Exaktheit eines Indikators ergibt sich allerdings nicht nur aus der *Qualität der Messung* (also wie genau die angestrebten Sachverhalten zu messen sind), sondern auch aus der *Qualität seiner Operationalisierung* (also wie genau die gemessenen Sachverhalte das theoretische Konstrukt abbilden). Einen dieser Ansprüche (geschweige den beide gemeinsam) perfekt zu verwirklichen ist in der Praxis unmöglich, weshalb der Dokumentation der Grenzen der Aussagekraft von Indikatoren große Bedeutung zukommt. Ob das Ergebnis eines Indikators Zahlenwerte enthält oder nicht, ist dagegen für die Messgenauigkeit irrelevant.

Zahlen reduzieren komplexe Inhalte auf einen einfachen Wert und geben dabei häufig eine Exaktheit der Messung vor, die überhaupt nicht gegeben ist. Sie können die inhaltliche Interpretation und Bewertung der gemessenen Sachverhalte nicht ersetzen, im schlimmsten Fall sogar fehlleiten.

13 EINFACHE INDIKATOREN ...

Einfache Indikatoren bestehen aus der Messung eines einzigen Sachverhaltes, den sie dann einem theoretischen (d.h. nicht unmittelbar messbaren) Konstrukt zuordnen. Wenn z.B. monatlich die Menge der bei den Arbeitsämtern arbeitslos gemeldeten Personen gezählt wird, so soll diese Messung etwas über den Ausgleich von Angebot und Nachfrage am Arbeitsmarkt aussagen. Die Annahme ist, dass je höher die Zahl der Arbeitslosen ist desto schlechter dieser Ausgleichsprozess funktioniert.

Mindestvoraussetzung für einen einfachen Indikator ist, dass er – wenigstens theoretisch - eine *Variable* ist. Ein konstanter Indikator, der also nur einen einzigen Zustand für alle Beobachtungen anzeigt und für den eine Veränderung *grundsätzlich* ausgeschlossen wird, wäre vollkommen nutzlos. Als Minimalanforderung an einen Indikator ist somit die theoretische Definition von zumindest zwei unterschiedlichen Ausprägungen auf der gemessenen Skala (*Dichotomie*). Ein Beispiel für einen einfachen Indikator ist etwa die Frage, ob eine Person arbeitslos gemeldet ist, eine Frage, die immer mit „Ja“ Ausprägung 1 oder „Nein“ Ausprägung 2 zu beantworten ist. Soweit es die *potentielle Möglichkeit* einer Arbeitslosigkeitsmeldung gibt, macht dieser Indikator Sinn – selbst wenn bei den Messungen kein konkreter Fall (bisher) aufgetreten ist. Grundvoraussetzung für die Messung von Dichotomien ist die *klare Abgrenzung* zwischen beiden Kategorien, die eine *eindeutige Zuordnung* zulässt.

Die Definition einer größeren Anzahl von Kategorien, zwischen denen die Angaben variieren können, macht allerdings aus einem einfachen noch keinen „komplexen“ Indikator. Wird dabei eine Zuordnung zwischen klar abgegrenzten, in seiner Zahl begrenzten Anzahl von Kategorien (im Beispiel etwa: Erwerbstätig – In Ausbildung – Arbeitslos gemeldet – Nichterwerbstätig) zugelassen, so spricht man von einer *diskreten Variablen*. Wiederum ist die Eindeutigkeit der Zuordenbarkeit zu den Kategorien eine wichtige Anforderung an die Bildung solcher Variablen, insbesondere wenn diese als Indikator dienen soll.

Als „*stetig*“ wird eine Variable dann bezeichnet, wenn sie jeden beliebigen (Zahlen-)Wert annehmen kann. Grundvoraussetzung ist hier eine gemeinsame Dimension (in der Regel die Menge der reellen Zahlen), auf der diese Zahlen abgebildet werden können. Prinzipiell könn-

te es sich bei einer stetigen Variablen auch um qualitative Aspekte (etwa das Niveau des Glücksgefühls) handeln, wobei hierzu allerdings die korrekte Zuordenbarkeit auf einer wie auch immer definierten Skala gewährleistet sein müsste – ein zumeist unerfüllbares Kriterium. Als Beispiel könnte etwa der Zeitraum der Meldung einer Arbeitslosigkeit gemessen werden.

Einfache Indikatoren haben häufig den Vorteil, dass sie selbst Laien leicht verständlich zu machen sind. Sie haben aber oft auch den Nachteil, dass ihre Aussagen von eben diesen Laien falsch interpretiert werden oder ihre Aussagekraft generell sehr limitiert ist. So sagt z.B. die Absolutzahl der arbeitslos gemeldeten Personen u.a. aufgrund der z.T. jahreszeitlich bedingten Schwankungen in der Zahl der potentiell an einer Erwerbsarbeit interessierten Personen nur relativ wenig über den Ausgleichsprozess am Arbeitsmarkt aus.

Einfache Indikatoren bilden direkt messbare Sachverhalte ab und ordnen diese einem nicht-messbaren theoretischen Konstrukt zu. Sie müssen zumindest theoretisch wenigstens zwei unterschiedliche und klar voneinander abgrenzbare Ausprägungen besitzen, denen die beobachteten Sachverhalte eindeutig zugeordnet werden können.

14 ... RELATIVE MAßZAHLEN ...

Im Unterschied zu einfachen Indikatoren werden bei relationalen Maßzahlen mindestens zwei unterschiedliche Indikatoren über eine einfache Rechenregel miteinander in Verbindung gesetzt. Ziel dieser Maßnahme ist eine *Standardisierung* des Ergebnisses, wodurch die Vergleichbarkeit erhöht und die Interpretation vereinfacht wird. Sicherlich die am häufigsten verwendete Form einer solchen Transformation ist die *Prozentuierung*, wobei eine Teilpopulation durch die Gesamtpopulation dividiert und das Ergebnis mit Hundert multipliziert wird. Durch diese Rechenregel streut der Anteil immer zwischen 0 und 100%. Auch die Arbeitslosenquote stellt eine solche Maßzahl dar, wobei die Arbeitslos gemeldeten Personen mit der Gesamtzahl der Erwerbspersonen ins Verhältnis gesetzt wird (wobei allerdings die Basis dieser Berechnungen heftig umstritten ist und bei verschiedenen Berechnungsformen der Arbeitslosenquote variiert).

Neben der trivialen und im Alltag weiterverbreiten Standardisierungsform der Prozentuierung gibt es allerdings noch eine Vielzahl anderer Formen, die auf diesem Prinzip aufbauen. Als Beispiele seien hier die Berechnungen von Gewichtungsfaktoren zum Inflationsausgleich in der Volkswirtschaftlichen Gesamtrechnung sowie die bei multivariaten statistischen Analysen zur Bereinigung von Skaleneffekten notwendigen Standardisierungen durch z-Transformation (von einem realen Wert wird der Mittelwert der Verteilung abgezogen und das Ergebnis durch die Standardabweichung dividiert, wodurch sich eine einheitliche Skala in Standardabweichungen und ein gemeinsamer Nullpunkt mit dem Mittelwert ergibt) genannt. Weitere bekannte Beispiele sind die Bevölkerungsdichte, die Erwerbsquote, die Scheidungsrate oder die Geburtenziffer, wobei die verwendeten Anhänge der Maßzahlen (–dichte, -quote, -rate oder –ziffer) jeweils auf unterschiedliche Berechnungsformen der Verhältniszahlen hinweisen.

Relationale Maßzahlen haben den Vorteil, dass sie standardisierte Vergleiche zwischen höchst unterschiedlichen Phänomenen und Entwicklungen ermöglichen. Sie haben aber zugleich den Nachteil, dass die Absolutwerte aus dem Blick geraten und in der Regel eine ganze Reihe unterschiedlicher Verhältniszahlen berechnet werden können. Ein Beispiel ist die Arbeitslosenquote, deren Berechnungsgrundlage entweder alle Erwerbspersonen oder nur die abhängig Beschäftigten (also ohne Einbeziehung der Selbständigen) sein können. Aber auch diese Berechnungen werden heftig diskutiert, da z.B. auch die nicht arbeitslos gemeldeten, aber dennoch Arbeit suchenden Personen (die „Stille Reserve“) eventuell in die Berechnung mit aufgenommen werden könnten. Je nach Definition der Populationen im Zähler und/oder Nenner verändern sich natürlich die Zahlenwerte und würden bei – unberechtigten – vergleichen mit der derzeitigen Berechnungsform zu günstigeren oder ungünstigeren Einschätzungen der Lage führen. Der Streit um die „richtige“ Berechnung der Arbeitslosenquote zeigt aber vor allem, dass eine solche, der Realität *und* den Erwartungen der stakeholder an den Indikator genügende Berechnungsform schwierig ist.

Relationale Indikatoren verknüpfen zwei Maßzahlen durch eine einfache mathematische Rechenregel. Ziel ist in der Regel die Standardisierung der Ergebnisse zum Zweck der Vereinfachung von Vergleichen. Zumeist werden Anteilswerte (Prozente) oder Verhältniszahlen (Quoten) berechnet.

15 … UND KOMPLEXE INDIZES

Neben einfachen Indikatoren und relationalen Maßzahlen gibt es eine Vielzahl wesentlich aufwendig berechneter Indizes, die Aussagen über hoch komplexe Sachverhalte geben sollen. Zumeist besteht hier das Bestreben, eine ganze Reihe von Indikatoren möglichst durch eine einzige aussagekräftige Maßzahl zu ersetzen. Vorbild und bekanntestes Beispiel ist sicherlich das *Bruttosozialprodukt* (BSP), welches einen Aufschluss über die wirtschaftliche Leistungsfähigkeit geben soll. Obwohl in die sehr aufwendige Berechnung des Bruttosozialprodukts eine Vielzahl von Einzelaspekten aus der Volkswirtschaftlichen Gesamtrechnung einfließen (z.B. die Summe des privaten und des staatlichen Konsums, die Investitionen, die Außenhandelsbilanz etc.), berücksichtigt das BSP keineswegs alle wohlfahrtsrelevanten Aspekte (z.B. fehlen ökologische Tatbestände vollständig). Die Stärke dieses Indikators ist allerdings darin zu sehen, dass er mit den meisten anderen Wohlfahrtsindikatoren hoch korreliert und dementsprechend weit über seine eigentliche Berechnungsgrundlage hinaus Auskunft über die Wohlfahrt eines Landes gibt.

Insbesondere in der Entwicklungszusammenarbeit gibt es seit vielen Jahren Bemühungen, einen entsprechenden sozialen Indikator zur Bewertung von Entwicklungsfortschritten zu entwickeln. Am ehesten kann der *Human Development Index* (HDI) der UNDP diesem hohen Anspruch genügen. Die Berechnung basiert auf der Lebenserwartung, dem Bildungsstand (Alphabetisierung Erwachsener und Gesamteinschulung auf der Primär-, Sekundar- und der tertiären Bildungsstufe) sowie dem bereinigten Pro-Kopf-Einkommen als Indikator für einen angemessenen Lebensstandard. Ziel des HDI ist es, auf der Grundlage weniger, weltweit

verfügbarer Indikatoren Aussagen über die Veränderungen speziell in Entwicklungsländern treffen zu kennen (vgl. UNDP 2001; Zur Kritik des HDI siehe z.B. García-Verdú 2002)

Generell ist festzustellen, das komplexe Maßzahlen wie das Bruttosozialprodukt oder der Human Developmentindex einen beträchtlichen Aufwand sowohl bei der Berechnung als auch der Datenerhebung für die verwendeten Indikatoren erfordern und somit zumeist von größeren Organisationen (in der Regel staatliche Stellen) bereit gestellt werden. Eine Ausnahme aus dem akademischen Umfeld stellen psychologische Skalen dar, die komplexe psychologische Konstrukte wie etwa Glück oder Neid mit Hilfe einer Vielzahl von Einzelindikatoren in Befragungen zu erfassen suchen. Auch hier gilt, dass der Aufwand zur Validierung und Prüfung der Skalen erheblich ist. Durch Sammlungen erprobter Skalen wie z.B. in dem seit 1983 publizierten ZUMA-Skalenhandbuch stehen diese aber einer interessierten Öffentlichkeit zur Verfügung und erleichtern ihre Replikation. Eine Anwendung in Evaluationsstudien ist allerdings aufgrund des hohen Erhebungsaufwandes eher selten.

Komplexe Indizes besitzen den Vorteil, dass sie mit einem einzigen Wert schwierige Zusammenhänge und Entwicklungen darstellen können und entsprechende Bewertungen ermöglichen. Sie haben aber häufig den Nachteil, dass sie sehr abstrakt und wirklichkeitsfremd sind, weshalb die direkte Ableitung von Handlungsanweisungen für Steuerungsentscheidungen nicht möglich und ein Verständnis der Maßzahl nur äußerst schwer zu vermitteln ist. Außerdem ist der Aufwand zur Erhebung der benötigten Daten in der Regel so hoch, dass eine Anwendung im Zuge von Evaluationen (speziell bei Kurzzeitaufenthalten) kaum möglich ist.

Komplexe Indizes setzen sich aus einer Vielzahl unterschiedlicher Indikatoren zusammen, die häufig durch komplizierte Rechen- und Gewichtungsregeln verbunden werden. Das Ziel ist es, möglichst mit einer einzigen Maßzahl komplizierte Sachverhalte und Entwicklungen auszudrücken. Für die Anwendung bei Evaluationen ist der notwendige Aufwand zur Erhebung und Berechnung solcher Indizes meistens zu hoch.

16 DER NUTZEN VON INDIKATOREN: EINE ERSTE ZUSAMMENFASSUNG

Indikatoren können a) nicht messbare Sachverhalte messen, b) unvergleichbare Dinge vergleichen, c) komplizierte Zusammenhänge einfach darstellen und d) auf nicht erkannte Steuerungsmöglichkeiten hinweisen. Ihr Nutzen liegt somit in der Früherkennung von Problemfeldern, der Dauerbeachtung von (Fehl-) Entwicklungen, der Überwachung des Projektfortschritts und in der Unterstützung bei der Optimierung von Lösungen für wahrgenommene Probleme. Sie erfüllen diesen Nutzen in erster Linie dadurch, dass sie die gemessene Realität einer einzelnen Beobachtung anhand eines einheitlichen Zuweisungskriteriums einer einzigen Klasse innerhalb einer begrenzten und eindeutig festgelegten Auswahl von Möglichkeiten zuordnen sowie durch ihre einheitliche Formulierung und Messung Vergleiche zu unterschiedlichen Zeitpunkten oder mit anderen Beobachtungsobjekten sowie Normen und Zielen ermöglichen. Die Festlegung der Vergleichsdimensionen, der Zuweisungskriterien und der verschieden Auswahlkategorien entspringt nicht einer objektiv entscheidbaren Gesetzmäßig-

keit, sondern ist letztendlich das Ergebnis menschlicher Entscheidungen und abhängig vom Verwendungszusammenhang.

Indikatoren entbinden uns auch nicht der Arbeit der *Interpretation*, wobei sowohl die theoretischen als auch die methodischen Implikationen der Indikatorenbildung sowie der Messung berücksichtigt werden müssen. Indikatoren an sich sagen nichts über die notwendigen Konsequenzen aus, beantworten keine der uns interessierenden Fragen und sind weder „richtig“ noch „falsch“. Ohne Interpretation sind sie in der Tat nutzlos!

Allerdings ist die Interpretation eines Indikators (und dementsprechend auch sein Nutzen) nicht unabhängig von seiner *Qualität* zu sehen. Die Qualität eines Indikators kann aus methodischer und aus theoretischer Sicht beurteilt werden. Unter methodischen Gesichtspunkten geht es darum, dass die gewünschten Sachverhalte möglichst exakt gemessen werden und die hohe Messqualität auch jederzeit wiederholt werden kann. Mangelhafte Qualität von Messungen kann zu falschen (vom „wahren“ Wert abweichenden) Ergebnissen führen, Vergleiche verfälschen und unsere Interpretation in die Irre leiden. Unter theoretischen Gesichtspunkten entscheidet sich die Qualität eines Indikators an der Frage, ob er das nichtmessbare inhaltliche Konstrukt tatsächlich wie gewünscht abbilden kann oder nicht. Selbst ein perfekt gemessener Indikator kann inhaltlich am Ziel vorbei gehen. Umgekehrt hilft es uns auch nicht weiter, wenn ein theoretisch korrekter Indikator einfach nicht zu messen ist. Die Indikatorenqualität ist somit immer ein Produkt aus methodischen und theoretischen Anforderungen.

Keinen Einfluss auf die Qualität hat dagegen die Frage, ob ein Indikator *qualitativ oder quantitative Maßzahlen* verwendet. Generell haben bei gegebenem Sachverhalt quantitative Indikatoren mit metrischen Skalenwerten einen höheren Informationsgehalt, weil sie auch über die Rangfolge und die Abstände zwischen den Objekten Aufschluss geben. Dieser Informationsgewinn muss allerdings durch entsprechend höhere Aufwendungen bei der Datenerhebung „erkauft“ werden. Umgekehrt sparen qualitative Indikatoren diesen Mehraufwand bei der Erhebung, übertragen uns aber dann die Aufgabe der Rangordnung sowie der Festlegung von Abständen zwischen den Einheiten bei der Auswertung der Daten. Entsprechend unserer Vorkenntnisse kann sowohl das eine als auch das andere nützlicher sein. Auch der *Grad der Komplexität einer Maßzahl* sagt nichts über deren Qualität aus. Einfache Maßzahlen können mit geringerem Aufwand erhoben und berechnet werden, besitzen aber einen niedrigeren Bezug zu dem theoretischen Konstrukt und erhöhen den Interpretationsaufwand. Komplizierte Indizes haben den Vorteil, komplizierte Sachverhalte auf eine Zahl zu reduzieren, allerdings zu dem Preis erheblicher Aufwendungen zur Datenerhebung und zur Berechnung der Indizes. Wiederum ist die Frage des Nutzens weniger mit dieser technischen Seite als mit dem Verwendungszusammenhang verknüpft.

Teil II: Anforderungen an Indikatoren

17 WANN IST EIN INDIKATOR EIN GUTER INDIKATOR?

„Die Konstruktion guter sozialer Indikatoren ist zuallererst eine Frage der sozialwissenschaftlichen Phantasie“ (Wolfgang Zapf 1977)

Da es keinen allgemeingültigen Regeln für die Erstellung von Indikatoren gibt und auch die Möglichkeiten zur Bewertung der Operationalisierung begrenzt sind, ist es in der Tat erstmal ein kreativer Akt, geeignete Maßzahlen und Messverfahren zu entwickeln. Es ist auch keineswegs gewährleistet, dass ein in einem bestimmten Kontext entworfener Indikator, der sich dort auch in der Praxis bewährt hat, unbedingt immer auf andere Kontexte übertragbar ist. Insbesondere die Anforderungen der Praxis (z.B. die Möglichkeiten zur Umsetzung von Messregeln der Indikatoren) können stark variieren und schnell aus einem „guten“ einen „schlechten“ Indikator machen.

Für die UN ist ein guter Indikator „SMART“ (S=Specific, M=Measurable, A=Attainable, R=Relevant, T=Trackable), das heißt, er trifft eindeutig den angestrebten theoretischen Sachverhalt, ist messbar und für den gewünschten Arbeitskontext anwendbar, zur Lösung der anstehenden Probleme relevant und für die Beteiligten nachvollziehbar (UNDP 1997, 2002b,c; der „SMART“ Anspruch wird aber auch von anderen Organisationen für Indikatoren erhoben siehe z.B. CEFIC 2000, S. 17). Ein guter Indikator entspricht dementsprechend sowohl *theoretischen, methodischen, praktischen und politischen Anforderungen*.

Die Konstruktion von Indikatoren erfordert theoretisches Fachwissen während der Operationalisierung der zumessenden Sachverhalte. Dies bedeutet, dass zur Indikatorenentwicklung inhaltliche Experten hinzugezogen werden müssen, um diesen Anspruch realisieren zu können. Vielfach fehlt diesen inhaltlichen Experten aber auf der anderen Seite der Einblick in die Problematik der Datenerhebung, weshalb zusätzlich methodische Fachleute zur Klärung der damit verbundenen Fragestellungen zu beteiligen sind. Die Frage der Praxis berührt – insbesondere bei Projektevaluationen – die vor Ort gegebenen Kapazitäten und spezifischen Rahmenbedingungen, die von Fachexperten nur unzureichend beurteilt werden können und die Einbindung der Praktiker notwendig macht. Schließlich können Indikatoren politisch sensibel sein und die Datenerhebung bei verschiedenen Personengruppen (z.B. die befragten Zielgruppen, die den Sinn der Datenerhebung nicht einsehen oder die Auftraggeber, welche zur Bereitstellung der notwendigen finanziellen Mittel nicht bereit sind) auf Widerstand stoßen. Eine Einbindung der Stakeholdergruppen in den sozialen (und nicht nur technischen) Prozess der Indikatorenentwicklung ist dementsprechend ebenfalls notwendig.

Ein „guter“ Indikator ist somit nicht nur theoretisch und messtechnisch optimiert, er muss außerdem auch noch praktikabel und bei allen Beteiligten akzeptabel sein. Es bedarf dementsprechend nicht nur „sozialwissenschaftlicher Kreativität“ sondern auch profundes Fachwissen, umfangreiche praktische Erfahrungen, sowie viel Geduld und Verhandlungsgeschick, um einen „guten“ Indikator zu entwerfen.

Indikatoren müssen gleichzeitig theoretischen, methodischen, praktischen und politischen Anforderungen genügen. Die Entwicklung guter Indikatoren ist deshalb nicht durch eine Person allein zu bewerkstelligen, sondern erfordert die Beteiligung von Fachexperten, Praktikern und Stakeholdern.

18 ES GIBT NICHTS PRAKTISCHERES ALS EINE GUTE THEORIE ...

(Immanuel Kant 1724-1804)

Häufig ist außerhalb der Hochschulen eine gewisse Abneigung gegen Theorien und ein erheblicher Zweifel am praktischen Nutzwert theoretischen Denkens zu spüren. Es herrscht vielfach die Vorstellung, dass es eine Dichotomie zwischen Theorie und Praxis gibt, die quasi unvereinbar neben einander stehen. Auch in der Evaluation gibt es Verfahren, die hauptsächlich auf Grundlage der Erfahrungswerte von Praktikern vorgehen und die Verwendung abstrakter Theorien vehement ablehnen. Hierbei wird allerdings übersehen, dass eine solche Vorgehensweise keineswegs a-theoretisch ist, sondern lediglich an Stelle wissenschaftlich formulierter Aussagen *Alltagshypothesen und –theorien* verwendet werden.

Wissenschaftliche Theorien unterscheiden sich vorrangig in zweierlei Hinsicht von Alltagstheorien (vgl. hierzu ausführlich Bortz/Döring 2002, S. 34ff.; Turner 2003).

- ✓ die Aussagen in wissenschaftlichen Theorien werden *explizit formuliert* und als *keineswegs gesicherte Annahmen* behandelt, die mit Hilfe von Daten und Informationen überprüft werden sollen. Alltagshypothesen werden dagegen häufig als selbstverständliche Wahrheiten dem eigenen Handeln (auch der Informationssammlung) zugrunde gelegt und können deshalb zu „Betriebsblindheit“ führen.
- ✓ Wissenschaftliche Theorien *leiten und begrenzen das Forschungsinteresse* (und damit verbunden auch das Sammeln von Informationen). Da Alltagshypothesen in der Regel nicht formuliert werden, fehlt ihnen diese leitende Funktion und sie verführen dazu, sich in der unendlichen Menge eventuell interessanter Informationen zu verlieren.

Wissenschaftlich formulierte Theorien sind dementsprechend für die Sammlung von Informationen und die Entwicklung von Messinstrumenten ungemein praktisch: sie sagen uns, auf welche Fragen wir uns zu konzentrieren haben und warum es wichtig ist, diese zu beantworten. Da die Menge an Informationen grundsätzlich unendlich groß ist, können uns nur Theorien helfen, einen Fokus zu bilden und entsprechend zu selektieren. Das bedeutet natürlich nicht, dass diese Entscheidungen immer richtig sind – im Gegenteil: Theorien weisen uns darauf hin, dass sie falsch sein können und eben nicht Wahrheiten darstellen!

Erfahrungswissen ist die wichtigste Grundlage für die Hypothesen- und Theorienbildung! Ohne Praxiskenntnisse bleibt es schwierig, die richtigen Fragen zu stellen. Man kann es auch drastischer formulieren:

„Wenn die Fakten nicht ausreichen, sind Theorien bestenfalls lächerlich und schlimmstenfalls irreführend.“ *Poul Anderson "Herrin der Luft und der Dunkelheit" 1971*

Theorie und Praxis sind nicht Gegner, sondern untrennbare Partner des Erkenntnisfortschritts!

Dies gilt natürlich im besonderen Masse für die Indikatorenentwicklung: Bevor ein Indikator für die Datenerhebung eingesetzt wird, sollte gefragt werden, warum die so erhobenen Informationen wichtig sind. Nur wenn es hierfür eine befriedigende theoretische Antwort gibt, sollte er aufgenommen werden. Umgekehrt sollten Indikatoren, die zwar theoretisch wichtig erscheinen, sich aber praktisch nicht bewährt haben, vermieden werden.

Je besser sich ein Indikator theoretisch begründen lässt, umso geeigneter ist er. Je besser sich ein Indikator in der Erhebungspraxis bewährt hat, umso eher sollte er eingesetzt werden.

19 ... UND NICHTS THEORETISCHERES ALS EINE GUTE PRAXIS!

(Peter Klotz, Harro Müller-Michaels, Jakob Ossner, Gerhard Rupp 1996)

Das Streben nach Perfektion in der praktischen Umsetzung von Indikatoren ist sicherlich wünschenswert, ein perfektes Indikatorensystem (im Sinne einer perfekten Auswahl messtechnisch optimaler Indikatoren) letztlich unerreichbar. Ein häufig in Diskussionen zur Bildung von Indikatoren vorgebrachter Vorwurf besteht darin, dass ein Vorschlag zu „akademisch“ sei (womit „weltfremd“ und „unpraktisch“ gemeint ist). In der Regel geht es dann darum, dass der vorgeschlagene Indikator zwar unter theoretischen Annahmen wünschenswert, seine praktische Umsetzung aber zu kompliziert und der zur Messung notwendige Aufwand unvertretbar hoch sei.

In der Tat stößt das theoretisch wünschenswerte *immer* an Grenzen des praktisch machbaren. Dies ist aber – entgegen der landläufigen Meinung – nicht ein Problem der Theorie, sondern ein Problem der Praxis! Fehlende Ressourcen, schwierige Rahmenbedingungen oder kulturelle Barrieren sind nicht das Problem einer „schlechten“ Theorie, die nun diesen Gegebenheiten anzupassen ist, sondern sind umgekehrt praktische Schwierigkeiten, die einer perfekten Umsetzung theoretisch wünschenswerter Aspekte im Wege stehen. Dementsprechend sollte die Suche nach einer Annäherung des Machbaren an die theoretischen Anforderungen und nicht etwa der umgekehrte Weg – eine praxisorientierte Reformulierung von Theorien - beschritten werden.

Dies ist übrigens ein Problem, welches sich in der Grundlagenforschung (und dort in allen Fachgebieten) genauso stellt: ein optimales Experiment ohne jeglichen störenden Einfluss ist leider praktisch unmöglich durchführbar! Selbstverständlich ist ein Erkenntnisfortschritt aber trotz dieser Unzulänglichkeiten der Forschungspraxis möglich gewesen und wird auch weiterhin möglich sein. Letztlich wird auch in der Grundlagenforschung immer die Frage gestellt werden müssen, wie mit einem vertretbaren Aufwand möglichst präzise gemessen werden kann. Und die für Forschungsarbeiten zur Verfügung stehenden Ressourcen sind keineswegs so unbegrenzt, wie dies vielfach Praktiker glauben.

Dementsprechend zeichnet sich die Praxis – auch die akademische – durch Pragmatismus aus. Bei der Indikatorenbildung geht es somit um die Optimierung der eingesetzten Instrumente mit Blick auf die theoretisch formulierten Ansprüche an die Indikatoren unter zu Grundlegung der gegebenen Ressourcen und Rahmenbedingung. Das heißt aber auch, Abschied zu nehmen von dem Wunsch nach einer perfekten Messung und die Fehleranfälligkeit der Instrumente als unabdingbar hinzunehmen und bei der Interpretation zu berücksichtigen.

Ein guter Indikator muss dementsprechend zwei konkurrierenden und niemals perfekt miteinander zu vereinbarenden Ansprüchen genügen:

- ✓ der Indikator muss möglichst optimal theoretisch bestimmbare Sachverhalte messen (**Effektivität des Indikators**)

- der Indikator muss mit den vorhandenen Möglichkeiten realisierbar sein (**Effizienz des Indikators**)

Je besser ein Indikator einem theoretischen Konstrukt entspricht, umso effektiver misst er den angestrebten Sachverhalt. Je geringer die zur praktischen Umsetzung eines Indikators notwendigen Aufwendungen (bei gegebener Qualität der Messung), umso effizienter misst er den angestrebten Sachverhalt. Bei der Konstruktion von Indikatoren ist immer eine möglichst optimale Übereinstimmung mit den theoretischen Anforderungen unter Beachtung der Grenzen praktischer Umsetzbarkeit anzustreben.

20 DAS UNMESSBARE MESSEN: OPERATIONALISIERUNG UND INDIKATORENBILDUNG

Die Anforderungen, die an Indikatoren gestellt werden, sind genauso einfach zu beschreiben wie schwer zu erfüllen: Indikatoren sollen letztlich Dinge messen, die wir für unmessbar halten. Wir wollen uns stattdessen dafür entscheiden, gemessene Dinge an Stelle des Unmessbaren zu verwenden – wobei wir mehr oder weniger begründet annehmen, dass zwischen dem Gemessenen und dem Unmessbaren ein wesentlicher und konstanter Zusammenhang besteht. In der Regel haben wir nur wenige Möglichkeiten, diese Annahme zu überprüfen, da wir das Unmessbare nun einmal nicht messen können. Die einzige Chance zum Testen dieser *operationalen Definition* besteht in der Konstruktion alternativer Indikatoren und einem Vergleich derer Werte miteinander (solche Tests werden z.B. zur Skalenbildung in der Psychologie verwendet).

Indikatoren messen somit immer künstliche Konstrukte, niemals Realität. Sie *operationalisieren* theoretische Überlegungen und stellen damit selbst theoretische Konstrukte dar. Die Beurteilung der Qualität eines Indikators bleibt somit immer hypothetisch, da das eigentliche Konstrukt niemals direkt gemessen werden kann und dadurch eine empirische Überprüfung unserer operationalen Annahmen unmöglich ist. Daraus folgt allerdings nicht, dass jeder Indikator fiktiv und beliebig austauschbar ist. Genauso, wie es einer logisch begründeten Annahme bedarf, warum ein bestimmter Indikator das angestrebte Konstrukt messen soll, müssen zur Kritik eines Indikators auch plausible Annahmen gemacht werden, warum ein anderer Indikator besser geeignet sein soll als der bisher verwendete. Darüber hinaus ist es notwendig, *Entscheidungsregeln* festzulegen, anhand derer die Messqualität konkurrierender Indikatoren evaluiert werden sollen. Dementsprechend ist der Prozess der Indikatorenkonstruktion und –prüfung sowohl theoretisch als auch empirisch ein sehr anspruchsvolles Vorgehen.

Beim Prozess der Operationalisierung können vier Schritte unterschieden werden, bei denen jeweils spezifische Fragen zu beantworten sind (vgl. Laatz 1993, S. 30ff., Wottawa/Thierau 1990, S. 85ff.) :

- Bestimmung des zu beobachtenden theoretischen Konstrukts und eines potentiellen Indikators (**was soll wie operationalisiert werden?**)

- ✓ Erstellung einer logisch abgeleiteten, allgemeingültigen und durch die Messung selbst nicht zu beeinflussenden Korrespondenzregel zur Zuordnung des Konstrukts zu einem gemessenen Sachverhalt (**warum misst der Indikator das Konstrukt?**)
- ✓ Festlegung von Falsifikatoren dieser Korrespondenzregel (**wann ist diese Zuordnung zwischen Indikator und Konstrukt als unzureichend zu bewerten?**)
- ✓ Prüfung der Korrespondenzregel (zumeist durch Vergleich mit alternativen und/oder bewährten Indikatoren, da eine unmittelbare Messung des theoretischen Konstrukts in der Regel nicht möglich ist) (**wie kann eine Prüfung der korrekten Zuordnung zwischen Indikator und Konstrukt erfolgen?**)

Unter Operationalisierung wird die Angabe einer Beobachtungsoperation zur Messung eines theoretischen Konstrukts mittels eines Indikators verstanden. Die zur Verknüpfung zwischen Indikator und Konstrukt notwendige Korrespondenzregel muss den Anforderungen einer theoretisch korrekt formulierten Hypothese soweit möglich entsprechen und zumindest Bewertungskriterien sowie Testprozeduren zur Selektion von Indikatoren enthalten.

21 HAUPTSACHE GUT GEMESSEN …

An die Messqualität eines Indikators sind zwei methodische Anforderungen zu stellen, die zumindest ansatzweise auch mathematisch zu überprüfen sind. Dementsprechend sollte ein Indikator

- ✓ das messen, was er vorgibt zu messen (*Validität* oder Gültigkeit der Messung) und
- ✓ bei wiederholten Messungen unter exakt gleichen Bedingungen die gleichen Ergebnisse produzieren (*Reliabilität* oder Zuverlässigkeit der Messung)

Entsprechend der klassischen Testtheorie lässt sich die Reliabilität als der Quotient der Varianz der wahren Werte und der Varianz der beobachteten Werte definieren. Unter der Annahme, dass sich Messfehler statistisch zufällig verteilen gibt es eine Reihe statistischer Verfahren zur Prüfung der Reliabilität, die sich entweder auf Re-Test oder Paralleltests beziehen (vgl. hierzu ausführlich Feldt & Brennan 1988: 109ff., Bortz & Döring 2002: 195ff.).

Im Unterschied zur Reliabilität sind Prüfungen der Validität enge Grenzen gesetzt (vgl. z.B. Crano 2003). Zu unterscheiden ist hier zwischen:

- ✓ *Inhaltsvalidität* (das theoretische Konstrukt wird durch das verwendete Messinstrumentarium komplett erfasst)
- ✓ *Kriteriumsvalidität* (das verwendete Messinstrumentarium misst das theoretische Konstrukt in gleicher Qualität wie andere, als valide erkannte Messinstrumentarien)
- ✓ *Konstruktvalidität* (die verwendeten Messinstrumentarium ergeben bei der Überprüfung inhaltlicher Hypothesen die theoretisch zu erwartenden Ergebnisse).

Lediglich für die Konstruktvalidität gibt es mit den Multitrait-Multimethod-Matrizen (MTMM) ein geeignetes Prüfverfahren (vgl. hierzu Schnell, Hill & Esser 1999: 149ff.). Die Inhaltsvalidität ist eher eine Frage der Theoriebildung und der Indikatorenkonstruktion denn der Messqualität und die Kriteriumsvalidität benötigt einen als valide erkannten Vergleichsindikator, wobei sich dann schnell die Frage stellt, für was in diesem Fall ein neues Instrument entwickelt werden soll und nicht einfach der bewährte Indikator verwendet wird (bei Monitoring-Systemen oder Evaluationen kann allerdings der mit den Indikatoren verbundene Erhebungsaufwand ein wichtiges Argument sein). Die Kriteriumsvalidität ist allerdings ein wichtiges Selektionsinstrument, wenn es zur Auswahl von Indikatoren bzw. zur Einführung eines neuen, alternativen Indikators kommt.

Generell sollte immer versucht werden, die entwickelten Indikatoren im Rahmen eines Indikatorentableaus hinsichtlich ihrer Validität und Reliabilität zu optimieren sowie bei Auswahlentscheidungen diese Kriterien angemessen zu berücksichtigen.

Die Güte von Messinstrumenten ergibt sich aus ihrer Validität (Gültigkeit) und Reliabilität (Zuverlässigkeit). Ein Indikator ist dann valide, wenn er genau das misst was er messen soll. Er ist dann reliabel, wenn er bei wiederholten Messungen unter genau gleichen Bedingungen zu exakt denselben Ergebnissen gelangt.

22 ... BIS ZU DEN GRENZEN DER MACHBARKEIT

In der Praxis ist es allerdings nur in den seltensten Fällen möglich, die hohen methodischen Ansprüche an eine korrekte Messung von Indikatoren bestmöglich zu erfüllen. Neben den technischen Restriktionen, die eine perfekte Messung bereits unter optimalen Randbedingungen zu Messfehlern führen, ergeben sich mehr oder weniger starke Effekte auf die Messqualität, die aus dem sozialen Umfeld der Messung resultieren. Zu unterscheiden ist dabei zwischen sozialen Effekten, die sich aus der:

- ✓ *Akzeptanz* der Messung
- ✓ *Durchführung* der Messung und der
- ✓ *Kontrollmöglichkeit* von Störfaktoren

ergeben. Akzeptanzprobleme können an unterschiedlicher Stelle bei verschiedenen sozialen Gruppen auftreten und in mehr oder weniger erheblichem Maße die Messqualität beeinflussen. So können z.B. Auftraggeber das verwendete Messverfahren als unzureichend ablehnen oder die Zielgruppe eine Datenerhebung boykottieren, weil sie keinen Nutzen für sich erkennen. Die eigentliche Durchführung einer Messung ist u.a. von den verwendeten Messinstrumenten und Erhebungsverfahren sowie den damit verbundenen Stärken und Schwächen der Methode abhängig. Unter Umständen können die für eine korrekte Messung benötigten Ressourcen die vorhandenen Möglichkeiten sprengen und dementsprechend zu einer erhöhten Fehlerquote führen. Schließlich lassen sich in bestimmten sozialen Kontexten (z.B. in der Entwicklungszusammenarbeit) ein Teil der gegebenen Randbedingungen der Messung nur unzureichend kontrollieren. So ist es z.B. in einer Diktatur kaum möglich, die kor-

rekte politische Einstellung der Bevölkerung zu erfassen, da dies voraussichtlich durch das politische Regime mittels gezielter Eingriffe verhindert werden würde.

Aufgrund dieser Probleme werden an Indikatoren nicht nur hohe theoretische und methodische Anforderungen gestellt, sondern auch zusätzlich pragmatische Ansprüche, die sich auf die erfolgreiche Umsetzung des sozialen Prozesses einer Messung beziehen. Wie jede andere Maßnahme, die im Rahmen eines Projekts umgesetzt wird, müssen auch Indikatoren einer sorgfältigen Prüfung hinsichtlich ihrer Umsetzbarkeit angesichts der vorhandenen Ressourcen, den Randbedingungen der Erhebung, Kosten-Nutzen-Überlegungen u.a. unterziehen. Vielfach sind deshalb Idealvorstellungen einer Messung nicht oder nur unzureichend zu erfüllen und es müssen *Proxy-Indikatoren* entworfen werden. Proxy-Indikatoren sind Maße, die zwar nicht die angestrebte Abbildung des theoretischen Konstrukts, dafür aber ein anderes, verknüpftes Objekt messen können. Ein Beispiel ist die Vorgehensweise bei einer Volkszählung in einem afrikanischen Staat, der (aufgrund des hohen Anteils an Nomaden und der gegebenen Weite des Landes) auf die Idee kam, nachts von Flugzeugen aus die Feuerstellen zu zählen (Proxy-Indikator) und anhand der durchschnittlichen Größe der Nomadengruppen die Bevölkerungszahl schätzte. Schwierigkeiten bei der Messung erhöhen primär den Anspruch an die mit der Entwicklung von Indikatoren beauftragten Experten. Es gilt nämlich auch der alte Managementspruch:

„If you can measure it, you can manage it“
(Centre for International Development and Training 2002)

Die Messqualität eines Indikators ist nicht nur nach theoretischen und methodischen (und damit technischen) Gesichtspunkten, sondern auch nach sozialen Kriterien zu beurteilen. Nur ein Indikator, der mit den vorhandenen Ressourcen in dem gegebenen sozialen Umfeld gemessen und von den beteiligten Personen akzeptiert wird, kann für die Steuerung relevante Ergebnisse liefern. Eventuell müssen deshalb Proxy-Indikatoren eingesetzt werden, die zwischen den messtechnischen Anforderungen und den realen Gegebenheiten vermitteln sollen.

23 ANFORDERUNGEN AN INDIKATOREN: EINE ZWEITE ZUSAMMENFASSUNG

Indikatoren müssen *a) theoretischen, b) methodischen, c) praktischen und d) politischen Anforderungen an ihre Qualität* genügen. Aus *theoretischer Sicht* entscheidet die *Güte der Operationalisierung* des theoretischen Konstrukts über die Indikatorqualität. Es müssen Aussagen über den Zusammenhang zwischen dem vom Indikator gemessenen Sachverhalt und dem eigentlich angestrebten, nicht-messbaren Konstrukt formuliert werden, die eindeutig abgrenzbar und durch geeignete, benennbare Prüfverfahren zu testen sind. Dementsprechend müssen Kriterien zur Entscheidung, ob eine Operationalisierung als gelungen oder nicht gelungen anzusehen ist, gefunden werden.

Unter *methodischen Gesichtspunkten* ist primär die *Güte der Messung* das Maß, an dem die Qualität eines Indikators gemessen werden sollte. Die Anforderung an einen Indikator besteht zum einen darin, dass er messen soll, wofür er konstruiert wurde (Validität) und dies

bei wiederholten Messungen immer auf die selbe Weise zu realisieren hat (Reliabilität). Wiederum sind Kriterien darüber zu erstellen, wann eine Messung als ausreichend valide und reliabel anzusehen ist und wann dies nicht der Fall sein soll.

Für die Praxis stellt sich die Frage, inwieweit bei einem Indikator eine den theoretischen und methodischen Anforderungen entsprechende Erhebung überhaupt realisierbar ist. Die *praktische Perspektive* verweist somit auf die *Güte der Durchführung* als Maßstab zur Beurteilung der Indikatorqualität. Auch hierzu sind Kriterien über die Grenzen der Machbarkeit festzulegen, anhand derer ein Indikator abgelehnt oder angenommen werden kann.

Schließlich stellt sich die Frage nach der „politischen Korrektheit“ eines Indikators, d.h. inwieweit seine Messergebnisse als Entscheidungsgrundlage akzeptiert werden, die Bereitschaft zur Beteiligung an der Datenerhebung gegeben ist und die Interpretation mit Bezug auf das theoretische Konstrukt anerkannt wird. Aus dieser *politischen Sichtweise* ergibt sich eine Bewertung der Indikatorqualität hinsichtlich der *Güte der Anerkennung*, für die wiederum Toleranzgrenzen für die Beteiligtengruppen zu definieren sind.

Nur ein Indikator, der gleichzeitig diese vier Ansprüche erfüllt, kann eine entscheidungsrelevante Datenbasis liefern. Je besser sich ein Indikator theoretisch begründen und operationalisieren lässt, umso besser kann er das gewünschte nicht messbare Konstrukt abbilden. Je höher seine Messqualität, umso genauer wird diese Abbildung sein. Je einfacher die mit seiner Messung verbundenen Anforderungen im gegebenen sozialen Umfeld umsetzbar sind, umso eher wird diese Abbildung auch zu erstellen sein. Und schließlich: je größer die Überzeugung der am Entscheidungs- und am Messprozess beteiligten Personen ist, dass dieser Indikator entscheidungsrelevante Informationen bieten kann, umso eher werden seine Ergebnisse auch in die tatsächlichen Entscheidungsprozesse einfließen.

Durch diesen letzten Punkt wird deutlich, dass die Entwicklung von Indikatoren nicht nur eine rein technische Prozedur ist, sondern ein *sozialer Prozess*, der verschiedene Expertise, Erfahrungen und Auffassungen für eine erfolgreiche Erstellung „guter“ Indikatoren mit einbinden muss. Die Bewertung der *Effektivität* (Übereinstimmung der Messergebnisse mit dem nicht-messbaren theoretischen Konstrukt) und der *Effizienz* (Verhältnis von Erhebungsaufwand und Informationsgewinn) von Indikatoren ist nicht anhand eindeutiger technischer oder logisch ableitbarer Normen möglich. Die *Qualität von Indikatoren* ergibt sich zu einem großen Teil auch aus dem *Vertrauen*, welches die Beteiligten in die gewonnenen Informationen setzen. Dies bedeutet aber wiederum nicht, dass die Verwendung von Maßzahlen beliebig unter Ignorierung sämtlicher theoretischer und methodischer Anforderungen frei auszuhandeln ist. Vertrauen bei den Beteiligten erzeugt vor allem, dass die höchstmöglichen methodischen Standards überprüfbar eingehalten werden.

Teil III: Entwicklung von Indikatoren

24 DIE UNENDLICHE GESCHICHTE: INDIKATORENENTWICKLUNG ALS SOZIALER PROZESS

"Es gibt Menschen, die können nie nach Phantasien kommen und es gibt Menschen, die können es, aber sie bleiben für immer dort. Und dann gibt es noch einige, die gehen nach Phantasien und kehren wieder zurück. Und die machen beide Welten gesund" (Michael Ende, Die unendliche Geschichte,)

Das soziale Umfeld ist nicht nur für die Messung von Indikatoren, sondern bereits in ihrem Entwicklungsprozess von entscheidender Bedeutung. Da insbesondere in einem Projektumfeld die Konstruktion geeigneter Indikatoren die Einbeziehung fachlicher und methodischer Experten sowie von Praktikern und Vertretern der Stakeholder-Gruppen erforderlich ist, ergibt sich eine komplizierte soziale Situation, die sich schnell zu einer konfliktreichen unendlichen Geschichte entfalten kann.

Damit der „Rückweg aus Phantasien" gelingt, bedarf es vielfach spezieller Kenntnisse in *Mediationstechniken*, die insbesondere von Evaluatoren erwartet wird (als Einführung siehe HTD 2001). Ein erfolgreicher Interaktionsverlauf in Kleingruppen kann mit dem Phasen-Modell von Tuckman & Jensen 1977) beschrieben werden und ist auch bei der Indikatorenentwicklung nachweisbar (siehe z.B. Fietkau & Trénel 2002):

1. *Forming Phase* (**Prozess der Gruppenfindung**)

   Zu Beginn erfolgt zumeist ein vorsichtiges Abtasten und gegenseitiges Kennenlernen, bei dem die Gemeinsamkeiten gesucht und hervorgehoben werden

2. *Storming Phase* (**Kontroverse Auseinandersetzungen**)

   Bei kontroversen Interessenlagen werden in einer zweiten Phase zunehmend die unterschiedlichen Positionen deutlich gemacht und Meinungsverschiedenheiten (auch emotional) ausgetragen.

3. *Norming Phase* (**Rollenfindungsprozess**)

   Mit Klärung der unterschiedlichen Positionen ergeben sich zunehmend spezifische Rollenmuster, deren adäquate Berücksichtigung bei Kompromissen erfolgt.

4. *Performing Phase* (**Entwicklung gemeinsamer Normen**)

   Am Ende werden diese unterschiedlichen Positionen stillschweigend in impliziten Gruppennormen integriert und damit zu Teilen der gemeinsamen Position.

Ein solcher Verlauf ist keineswegs „naturgegeben", sondern bedarf – mit zunehmender Konfliktstärke in erhöhtem Masse – einer erfolgreichen Moderation und Mediation. Insbesondere der Übergang von Phase 3 in Phase 4 kann den Mediator vor Probleme stellen. Im interkulturellen Kontext sind dabei zusätzliche Schwierigkeiten vorauszusehen (siehe dazu als Einführung: Liebe & Gilbert 1996).

Ziel des sozialen Prozesses der Indikatorenentwicklung ist es, einen möglichst von allen Beteiligten getragenen, praktisch durchführbaren und möglichst optimal die gewünschten theoretischen Konstrukte messenden Indikatorenkatalog zu erstellen, der dann im Zuge einer praktischen Erprobung kontinuierlich weiterentwickelt und verbessert werden kann. Dieses Ziel erfordert einer geschickten Zusammenführung der unterschiedlichen Kompetenzen und Interessen der Beteiligten.

Auch die Entwicklung von Indikatoren stellt einen komplexen sozialen Prozess dar, an dem eine Reihe von Personen mit unterschiedlichem Fachwissen und abweichenden Einstellungen gegenüber dem Messverfahren zu beteiligen sind. Damit ein solcher Prozess erfolgreich in ein qualitativ hochwertiges Indikatorentableau münden kann, muss er professionell moderiert werden.

25 AUF DEN RICHTIGEN RAHMEN KOMMT ES AN: DAS KONZEPT DER LOGICAL FRAMEWORKS

Indikatorenentwicklung ist in den seltensten Fällen ein isolierter Prozess. Speziell im Rahmen der Projektplanung wird die Festlegung von Messinstrumenten bereits in den *Zielentwicklungsprozess* integriert, um später die sachgerechte Überprüfung der Zielerreichung zu ermöglichen. Zwischen- und ex-post-Evaluationen müssen sich häufig an diesen Entscheidungen in der Planungsphase orientieren und können auf die mit den damals entwickelten Indikatoren gewonnenen Daten und Informationen zurückgreifen (wobei natürlich nicht gewährleistet ist, dass sämtliche relevanten Einflussfaktoren damals erkannt und entsprechend ausreichend berücksichtigt wurden).

In der Entwicklungszusammenarbeit werden von den unterschiedlichen Trägern als zentrales Planungsinstrument zur Vorbereitung von Entwicklungsprojekten mehr oder weniger stark voneinander abweichende Variationen des „Logical Framework“ verwendet. So sind z.B. die bei der GTZ im Rahmen der Zielorientierten Planung von Projekten (ZOPP) eingesetzten Projektplanungsübersichten (PPÜ) und das von der KfW eingesetzt LogFrame als solche Variationen zu verstehen (vgl. Werner 2000).

Der „*Logical Framework*“ Ansatz ist ein analytisches, präsentationsorientiertes Managementwerkzeug, welches bei der Projektplanung zur systematischen Analyse des Ist-Zustands, zum Entwicklung einer logisch aufeinander aufbauenden und empirisch überprüfbaren Zielhierarchie, zur Identifikation möglicher Risiken und Hindernisse bei der Projektdurchführung, zur komprimierten und standardisierten Zusammenfassung des Projekts sowie zur Festlegung, wie Ergebnisse und Wirkungen am besten evaluiert werden können und welche Maßnahmen und Instrumente hierzu im Rahmen eines Monitoring und Evaluationssystems zu implementieren sind (vgl. AusAID 2000).

Im Wesentlichen geht es darum, während der Planungsphase die folgende Projektmatrix auszufüllen, wobei die einzelnen Schritte hier kurz angesprochen werden sollen.

	Summary (Zusammenfassung)	Indicators (Indikatoren)	Verifications (Quellen der Nachweisbarkeit)	Assumptions (Annahmen)
Goal (Oberziel)				
Purpose (Projektziel)	1	4		3
Outputs (Ergebnisse)				
Activities (Aktivitäten)	2			

Quelle: Centre for International Development and Training 2002

Erster Schritt: Zieldefinition (Oberziele, Projektziele, angestrebte Ergebnisse)

Zunächst werden bei der Erarbeitung des „Logical Frameworks“ die Ziele in der Reihenfolge ihrer Positionierung innerhalb der Zielhierarchie (vom Oberziel über die Projektziele bis zu den konkreten Projektergebnissen) definiert.

Zweiter Schritt: Definition der zur Zielerreichung notwendigen Aktivitäten und Überprüfung der logischen Konsistenz

Nachdem die Ziele „von oben nach unten“ festgelegt wurden, erfolgt nun eine Bestimmung der notwendigen Maßnahmen zur Erzielung der gewünschten Projektergebnisse und eine Überprüfung des logischen Zusammenhangs zwischen den verschiedenen Zielebenen, d.h. es wird die Frage beantwortet, ob die Ergebnisse wirklich aus den Projektzielen und diese wiederum aus den Projektergebnissen ableitbar sind oder eventuell logische Denkfehler im Entwicklungsprozess aufgetreten sind.

Dritter Schritt: Formulierung von Annahmen über Einflussfaktoren (Risiken) auf jeder Zielebene

In der nächsten Entwicklungsphase werden die mit der Durchführung von Maßnahmen sowie den jeweils auf den einzelnen Zielebenen erzielten Ergebnisse und Wirkungen verknüpften Risiken und Widerstände zu ergründen gesucht. Es geht dabei im Wesentlichen um Projektexterne Einflussfaktoren, die nicht oder nur ungenügend durch Steuerungsmaßnahmen in den Griff zu bekommen sind.

Vierter Schritt: Entwicklung von Indikatoren und Definition der Nachweisbarkeitsquellen

Nun erfolgen die Erstellung geeigneter Indikatoren und die Festlegung der Quellen ihrer Nachweisbarkeit. Die Indikatoren orientieren sich dabei sowohl an den definierten Zielen (auf den verschiedenen Ebenen) als auch an den festgestellten potentiellen Risiken der Durchführung, die Maßnahmenwirkungen eventuell negativ beeinflussen können.

Fünfter Schritt: Vorbereitung des Durchführungsbudgets, logische Prüfung des Frameworks anhand theoretischer Anforderungen an das Projektdesign und anhand praktischer Erfahrungen

Der letzte Schritt schließlich umfasst die finanzielle Kalkulation des Durchführungsbudgets sowie eine abschließende logische Überprüfung des Konzepts an den theoretisch festzulegenden Anforderungen an ein (idealtypisches) Projektdesign sowie anhand der aus ähnlichen Projekten gewonnenen praktischen Erfahrungen.

Die Entwicklung von Indikatoren stellt somit im Rahmen des „Logical Frameworks“ lediglich einen – allerdings für das spätere Monitoring und Evaluation System entscheidenden – Baustein dar. Es kann gesagt werden, dass praktisch gleichzeitig zwei unterschiedliche Projekte begonnen werden: das eigentliche inhaltliche Projekt sowie das Projekt der Messung seiner Wirkungen. Durch die korrekte Einbindung der Indikatoren in den Rahmen der Projektplanung sowie dem davon ausgehenden Zwang, Ziele möglichst konkret und messbar (z.B. durch die Definition überprüfbarer Zielvorgaben) zu bestimmen, können die notwendigen Instrumente frühzeitig erstellt und gemessen werden. Die Durchführung einer Ausgangsmessung ist ferner eine Grundvoraussetzung zur korrekten Erfassung von Veränderungen (seien sie durch die Projektmaßnahmen verursacht oder nicht). Wenn das „Logical Framework“ tatsächlich seinem Anspruch gerecht werden kann, so sind alle wesentlichen Störgrößen berücksichtigt und die kausalen Verbindungen zwischen den Maßnahmen und den an-

gestrebten Ergebnissen sowie ihre logisch ableitbaren Wirkungen auf die weiterführenden Zielsetzungen festgelegt worden. Darüber hinaus wurden geeignete Messinstrumente (Indikatoren) für die Beobachtung der Projektentwicklung bestimmt worden. Dieses positive Ergebnis ist allerdings nur unter Einbeziehung einer Vielzahl unterschiedlicher Experten erreichbar – das „Logical Framework“ ist nicht nur ein Management Werkzeug (so sein eigentlicher Anspruch) sondern auch ein schwieriges Verfahren zur Einbindung unterschiedlicher Expertisen in den Planungsprozess.

Das Logical Framework ist ein Planungsrahmen zur Vorbereitung komplexer Vorhaben, wie sie z.B. in der Entwicklungszusammenarbeit erfolgen. Durch dieses Verfahren wird die Einbindung der Indikatorenentwicklung in den Planungsprozess ermöglicht und zugleich die frühzeitige Berücksichtung der messtechnischen Ansprüche von M&E-Systemen an die Zielformulierung sowie die potentiellen externen Störfaktoren erreicht. Nachteile sind der hohe Aufwand sowie der umfangreiche Bedarf an unterschiedlicher Expertise für die korrekte Durchführung des Verfahrens.

26 VORGEHENSWEISE BEI DER INDIKATORENENTWICKLUNG – 10 SCHRITTE ZUM INDIKATOR

Die Vorgehensweise bei der Indikatorenentwicklung lässt sich idealtypisch wie folgt beschrieben (vgl. für ähnliche Muster Shavelson u.a. 1991; Hulliger 2002):

1. *Entwurf des theoretischen Konstrukts und Festlegung von Messzielen*

   In einem ersten Schritt muss geklärt werden, wie das theoretische Konstrukt, welches der Indikator abbilden soll, zu bestimmen ist. Insbesondere ist festzulegen, welche der verschiedenen *Dimensionen* dieses Konstrukts der Indikator abbilden soll (*Definition von Messzielen*). Außerdem müssen die Kriterien bestimmt werden, anhand derer die Güte eines Indikators bewertet werden soll. (Beispiel: Wenn Indikatoren zur Bewertung der Qualität dieses Aufsatzes entwickelt werden sollen, so muss zunächst geklärt werden, was unter „Qualität“ in diesem Zusammenhang zu verstehen ist, welche verschiedenen Facetten der Qualitätsbegriff umfasst und über welchen Teilaspekt jeder einzelne der zu verwendenden Indikatoren Auskunft geben soll).

2. *Ableitung beobachtbarer Sachverhalte und Entwicklung von Korrespondenzregeln*

   Den zweiten Schritt stellt die Operationalisierung dar, d.h. den vorgegebenen Messziele werden beobachtbare Sachverhalte zugeordnet und über geeignete Korrespondenzregeln (Aussagen über die Art und die Stärke des Zusammenhangs) miteinander verknüpft. Darüber hinaus ist zu erarbeiten, wie diese (vermuteten) Zusammenhänge zwischen Indikator und theoretischem Konstrukt überprüft werden können. (Beispiel: Ein Qualitätskriterium ist die Verbreitung des Aufsatzes unter der Zielgruppe. Nun muss geklärt werden, wie die „Verbreitung unter der Zielgruppe“ beobachtet werden kann. Z.B. könnte eine Operationalisierung über die Messung der „downloads“ des Beitrags von der Webseite erfolgen, wobei die Annahme gemacht wird, dass die Zahl der „downloads“ mit der Verbreitung in der Zielgruppe direkt zusammenhängt. An dieser Stelle muss nun ge-

klärt werden, ob und wie stark dieser Indikator mit dem Messziel der Verbreitung vermutlich zusammenhängt und wie diese Vermutung über den Zusammenhang eventuell überprüft werden kann).

3. *Entwurf idealer Indikatoren, die diese Sachverhalte optimal messen würden*

   Auf der Basis des messbaren Gegenstandes werden zunächst ideale Indikatoren entworfen, also Maßzahlen, die diese Sachverhalten theoretisch am besten messen könnten. Gleichzeitig sind die Probleme, die mit einer Umsetzung dieser optimierten Messinstrumente bei der praktischen Umsetzung auftreten würden, systematisch zu erarbeiten und ihre Lösungsmöglichkeiten kritisch zu hinterfragen. (Beispiel: An dieser Stelle ist zu klären, wie ein Indikator theoretisch aussehen müsste, der optimal der Operationalisierung des Messziels entspricht. Ein idealer Indikator für die Verbreitung des Aufsatzes in der Zielgruppe würde genau die Zahl der „downloads“ der Zielgruppenmitglieder von der Internetseite messen - ohne jegliche Doppeltmessung und ohne Berücksichtigung von Personen, die nicht der Zielgruppe angehören).

4. *Ableitung realisierbarer Indikatoren, die trotz erkannter Probleme umsetzbar erscheinen*

   Der nächste Schritt besteht in der Konstruktion realisierbarer Indikatoren, die den gegebenen Sachverhalt und die damit verbundenen Messziele ebenfalls messen könnten. Dabei sollten mehrere Alternativen zur Umsetzung eines idealen Indikators entwickelt und die mit ihnen verbundenen praktischen, aber auch die theoretischen und methodischen Probleme erfasst werden. (Beispiel: Messbar ist z.B. die Zahl der Zugriffe auf die Webseite, die Zahl der „downloads“ insgesamt oder die Häufigkeit der Anfragen aus der Zielgruppe, die sich auf den Aufsatz beziehen. Zu klären ist, inwieweit diese realisierbaren Indikatoren dem idealen entsprechen und in welcher Hinsicht nicht).

5. *Festlegung der Skalen und Instrumentenentwicklung*

   Für die realisierbaren Indikatoren muss nun geklärt werden, auf welcher Skala und mit welcher Genauigkeit sie eine Zuordnung gewährleisten können. Es muss entschieden werden, wie die Daten für den Indikator zu erheben sind, welche Arten von Messinstrumente zur Informationsgewinnung eingesetzt werden können, wie exakt die Messungen voraussichtlich sind, wie die Messinstrumente eventuell zur Verbesserung der Messqualität für den Indikator optimiert werden können und welche statistischen sowie datentechnischen Aufbereitungen zur Gewinnung der Maßzahlen notwendig sind (z.B. zur Erstellung von Indizes oder Verhältniszahlen). Es geht mit anderen Worten um die Definition des Messprozesses und die Bestimmung des notwendigen Aufwands. (Beispiel: der Indikator „Zahl der downloads“ könnte mit einem Counter auf der Webseite relativ einfach erhoben werden. Zu fragen wäre, in welchem Rhythmus eine Erhebung erfolgen sollte, ob eine metrische Darstellung tatsächlich notwendig und sinnvoll ist oder ob vielleicht ein Relativzahl – z.B. Verhältnis der downloads des Aufsatzes zu allen downloads innerhalb einer Zeitspanne – aufschlussreicher wäre. Zusätzlich ist der Aufwand – personell, finanziell und technisch – zur Erhebung dieser Daten in Betracht zu ziehen).

6. *Vergleich verschiedener Indikatoralternativen*

   Anschließend werden die verschiedenen Indikatoralternativen sowohl miteinander als auch mit dem idealen Indikator hinsichtlich ihres Bezugs zum theoretischen Konstrukt (Realisierung des Messziels), der zu erwartenden Messqualität (erreichbare Validität und

Reliabilität) sowie der Praktikabilität (notwendiger Aufwand und Durchführbarkeit der Messung) verglichen und bewertet. Möglichkeiten zur Optimierung der besten Kandidaten sollten ebenfalls diskutiert werden. (Beispiel: Bei den drei oben genannten Indikatoren ist abzuwägen, ob die einfach und ohne viel Aufwand zu erhebende Zugriffszahl ausreichend ist oder eher die Aufwendigere Erfassung der downloads selbst in Angriff genommen werden muss. Die Rückmeldungen mögen genauere Informationen über die Nutzung des Beitrags durch die Zielgruppe geben, sind aber aufgrund der Selbstselektion weniger exakt als die Zugriffzahlen usw. Am Ende entsteht so eine Liste der Pro's und Con's einzelner Indikatoren als Grundlage für eine Auswahlentscheidung).

7. *Datenerhebung und Berechnung der Indikatoren (Pretest)*

   Da auch bei der Indikatorenentwicklung nicht alle Probleme im Vorfeld vorausgesehen werden können, erfolgt nun eine praktische Erprobung der aussichtsreichsten Alternativen. Wichtig ist in dieser Erprobungsphase, dass nicht nur der in der Diskussion als beste Lösung identifizierte Indikator, sondern auch eine Reihe möglichst gleichwertiger Alternativen „auf den Prüfstand" gelangen. Während der Datenerhebung und der Berechnung der Indikatoren müssen sämtliche auftretenden Schwierigkeiten und die hierfür gewählten Lösungen dokumentiert werden. (Beispiel: In einer Testphase werden alle drei Indikatoren erhoben und hinsichtlich ihrer Veränderungen und ihres gegenseitigen Zusammenhangs beobachtet. Im konkreten Beispiel können unterschiedliche Nutzerverhalten z.B. sogar simuliert und die gewonnenen Daten nach einem solchen „offline" Test ausgewertet werden).

8. *Evaluation des Messverhaltens und der Schwierigkeiten bei der Datenerhebung*

   Die während der Erhebung und Berechnung von Indikatoren aufgetretenen Probleme und gewählten Lösungen müssen schließlich einer Bewertung unterzogen werden, die Aufschluss über die Güte der einzelnen Maßzahl sowohl bezüglich des Messziels als auch im Vergleich zu den untersuchten Alternativen Aufschluss gibt. Es sollte außerdem diskutiert werden, inwieweit eine Optimierung des Vorgehens möglich ist und wie erkannte Schwachstellen zu beseitigen sind. (Beispiel: Im Pretest kann sich z.B. herausstellen, dass eine Erfassung der Zugriffsdaten aufgrund häufiger Abstürze des Computersystems nur unvollständig oder fehlerhaft gelingt. Zu Fragen ist dann, ob geeignete Maßnahmen zur Stabilisierung des Systems möglich sind oder Indikator wegen dieser unwägbaren technischen Probleme sowie den damit verbundenen Einschränkungen der Datenqualität verworfen werden muss).

9. *Ergebnisinterpretation und Evaluation des Vermittlungsprozesses*

   Die Bewertung von Indikatoren sollte nicht allein auf die messtechnischen und praktischen Fragen beschränkt bleiben, sondern unbedingt auch den Vermittlungsprozess der Ergebnisse an die stakeholder sowie deren Interpretation der Daten berücksichtigen. Untersucht werden sollte die Frage, ob die Resultate der Indikatoren überhaupt verständlich sind und ob sie von den stakeholdern entsprechend der Messziele interpretiert werden. Wenn Missverständnisse auftreten, müssen Strategien zu deren Vermeidung entwickelt, Interpretationshilfen erstellt oder Vereinfachungen der Maßzahlen in Betracht gezogen werden. (Beispiel: Die ermittelten Zahlenwerte über den download des Beitrags können eventuell wegen fehlender Vergleichswerte oder aufgrund der Verwendung technischer Begriffe für bestimmte Personengruppen schlecht verständlich sein. Dies ist

in dieser Phase festzustellen und z.B. durch Bereitstellung von Vergleichswerten und einem Glossar entgegenzuarbeiten).

10. *Entscheidung über die Beibehaltung, Weiterentwicklung oder die Ablehnung eines Indikators*

    Im letzten Schritt ist schließlich über die Beibehaltung eines oder einer Reihe von Indikatoren zur Messung des angestrebten Konstrukts zu treffen. Dabei sollten auch die Weiterentwicklungsmöglichkeiten eines Indikators mit berücksichtigt werden. Ein Indikator sollte nur dann entgültig aus dem Tableau entfernt werden, wenn sich seine Nützlichkeit als zu begrenzt erwiesen hat oder bessere Alternativen vorliegen. Sollte der Indikator bereits über einen längeren Zeitraum erhoben worden sein, so sind gleichzeitig Umrechnungsroutinen auf die neuen Maßzahlen zu entwickeln, damit eine dauerhafte Vergleichbarkeit mit den früheren Messwerten gewährleistet werden kann. Eine andere Möglichkeit ist die Eliminierung eines Indikators (insbesondere bei der Messung der Zielerreichung), weil eine Sättigungsgrenze erreicht wurde. Z.B. wurde in der Bundesrepublik jahrelang der Indikator „Anteil der Wohnungen mit Innen WC“ erhoben, um das wohnungspolitische Ziel eines vollkommenen Anschlusses aller Haushalte an das Abwassernetz zu überprüfen. Nachdem mit einem dauerhaft stabilen Wert von 99% die Sättigungsgrenze erreicht wurde, sollte dieser Indikator aus dem Tableau gestrichen werden (durch die deutsche Einheit erhielt der Indikator dann allerdings wieder Bedeutung und wird weiterhin erhoben). Werden Zielvorgaben dauerhaft erfüllt, so werden in der Regel neue Zielwerte mit anspruchsvolleren Vorgaben gesetzt (z.B. bei der Abgasverordnung für Automobile ersetzt die anspruchsvollere EURO 3 Regelung die mittlerweile standardmäßig umgesetzte EURO 2 Regelung). Ein Austausch der Indikatoren findet normalerweise erst dann statt, wenn eine weitere positive Veränderung durch diese Maßzahl nicht mehr zu erfassen ist und ihr Wegfall aufgrund der stabilen Überschreitung der Zielwerte keinen Informationsverlust darstellt.

Die Indikatorenentwicklung ist ein „Projekt im Projekt“ und bedarf einer genauso gewissenhaften Schrittweisen Vorgehensweise bei der Planung, Implementierung, Evaluierung und Nachbetreuung der Indikatoren. Die verschiedenen Anforderungsebene (theoretisch, methodisch, praktisch und politisch) müssen stetig im Auge behalten und ausreichend berücksichtigt werden.

27 MESSEN UM DES MESSENSWILLEN? KRITERIEN DER INDIKATORENAUSWAHL

Prinzipiell ist bei der Zusammenstellung von Indikatoren natürlich eine Maximallösung immer möglich, indem alle denkbaren Maßzahlen für ein theoretisches Konstrukt erhoben und schließlich in umfassenden Reports dargestellt werden. Sonderlich sinnvoll ist eine solche Vorgehensweise aber aus verschiedenen Gründen nicht: da jede Messung einen gewissen Aufwand erfordert, steigt bei einer solchen Vorgehensweise der Messaufwand extrem an, was nicht nur die Kosten erhöht, sondern gleichzeitig auch die Akzeptanz bei den mit der Datenerhebung betrauten Personen senken kann. Dieser Akzeptanzverlust der Messungen

wird zusätzlich noch verstärkt durch die Wahrnehmung, dass ein solch umfangreiches Indikatorenbündel und die hierdurch bedingten umfangreichen Berichte nur rudimentär und hochgradig selektiv genutzt werden. Eine Vielzahl von Indikatoren wird dementsprechend nie oder nur sehr selten in die Steuerungsentscheidungen mit einfließen.

Dementsprechend ist ein Anspruch an ein Indikatorentableau, dass es möglichst effektiv und effizient die benötigten Informationen bereitstellen kann. Dies führt zur Frage, wie eine *Indikatorenselektion* geschehen kann und welchen *Kriterien* sie dabei zu folgen hat. Erschwert wird die Auswahl geeigneter Indikatoren aus einer Reihe von Vorschläge zum einen durch die unterschiedlichen Anforderungen theoretischer, methodischer, praktischer und politischer Art sowie durch die Beteiligung von unterschiedlichen Experten mit z.T. kontroversen Einstellungen gegenüber der Notwendigkeit und den Möglichkeiten der Messung des theoretischen Konstrukts.

Eine Möglichkeit zur Versachlichung der Indikatorenauswahl ist die Erstellung einer *Indikatoren Checkliste*, anhand derer die einzelnen zur Diskussion stehenden Indikatoren durch die Beteiligten an der Indikatorenentwicklung bewertet werden können und somit eine Versachlichung der Diskussion zu erzielen ist. Die Checklisten können – wie im folgenden Beispiel - in Form eines Fragebogens gehalten sein und durch metrische Bewertungsskalen eine Abstufung zwischen den verschiedenen Alternativen ermöglichen Sie lassen dann die Ermittlung von Mittelwerten und Streuungsmaße (als Indikator für abweichende Auffassungen über bestimmte Aspekte eines Indikators zwischen den verschiedenen Beteiligtengruppen) sowie durch die Erstellung von Profilen den Vergleich zwischen einzelnen Indikatoren hinsichtlich ihrer wahrgenommenen Stärken und Schwächen zu.

Eine solche Indikatoren Checkliste ist wiederum nichts anderes als ein Indikatorentableau zur Bewertung der Qualität von Indikatoren und soll als Entscheidungsgrundlage dienen. Dementsprechend kann eine solche Checkliste die Interpretation und Diskussion der Ergebnisse nicht durch eine „mechanische“ Auswahl (etwa nach dem höchsten „Gesamtwert“) ersetzen. Sie ist – wie bei allen anderen Indikatoren – lediglich der zusammenfassende und bilanzierende Ausgangspunkt für den eigentlichen Entscheidungsprozess. Sie ermöglicht einen Ordnung der Indikatoren hinsichtlich verschiedener Merkmale und einen schnellen Überblick zu deren Vor- und Nachteilen.

Das folgende Beispiel orientiert sich an den hier vorgestellten Anforderungen (theoretischer, methodischer, praktischer und politischer Art) sowie den verschiedenen Subdimensionen, die in diesem Beitrag aufgeführt worden sind. Selbstverständlich ist eine Erweiterung oder Modifikation dieser Liste denkbar.

Zur Unterstützung der Auswahlentscheidung über potentielle Indikatorenkandidaten ist eine Ermittelung ihrer Vor- und Nachteil sowie dieser Einschätzungen durch verschiedene Beteiligtengruppen an der Indikatorenentwicklung die Erstellung einer Indikatorencheckliste sinnvoll. Diese Checkliste muss alle Anforderungen und Bewertungsaspekte für die Auswahl von Indikatoren umfassen.

Indikatoren Checkliste		
I. Theoretische Anforderungen		
	Welches **theoretische Konstrukt** soll der Indikator abbilden?	.....
I-1	Inwieweit ist dieses theoretische Konstrukt für die Fragestellung/das Projekt/die Aufgabe relevant?	**10 9 8 7 6 5 4 3 2 1 0**
	Welche **Dimensionen** dieses Konstruktes soll der Indikator abbilden?	.....
I-2	Inwieweit ist diese Dimension für die Beschreibung des Konstrukts relevant?	**10 9 8 7 6 5 4 3 2 1 0**
	Durch welches beobachtbare und durch den Indikator messbare Objekt wird diese Dimension des Konstrukts **operationalisiert**?	.....
I-3	Inwieweit ist der Zusammenhang zwischen dem beobachtbaren Objekt und dem theoretischen Konstrukt plausibel?	**10 9 8 7 6 5 4 3 2 1 0**
	Durch welche Maßnahmen ist diese theoretische Annahme des Zusammenhangs zwischen Objekt und Konstrukt eventuell **prüfbar**?	.....
I-4	Inwieweit ist eine Prüfung der messtheoretischen Annahme des Zusammenhangs zwischen Objekt und Konstrukt möglich?	**10 9 8 7 6 5 4 3 2 1 0**
I-5	Inwieweit bildet der Indikator voraussichtlich das Konstrukt ab?	**10 9 8 7 6 5 4 3 2 1 0**
Gesamtbewertung der theoretische Anforderungen		**10 9 8 7 6 5 4 3 2 1 0**

Legende zu den Skalenwerten: 10 heißt voll und ganz, 0 heißt überhaupt nicht.

II. Methodische Anforderungen		
	Welche **Instrumente** zur Messung des Indikators können verwendet werden?	.....
II-1	Inwieweit entspricht die durch diese Instrumente zu erreichende Messqualität den Erfordernissen?	10 9 8 7 6 5 4 3 2 1 0
	Welche Möglichkeiten zur **Bewertung von Güte und Zuverlässigkeit** der Messung sind potentiell denkbar?	.....
II-2	Inwieweit entspricht die Güte der Messung des Indikators der Qualität eines denkbaren optimalen Indikator?	10 9 8 7 6 5 4 3 2 1 0
II-3	Inwieweit entspricht die Güte der Messung des Indikators der Qualität bewährter Vergleichsindikatoren?	10 9 8 7 6 5 4 3 2 1 0
II-4	Inwieweit ergeben wiederholte Messungen voraussichtlich für den Indikator immer wieder die selben Werte?	10 9 8 7 6 5 4 3 2 1 0
	Welche Möglichkeiten zur **Erprobung des Messverhaltens** des Indikators gibt es?	.....
II-5	Inwieweit können die Erwartungen an die Genauigkeit der Messergebnisse überprüft werden?	10 9 8 7 6 5 4 3 2 1 0
Gesamtbewertung der methodischen Anforderungen		10 9 8 7 6 5 4 3 2 1 0

Legende zu den Skalenwerten: 10 heißt voll und ganz, 0 heißt überhaupt nicht.

III. Praktische Anforderungen		
	Welche **Ressourcen** stehen zur Datenerhebung für diesen Indikator zur Verfügung?	.....
III-1	Inwieweit reichen voraussichtlich die vorhandenen finanziellen Mittel für eine Umsetzung der Messung aus?	10 9 8 7 6 5 4 3 2 1 0
III-2	Inwieweit reichen voraussichtlich die technischen Möglichkeiten für eine Umsetzung der Messung aus?	10 9 8 7 6 5 4 3 2 1 0
III-3	Inwieweit reichen voraussichtlich die Personalkapazität (Zeit und Wissen) für eine Umsetzung der Messung aus?	10 9 8 7 6 5 4 3 2 1 0
	Welche **Anforderungen an die Ressourcen** stellt die Messung des Indikators an die Durchführungsorganisation?	.....
III-4	Inwieweit ist die Belastung der Durchführungsorganisation durch die Datenerhebung zumutbar?	10 9 8 7 6 5 4 3 2 1 0
	Welche **Möglichkeiten zur Kapazitätserhöhung** für die Messung des Indikators sind denkbar?	.....
III-5	Inwieweit ist eine effektive und effiziente Durchführung der Messung vorstellbar?	10 9 8 7 6 5 4 3 2 1 0
Gesamtbewertung der praktischen Anforderungen		10 9 8 7 6 5 4 3 2 1 0

Legende zu den Skalenwerten: 10 heißt voll und ganz, 0 heißt überhaupt nicht.

IV. Politische Anforderungen		
	Welche **Stakeholdergruppen** müssen bei der Interpretation des Indikators berücksichtigt werden?	.....
	Welche **Verständigungsprobleme** kann es bei der Ergebnisvermittlung bezüglich dieses Indikators geben?	.....
IV-1	Inwieweit kann von einem einheitlichen Verständnis der Messergebnisse ausgegangen werden?	10 9 8 7 6 5 4 3 2 1 0
IV-2	Inwieweit kann von einer einheitlichen Interpretation der Messergebnisse ausgegangen werden?	10 9 8 7 6 5 4 3 2 1 0
	Welche **Widerstände** sind von den einzelnen Stakeholdergruppen gegen den Indikator zu erwarten?	.....
IV-3	Inwieweit ist mit einer allgemeinen Akzeptanz dieses Indikators als geeignetes Messinstrument zu rechnen?	10 9 8 7 6 5 4 3 2 1 0
IV-4	Inwieweit kann vermieden werden, dass das Verhalten von Stakeholdergruppen die Messqualität des Indikators negativ beeinflusst?	10 9 8 7 6 5 4 3 2 1 0
	Welche Möglichkeiten zur **Kontrolle manipulativer Eingriffe** einzelner Stakeholdergruppen sind denkbar?	.....
IV-5	Inwieweit kann eine Unabhängigkeit der Messung dieses Indikators gewährleistet werden?	10 9 8 7 6 5 4 3 2 1 0
Gesamtbewertung der politischen Anforderungen		10 9 8 7 6 5 4 3 2 1 0

Legende zu den Skalenwerten: 10 heißt voll und ganz, 0 heißt überhaupt nicht.

28 WENIGER IST HÄUFIG MEHR…

Eine der am häufigsten gestellten Fragen bei Einführungsveranstaltungen zur Indikatorenentwicklung ist diejenige nach der *Anzahl der Indikatoren*, die für eine wissenschaftlich exakte Messung erforderlich sind. Die Antwort hierauf ist genauso simpel wie unbefriedigend: am besten ist es aus Effizienzgründen natürlich, wenn nur ein einziger präzis messender Indikator mit einem klar verständlichen und eindeutig zu interpretierenden Ergebnis einzusetzen ist. Der Pferdefuss an der Sache: wenn es dieses Messinstrument gäbe, bräuchte ich keinen „Indikator", denn ich würde ja den Sachverhalt zufriedenstellend direkt messen. Da Indikatoren immer (mehr oder weniger) unbefriedigende „Krücken" zur Messung nicht-messbarer Dinge darstellen, ist die Zahl der Indikatoren eng verknüpft mit dem Anspruch an die Messgenauigkeit dieses (unmessbaren) theoretischen Konstruktes. *Je schlechter die von mir ausgewählten Indikatoren eine Dimension eines bestimmten Tatbestandes messen, umso mehr verschiedene Indikatoren für dieses Konstrukts muss ich zur Erreichung einer vertretbaren Messgenauigkeit verwenden*.

Dementsprechend gilt natürlich auch umgekehrt: *je besser ein Indikator das theoretische Konstrukt abbildet umso weniger notwendig ist die Aufnahme eines zweiten Indikators zur Absicherung („Validierung") der Ergebnisse*. Da jede Messung einen bestimmten Aufwand bedeutet, ist also bei der Indikatorenentwicklung dieser zusätzliche Messaufwand immer in Relation zu dem Zusatzaufwand für eine Präzisierung der Messung bestehender Instrumente zu sehen und in der Regel wird die Verbesserung der Messgenauigkeit der Konstruktion eines neuen Indikators und der Implementation von Messprozeduren zu seiner Erhebung vorzuziehen sein. Beim Einsatz von Indikatoren gilt sicher nicht, dass „viel auch viel hilft" – im Gegenteil: Indikatoren sollen uns einen schnellen Überblick zur Lage ermöglichen und schon zwei verschiedene, eventuell unterschiedliche Tendenzen anzeigende Maßzahlen können mehr verwirren als unterstützen.

Auf der anderen Seite ist es aber gefährlich, Maßzahlen unterschiedlicher Güte gleichberechtigt nebeneinander bei der Interpretation zu verwenden. Wenn ein bestimmtes theoretisches Konstrukt schlechter messbar ist als ein anderes werden häufig mehrere Indikatoren zu seiner Abbildung in vergleichbarer Güte benötigt. Damit nun nicht ein „guter" Indikator mit mehreren „schlechten" verglichen werden muss, ist es möglich, aus einer Reihe von Indikatoren einen *Index* zu bilden. Voraussetzung für die Indexbildung ist die *Eindimensionalität der Messung*: sämtliche Indikatoren müssen einen gemeinsamen Aspekt des Konstrukts abbilden. Die Eindimensionalität von Indikatoren lässt sich durch das Verfahren der Faktoranalyse (vgl. Kim/Mueller 1987) überprüfen – wenn diese Zahlenwerte mit einem metrischen Skalenniveau besitzen. Die über ein solches Verfahren erzeugten *Faktorvariablen* stellen einen gemeinsamen Indikator für den theoretischen Sachverhalt dar und gleichen die Messfehler der einzelnen in die Faktorenbildung einbezogener Indikatoren aus. Hierfür ist die Voraussetzung (und so bildet die Faktoranalyse auch ihre Faktoren), dass diese Indikatoren untereinander und mit der Faktorvariablen hoch korrelieren. Indikatoren, die diesen Anspruch nicht erfüllen, bilden entweder eine andere Dimension des theoretischen Konstrukts ab (auch dieses kann mit der Faktoranalyse festgestellt werden) oder sie sind schlicht aus messtechnischen Gründen für die Abbildung dieses Konstrukts ungeeignet. Die meisten psychologischen Skalen zur Messung komplexer Konstrukte wie „Glück" oder „Intelligenz" basieren auf solchen faktoranalytisch erstellten komprimierten Indikatoren. Die Validität der Messung wird in diesen Fällen durch Parallelmessung verbessert.

Die Zahl der verwendeten Indikatoren hängt von der gewünschten Messqualität ab. Als Ziel sollten so wenig wie möglich „gute“ Indikatoren verwendet werden. Durch die Zusammenfassung „schlechter“ Indikatoren in einen Index kann die Messqualität aufgrund der Nivellierung der Messfehler verbessert werden.

29 … MANCHMAL ABER AUCH NICHT

Mehrere Indikatoren werden aber nicht nur ausschließlich aufgrund der mangelhaften Qualität der einzelnen Maßzahlen benötigt. Häufig ist die Abbildung eines komplexen theoretischen Konstrukts mit einem Indikator aufgrund dessen *Mehrdimensionalität* nicht praktikabel. So ist es z.B. nicht möglich, einen einzigen sinnvollen „Qualitätsindikator“ zu entwickeln, da „Qualität“ ein subjektiver Bewertungsbegriff ist, der verschiedene Aspekte umfassen kann. Die Produktqualität kann z.B. anhand technischer Kriterien (der Funktionalität des Produktes), anhand der Nutzung durch die Kunden (dem Einsatz des Produktes) oder anhand des öffentlichen Images (dem Ansehen des Produktes) gemessen werden.

Es ist selbstverständlich technisch auch möglich, alle drei Aspekte in einen gemeinsamen Index zu fassen – dies setzt aber eine perfekte Korrelation der einzelnen Elemente untereinander voraus. Wenn stattdessen die einzelnen Teile (z.B. das Image und die technische Funktionalität) im Widerspruch stehen, kann dies zu Verzerrungen der Messwerte und fehlerhaften Interpretation aufgrund der „Pseudo-Eindimensionalität“ des Indikators führen. In dieser Hinsicht können z.B. Verfahren zum Hochschulranking kritisiert werden, da sie häufig künstlich eine einheitliche Skala produzieren, in der inhaltlich unterschiedliche und sich z.T. widersprechende Teilbereiche unberechtigterweise vereint werden. Die Bewertung der Qualität einer Hochschule durch die Studierenden basiert auf vollständig anderen Kriterien als die durch die Professorinnen und Professoren, weshalb deren Ergebnisse häufig im Widerspruch zueinander stehen. Aus diesen Gründen lehnt z.B. das CHE die Erstellung eines Gesamtindexes und ein Ranking der Hochschulen anhand einer komprimierten Skala ab (Berghoff u.a. 2003: 4ff.).

In allen Fällen, in denen ein Konstrukt mehrdimensional ist, reicht ein einziger Indikator für seine korrekte Abbildung selbst dann nicht aus, wenn dieser perfekt messen würde und die theoretischen sowie praktischen Anforderungen optimal erfüllen könnte. *Zur Abbildung mehrdimensionaler Konstrukte werden mindestens genau so viele Indikatoren benötigt wie es unterschiedliche und voneinander unabhängige Dimensionen gibt*.

Ein *Gesamtindikator* müsste zudem in korrekter Weise die Gewichtung der einzelnen Dimensionen widerspiegeln – häufig eine theoretisch nicht zu klärende und deshalb praktisch nicht umsetzbare Forderung. Die meisten Indizes verknüpfen die einzelnen Indikatoren gleichgewichtig durch einfache Rechenregeln (zumeist Additionen oder Multiplikationen). Allein durch diese Operation steht ein mehr oder weniger großer Messfehler, der zu Verzerrungen der Ergebnisse führt und der selbst wiederum bei der Bewertung der Indikatoren und ihrer Messqualität berücksichtigt werden muss. *Ein Gesamtindikator erfordert eine perfekte Annahme über die Rechenoperation, die für eine Verknüpfung der einzelnen Indikatoren zur korrekten Abbildung des theoretischen Konstrukts notwendig ist.*

Dementsprechend ist es in der Regel angemessener, einen Set von Indikatoren zu verwenden, der mehrere unterschiedliche, theoretisch aber präzise abgegrenzte Dimensionen eines theoretischen Konstrukts abbildet (vgl. Sullivan/Feldman 1979). Eine solche Sammlung einzelner Indikatoren für einen bestimmten Teilbereich wird *Indikatorentableu* genannt.

Mehrdimensionale theoretische Konstrukte können nur dann durch einen einzigen Indikator adäquat abgebildet werden, wenn die Rechenoperationen zur Verknüpfung der einzelnen Dimensionen bekannt und praktisch realisierbar sind. Da dies in der Praxis selten der Fall ist, müssen bei mehrdimensionalen theoretischen Konstrukten mindestens genau so viele Indikatoren entwickelt werden wie es unterschiedliche Dimensionen des Konstrukts gibt. Eine unzulässige Verknüpfung einzelner Indikatoren führt zu zusätzlichen Messfehlern und vermindert dadurch die Qualität des Indexes.

30 DIE UNENDLICHE GESCHICHTE TEIL II: MIT INDIKATOREN LEBEN

„The development of even a single indicator is an iterative process that takes about ten years to complete“ (Shavelson et. al. 1991).

Diese Aussage ist leider zu optimistisch – selbst nach zehn Jahren ist kein Indikator wirklich vollendet. Da ein perfekter Indikator undenkbar ist (er würde das unmessbare Konstrukt messen) und die Erhebungsbedingungen über die Zeit niemals vollständig konstant bleiben, ist eine stetige Überwachung, Anpassung und Weiterentwicklung der Indikatoren mit dem Ziel einer Verbesserung der Messqualität notwendig.

Dem steht allerdings bei Langzeitbeobachtungen (z.B. im Rahmen von Monitoring-Systemen) das Ziel entgegen, möglichst dauerhaft vergleichbare Werte zu erhalten. Dies bedeutet, dass Eingriffe in die Messprozedur – selbst wenn Verbesserungen möglich sind – vermieden werden sollten, damit es nicht aufgrund der veränderten Messung zu zusätzlichen Messfehlern kommt. Soweit es sich „nur“ um eine Verbesserung handeln würde, wäre dies unproblematisch (in diesem Fall würde lediglich die Streuung der gemessenen Werte um den „wahren Wert“ geringer werden). In der Regel sind allerdings die nicht-intendierten Folgen einer Veränderung des Messens unabsehbar, d.h. indem wir den einen Fehler beseitigen können wir einen anderen produzieren. Indikatoren sind dementsprechend in zweifacher Hinsicht eine „unendliche Geschichte“: *Indikatoren müssen zum einen dauerhaft den gegebenen Rahmenbedingungen angepasst und in ihrer Messqualität optimiert werden. Zum anderen dienen Indikatoren der Dauerbeobachtung und müssen dem Anspruch der Reliabilität genügen, weshalb sie fortlaufend auf die selbe Weise gemessen werden sollten.*

Ähnliches gilt auch, wenn der Blick nicht nur auf einen einzelnen Indikator, sondern auf ein ganzes Tableau gerichtet wird. Die Aufnahme von neuen, das selbe Konstrukt besser als die vorhandenen Instrumente messenden Indikatoren sollte immer möglich sein. Da kein Indikator ein Konstrukt perfekt misst und die Entwicklung von Indikatoren letztendlich ein kreativer Akt ist, kann jederzeit ein besserer Indikator „erfunden“ werden. Auf der anderen Seite ist aus den Gründen der Dauerbeobachtung die Entfernung alter Indikatoren aus dem Tableau wegen dem Verlust der langfristigen Vergleichbarkeit und den damit gegebenen Möglichkei-

ten, Entwicklungen zu verfolgen, problematisch. *Dementsprechend ist bei der Aufnahme neuer Indikatoren ein progressives Vorgehen, bei der Entfernung alter Indikatoren dagegen eine konservative Einstellung zur Pflege von Indikatorentableaus erforderlich*.

Dies führt aber mittelfristig dazu, dass ein Indikatorentableau stetig wächst und der damit verbundene Informationsgewinn in ein immer schlechteres Verhältnis zu dem hierfür notwendigen Erhebungsaufwand gerät (die neuen Indikatoren messen schließlich keine neuen Tatbestände, sondern ersetzen lediglich vorhandene Messungen, die aber aus Gründen der Vergleichbarkeit trotzdem weitergeführt werden). Das hierdurch entstehende Dilemma ist nur zu lösen, wenn es gelingt, die Zahl der Indikatoren wieder zu reduzieren. *Generell gilt, dass ein Indikator dann aus einem Tableau entfernt werden kann, wenn hierdurch kein wesentlicher Informationsverlust entsteht*. Der Informationsgehalt der „alten" Indikatoren wächst aufgrund der Vergleichbarkeit zu früheren Messungen mit jedem Messintervall an. Dieser Informationsgewinn kann allerdings durch neue Indikatoren kompensiert werden, wenn eine *Transformationsregel* zu erstellen ist, d.h. eine Umrechnung von alten auf neue Werte oder eine gute Schätzung der neuen Werte für vergangene Zeiträume ermöglicht wird.

Indikatoren sind niemals „fertig", sie müssen ständig hinsichtlich ihrer Messqualität optimiert werden. Diesem Anspruch steht aber das Ziel der langfristigen Vergleichbarkeit entgegen. Beim Austausch von Indikatoren muss dementsprechend der hierdurch bedingte Informationsverlust mit dem Informationsgewinn bilanziert werden.

31 DIE ENTWICKLUNG VON INDIKATOREN: EINE DRITTE ZUSAMMENFASSUNG

Die Indikatorenentwicklung stellt a) einen sozialen Prozess dar, der b) nicht unabhängig von dem Zweck der Indikatorennutzung erfolgen darf, sowohl bei der c) Erstellung als auch d) der Auswahl von Indikatoren eine systematische und schrittweise Vorgehensweise erfordert und eine e) Minimierung der Indikatorenzahl unter f) Berücksichtigung der für eine korrekte Abbildung des theoretischen Konstrukts mindestens notwendigen Größe des Indikatorentableaus g) dauerhaft unter Berücksichtigung der gegebenen Erhebungsbedingungen und des erreichten Informationsgehalts gewährleisten soll. Selbstverständlich variiert der Aufwand, der zur Entwicklung von Indikatoren notwendig ist, mit den Anforderungen, die an diese gestellt werden. Während z.B. beim Aufbau eines landesweiten statistischen Systems eine Vielzahl von Experten dauerhaft beschäftigt sind und eine entsprechende Institutionalisierung der Indikatorenentwicklung im Rahmen von staatlichen Großorganisationen erfolgt, ist dies bei der Entwicklung von Indikatoren zur Durchführung einer Zwischenevaluierung während eines Kurzzeitaufenthalts vor Ort sicher nicht möglich. Aber: selbst in diesem Fall ist eine Offenlegung der Bewertungskriterien und der für ihre Messung eingesetzten Indikatoren sowohl gegenüber dem Auftraggeber als auch gegenüber den Betroffenen mehr als nur ein Akt der Höfflichkeit, sondern ein für die Akzeptanz und Nutzung der Evaluationsergebnisse zwingend notwendige Voraussetzung.

Der *soziale Prozess* der Indikatorenentwicklung erfordert die Einbindung unterschiedlichster Fachexpertise und Interessenlagen, die insbesondere bei längerfristigen Vorhaben (z.B. dem Aufbau eines Monitoringsystems) schnell zu einer komplizierten und konfliktträchtigen Ange-

legenheit werden kann. An Evaluatorinnen und Evaluatoren wird dabei die Anforderung einer zielführenden Moderation dieses Aushandlungsprozesses gestellt. Gleichzeitig ist zu beachten, dass die Indikatorenentwicklung kein isoliertes Vorhaben, sondern in der Regel *in einen übergeordneten Managementprozess einzubinden ist* (wie z.B. als Teilelement des Logical Frameworks in der Planungsphase darstellt). Da sich in diesem Rahmen der Verwertungszusammenhang der Indikatorenergebnisse konstituiert, ist eine erfolgreiche Vermittlung zwischen Instrumenten und Erhebungsverfahren einerseits und der Interpretation und Nutzung der erzielten Ergebnisse von großer Bedeutung.

Die wahrgenommene Qualität der erstellten Indikatoren ist die zentrale Voraussetzung für ein (erst herzustellendes) Vertrauen in die verwendeten Messinstrumente. Zu den „vertrauensbildenden" Maßnahmen gehört eine sorgfältige, schrittweise, systematische und wissenschaftlich korrekte Vorgehensweise bei der *Erstellung und Auswahl von Indikatoren*. Je besser die verschiedenen Beteiligtengruppen in diesen Prozess eingebunden sind und je transparenter die Entscheidung für und gegen bestimmte Indikatoren begründet werden kann, desto größer die Bereitschaft, sich korrekt an der Datenerhebung zu beteiligen und desto höher schließlich die Glaubwürdigkeit der durch diese Indikatoren erzielten Ergebnisse. Die Zielsetzung beim Aufbau eines Indikatorentableaus muss es sein, *eine möglichst geringe Zahl an Indikatoren die möglichst exakt das theoretische Konstrukt abbilden* für die Messung auszuwählen. Die zusätzliche Aufnahme weiterer Indikatoren begründet sich stets durch den dadurch *zu erzielenden Informationsgewinn*, der in einem positiven Verhältnis zu dem hierdurch *benötigten Erhebungsaufwand* stehen muss. *Insbesondere bei mehrdimensionalen theoretischen Konstrukten oder bei zu erwartender schwacher Messqualität ist die Aufnahme weiterer Indikatoren unabdingbar notwendig*.

Die in der Natur von Indikatoren liegenden Schwachpunkte der Messung (Indikatoren messen die eigentlich angestrebten Sachverhalte lediglich indirekt) führen dazu, dass ein einmal erstelltes Indikatorentableau keinesfalls „perfekt" sein kann. Da sich zudem die Erhebungsbedingungen stetig verändern und hierdurch Messergebnisse verfälschen können, ist eine *kontinuierliche Beobachtung und Weiterentwicklung* von Indikatorentableaus nötig.

32 LESSONS LEARNT

Die wichtigsten Aspekte der einzelnen Abschnitte dieses Beitrags sind im folgenden nochmals im Überblick zusammengestellt. Dabei entsprechen die Ordnungs- den Kapitelnummern.

1. Indikatoren sind Kenngrößen, die über einen festgelegten, nicht oder nur sehr schwer messbaren Tatbestand Auskunft geben sollen. (**Definition**)
2. Indikatoren stellen den Ausgangspunkt und nicht den Endpunkt einer Evaluierung dar: sie geben Anhaltspunkte für eine Bewertung, ersetzen diese aber nicht. (**Bewertung**)
3. Wirkungsindikatoren beziehen sich auf die einzelnen Ebenen des Zielssystems und können dementsprechend mit Zielindikatoren, aber auch mit Input- und Bedarfsindikatoren in Bezug gesetzt werden. (**Wirkungsindikatoren**)

4. Durch eine möglichst einheitliche Formulierung von Indikatoren mit einer gemeinsamen Vergleichsdimension wird eine Bewertung von Sachverhalten durch Ergebnisvergleich der Indikatoren ermöglicht. (**Einheitliche Formulierung**)
5. Indikatoren klassifizieren Merkmale anhand eines Ordnungsprinzips, welches künstlich festgelegt werden muss. (**Klassifizierung**)
6. Indikatoren sind Vergleichsinstrumente, die Unterschiede und Gemeinsamkeiten der beobachteten Objekte auf einer gemeinsamen Skala mit mindestens zwei Skalenwerten durch eindeutige Zuordnung jedes einzelnen Objekts anzeigen. (**Vergleich**)
7. Ob zwei Objekte einer gemeinsamen oder zwei unterschiedlichen Klassen zugeordnet werden ist eine Frage des gewählten Aggregationsniveaus sowie der Differenziertheit des gewählten Vergleichsmaßstabs. (**Aggregierung**)
8. Während qualitative Indikatoren Objekte ausschließlich nach den Kriterien der Gleichheit (bzw. des Unterschieds) einer Klasse zuordnen, geben quantitative Indikatoren auch Aufschluss über die Rangfolge der Kategorien und die Größe der Abstände dazwischen. (**Qualitativ vs. Quantitativ**)
9. Zahlenwerte vereinfachen die Resultate von Indikatoren mit dem Ziel, vergleichende Bewertungen zu ermöglichen oder zumindest zu beschleunigen. Durch Codierungsverfahren ist eine Übertragung qualitativer Informationen in Zahlen möglich. (**Nutzen von Zahlen**)
10. Zahlenwerte geben häufig eine Präzision von Messungen vor, die so nicht gegeben ist. Dies gilt allerdings in ähnlicher Forma auch für qualitative Beschreibungen. Deshalb ist es immer wichtig, die Grenzen der Präzision deutlich zu machen und diese bei der Bewertung hinreichend zu berücksichtigen. (**Grenzen der Präzision**)
11. Einfache Indikatoren bilden direkt messbare Sachverhalte ab und ordnen sie unmittelbar einem theoretischen Konstrukt zu. Sie sind deshalb leicht verständlich (leider manchmal auch falsch), bedingt durch die simplifizierte Verknüpfung mit dem Konstrukt aber häufig schwer zu interpretieren. (**Einfache Maßzahlen**)
12. Durch mathematische Rechenoperationen können verschiedene Indikatoren mit dem Ziel einer Standardisierung des Ergebnisses und damit einer Erleichterung des Vergleichs verknüpft werden. (**Standardisierung**)
13. Über aufwendig erhobene und berechende Indizes wird versucht, einen komplizierten Sachverhalt mit möglichst nur einer einzigen, leicht interpretierbaren Maßzahl wiederzugeben und damit einen umfassenden Überblick zu gewährleisten. (**Indizes**)
14. Der Nutzen von Indikatoren besteht darin, dass sie nicht messbare Sachverhalte messen, unvergleichbare Dinge vergleichbar machen, komplizierte Zusammenhänge vereinfacht darstellen und auf einen nicht erkennbaren Handlungsbedarf hinweisen. Ohne angemessene Interpretation sind sie allerdings nutzlos. (**Nutzen von Indikatoren**)
15. Indikatoren müssen gleichzeitig theoretischen, methodischen, praktischen und politischen Anforderungen genügen. Für eine erfolgreiche Entwicklung von Indikatoren müssen deshalb unterschiedliche Personengruppen mit entsprechenden Fachwissen und Entscheidungskompetenzen eingebunden werden. (**Anforderungen an Indikatoren**)

16. Theorien stellen die Grundlage jedes Indikators dar. Je wissenschaftlicher der theoretische Zusammenhang zwischen Indikator und Konstrukt formuliert wurde, um so sachlicher lässt er sich bewerten und umso korrekter leitet er den sozialen Prozess des Messens an. (**Theoretische Grundlage**)
17. Die Umsetzung des theoretisch Wünschenswerten stößt bei der Entwicklung von Indikatoren immer an die Grenzen des praktisch Machbaren. Bei der Indikatorenentwicklung geht es um die Optimierung der eingesetzten Instrumente in Hinsicht auf die theoretisch formulierten Ansprüche unter Zugrundelegung der gegebenen Möglichkeiten. (**Praxisorientierte Optimierung**)
18. Durch eine Korrespondenzregel wird ein nicht-messbares Konstrukt mit einem messbaren Sachverhalt theoretisch verknüpft und dadurch der für diese Messung eingesetzte Indikator operationalisiert. Die Korrespondenzregel ist selbst wiederum eine theoretische Annahme, die überprüfbar formuliert und soweit möglich auch überprüft werden sollte. (**Operationalisierung**)
19. Die Güte eines Indikators als Messinstrument ergibt sich daraus, wie genau er das misst, was er messen soll (Validität) und ob er dies bei wiederholten Messungen immer wieder auf die selbe Weise tut (Reliabilität). (**Messqualität**)
20. Die Randbedingungen einer Messung führen zu mehr oder weniger starken Messfehlern. Das soziale Umfeld stellt eine wesentliche „Störgröße“ für die Messqualität dar, die z.B. durch mangelnde Akzeptanz der Messung bei den Beteiligten, ungenügende Kapazitäten für ihre korrekte Durchführung oder fehlende Möglichkeiten zur experimentellen Kontrolle dieser oder anderer Störfaktoren Messergebnisse erheblich verzerren kann. (**Soziale Randbedingungen**)
21. Die Qualität eines Indikators ist das Produkt der Güte der Operationalisierung des theoretischen Konstrukts, der Güte der Messung des zu beobachtenden Objekts, der Güte der Durchführung dieser Messung sowie der Güte der Anerkennung der Messergebnisse durch alle Beteiligten. Sie verknüpft damit technische und soziale Aspekte, die gleichwertig zu berücksichtigen sind (**Indikatorqualität**)
22. Die Entwicklung von Indikatoren stellt einen komplexen sozialen Prozess dar, an dem Idealerweise Personengruppen mit spezifischem Fachwissen und unterschiedlichen Interessenlagen zu beteiligen sind. Er muss dementsprechend professionell moderiert werden. (**Indikatorenentwicklung als sozialer Prozess**)
23. Die Entwicklung von Indikatoren ist nicht unabhängig von ihrem Verwertungszusammenhang zu sehen und sollte deshalb über geeignete Verfahren wie z.B. das Logical Framework in die allgemeinen Managementaktivitäten eingebunden werden. (**Einbindung in Managementprozesse**)
24. Ähnlich der Vorgehensweise bei der Vorbereitung, Planung und Durchführung eines Projektes muss die Erstellung von Indikatoren schrittweise und systematisch erfolgen. (**Schrittweise Vorgehensweise**)
25. Die Auswahl zwischen verschiedenen Indikatorenkandidaten sollte anhand einer geeigneten Checkliste erfolgen, die sämtliche Anforderungen an Indikatoren ausreichend und angemessen berücksichtigt. (**Systematisches Auswahlverfahren**)

26. Die Zahl der verwendeten Indikatoren sollte so klein wie möglich gehalten werden. Dafür ist eine möglichst präzise Abbildung des theoretischen Konstrukts durch die ausgewählten Indikatoren erforderlich. (**Minimierung der Indikatorenzahl**)
27. Durch eine geringe Messqualität oder die Mehrdimensionalität theoretischer Konstrukte wird eine Ausweitung der Indikatorenzahl notwendig. Eine Verknüpfung von Indikatoren, die verschiedene Dimensionen des theoretischen Konstruktes abbilden sollen, ist nur bei theoretischer Festlegung von Verknüpfungsregeln möglich. (**Notwendigkeit von Indikatorentableaus**)
28. Indikatoren sind Hilfsmittel, die niemals optimal alle an sie gestellten Anforderungen erfüllen können. Eine kontinuierliche Weiterentwicklung unter Berücksichtigung des Informationsgehalts und des Erhebungsaufwands ist notwendig (**Indikatorenentwicklung als Daueraufgabe**)
29. Obwohl es erhebliche Unterschiede im Aufwand und im Umfang der für eine Indikatorenentwicklung zur Verfügung stehenden Ressourcen gibt, muss der soziale Entstehungsprozess eines Indikatorentableaus, seine Einbindung in den Verwertungszusammenhang, die systematische und transparente Vorgehensweise bei Erstellung und Auswahl von Indikatoren sowie die Qualität der Weiterentwicklung bei jeder Entwicklung von Indikatoren im Auge behalten werden. Dies gilt selbstverständlich auch für die von Evaluatorinnen und Evaluatoren während Kurzzeitaufenthalten zur Bewertung von Projekten eingesetzten Indikatoren. (**Indikatorenentwicklung als Element der Qualitätssicherung**)

Es gibt eine unüberschaubare Menge von Beispielen für Indikatorensysteme und auch Erfahrungsberichte über den Prozess der Indikatorenentwicklung lassen sich finden. Für das Literaturverzeichnis wurde folgende Auswahl zusammengestellt:

Stockmann (1996) beschreibt ein theoriegeleitetes Evaluationsverfahren zur Bewertung von Projekten und Programmen in der Entwicklungszusammenarbeit, welches sich mittlerweile auch in vielen anderen Bereichen als praktikables Instrument bewährt hat. Rennings (1994) und Hensling u.a. (1999) geben einen Überblick zum Stand der Diskussionen bei der Entwicklung von Indikatoren, die „nachhaltige Entwicklung“ abbilden sollen. Stärker auf die praktische Anwendung innerhalb der Entwicklungszusammenarbeit in diesem Bereich sind die Beiträge von Paulus & Haumacher (1996) sowie Segnestam (2002) ausgerichtet. Die Bedeutung von Indikatoren im Gesundheitswesen speziell in der Entwicklungszusammenarbeit erarbeitet an Hand vieler praktischer Beispiele Schirnding (2002). Und schließlich berichtet Segnestam (2000) über Erfahrungen bei der Indikatorenentwicklung während eines Aufenthalts in Zentralamerika.

LITERATUR:

Australian Agency for International Development (AusAID) (2000): AusGUIDElines. 1. The Logical Framework Approach, Commonwealth of Australia: Canberra (online: http://www.ausaid.gov.au/ausguide/ausguidelines/1.cfm - Version vom 21.06.02)

Berghoff, Sonja, Federkeil, Gero, Giebisch, Petra, Hachmeister, Cort-Denis und Detlef Müller-Böling (2003): Das Hochschulranking. Vorgehensweise und Indikatoren. Gütersloh: CHE (CHE-Arbeitspapier Nr. 46; April 2003).

BMZ-GTZ (2003): Handreichung zur Erstellung von AURA-Angeboten an das BMZ (nicht publizierte Fassung vom Frühjahr 2003).

Bortz, Jürgen & Döring, Nicola (2002): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, Berlin u.a.: Springer (3. Auflage)

CEFIC – European Chemical Industry Council (2000): ISO 9001: 2000. CEFIC Guidelines for use by the chemical industry, Brüssel: CEFIC (online: http://www.cefic.be/files/Publications/ISO.CEFIC.PDF).

Centre for International Development and Training (2002): A Guide for Developing a Logical Framework, Wolverhampton: University of Wolverhampton (online: http://www.dfid-engkar.org.uk/apps/logical_framework.pdf).

Crano, William D. (2003): Theory-Driven Evaluation and Construct Validity. In: Donaldson, Stewart I. & Scriven, Michael (eds.): Evaluating Social Programs and Problems. Visions for the New Millennium, Mahwah/London: Lawrence Erlbaum Associates, S. 145-157.

Diekmann, Andreas (1995): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, Reinbek: rororo (6. Auflage).

EC - European Commission (2002): Guidelines for the use of Indicators in country performance assessment, Brüssel: EC.

Feldt, L. S. & Brennan, R. L. (1988): Reliability. In: Linn, R.L. (ed.) Educational Measurement, New York: London (3. Auflage), S. 105-146.

Fietkau, Hans-Joachim und Trénel, Matthias (2002): Interaktionsmuster in einem Diskursverfahren zur Indikatorenentwicklung im Umweltbereich, Berlin: Wissenschaftszentrum (Discussion Paper No. FS II 02-301).

García-Verdú, Rodrigo (2002): The Human Development Index and its Application to States in Mexico, Mexico: Banco de México (online: http://home.uchicago.edu/~rgarciav/research/development/rgv-hdi.pdf).

Glatzer, Wolfgang (2002): Indikatoren, soziale. In: Endruweit, Günter & Trommsdorff, Gisela (Hrsg.), Wörterbuch der Soziologie, Stuttgart: Lucius & Lucius, S. 225-227.

GTZ (2003): Die Begriffswelt der GTZ, Eschborn: GTZ (online: http://www.gtz.de/glossar/).

Hartmann, Petra (2002): Indikator. In: Endruweit, Günter & Trommsdorff, Gisela (Hrsg.), Wörterbuch der Soziologie, Stuttgart: Lucius & Lucius, S. 223-224.

Hensling, C., Eberle, U. & Grießhammer, R. (1999): Soziale und ökonomische Nachhaltigkeitsindikatoren, Freiburg: Öko-Institut.

HTD - Hill Taylor Dickson (ed., 2001): Mediation at a glance, London u.a.: HTD.

Hullinger, Beat (2002): Werkstatt Methodische Probleme. Sozialberichterstattung und politisches Monitoring 26-28.06.2001 Neuchâtel, Bundesamt für Statistik der Schweiz: Bern.

Kim, J.O. & Mueller, C. (1987): Factor Analysis. Statistical Methods and Practical Issues, Beverley Hills: Sage.

Kluge, S. (1999): Empirisch begründete Typenbildung. Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung, Opladen: Leske+Budrich.

Krämer, Walter (1991): So lügt man mit Statistik, München: Piper.

Kromrey, Helmut (2002): Empirische Sozialforschung, Opladen: Leske+Budrich (10. Auflage)

Laatz, Wilfried (1993): Empirische Methoden. Ein Lehrbuch für Sozialwissenschaftler, Thun/Frankfurt: Deutsch.

Liebe, Frank & Gilbert, Nadja (1996): Interkulturelle Mediation – eine schwierige Vermittlung. Eine empirisch-analytische Annährung zur Bedeutung von kulturellen Unterschieden, Berlin: Berghofzentrum für konstruktive Konfliktbearbeitung (Berghof Report Nr. 2).

Meyer, Wolfgang (2000): Ratgeber Umweltberatungsprojekte, Saarbrücken: Universität des Saarlandes (herausgegeben vom Umweltbundesamt und vom Bundesverband deutscher Umweltberater).

Nagarajan, Nigel & Vanheukelen, Marc (1997): Evaluating EU Expenditure Programmes: A Guide. Ex-post and Intermediate Evaluation, European Commission XIX/02 – Budgetary Overview and evaluation, Luxemburg: EU.

Paulus, S. & Haumacher, W. (1996): Indikatoren der Institutionenentwicklung im Umweltbereich. Anregungen und Beispiele für Projektplanung und –management, Eschborn: GTZ (AP 402/92-23d – PVI)

Picciotto, Robert (2002): Development Cooperation and Performance Evaluation – The Monterrey Challenge, Washington: Worldbank (OED Working Paper).

Rennings, Klaus (1994): Indikatoren für eine dauerhaft-umweltgerechte Entwicklung, Stuttgart: Metzler-Poeschel.

Schirnding, Yasmin von (2002): Health in Sustainable Development Planning: The Role of Indicators, Genf: WHO.

Schnell, Rainer, Hill, Paul B. & Esser, Elke (1999): Methoden der empirischen Sozialforschung, München/Wien: Oldenbourg Verlag (6. Auflage).

Segnestam, Lisa (2000): Developing Indicators. Lessons Learnt from Central America, Washington: Worldbank.

Segnestam, Lisa (2002): Indicators of Environment and Sustainable Development. Theories and Practical Efficency, Washington: Worldbank (Environmental Economic Series 89)

Shavelson, R., Mc Donnell, L.M. & Oakes, J. (1991): Steps in designing an indicator system. In: Practical Assessment, Research & Evaluation 2 (12); online: http://erical.net/pare/getvn.asp?v=2&n=12.

Stockmann, Reinhard (1996): Die Wirksamkeit der Entwicklungshilfe. Eine Evaluation der Nachhaltigkeit von Programmen und Projekten, Opladen: Westdeutscher Verlag.

Sullivan, John L. & Feldman, Stanley (1979): Multiple Indicators. An Introduction, Thousand Oaks u.a.: Sage (Quantitative Applications in the Social Sciences 15).

Tuckman, B.W. & Jensen, M. A. (1977): Stages of small group development revisited. In: Group and Organizational Studies, 2, S. 419-427.

Turner, Jonathan H. (2003): The Structure of Sociological Theory, Belmont: Wadsworth (7th).

UNDP – Office of Evaluation and Strategic Planning (1997): Results-oriented Monitoring and Evaluation. A Handbook For Programme Managers, New York: UNDP (online: http://www.undp.org/eo/documents/mae-toc.htm).

UNDP – Evaluation Office (2002a): Handbook on Monitoring and Evaluating for Results. New York: UNDP.

UNDP – Evaluation Office (2002b): RBM in UNDP: Selecting Indicators, New York: UNDP (online: http://www.undp.org/eo/documents/Indicators.pdf).

UNDP – Evaluation Office (2002c): Guidelines for Outcome Evaluators. Monitoring and Evaluation Companion Series 1, New York: UNDP.

UNDP – Human Development Report Office (2001): Human Development Reports: Measuring Development and Influencing Policy. Fast Facts Dec. 2001. New York: UNDP (online: http://mobile.undp.org:8100/dpa/publications/ffNHDRe101201.pdf).

USAID (1998): Handbook of Democracy and Governance Program Indicators. Center for Democracy and Governance. The Technical Publications Series 8/1998; online: http://www.usaid.gov/democracy/pdfs/pnacc390.pdf.

VENRO – Verband Entwicklungspolitik deutscher Nicht-Regierungsorganisationen (2000): Prüfen und Lernen. Praxisorientierte Handreichung zur Wirkungsbeobachtung und Evaluation, Bonn: VENRO.

Werner, Heinecke (2000): Entwicklung von Indikatoren zur Feststellung der Zielerreichung in der bilateralen FZ und TZ, Bonn: Deutsches Institut für Entwicklungspolitik.

World Bank – OED (2002): Monitoring and Evaluation. Some Tools, Methods and Approaches, Washington: World Bank.

Wottawa, Heinrich & Thierau, Heike (1990): Lehrbuch Evaluation, Bern u.a.: Huber.

Zapf, Wolfgang (1977): Soziale Indikatoren – Eine Zwischenbilanz. In: Krupp, H.J. & Zapf, W. (Hrsg.), Sozialpolitik und Sozialberichterstattung, Frankfurt/New York: Campus, S. 231-246.

CEval-Arbeitspapiere:

Nr. 0 Selbstdarstellung des CEval (Reinhard Stockmann)

Nr. 1 Evaluation als integriertes Lehr- und Forschungsprogramm (Reinhard Stockmann)

Nr. 2 Soziologie im Abwärtstrend – Eine empirische Untersuchung zur Situation der Soziologie an den bundesdeutschen Hochschulen (Thomas Knoll, Wolfgang Meyer, Reinhard Stockmann)

Nr. 3 Qualitätsmanagement und Evaluation – Konkurrierende oder sich ergänzende Konzepte? (Reinhard Stockmann)

Nr. 4 Konzept zur Evaluation von E-Learning Angeboten im Rahmen von VISU (Virtuelle Saar-Universität) (Reinhard Stockmann, Erik Schäffer)

Nr. 5 Was ist Evaluation? (Wolfgang Meyer)

Nr. 6 Sociological Theory and Evaluation Research. An Application and its Usability for Evaluating Sustainable Development (Wolfgang Meyer)

Nr. 7 Evaluation der Umweltberatungsprojekte des Bundesumweltministeriums und des Umweltbundesamtes / Evaluation of Environmental Consulting Projects sponsored by BMU and UBA (Wolfgang Meyer, Klaus-Peter Jacoby, Reinhard Stockmann)

Nr. 8 Measuring the Impact of Vocational Training Projects – An Evaluation Method and its Theoretical Foundations (Reinhard Stockmann, Wolfgang Meyer)

Nr. 9 Was ist eine gute Evaluation – Einführung zu Funktionen und Methoden von Evaluationsverfahren (Reinhard Stockmann)

Centrum für Evaluation (CEval)

Universität des Saarlandes
Philosophische Fakultät III
Empirische Humanwissenschaften
Lehrstuhl für Soziologie

Postfach 15 11 50
66041 Saarbrücken

info@ceval.de
www.ceval.de

LEITUNG:

Prof. Dr. Reinhard Stockmann
Tel.: +49 (0)681 – 302 3372
Fax: +49 (0)681 – 302 3899
Email: r.stockmann@mx.uni-saarland.de

BEREICHSKOORDINATION ENTWICKLUNGSZUSAMMENARBEIT + BILDUNG:

Dipl.-Soz. Stefanie Krapp
Tel.: +49 (0)681 – 302 4509
Fax: +49 (0)681 – 302 3899
Email: s.krapp@ceval.de

BEREICHSKOORDINATION UMWELT:

Dr. Wolfgang Meyer
Tel.: +49 (0)681 – 302 4358
Fax: +49 (0)681 – 302 3899
Email: w.meyer@mx.uni-saarland.de