
Einführung in die Grundlagen der Entwicklung von Indikatoren

Wolfgang Meyer

1 Einleitung

Indikatoren spielen im Alltag eine wichtige Rolle und werden immer wieder mehr oder weniger bewusst zum Einsatz gebracht. Dazu gehören Rituale wie z. B. der Blick aus dem Fenster zur Einschätzung des Wetters, bevor die Kleidung gewählt wird, oder die Interpretation des Mienenspiels eines Gegenübers zur Abschätzung seiner Stimmungslage. Kaum jemand betrachtet eine solche Verwendung von Zuständen oder Signalen für weiterführende Rückschlüsse und zur Grundlegung von Entscheidungen als einen Einsatz von „Indikatoren“, und schon gar nicht werden hier verschiedene „Indikatorenalternativen“ kritisch gegeneinander abgewogen.

Dies geschieht schon eher beim Blick in die Tageszeitung, wenn dort über Maßzahlen berichtet und diese im Sinne der Autorinnen und Autoren bewertet werden. Zu solchen häufig diskutierten Indikatoren gehört z. B. das Ergebnis einer Befragung hinsichtlich einer (fiktiven) Wahlentscheidung am kommenden Sonntag, die als Wiedergabe der politischen Stimmungslage und Bewertung der Regierungsleistung interpretiert wird. Auch die allseits beliebten (pseudo-)psychologischen Tests in Illustrierten verwenden Indikatoren zum Rückschluss auf persönliche Eigenschaften oder Einstellungen.

W. Meyer (✉)

Centrum für Evaluation, Universität des Saarlandes, Saarbrücken, Deutschland
E-Mail: w.meyer@mx.uni-saarland.de

Die wissenschaftlichen Grundlagen eines solchen Einsatzes von Indikatoren in Alltag und Massenmedien lassen sich sicherlich häufig anzweifeln: Oft leiten die hier gezogenen Schlussfolgerungen in die Irre, die dargestellten Ergebnisse sind diffus und können sehr unterschiedlich interpretiert werden. Meist geht es gerade darum, eine Kontroverse zu provozieren, oder einfach um Unterhaltung – und nicht um einen wissenschaftlichen Erkenntnisfortschritt oder gar um die Steuerungsanforderungen einer „evidence-based policy“.

Dies bedeutet jedoch nicht, dass wissenschaftliche Indikatoren nur hinter den verschlossenen Türen des „Elfenbeinturms“ von Bedeutung sind. So können medial verarbeitete Test- und Befragungsergebnisse durchaus auf seriösen Forschungsarbeiten beruhen und die verwendeten Indikatoren höchsten Qualitätsanforderungen genügen. Ein hohes Maß an Wissenschaftlichkeit schützt aber nicht vor politisch motivierten Diskussionen, und dabei geraten die eingesetzten Indikatoren zumeist ins Zentrum der Kritik.

Gerade in emotional aufgeladenen politischen Diskussionen sind deshalb der kompetente und korrekter Umgang mit Indikatoren sowie ein genaues Wissen um deren Stärken und Schwächen notwendig, um eine Versachlichung der Auseinandersetzungen und ein zielgerichtetes Handeln zu ermöglichen. Die Gleichstellung von Männern und Frauen gehört sicherlich zu den Politikfeldern, in denen es emotionale und von Vorurteilen geprägte Auseinandersetzungen gab und immer noch gibt – und in dem der Einsatz von Indikatoren schon häufig zu einer faktenbasierten Diskussion und besseren Fundierung von Entscheidungen beigetragen hat.

Damit sind bereits die wichtigsten Aspekte von Indikatoren angesprochen, die im Rahmen dieses Beitrags noch systematischer ausgearbeitet und diskutiert werden sollen. Der nächste Abschnitt beschäftigt sich dabei zunächst mit einer allgemein gehaltenen Definition des Begriffs „Indikator“ und erläutert die wichtigsten Eigenschaften von Indikatoren an einer Vielzahl unterschiedlicher Beispiele. Ein besonderer Fokus der Darstellung liegt dabei auf „Sozialindikatoren“, die sich zwar nicht grundsätzlich von naturwissenschaftlich-technischen Indikatoren unterscheiden, trotzdem aber einige wichtige messtechnische und politische Eigenheiten haben, auf die hier eingegangen werden soll. Hieraus lassen sich grundsätzliche Anforderungen ableiten, die im dritten Abschnitt auf die Nutzung von Indikatoren bezogen werden. Als Ergebnis werden vier zentrale Qualitätskriterien zur Bewertung von Indikatoren vorgestellt. Der letzte Abschnitt skizziert schließlich die Vorgehensweise bei der Entwicklung von Indikatoren.

2 Was sind Indikatoren?

Der Begriff „Indikator“ kommt aus dem Lateinischen und ist von dem Wort „indicare“ abgeleitet, das so viel wie „anzeigen“ bedeutet. Ein Indikator ist also ein „Anzeiger“ und setzt in dieser Funktion zwei Dinge miteinander in Beziehung: Ein wahrnehmbares Element und ein nicht wahrnehmbares, durch das wahrgenommene Element aber abbildbares und damit mit diesem verknüpftes Element. Der Indikator steht somit niemals für sich selbst, sondern für einen mit ihm verbundenen Sachverhalt.

Diese „Anzeigerfunktion“ eines Indikators wird in allen gängigen Definitionen hervorgehoben (Meyer 2007a, 2004). Auch wenn in Lexika unterschiedliche Beispiele und Formulierungen verwendet werden, so geht es doch immer um diese Verknüpfung zwischen „Messbarem“ und „Nicht-Messbarem“.

Entsprechend diesem Grundverständnis von Indikatoren lassen sich drei zentrale Elemente zur Bestimmung eines Indikators unterscheiden:

- Der *Anzeiger* selbst, welcher in irgendeiner Form wahrnehmbar sein muss,
- das nicht wahrnehmbare, durch den Indikator abzubildende *Konstrukt* und
- die Anzeiger und Konstrukt verbindende Verknüpfung (*Korrespondenzregel*).

Diese universalen Elemente des Indikators sind vollkommen unabhängig vom Gebrauch in den verschiedensten fachlichen Kontexten: So ist es z. B. lediglich der Gegenstandsbereich, der Soziale Indikatoren von anderen Indikatoren abhebt. Jeder Indikator muss also die genannten drei Elemente beinhalten. Dementsprechend sind weder präzise technische Messungen noch charakterisierende Sachbeschreibungen per se bereits „Indikatoren“, sie können es aber werden, wenn sie eine Korrespondenzregel mit dem Konstrukt verbinden, das abgebildet werden soll. Der Einkommensunterschied zwischen Männern und Frauen wird also erst dann zum Indikator, wenn durch ihn etwas anderes angezeigt werden soll. Umgekehrt wird ein nicht sinnlich fassbarer Gegenstand nur dann durch eine Messung oder Beschreibung korrekt abgebildet, wenn eine entsprechende Korrespondenzregel angenommen wird. Soll der Einkommensunterschied als Indikator die Gleichstellung von Mann und Frau durch ArbeitgeberInnen anzeigen, so ist damit zugleich die Annahme über einen Zusammenhang zwischen Gleichstellung und Gehalt enthalten. Der Nutzen eines Indikators hängt entsprechend nicht nur von der Präzision einer Messung, sondern auch von der Beschreibung des abzubildenden Konstrukts und der Qualität der theoretischen Verknüpfung zwischen Indikator und Konstrukt ab.

Dieser Zusammenhang lässt sich am besten durch Beispiele für den praktischen Einsatz von Indikatoren näher beleuchten. Indikatoren werden in allen wissenschaftlichen Disziplinen – und wie bereits gesagt auch in vielen Alltagssituationen – verwendet. Dieser Einsatz ist dementsprechend sehr unterschiedlich und der Kreativität in der Indikatorenentwicklung sind grundsätzlich keine Grenzen gesetzt. Beispiele für Indikatoren in den verschiedenen Disziplinen sind in Tab. 1 zusammengestellt, die damit gleichzeitig als „Anzeiger“ für die konzeptionelle Vielfalt von Indikatoren dient. Auf diese Beispiele soll bei der folgenden Beschreibung von Indikatoreigenschaften zurückgegriffen werden.

Zunächst soll der Fokus auf den abzubildenden Gegenstand, auf das häufig nur theoretisch bestimmbare *Konstrukt*, gelegt werden. Nicht jeder der in der Übersicht aufgeführten Gegenstände ist *grundsätzlich* nicht messbar und sie sind auch nicht *ausschließlich* durch den aufgeführten Indikator abzubilden. Generell geht es um eine *Übertragung* in die sinnliche Wahrnehmung des Menschen und hier kann auch die *Effizienz* dieser Übertragung ein wichtiger Grund für den Indikatoreinsatz sein. Die Leuchtdiode ist z. B. sicher nicht die einzige Methode zur Messung des Stromflusses, das sichtbare Leuchten der Diode erzeugt allerdings schnell und unkompliziert Aufmerksamkeit bei NutzerInnen – auf jeden Fall schneller als andere – möglicherweise genauere – Messinstrumente.

Manche der mit Indikatoren abgebildeten Konstrukte sind *reale Phänomene*, wie z. B. der Stromfluss, der Druck im Kessel oder der Säuregehalt einer Flüssigkeit. Sie lassen sich hinsichtlich ihrer Eigenschaften, Verhaltensweisen, Formen oder ihres Auftretens etc. exakt definieren und eingrenzen. Ihre Existenz ist unstrittig und nachweisbar. Zumeist gilt dies für naturwissenschaftliche Gegenstände, aber nicht nur: Beispielsweise ist auch die Tat eines Angeklagten – unabhängig vom Schuldspruch eines Gerichts – *objektiv* gegeben oder nicht-gegeben.¹ Im Unterschied zum Stromfluss, zum Druck oder zum Säuregehalt fällt hier allerdings häufig die theoretische *Abgrenzung des Konstrukts* gegenüber dem Kontrafaktischen (also im Beispiel dem Zustand der „Nichttat“) schwerer. Die „Messprobleme“ in den Sozialwissenschaften sind also oft weniger mit dem eigentlichen Vorgang des Messens verbunden, sondern mit der wenig präzisen, manchmal sogar recht diffusen *Konstruktdefinition*. Es liegt in der Natur der

¹In der Frühzeit der Sozialindikatorenbewegung wurde versucht, nur objektivierbare Sachverhalte wie den hier im Beispiel aufgeführten Wohnstandard zur Sozialberichterstattung zu verwenden. Subjektive Indikatoren wie z. B. Glück und Zufriedenheit wurden erst später in die Tableaus aufgenommen (eine Sammlung von subjektiven Indikatoren zur Lebensqualität findet sich bei Hagerty 2001; zur Entstehungsgeschichte siehe Zapf 2002).

Tab. 1 Beispiele für Indikatoren

Bereich	Indikator	Was wird gemessen?	Was soll abgebildet werden?
Biologie	Flechtenvegetation	Artenzahl und Verteilung	Allgemeine Belastung von Ökosystemen
Chemie	Lackmустest	Verfärbung eines Teststreifens	Säure- bzw. Basengehalt einer Flüssigkeit
Medizin	Enzyme-linked Immunosorbent Assay (ELISA)	Verfärbung des Patientenserums in Interaktion mit Virus-Antigen (mikrobiologischer Test)	Virusinfektion
Physik	Lumineszenz-Diode (LED)	Leuchten in einer bestimmbar Farbe und Stärke (elektrische Schaltung)	Stromfluss und Durchlassspannung
Psychologie	IQ-Test	Lösung verschiedener standardisierter Aufgaben in vorgegebener Zeit	Intelligenz
Recht	Indizienbeweis	Personenbezogene Zuordnung belegbarer Sachverhalte	Schuld oder Unschuld einer Person
Sozialwissenschaft	Wohnstandard	Antworten in einer Repräsentativbefragung zur Anzahl der Wohnungen mit Balkon, Badezimmer, Sammelheizung, Barrierefreiheit etc.	Wohnqualität der Wohnung einer bestimmten Personengruppe
Technik	Druckmessung	Aufzeichnung eines Verlaufsdiagramms mittels mechanischer Übertragung	Druck in einem Zylinder
Wirtschaft	Bruttoinlandsprodukt	Waren- und Dienstleistungsproduktion im Inland abzüglich Vorleistungen (amtliche Zählung)	Gesamte wirtschaftliche Leistungsfähigkeit einer Volkswirtschaft

Sache, dass kein noch so gut formulierter Indikator einen nebulösen und subjektiv interpretierbaren Gegenstand exakt abbilden kann (Tab. 1).

Dementsprechend ist in vielen Fällen eine *Einigung der Beteiligten* über den mit Indikatoren abzubildenden Gegenstand von zentraler Bedeutung. Theoretische Konstrukte wie z. B. „Demokratie“, „Frieden“ oder eben auch „Gleichstellung“ werden von verschiedenen Menschen höchst unterschiedlich gefasst und in der subjektiven Wahrnehmung wird der Schwerpunkt gegebenenfalls auf Eigenschaften gelegt, die andere nicht als wesentlich ansehen. Dies ist häufig vom Kontext abhängig: In Bürgerkriegsgebieten, wie etwa gegenwärtig in Syrien, existieren sicher andere Vorstellungen von „Frieden“ als in Deutschland, wo bewaffnete Auseinandersetzungen nicht zum Alltag gehören. Der Versuch einer „objektiven“, anhand theoretischer Kriterien erarbeiteten Definition von „Frieden“ gerät hier in Widerspruch mit einem „subjektiven“, anhand eigener Erfahrungen geprägten Verständnis der Betroffenen. Probleme dieser Art sind nicht durch „bessere Indikatoren“ zu lösen, sondern müssen auf theoretischer Ebene der Konstruktbestimmung in einen *tragbaren und allgemein akzeptierten Kompromiss* überführt werden, bevor geeignete Indikatoren zur Abbildung dieses Konstrukts entwickelt werden können.

Beispiele für solche sozialen Kompromisse finden sich auch in der obigen Indikatorenliste: „Intelligenz“, „Wohnqualität“, „wirtschaftliche Leistung“ oder „Belastung von Ökosystemen“ sind theoretische Konzepte; ihre reale und objektiv belegbare Existenz sowie ihre subjektive Bedeutung für die untersuchten Elemente (in den Beispielen Personen, Räume, Waren- und Dienstleistungsprodukte oder Flechten) sind schwer nachweisbar und deshalb wissenschaftlich umstritten. Sie sind *wissenschaftliche Artefakte*, die sich teilweise erst durch die messtechnischen Anweisungen in den Korrespondenzregeln fassen lassen: Bereits 1923 konstatierte z. B. der US-Psychologe Edwin Boring: „Intelligenz ist, was Intelligenztests messen“ („Intelligence is what the tests test“, Boring 1923, S. 35), und brachte damit die Grundlage vieler psychologischer Tests (aber auch theoretischer Konstrukte anderer Fächer) auf den Punkt.

Eine solche *definitivische Verknüpfung* zwischen Messung und theoretischem Konstrukt impliziert eine Reihe von Anforderungen an die verwendeten Messinstrumente. Im Fall des Intelligenztests bedeutet dies z. B., dass die zur Bewertung verwendeten Einzelaufgaben unter gleichen Bedingungen immer zum selben Ergebnis führen (Reliabilität), unabhängig von den TestleiterInnen und deren Einstellungen gegenüber den Testpersonen sind (Objektivität) und gleichzeitig mit dem Gesamtergebnis der Testskala sowie den anderen im Test verwendeten Aufgaben hoch korrelieren (Konvergenzvalidität). Durch Sicherstellung einer in dieser Hinsicht hohen Messqualität kann dann gewährleistet werden, dass die Tests

homogen immer das Gleiche messen – ob dies allerdings „Intelligenz“ ist, kann auf der Konstruktebene weiterhin bezweifelt werden (was bei den Intelligenztests auch bis zum heutigen Tag der Fall ist). Letztlich bleibt somit die Nutzung selbst qualitativ hochwertiger Intelligenztests zur Messung von Intelligenz eine durch objektive Gegebenheiten nicht begründbare Entscheidung der Beteiligten.

Indikatoren können bei solchen Entscheidungen nur dann eindeutige Hinweise geben, wenn eine *deterministische Verknüpfung* zwischen dem Anzeiger und dem abzubildenden Konstrukt gegeben ist. Dies bedeutet, dass jede Veränderung des Konstrukts zu jeder Zeit auch zu Veränderungen des Indikators in die gleiche Richtung und im gleichen Umfang führt. Manche naturwissenschaftlich-technische Indikatoren erfüllen diese Voraussetzungen in hohem Maße: das Lackmuspapier, die Leuchtdiode oder auch der unter anderem zur HIV-Diagnostik eingesetzte ELISA-Test sind sehr stark mit dem von ihnen abzubildenden Konstrukt verknüpft. Allerdings ist der weitverbreitete Glaube falsch, dass es sich in solchen Fällen um perfekte Zusammenhänge handelt. Selbstverständlich gibt es auch bei naturwissenschaftlich-technischen Messungen Messfehler und Schwankungen in den Rahmenbedingungen, welche die Messergebnisse beeinflussen und damit auch den Zusammenhang zwischen Konstrukt und Messung. Ob dies für die Nutzung der Messdaten als Indikator problematisch ist oder nicht, bleibt eine Abwägungsfrage und muss im Kontext des Verwendungszwecks entschieden werden.

Deutlich mehr Diskussionen lösen Indikatoren dann aus, wenn nur ein *probabilistischer Zusammenhang* zwischen Indikator und abzubildendem Gegenstand besteht. Hier führen Veränderungen im Konstrukt nur mit einer *angebbaren Wahrscheinlichkeit* zu Veränderungen im Anzeigeverhalten des Indikators, d. h., mit jeder Messung ist zugleich ein gewisses Risiko des Irrtums verbunden.² Dieses Risiko wird berechenbar, wenn entsprechende Kenntnisse über den probabilistischen Zusammenhang zwischen Konstrukt und Indikator vorhanden sind. Dies ist z. B. bei Stichprobenziehungen der Fall: Eine repräsentative Stichprobe von Untersuchungselementen setzt ein Auswahlverfahren voraus, welches jedem Element der Grundgesamtheit nicht nur die Möglichkeit bietet, in die Stichprobe zu

²Dies ist eine Frage der Messgenauigkeit, welche die möglichst präzise Wiedergabe des exakten, „wahren“ Wertes bezeichnet. Abweichungen hiervon können sich sowohl aufgrund der verwendeten Instrumente und Messverfahren (Messfehler) als auch aufgrund anderer, externer Störeinflüsse ergeben. Dies beeinflusst dann ebenfalls den Zusammenhang zwischen Konstrukt und Messung: Eine bessere Messqualität durch Steigerung von Validität, Reliabilität und/oder Objektivität muss nicht unbedingt zu einer höheren Messgenauigkeit führen und umgekehrt bedeutet eine präzisere Messung nicht gleichzeitig, dass sich dadurch auch die Validität, Reliabilität und/oder Objektivität von Messungen und damit die Messqualität verändert.

gelangen, sondern dies auch mit einer mathematisch bestimmbarer Wahrscheinlichkeit sicherstellt (z. B. durch den Einsatz von Zufallszahlen bei der Adressenauswahl für einen Survey). Dies ermöglicht dann die Verwendung der Stichprobenergebnisse zur Abschätzung der „wahren“ Werte in der Population, aus der die Stichprobe gezogen wurde, und reduziert den Messaufwand erheblich (Meyer 2007b, S. 233).

Leider gibt es in der Praxis häufig auch Verknüpfungen zwischen Indikatoren und Konstrukten, die auf *willkürlichen Annahmen* über deren Zusammenhang basieren. Es gibt in diesen Fällen keine wissenschaftlich begründbare Grundlage für den Schluss vom Anzeiger auf sein Abbild, was zum einen die Gefahr von Fehlschlüssen enthält und zum anderen Tür und Tor für Diskussionen über den Nutzen des Indikators öffnet. Solche fragwürdigen Indikatoren sollten aufgrund der Mängel in der Anzeigefunktion vermieden werden, auch wenn die eingesetzten Messinstrumente noch so präzise messen oder einfach einsetzbar sind.

Der in der Praxis vorzufindende Zusammenhang zwischen Indikatoren und Konstrukten wird im Folgenden am Beispiel von Sozialindikatoren näher erläutert und diskutiert. Das Zentrum für Sozialindikatorenforschung (ZSI) des Leibniz-Institut für Sozialwissenschaften (GESIS) in Mannheim pflegt eine Liste von insgesamt 89 Indikatoren aus 14 Lebens- und Politikbereichen, die im Rahmen der Sozialindikatorenbewegung in Deutschland von den Vorläuferprojekten SPES („Sozialpolitische Entscheidungs- und Indikatorensystem für die Bundesrepublik Deutschland“) und dem Sonderforschungsbereich 3 („Mikroanalytische Grundlagen der Gesellschaftspolitik“) begründet und seit 1972 kontinuierlich weiterentwickelt wurden.³

Tab. 2 gibt einen Überblick zum ZSI-Indikatorensystem und stellt aus jedem der 14 von GESIS ausgewiesenen Bereiche ein Beispiel eines Indikators vor. Im Indikatorensystem sind die Bereiche hierarchisch mit einer unterschiedlichen Anzahl von Zielbereichen verknüpft, die sich dann wiederum in Zieldimensionen ausdifferenzieren lassen. Die Grundlage dieser Einteilung ist eine ausführliche Analyse der politischen Zielsetzungen von Parteien, Verbänden und gesellschaftlichen Gruppen, die in diesen Lebens- und Politikfeldern relevant sind. Das Indikatorensystem folgt somit nicht einem objektiven, aus theoretischen Überlegungen abgeleiteten Ordnungsprinzip, sondern orientiert sich an den Vorstellungen der Gesellschaft und einiger ihrer zentralen Akteure. Dementsprechend ist es nicht

³Zur Geschichte der Sozialindikatorenbewegung in Deutschland siehe Zapf (2002); zur Entstehung der Sozialindikatorenbewegung insgesamt siehe Bauer (1966), Duncan (1969) und Miles (1985); zum ZSI-Indikatorensystem siehe Noll (2014).

Tab. 2 Übersicht zum ZSI-Indikatorensystem 2015. (Quelle: SIMon – Social Indicators Monitor (www.gesis-simon.de); eigene Auswahl und Ergänzung)

Nr.	Bereich	Beispiel Indikator	Zielbereich	Konstrukt (Beispiel)
1	Bevölkerung	Anteil nicht ehelicher Lebensgemeinschaften	Formen des Zusammenlebens	Lebensform
2	Sozioökonomische Gliederung und SchichtEinstufung	Anteil klassenhomogener Ehen in der oberen Dienstklasse	Soziale Mobilität	Soziale Ungleichheit
3	Arbeitsmarkt und Beschäftigung	Befürchtung, arbeitslos zu werden	Arbeitsmarktchancen und -risiken	Subjektiv wahrgenommene Beschäftigungssicherheit
4	Einkommen und seine Verteilung	Armutsquote relative Armut (40 %-Schwelle)	Armut in der Einkommensdimension	Ressourcenmangel in Privathaushalten
5	Einkommensverwendung und Versorgung	Kosten der Versorgungssicherung	Sicherheit der Versorgung	Versorgungsausgaben
6	Verkehr	Anteil des Verkehrs am Endenergieverbrauch	Belastung durch das Verkehrssystem	Umweltbelastung
7	Wohnung	Wohnungen ohne Standardausstattung	Qualität der Wohnungsausstattung	Wohnbedingungen
8	Gesundheit	Sterbefälle durch Kreislauferkrankungen	Todesursache	Gesundheitsrisiken
9	Bildung	Studienanfängerinnenquote (Frauen)	Chancengleichheit im Tertiärbereich	Bildungschancen
10	Partizipation	Anteil politisch stark Interessierter	Politisches Interesse	Politische Beteiligung
11	Umwelt	Staubemission pro EinwohnerIn	Qualität der Umweltbedingungen	Luftverschmutzung
12	Öffentliche Sicherheit und Kriminalität	Opferisiko bei Mord und Totschlag	Belastung durch Kriminalität	Öffentliche Sicherheit
13	Freizeit	Reichweite der tagesaktuellen Medien	Medien	Mediennutzung
14	Globale Wohlfahrtsmaße	Lebenszufriedenheit	Subjektives Wohlbefinden	Kognitive Wohlfahrt

unveränderlich und muss im Gegenteil entsprechend dem sozialen Wandel hinsichtlich seiner Maßzahlen angepasst werden. Bezüglich der Auswahl der Indikatoren wird ein Kompromiss aus logischer Ableitung und pragmatischer Umsetzung gewählt, der zunächst ideale, den Zieldimensionen möglichst exakt entsprechende Indikatoren erarbeitet und dann nach realisierbaren, diesen idealen Indikatoren möglichst nahekommenden Maßzahlen sucht. Dank der Unterstützung verschiedener Forschungsförderinstitutionen wie dem GESIS, dem Wissenschaftszentrum Berlin (WZB) und dem Deutschen Institut für Wirtschaftsforschung (DIW) konnten vor allem im Bereich der subjektiven Einschätzungen durch regelmäßig durchgeführte Bevölkerungsumfragen wie den Wohlfahrtssurvey oder das Sozio-oekonomische Panel Datenlücken geschlossen und Entwicklungstrends abgebildet werden (vor allem in den vom Statistischen Bundesamt gemeinsam mit WZB und DIW im zweijährigen Rhythmus herausgegebenen Datenreports, siehe für die neuste Ausgabe DESTATIS, WZB und DIW 2013; zur Geschichte und den theoretischen Grundlagen siehe Zapf 1974 ff., 1976, 2002) (Tab. 2).

Die in der letzten Spalte aufgeführten Beispiele sind Konstrukte, die mit den hier vorgestellten Indikatoren abzubilden sind. Der erste Indikator – Anteil nicht-ehelicher Lebensgemeinschaften – ist ein gutes Beispiel für einen „objektiven“ Sozialindikator: Ohne Zweifel existieren jenseits der Ehe unterschiedliche Formen des Zusammenlebens von Menschen, die unterschiedliche Ursachen haben und unterschiedliche Zielsetzungen verfolgen. Der Indikator soll den Umfang dieser von der gesellschaftlichen Norm abweichenden Lebensformen abbilden, indem er auf der Grundlage von Daten der amtlichen Statistik diese Arten des Zusammenlebens zum ehelichen Zusammenleben ins Verhältnis setzt. Im Unterschied zu den Eheschließungen, die aufgrund der Meldepflichten anhand „prozessproduzierter Daten“ vollständig und tagesgenau erfasst werden können, ist die Ausweisung von nicht-ehelichen Lebensgemeinschaften auf Befragungsdaten angewiesen, die in der amtlichen Statistik über den Mikrozensus erhoben werden. Hierdurch entstehen allein aufgrund der Dauer der Erhebung (und der hohen Fluktuation bei nicht-ehelichen Lebensgemeinschaften) sowie dem Stichprobencharakter mehr Ungenauigkeiten als bei der Zählung von Eheschließungen. Nichtsdestotrotz ist aber das Konstrukt einer Lebensform jenseits der Ehe noch vergleichsweise einfach und exakt zu erfassen, weil es den Befragten bekannt oder vermittelbar ist. Grenzfälle bei der Messung ergeben sich z. B. durch die Bestimmung des Begriffs „Zusammenleben“, welcher von einer rein zweckorientierten Wohngemeinschaft anders verstanden wird als von einem getrennt lebenden Liebespaar. Trotzdem dürften an dieser Stelle ebenso wie bei der Verknüpfung von Konstrukt und Indikator vergleichsweise wenig Fehler und Schwierigkeiten auftreten.

Die Situation ist bei den beiden folgenden hier ausgewählten Indikatoren deutlich anders. Die Messung sozialer Mobilität als wichtiges Element zur Abbildung sozialer Ungleichheit ist auf die Festlegung bestimmter Schichtungsmodelle mit entsprechenden Einheiten angewiesen, zwischen denen dann ein Austausch stattfinden bzw. nicht stattfinden kann. Diese Einheiten werden durch theoretische Modelle bestimmt, die für sich ein hohes Maß an Strukturstabilität selbst bei hoher individueller Fluktuation beanspruchen. Dieses muss bedingt durch den sozialen Wandel aber nicht unbedingt gegeben sein (z. B. Berger 1991). Bezüglich der Korrespondenzregel besteht somit das Problem, dass der Indikator implizit von einer Strukturstabilität ausgeht, die so nicht gegeben ist, und dementsprechend bei Strukturveränderungen fehlerhafte Werte bezüglich des Grads der sozialen Mobilität ausgeben kann. Diese Fehlerquelle ergibt sich nicht bei der Messung selbst (also in diesem Fall den Angaben, die Befragte in Interviews über ihre eigenen Lebensverhältnisse machen), sondern bei der Aufbereitung der Daten durch sozialwissenschaftliche ExpertInnen, die auf Grundlage verschiedener Variablen eine Zuweisung zu bestimmten festgelegten Kategorien vornehmen. Die Korrektheit dieser Zuweisung kann z. B. aufgrund der Verwendung der Ehe als prägende Form des Zusammenlebens infrage gestellt sein: Soziale Mobilität könnte sich theoretisch auch darin ausdrücken, dass die Institution Ehe als nicht mehr zeitgemäß angesehen wird und sich gerade in nicht-ehelichen Lebensgemeinschaften die soziale Mobilität in einer Gesellschaft ausdrückt. Beide Argumentationslinien haben gemeinsam, dass sie am Zusammenhang zwischen dem theoretischen Konstrukt „soziale Ungleichheit“ und der zu seiner Abbildung verwendeten Messung des Anteils „klassenhomogener Ehen“ ansetzen. Es handelt sich hier zwar auch um einen „objektiv“ messbaren Indikator, der aufgrund des ExpertInnen-Ratings zudem von den informationsgebenden Personen kaum manipuliert werden kann, aber die verwendete theoretische Grundlage ist im Unterschied zum ersten Beispiel deutlich stärker durch fachwissenschaftliche Diskurse sowie durch deren „subjektive“ Bewertung geprägt.

Das dritte Beispiel, die „Befürchtung, arbeitslos zu werden“, ist schließlich den „subjektiven“ Indikatoren zuzuordnen. Es handelt sich hier um Maßzahlen, die Einstellungen, Bewertungen und Einschätzungen der BürgerInnen wiedergeben und nicht gesellschaftlich definierte Gegenstände wie Lebensformen oder Schichtzugehörigkeit erfassen wollen. Die individuellen Einschätzungen der BürgerInnen basieren in der Regel auf sehr unterschiedlichen, subjektiv variierenden Kriterien, die zumeist über die Zeit nicht stabil sind. Der Zusammenhang zwischen „objektiven“ Lebenslagen und dem „subjektiven“ Wohlbefinden ist dementsprechend gering und hängt von einer Vielzahl von Störgrößen ab (z. B. der aktuellen Lebenssituation, der Stimmungslage, dem Vergleich mit individuellen

Ansprüchen, früheren Lebensphasen oder anderen Personen wie FreundInnen, Bekannten, KonkurrentInnen etc.). Dementsprechend bildet das von den Betroffenen selbst wahrgenommene Risiko eines Arbeitsplatzverlustes nicht die tatsächliche, durch die wirtschaftliche Situation der Arbeitgeberin/des Arbeitgebers oder deren Bewertungen der Arbeitsleistung der/des Beschäftigten gegebene Gefahr ab. Sie kann aber für den Alltag der Betroffenen und speziell deren Wohlbefinden deutlich höhere Relevanz als das „objektive“ Risiko haben. Inwieweit die in einem Fragebogen gemachten Angaben zu den eigenen Befürchtungen ihnen tatsächlich entsprechen, ist eine weitgehend unbeantwortbare Frage: Ob die Messung korrekt ist und ob die Maßzahlen wirklich die subjektive Wahrnehmung abbilden, entzieht sich einer Bewertung durch Dritte. Die theoretische Annahme ist hier, dass sich bei einer ausreichend großen Fallzahl die verschiedenen auftretenden Fehler zufällig verteilen und die Mittel- und Anteilswerte ein näherungsweise realistisches Bild der „wahren“ Einschätzung der BürgerInnen wiedergeben. Wie oben beschrieben wird somit ein probabilistischer Zusammenhang zwischen den gemessenen Angaben und den tatsächlichen Einstellungen angenommen, wobei solche messtheoretischen Annahmen natürlich Diskussionen auslösen können.

Dementsprechend werden Indikatorendiskussionen häufig vor allem mit Blick auf die „Messgenauigkeit“ geführt, während die eher pragmatisch ausgerichtete und für die Nutzung von Indikatoren viel wichtigere Anzeigefunktion vernachlässigt wird. Indikatoren werden dafür kritisiert, dass sie a) Konstrukte unvollständig messen, b) mit Messfehlern oder Schwankungen behaftet sind oder c) keine verlässlichen Daten liefern. Diese Diskussionen sind oft abstrakt und ohne Bezug zur geplanten Nutzung der Indikatoren. Hierauf soll im folgenden Abschnitt noch etwas näher eingegangen werden.

3 Wozu braucht man Indikatoren?

Der Einsatz von Indikatoren begründet sich ausschließlich durch ihren Nutzen. Dieser besteht – allgemein gesprochen – darin, dass sie Dinge für Menschen sinnlich wahrnehmbar machen, die ihnen ansonsten verborgen bleiben würden. Häufig ist der Einsatz von Indikatoren mit Handlungskonsequenzen verbunden: In den oben aufgeführten Indikatorenbeispielen werden Messungen in der Regel nicht ausschließlich aus einem Erkenntnisinteresse heraus durchgeführt, sondern zu einem *weiterführenden Zweck*. Der ELISA-Test in der Medizin dient z. B. der ärztlichen Diagnose einer HIV-Infektion, die wiederum die Therapie indiziert. Der Erfolg dieses Verfahrens und der eingesetzten Indikatoren für die praktische

Behandlung von HIV begründet sich dabei nicht nur aus der Messgenauigkeit, sondern auch aus anderen, mit der weiterführenden Nutzung verbundenen Anforderungen. So geht es z. B. auch darum, die Testergebnisse möglichst schnell und kostengünstig den behandelnden ÄrztInnen zur Verfügung stellen zu können.

Die NutzerInnen der von Indikatoren produzierten Ergebnisse sind wie in diesem Beispiel oft keine „MessexpertInnen“, sie können den Messvorgang also nicht selbst nachvollziehen (und sind auch selten daran interessiert), sondern erwarten verlässliche und eindeutige Anzeigen als Grundlage für ihr eigenes Handeln. Dementsprechend sind in vielen Fällen differenzierte Angaben auf kleinteilig abgestuften Skalen weniger hilfreich als eine grobe dichotome Ja/Nein-Rückmeldung. Hierfür ein Alltagsbeispiel: Wer in sein Auto steigt und den Zündschlüssel dreht, sieht eine Reihe von Kontrolllampen aufleuchten, die nach dem Start des Fahrzeugs sofort wieder ausgehen. Dies dient ausschließlich zum Beleg ihrer Funktionstüchtigkeit (die eingesetzten Leuchtdioden sind nicht defekt) und manche der Kontrollleuchten (z. B. die Öldruckkontrolle) bleiben danach während der gesamten Fahrt dunkel. Wenn allerdings die Öldruckleuchte angeht, bedeutet dies „sofort anhalten – es droht unmittelbar ein akuter Motorschaden“. Kein Fahrer und keine Fahrerin ist in dieser Situation an dem exakten Messwert des Öldrucks interessiert und die meisten wären vermutlich mit der Interpretation des Skalenwerts auf einem Manometer überfordert. Selbst wenn die Messsonde im Motor äußerst präzise arbeitet, ist es für die NutzerInnen dieser Information wichtiger, eine knappe und leicht verständliche Rückmeldung zu erhalten.

Die Anforderung an die Kontrolllampe als Indikator für den Öldruck ist demnach, dass sie *immer und nur dann* leuchtet, wenn Gefahr für den Motor besteht (mit Ausnahme des Kontrollleuchtens zur Signalisierung ihrer Funktionsfähigkeit zu Beginn der Fahrt). Die Genauigkeit der Messung des Öldrucks spielt dabei eine untergeordnete Rolle und muss in der Realität nicht sonderlich hoch sein, weil ein Schaden bei Verbrennungsmotoren nicht unmittelbar bei Unterschreitung einer engen Toleranzgrenze eintritt. Bezüglich der benötigten *Messgenauigkeit* genügt es, wenn ein bestimmter Grenzwert grob eingehalten wird, und dabei können sogar vergleichsweise hohe Messschwankungen toleriert werden.

Die Anforderungen an die *Anzeigequalität* sind dagegen sehr hoch: das Lämpchen soll nicht „flackern“ (weil der Grenzwert bei wiederholten Messungen mal überschritten wird, mal nicht), nicht zu früh leuchten (keinen teuren und lästigen „Fehlalarm“ auslösen), aber auch nicht zu spät oder gar nicht, wenn ein Defekt vorliegt. Die Folgen solcher Fehlfunktionen können dramatisch sein, wie sich an einem anderen, sehr traurigen Beispiel zeigen lässt: Im Juni 1998 kamen bei einem ICE-Unglück in der Nähe von Eschede mehr als 100 Menschen ums

Leben. Auslöser der Katastrophe war der Bruch eines Radreifens bei hoher Geschwindigkeit. Dieser Radreifen wurde – wie aus Sicherheitsgründen üblich – am Morgen des Unfalltages noch technisch auf Haarrisse geprüft. Bei den Ermittlungen ergab sich, dass drei unabhängige Tests den besagten Radreifen als schadhaft identifizierten, ohne dass ein Austausch erfolgte. Der Grund für die fehlende Reaktion der TechnikerInnen lag darin begründet, dass die eingesetzten Testverfahren aufgrund ihrer präzisen Messung viele als ignorierbar eingeschätzte Fehlmeldungen produzierten und dadurch ihre Anzeigefunktion im Hinblick auf zu behebende Schäden schlecht war. Die TechnikerInnen hätten jeden Tag Tausende von Radreifen mit verschwindend geringem Schadensrisiko austauschen müssen, um die eine (unvorhersehbare) Katastrophe zu verhindern.

In diesem Fall hat also eine zu genaue Messung durch ihre häufig falschen Anzeigen zu einer Fehlentscheidung mit dramatischen Folgen geführt. Das menschliche Versagen besteht aber weniger in dieser Fehlentscheidung als in dem Einsatz eines Indikators, der aufgrund seiner mangelnden Anzeigequalität (trotz oder sogar gerade wegen der hohen Messgenauigkeit) für das Monitoring der Züge ungeeignet war. Dies hätte bei einer kritischen Prüfung in dieser Richtung frühzeitig erkannt und problemlos zur Entwicklung besser geeigneter Indikatoren genutzt werden können. Die Katastrophe von Eschede hat zu einer Vielzahl von Konsequenzen geführt, eine regelmäßige systematische Überprüfung der eingesetzten Indikatoren gehört, soweit bekannt, allerdings nicht dazu.

Der Zusammenhang zwischen Aufwand und Nutzen von Indikatorenmessungen lässt sich theoretisch wie folgt darstellen (Abb. 1): Mit dem Anspruch an die Genauigkeit einer Messung steigt bei gleichbleibenden Messinstrumenten und gleicher Messqualität auch das Risiko einer *fehlerhaften Messung* (vgl. die nach rechts aufsteigende Kurve in der Abbildung). Ein Beispiel: Das Risiko, dass eine normale Uhr innerhalb eines Tages um eine Stunde vor- oder nachgeht, ist sehr gering. Wird der Anspruch an die Messgenauigkeit auf eine Abweichung von weniger als eine Minute pro Tag erhöht, so wird es vermutlich innerhalb eines Jahres bei derselben Uhr schon zu einer Reihe von Fehlmessungen kommen. Erhöht sich der Anspruch auf eine Messgenauigkeit von weniger als eine Sekunde, sind wahrscheinlich im Zeitraum von einem Jahr kaum Tage ohne mehr oder weniger geringfügige Abweichungen von diesem Messziel zu erwarten. Soll die Messgenauigkeit gleich bleiben, d. h. die Zahl der vom Messziel abweichenden Messungen konstant sein, kann dies nur durch einen exponentiell steigenden Aufwand zur Steigerung der Messqualität der verwendeten Instrumente und Verfahren erreicht werden. Im Beispiel: Bei einer tolerierten Abweichung von einer Minute genügt eine normale Quarzuhr. Für sekundengenaue Messungen muss diese Uhr schon durch Atomuhren getaktet werden. Sollen aber Zeitmessungen

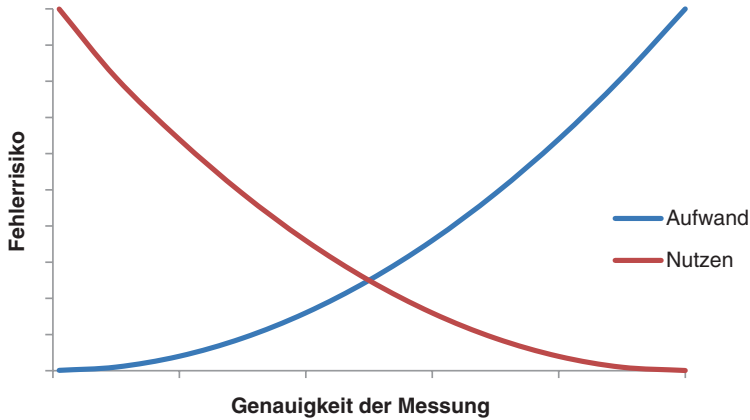


Abb. 1 Zusammenhang zwischen Aufwand und Nutzen von Indikatorenmessungen

auf tausendstel Sekunden oder noch präziser sein, so erfordert dies teure und aufwendig zu justierende Chronometer, deren Messgenauigkeit weit über die von handelsüblichen Geräten hinausgeht.⁴

Auf der *Nutzenseite* (die nach rechts abfallende Kurve in Abb. 1) steht weniger das Risiko einer Fehlmessung denn das Risiko einer *fehlerhaften Entscheidung* auf der Basis dieser Fehlmessung im Zentrum der Betrachtung.⁵ Dementsprechend ist der Zusammenhang beim Risiko der Fehlentscheidung genau umgekehrt wie beim Risiko eines Messfehlers: Mit steigender Messgenauigkeit nimmt bei gleichbleibender Nutzung in der Regel das Risiko einer Fehlentscheidung auf der Grundlage dieser Messergebnisse ab. Der Gewinn an Entscheidungssicherheit (und damit der Rückgang des Fehlrisikos der Entscheidung) verringert sich mit

⁴Die Prüftoleranz der deutschen Prüfstelle für Chronometer beträgt derzeit 0,3 s Abweichung pro Tag.

⁵Natürlich gibt es – wie z. B. in der Wissenschaft – Fälle, in denen die exakte Messung selbst das Ziel ist und keine weitere Nutzung angestrebt wird. Dies sind aber seltene Ausnahmen: Zumeist wird etwas gemessen, um daraus Konsequenzen für das Handeln zu ziehen. Der Blick auf die Uhr erfolgt im obigen Beispiel nicht wegen der Genauigkeit ihrer Messung, sondern weil damit Termine eingehalten, Arbeitsabläufe überwacht oder Entscheidungen in Wettbewerben getroffen werden sollen. Die zunehmende Messgenauigkeit führt dann dazu, dass Fehler (z. B. versäumte Termine, unkoordinierte Abläufe oder Siegerehrungen für VerliererInnen) abnehmen.

jeder Qualitätsverbesserung der Messung. Im Beispiel reicht für die Verabredung zu Geschäftsterminen eine Minutengenauigkeit sicher aus, der zusätzliche Nutzen einer genaueren Messung nimmt stark ab und niemand muss einen teuren Chronometer zur sekundenbruchteilgenauen Messung für die Einhaltung seiner Termine bei sich tragen. Dies gilt allerdings z. B. beim Rennrodeln nicht, da die Fahrzeiten bei Wettkämpfen häufig nur noch tausendstel Sekunden auseinanderliegen. Der Anspruch an die Messgenauigkeit und damit die Reduzierung von Messfehlern hängt somit von der Nutzung und dem Nutzwert genauerer Messungen ab.

Dementsprechend stellt der Schnittpunkt zwischen beiden Kurven (Aufwand und Nutzen) die optimale, im konkreten Anwendungsfall benötigte Messgenauigkeit dar. Je höher die mit einer Nutzung verbundenen Risiken einer Fehlentscheidung auf Basis falscher Messergebnisse sind, umso höher ist der Anspruch an die Messgenauigkeit und umgekehrt. Ein zu hoher Anspruch an die Messgenauigkeit schadet insofern, weil hierdurch unnötig Ressourcen für die Messung gebunden und Fehlanzeigen produziert werden. Ein zu geringer Anspruch wiederum erhöht das Risiko von Fehlentscheidungen auf Grundlage der Indikatorenwerte.

Durch die Nutzung ergeben sich somit unterschiedliche Anforderungen an die Messung von Indikatoren. Nur in wenigen Fällen ist eine Optimierung der Messgenauigkeit die beste Strategie bei der Entwicklung von Indikatoren, es empfiehlt sich vielmehr den Fokus auf die Anzeigequalität zu legen und eine sorgfältige Risikoabwägung des Indikatoreinsatzes vorzunehmen. Ein solcher Prozess ist von einer Reihe von *Rahmenbedingungen* abhängig und lässt sich deshalb nicht generell und mittels eines eindeutigen, objektiven Verfahrens entscheiden. Werden beispielsweise Indikatoren im Projektkontext eingesetzt, so treffen hier je nach Projekthalt und Akteurskonstellation unterschiedliche „Stakeholder-Gruppen“ aufeinander, die Indikatoren unterschiedlich nutzen und interpretieren möchten. Bei der Festlegung der zu verwendenden Indikatoren müssen diese abweichenden Nutzungsinteressen berücksichtigt und in einem Verhandlungsprozess abgestimmt werden. Hier gewinnt eine *subjektive Komponente* bei der Nutzung von Indikatoren an Bedeutung, weil es bei solchen Prozessen nicht nur um wissenschaftliche Aspekte, wie die theoretische Konzeption der Operationalisierung oder die methodische Qualität der Messung, sondern eben auch um die Praktikabilität des Einsatzes für eine bestimmte Nutzung und die mit dieser Nutzung verbundenen politischen Absichten der AkteurInnen geht. Weil Indikatoren durch ihre Anzeigefunktion einen Nutzen stiften sollen, ist die Berücksichtigung dieser „sozialen“ Komponenten neben den wissenschaftlichen Ansprüchen an Theorie und Messmethodik unabdingbar notwendig.

Die wichtigste soziale Eigenschaft des Indikators ist dementsprechend seine *Akzeptanz bei den NutzerInnen*. Die subjektive Bewertung des Indikators (und

seines Nutzwertes) stellt häufig ein schwierig zu überwindendes Hindernis für eine gemeinsame konstruktive Arbeit mit Indikatoren dar, und vielfach werden Indikatoren erst im Nachhinein (also nach ihrer Messung und ihrem Einsatz als Indikator) von einem Teil der Beteiligten grundsätzlich infrage gestellt. Dies ist wiederum bei jedem Indikator problemlos möglich, weil dieser a) nur im Falle einer definitorischen oder perfekten Beziehung zum Konstrukt fehlerfrei anzeigen kann, b) jede Messung mit Messfehlern behaftet ist und c) die Interpretation der Anzeige des Indikators durch diese Fehlerhaftigkeit immer einen Interpretationsspielraum eröffnet, mit dem sich auch ein zur Anzeige und der damit verbundenen Handlungserwartung abweichendes Verhalten rechtfertigen lässt (das Gegenteil des vom Indikatorergebnis nahegelegten „wahren“ Werts ist „nur“ unwahrscheinlich und nicht unmöglich, weshalb eine zwingende Handlungsempfehlung nicht abgeleitet werden kann). Bei der Indikatorenentwicklung geht es deshalb häufig weniger um die Erarbeitung eines theoretisch und methodisch perfekten Instruments denn um die Vereinbarung eines Konsenses bezüglich der Interpretation der Anzeigen.

Diese politische Seite der Indikatorenentwicklung wird aber häufig zugunsten einer *pragmatischen Vorgehensweise* zurückgestellt und irgendeine vorliegende, leicht durchzuführende Messung trotz oft eingeschränktem Bezug zum abzubildenden Konstrukt als Indikator genutzt. Diese Vorgehensweise ist aufgrund des hohen Aufwands der Entwicklung eines allgemein akzeptierten, theoretisch plausiblen und methodisch sauberen Indikators verständlich – allerdings werden dabei die Risiken (und daraus resultierende Kosten) fehlerhafter Anzeigen und ungeklärter Nutzung selten ins Gesamtkalkül mit einbezogen. Ein guter Anzeiger muss nicht immer aufwendig sein, ein schlechter ist es dagegen sicher, weil er immer zusätzliche Kosten bei seiner Nutzung erzeugt (wie z. B. Fehlentscheidungen, unnötige Diskussionen, abweichende Interpretationen und Reaktionen, Unsicherheiten hinsichtlich der Bedeutung und Reichweite etc.). Dies betrifft nicht nur so extreme Fälle wie das Zugunglück in Eschede, sondern auch die endlosen, notwendige Entscheidungen und Handlungen blockierenden Indikatorendiskussionen, die leider zum Alltag vieler Programme und Projekte gehören und durch „billige“, zur Nutzung wenig geeignete Messungen ausgelöst werden.

Bei der Entwicklung von Indikatoren müssen also vier Qualitätskriterien zugleich berücksichtigt und gleichberechtigt im Sinne der Optimierung des Verwendungszwecks gegeneinander abgewogen werden (Meyer 2004, S. 30):

- *Abbildungsqualität* (theoretische Dimension): Inwieweit ist der Indikator in der Lage, den eigentlich interessierenden Sachverhalt in der für die Nutzung notwendigen Qualität abzubilden und für diesen Zweck sinnvolle Anzeigen zu liefern?

- *Messqualität* (methodische Dimension): Inwieweit misst der Indikator in einer der Nutzung angemessenen Güte das, was er messen soll, und führt auch bei Messwiederholungen im Rahmen tolerierbarer Fehlermargen immer wieder zu ähnlichen Ergebnissen?
- *Akzeptanz* (politische Dimension): Inwieweit wird der Indikator von den NutzerInnen als handlungsleitendes Instrument akzeptiert und seine Anzeige als Grundlage für Entscheidungen im benötigten Umfang verwendet?
- *Praktikabilität* (praktische Dimension): Inwieweit ist der Indikator mit den vorhandenen Mitteln für den angestrebten Nutzen in der benötigten Qualität umsetzbar?

4 Wie sollten Indikatoren entwickelt werden?

Die Entwicklung von Indikatoren sollte den in diesem Abschnitt vorgestellten Schritten folgen. Generell muss die Indikatorenentwicklung dabei mehrere Ziele anstreben:

- Erstens geht es um eine Abwägung theoretischer und methodischer Ansprüche mit den praktischen und politischen Bedingungen für eine Indikatorenutzung. Die Risiken fehlerhafter Messungen für die Verwendung der Indikatoren müssen berücksichtigt werden.
- Zweitens ist diese Abwägung als kontinuierlicher Prozess zu verstehen, der ein Monitoring der Messung und Indikatorenutzung enthalten sollte und in eine angemessene Weiterentwicklung des Indikatorensystems münden kann.
- Drittens sollte dabei insbesondere die Akzeptanz der Indikatoren und ihrer Messergebnisse bei allen beteiligten AkteurInnen im Fokus stehen. Es geht hier vor allem um die Gewährleistung eines Konsenses hinsichtlich der Bewertung von Messergebnissen und der sich daraus ergebenden Schlussfolgerungen für die Zusammenarbeit der Beteiligten im gegebenen Kontext. Rein methodentechnische Diskussionen zur Verbesserung der Abbildungs- oder Messgenauigkeit der Indikatoren sollten nur dann geführt werden, wenn sie für die gemeinsamen Zielsetzungen unerlässlich sind.

Der im Folgenden vorgestellte Ablauf ist idealtypisch zu verstehen und kann häufig in der Praxis nur bedingt in dieser Form umgesetzt werden. Es sollte jedoch versucht werden, die Kernelemente soweit möglich zu berücksichtigen und in den gegebenen Kontext zu integrieren.

Schritt 1: Schärfung des theoretischen Konstrukts und daraus Ableitung der Anforderungen an Indikatoren zu dessen Abbildung (Messziele)

Im ersten Schritt muss eine Auseinandersetzung mit dem abzubildenden theoretischen Konstrukt erfolgen. Wenn also z. B. die Gleichstellung am Arbeitsmarkt von Männern und Frauen abgebildet werden soll, so müssen hierfür die theoretisch denkbaren Einflüsse und ein zu untersuchendes Kausalmodell entwickelt werden. Welche verschiedenen Aspekte enthält das Konstrukt „Gleichstellung“ und was muss im Kontext des hier gegebenen Verwendungszusammenhangs (z. B. eines regionalen Projekts zur Förderung gleicher Löhne) durch die einzusetzenden Indikatoren angezeigt werden? Hierüber ist mit den ProjektpartnerInnen ein Konsens zu erzielen.

Schritt 2: Operationalisierung und daraus Ableitung von Korrespondenzregeln für ideale Indikatoren (optimale Messung)

Sobald Klarheit über die abzubildenden Elemente des Konstrukts herrscht, kann mit der Entwicklung geeigneter Maßzahlen begonnen werden. Dabei sollten zunächst „optimale“ Messungen hypothetisch diskutiert werden. Im Beispiel wäre dies etwa ein „Gender Pay Gap“, der ausschließlich die auf der Basis von Diskriminierungen entstandenen Einkommensunterschiede misst. Alternative Messkonzepte und alternative Indikatoren sollten mitgedacht und in die weiteren Erwägungen mit einbezogen werden. Ziel ist die von allen akzeptierte und für den Einsatz „optimale“ Messung (die Machbarkeit sollte bei diesem Schritt noch nicht erwogen werden).

Schritt 3: Entwurf idealer Indikatoren und Bestimmung ihrer zentralen Charakteristika (Evaluation der optimalen Messung)

Im nächsten Schritt müssen die wichtigsten Kennzeichen und Elemente des „idealen Indikators“ erarbeitet und mit den Beteiligten abgestimmt werden. Hier geht es primär um die Erzeugung eines Konsenses und einer Ableitung von Eigenschaften, die ein zu realisierender Indikator enthalten muss. Außerdem ist hier zu diskutieren, welche potenziellen Störfaktoren die Ermittlung eines geeigneten Indikators und die mit ihm verbundenen Messungen beeinflussen könnten. Beim „Gender Pay Gap“ ginge es z. B. um die Faktoren, die einen Einkommensunterschied jenseits der geschlechtsspezifischen Diskriminierung erzeugen und dementsprechend zu Fehlinterpretationen und falschen Schlussfolgerungen führen könnten. Hieraus kann gegebenenfalls die Notwendigkeit der Entwicklung zusätzlicher Indikatoren zur Kontrolle dieser Einflüsse deutlich werden.

Schritt 4: Ableitung realisierbarer Indikatoren, die den optimalen Messungen am nächsten kommen (realisierbare Messung)

Nachdem das theoretische Verständnis des abzubildenden Sachverhalts geklärt worden ist und auf dieser Basis gemeinsam die Erwartungen an die Indikatoren offengelegt worden sind, können die „idealen“ Indikatoren mit den zu erwartenden Schwierigkeiten bei ihrer Umsetzung konfrontiert werden. Zu diskutieren sind z. B. die potenziell auftretenden Defizite und Störungen sowie deren Auswirkungen auf die Nutzung der Anzeigefunktion und im Weiteren die Frage, wie diese Auswirkungen gegebenenfalls methodisch kontrolliert werden könnten und welche ergänzenden Maßnahmen (z. B. der Einsatz weiterer Indikatoren) zur Verbesserung des Nutzwertes zu ergreifen wären. Bei der Abwägung des „Gender Pay Gap“ sind z. B. die Besonderheiten der zu beobachtenden Berufe und Beschäftigungssituationen, die Anforderungen an die persönlichen Qualifikationen und Kompetenzen sowie die Unterschiede hinsichtlich des Beschäftigungsumfangs zu beachten und hinsichtlich ihrer Auswirkungen auf das Anzeigeverhalten des Indikators zu diskutieren.

Schritt 5: Festlegung der Mess- und der Anzeigegenauigkeit mit Blick auf die praktische Nutzung

Sobald ein Indikator als realisierbar festgelegt worden ist, sollte über die Mess- und Anzeigeeigenschaften gesprochen werden. Die zentrale Frage dabei ist, welche Informationen für die Nutzung des Indikators im gemeinsamen Vorhaben benötigt werden. Wenn es z. B. um die Veränderung des „Gender Pay Gap“ während eines Projekts geht, so stellt sich die Frage nach den Zielwerten, die mithilfe der Maßnahmen angestrebt werden sollen. Hierbei geht es auch um die Reichweite der Maßnahmen, ihren Umfang und die durch sie zu erwartenden Veränderungen. So sind beispielsweise durch lokale Maßnahmen kaum Effekte auf nationaler Ebene zu erwarten; dies würde entweder einen extrem hohen Interventionsaufwand auf lokaler Ebene oder eine sehr präzise Messung auf nationaler Ebene erfordern. Entsprechend ergibt sich aus diesen Erwägungen, welche Genauigkeit bei den Messungen anzustreben ist und wie diese bezüglich der zu erhebenden Objekte und/oder räumlichen Bezüge eingegrenzt werden sollte.

Schritt 6: Vergleich verschiedener Indikatoren-Alternativen (Auswahl der umzusetzenden Indikatoren)

Im gesamten Prozess sollte nicht nur auf einen einzigen Indikator für ein abzubildendes Konstrukt fokussiert, sondern sollten verschiedene Alternativen parallel verfolgt werden. Dies erhöht zum einen die Chance der Realisierung einer der

Möglichkeiten und verringert das Risiko, den gesamten Prozess im Fall des Scheiterns eines Indikators wieder von vorne beginnen zu müssen und ermöglicht zum anderen durch Vergleiche der Alternativen eine bessere Bewertung der vorhandenen Handlungsoptionen und der mit ihnen verbundenen Möglichkeiten und Grenzen.

Schritt 7: Praxistest des Indikators und probeweise Erhebung von Daten (Pretest)

Im nächsten Schritt sollte der Indikator hinsichtlich seiner Erhebungspraxis und der sich dabei gegebenenfalls ergebenden Schwierigkeiten getestet werden. So kann es z. B. bei der Erhebung des „Gender Pay Gap“ in einer bestimmten Region aufgrund der Weigerung von ArbeitgeberInnen oder ArbeitnehmerInnen, diese Daten offenzulegen, zu nicht vorhergesehenen Schwierigkeiten kommen. Eventuell müssen unterstützende Maßnahmen (hier z. B. Gespräche zur Förderung der Akzeptanz und zum Abbau von Ängsten), alternative Erhebungsmethoden (z. B. regionale Surveys) oder sogar die Verwendung alternativer Indikatoren in Erwägung gezogen werden.

Schritt 8: Evaluation des Messverhaltens und gegebenenfalls Justierung der Messungen auf Basis der Ergebnisse des Pretests (messtechnische Evaluation)

In diesem Schritt geht es darum, die Probleme bei der Umsetzung der Datenerhebung und der Durchführung der Messungen zu identifizieren und diese hinsichtlich der Möglichkeiten ihrer technischen Bearbeitung zu prüfen. Lassen sich z. B. die Daten von sich verweigernden Personen in ausreichender Güte auf der Grundlage der erhobenen Resultate schätzen? Können die erforderlichen Informationen aus anderen Informationsquellen gewonnen werden? Lassen sich Maßnahmen implementieren, die zu einer Verbesserung von Rücklauf oder Response führen? Gibt es andere Personen oder Organisationen, die eine bessere Qualität der Durchführung gewährleisten können? Am Ende dieses Schrittes sollte entschieden werden, ob der im Praxistest eingesetzte Indikator aus technischer Sicht geeignet ist, modifiziert werden sollte oder verworfen werden muss.

Schritt 9: Evaluation der Indikatorenutzung und gegebenenfalls Justierung des Anzeigeverhaltens auf Basis der Ergebnisse des Pretests (Evaluation der Nutzungspraxis)

Analog zum vorherigen Schritt sollte die praktische Nutzung der Messergebnisse für die gemeinsame Arbeit evaluiert werden. Können die Ergebnisse für die zuvor gesetzten Ziele verwendet werden? Setzen alle Beteiligten diese Befunde in der gewünschten Art um? Welche Schwierigkeiten entstehen an den verschiedenen

Stellen der Nutzung der Ergebnisse bei den unterschiedlichen ProjektpartnerInnen? Gibt es unvorhergesehene Akzeptanzprobleme und lassen sich diese lösen? Welche Schlussfolgerungen ergeben sich für die Indikatorenutzung – kann der Indikator in der gegebenen Form beibehalten werden, sind Modifikationen zur Verbesserung der Nutzung notwendig oder muss er verworfen und ersetzt werden?

Schritt 10: Praxiseinsatz und kontinuierliches Monitoring der Messungen und der Nutzung des Indikators

Wenn der Indikator sich als tauglich erwiesen hat und praktisch eingesetzt werden kann, so sollten Prozesse zur Kontrolle sowohl der Messung als auch der Nutzung implementiert werden. Sollte sich ergeben, dass die Messungen nicht in der benötigten Qualität erfolgen oder die Anzeigen des Indikators nicht in der benötigten Form genutzt werden, müssen Beratungen zur Abstellung dieser Defizite zeitnah erfolgen. Letztlich ist ein nicht genutzter oder nicht in ausreichender Qualität gemessener Indikator wertlos und es werden nur unnötig Ressourcen für ihn verschwendet.

5 Zusammenfassung und Schluss

In diesem Beitrag sind einige zentrale Elemente der Indikatorenentwicklung angesprochen worden, die auch bei der Entwicklung von Gender-Indikatoren Berücksichtigung finden sollten. Dabei ging es um drei Kernaspekte:

- Indikatoren sind „Anzeiger“, an die vier zentrale Anforderungen zu stellen sind. Sie müssen theoretisch den Anforderungen einer guten Abbildung eines nicht oder nur schwer messbaren Konstrukts genügen und dieses über eine geeignete Korrespondenzregel mit dem gemessenen Objekt verbinden. Sie müssen bei der Messung dieses Objekts eine hohe Messqualität sichern, die möglichst valide, reliable und objektive Ergebnisse liefert. Diese Messungen müssen mit den gegebenen Ressourcen und Möglichkeiten durchführbar und ihre Ergebnisse praktisch verwertbar sein. Und schließlich ist es notwendig, dass die Beteiligten sich auf eine dem gemeinsamen Ziel verpflichtete Nutzung einlassen und den Indikator hinsichtlich seiner Leitfunktion akzeptieren.
- Der Einsatz von Indikatoren sollte effizient erfolgen – dies bedeutet, dass die theoretischen und methodischen Ansprüche mit den praktischen und politischen Erwartungen in Einklang zu bringen sind. Die optimale Messgenauigkeit ergibt sich aus dem Bedarf an präziser Messung für die gemeinsame

Nutzung und den methodisch-technischen Möglichkeiten zur Optimierung der Messungen. Im Idealfall können die diskutierten zentralen Aspekte der Indikatorenmessung – Messgenauigkeit und Anzeigequalität – zur Deckung gebracht werden, d. h., die erreichte Anzeigequalität entspricht genau den Erfordernissen der NutzerInnen. Nicht nur mangelnde Genauigkeit von Messungen, sondern auch eine für die praktische Nutzung zu präzise Messung kann aber die Nutzung beeinträchtigen und Fehler bei der Verwendung von Indikatoren nach sich ziehen.

- Schließlich ist zu berücksichtigen, dass die Entwicklung von Indikatoren keineswegs nur ein technischer Prozess ist, der bei Einhaltung bestimmter Regeln zu einem optimalen Ergebnis führt. Aufgrund der notwendigen Nutzung von Indikatoren und der mit ihrer Hilfe produzierten Ergebnisse stellt die Indikatorenentwicklung im Gegenteil einen sozialen Prozess dar, der zumindest über den Zeitraum der Zusammenarbeit kontinuierlich weitergeführt werden sollte. Die Indikatorennutzung ist in der Regel keine einmalige Angelegenheit, sondern erfordert eine gewisse Kontinuität und Kooperationsbereitschaft. Dementsprechend sollten die Indikatoren unter Beobachtung stehen und sollte regelmäßig geprüft werden, ob sie den an sie gerichteten Anforderungen noch genügen. Diese Überprüfung kann keineswegs nur eine technische Überprüfung ihrer Messqualität sein, sondern muss auch soziale Komponenten wie die Akzeptanz, das gemeinsame Verständnis oder den Ablauf der Verwendung von Ergebnissen betreffen.

Literatur

- Bauer, Raymond A. (Hrsg.) (1966). *Social Indicators*. Cambridge: MIT Press.
- Berger, Peter A. (1991). Von Bewegungen in zur Beweglichkeit von Strukturen. Provisorische Überlegungen zur Sozialstrukturanalyse im vereinten Deutschland. *Soziale Welt* 42 (1), 68–92.
- Boring, Edwin G. (1923). Intelligence as the Tests Test It. *New Republic* 36 (1923), 35–37.
- DESTATIS [Statistisches Bundesamt], WZB [Wissenschaftszentrum Berlin für Sozialforschung] & DIW [Zentrales Datenmanagement & Das Sozio-oekonomische Panel (SOEP) am Deutschen Institut für Wirtschaftsforschung] (Hrsg.) (2013). *Datenreport 2013. Ein Sozialbericht für die Bundesrepublik Deutschland*, Bonn: Bundeszentrale für politische Bildung. <http://www.bpb.de/nachschlagen/datenreport-2013/>. Zugegriffen: 4. November 2015.
- Duncan, Otis D. (1969). *Toward Social Reporting: Next Steps*. New York: Russell Sage.
- Endrueit, Günter & Trommsdorff, Giesela (Hrsg.) (2002). *Wörterbuch der Soziologie*. 2. Aufl. Stuttgart: Lucius & Lucius.
- Hagerty, Michael R. (2001). Quality of life indexes for national policy: review and agenda for research. *Social Indicator Research* 55 (1), 1–96.

- Meyer, Wolfgang (2004). *Indikatorenentwicklung. Eine praxisorientierte Einführung*. CEval-Arbeitspapiere Nr. 10. 2. Aufl. Saarbrücken: Centrum für Evaluation. http://www.ceval.de/modx/fileadmin/user_upload/PDFs/workpaper10.pdf. Zugegriffen: 4. November 2015.
- Meyer, Wolfgang (2007a). Messen: Indikatoren – Skalen – Indizes – Interpretationen. In Reinhard Stockmann (Hrsg.), *Handbuch zur Evaluation, Eine praktische Handlungsanleitung* (S. 195–222). Münster: Waxmann.
- Meyer, Wolfgang (2007b). Datenerhebung: Befragungen – Beobachtungen – Nicht-Reaktive Verfahren. In Reinhard Stockmann (Hrsg.), *Handbuch zur Evaluation, Eine praktische Handlungsanleitung* (S. 223–277). Münster: Waxmann.
- Miles, Ian (1985). *Social Indicators for Human Development*. London: Francis Pinter.
- Noll, Heinz-Herbert (2014). German system of social indicators. In Alex C. Michalos (Hrsg.), *Encyclopedia of Quality of Life and Well-Being Research* (S. 2541–2544). Heidelberg: Springer.
- Zapf, Wolfgang (Hrsg.) (1974ff.). *Soziale Indikatoren. Konzepte und Forschungsansätze*. 4 Bde. Frankfurt: Herder & Herder (Bd. 1/2); Frankfurt, New York: Campus (Bd. 3/4).
- Zapf, Wolfgang (1976). *Sozialberichterstattung: Möglichkeiten und Probleme*. Kommission für wirtschaftlichen und sozialen Wandel 125. Göttingen: Verlag Otto Schwartz & Co.
- Zapf, Wolfgang (2002). *Modernisierung und Wohlfahrtsentwicklung*. WZB-Vorlesung 5. Berlin: WZB.



<http://www.springer.com/978-3-658-13236-1>

Gleichstellung messbar machen
Grundlagen und Anwendungen von Gender- und
Gleichstellungsindikatoren

Wroblewski, A.; Kelle, U.; Reith, F. (Hrsg.)

2017, VIII, 289 S. 21 Abb., 3 Abb. in Farbe., Softcover

ISBN: 978-3-658-13236-1