

Mehr Humor und keine Katze – Rankingprodukte Marke Bertelsmann

Meik Michalke, Oliver Naß & Anna Nitsche

aus: »Netzwerk der Macht – Bertelsmann«
Wernicke, J. & Bultmann, T. (Hg.)

2. erw. Aufl., Oktober 2007 (S. 111–139), BdWi-Verlag, Marburg.

Auch als gedruckte Ausgabe erhältlich:
ISBN 978-3-939864-02-8
<http://www.bdwi.de>



Dieses Dokument steht unter einer
Creative Commons BY-NC-ND 3.0 Deutschland Lizenz:



Namensnennung · Keine kommerzielle Nutzung · Keine Bearbeitung
<http://creativecommons.org/licenses/by-nc-nd/3.0/de/>

Meik Michalke, Oliver Naß, Anna Nitsche

Mehr Humor und keine Katze – Rankingprodukte Marke Bertelsmann

Mit ihren Rankingprodukten veröffentlicht die Bertelsmann Stiftung regelmäßig Studien, die Einfluss bis auf höchste politische Ebenen ausüben. Dabei beschreiben sie vermeintlich nicht nur einen Ist-Zustand, sondern formulieren ausführliche Handlungsanweisungen für die Zukunft. In diesem Beitrag erklären drei methodenerfahrene PsychologInnen wissenschaftliche Grundlagen der Ranking-Verfahren. An konkreten Beispielen aus aktuellen Rankings der Stiftung weisen sie nach, wo diese erhebliche methodische Mängel aufweisen. Als politischer Kompass weisen sie exakt ins Ungewisse.

Mit diesem Beitrag möchten die AutorInnen keinesfalls erklären, Rankings und ähnliche Methoden seien per se sinnfrei oder irreführend. Vielmehr möchten wir mit unseren Anmerkungen dazu beitragen, derartige Veröffentlichungen mit angebracht kritischem Bewusstsein betrachten zu können – es fehlt in der Öffentlichkeit am nötigen Grundlagenwissen, um die Qualität eines Rankings einschätzen und seine Interpretation beurteilen zu können. Wer dieses grundlegende Handwerkszeug nicht beherrscht, ist leicht durch Rankings manipulierbar. Nicht jede Tabelle mit Zahlen ist das Ergebnis gewissenhaft wissenschaftlicher Arbeit. In diesem Kapitel möchten wir uns daher mit den Fragen beschäftigen, was Rankings überhaupt sind, welche Aussagen sie ermöglichen (und welche nicht), und an welchen Stellen typische Stolperfallen lauern.

Das Problemfeld Ranking ist nicht Bertelsmann-spezifisch. Viele unserer kritischen Anmerkungen in diesem Kapitel lassen sich leicht auf Ranglisten und Umfrageergebnisse anderer Herkunft übertragen. Die Bertelsmann Stiftung ist allerdings Quell einiger ausgesprochen einflussreicher Rankings; dieser Einfluss erstreckt sich dabei mehr oder weniger transparent bis auf Regierungsebene.

Um hierfür nur ein Beispiel anzuführen: Der saarländische Ministerpräsident Peter Müller zitiert das Standortranking der Bundesländer, das die Bertelsmann Stiftung durchgeführt hat (Berthold et al. 2003), gerne unkritisch als Beleg für den

Erfolg der eigenen Politik. In seiner Regierungserklärung zu Beginn der 13. Legislaturperiode (Müller 2004) hieß es:

Wir haben unsere Position im Vergleich der Bundesländer deutlich verbessert: In der Arbeitslosenquote vom Platz 9 auf Platz 5 und im Standortranking vom Platz 10 auf Platz 6 aller Bundesländer.

In der 19. Sitzung der 13. Wahlperiode des Saarländischen Landtags (Landtag des Saarlandes 2005) zitiert er die Studie in längeren Passagen wörtlich und schließt mit:

Das ist der Befund der Bertelsmann Stiftung. Gewinner Saarland. Dieses Land ist Aufsteigerland. Dieses Land soll Aufsteigerland bleiben. Reden Sie deshalb dieses Land nicht schlecht!

Dem Vorwurf des Realitätsverlusts durch die Opposition begegnet er schließlich mit der Macht des Faktischen:

Es ist so: Wenn man es nicht sehen will, dann kann man es auch nicht sehen. Das ist das alte Prinzip: Da nicht sein kann, was nicht sein darf, ist es halt nicht, was Wirklichkeit ist. [...] Herr Kollege Jost, Zahlen lügen nicht. Auch wenn es Ihnen nicht gefällt: Diese Landesregierung ist erfolgreich.

In der vom Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst herausgegebenen Zeitschrift *aviso*, die vierteljährlich erscheint und kostenlos im Ministerium, an bayerischen Universitäten, Fachhochschulen sowie staatlichen Theatern und Museen ausliegt, erhält Detlef Müller-Böling, seit 1994 Leiter des Centrums für Hochschulentwicklung (CHE), sechs Seiten lang Gelegenheit, in eigenen Worten das hauseigene Hochschulranking als internationalen Tonangeber zu bewerben:

Mobile Studierende und Lehrende wünschen sich ein Informationsangebot, das mit ihnen Grenzen überschreitet. Mit der Ausdehnung des CHE-Hochschulrankings auf Österreich und die Schweiz ist ein erster Schritt zur Internationalisierung eines Angebots getan. Weitere Länder, die sich an dieser Methodik orientieren, werden sicherlich folgen, so dass sich der europäische Hochschulraum zukünftig vermehrt in Rankings abbilden wird. [...] Deutlich ist insofern erkennbar, dass Rankings auf absehbare Zeit ein spannendes Thema bleiben werden. Und noch eines ist sicher: Das CHE wird mit seinen Rankings einigen Anteil an dieser Entwicklung haben. (Müller-Böling 2006)

Im Oktober 2006 berief der Landtag Mecklenburg-Vorpommern den CHE-Chef direkt in seine Expertenkommission zur »Zukunft der Erziehung und Bildung unter Berücksichtigung des lebenslangen Lernens« (MVregio 2006), die dem Landtag u. a. konkrete Handlungsempfehlungen zur »Erhöhung von Autonomie und Flexibilität in Lehre und Forschung« sowie den »Finanzierungsperspektiven des Hochschulsystems« (Landtag Mecklenburg-Vorpommern 2006) unterbreiten soll.

Wir kommen im Folgenden immer wieder vom Allgemeinen zum Konkreten und werden die lauernden Stolperfallen in Rankings vornehmlich an Beispielen aus dem Bertelsmann-Portfolio exemplarisieren. Wenn Sie dieses Kapitel gelesen haben, hoffen wir, ihnen die grundlegenden Ideen hinter Rankings und Co. verdeutlicht zu haben, so dass sie beim nächsten sensationellen Fund eines Medienimperiums selbst mit kritischem Blick erfassen können, ob seriöse Information oder konzerninterne Motive ausschlaggebend für die Veröffentlichung waren.¹

Rankings sind ›in‹. Aus der sogenannten Neuen Sozialen Marktwirtschaft (selbst ein *Buzz-Word* mit Bertelsmann-Verbindungen)² sind sie nicht wegzudenken, dienen sie dort doch als Evaluationsinstrument und Stimmungsbarometer der Bereiche, in die durch mehr Wettbewerb Bewegung kommen soll. Rankings, Ratings und Indizes sollen den Stand der Dinge quantifizieren, sie bringen Ordnung in das Chaos unendlicher Wahlmöglichkeiten. Dabei versprechen uns ihre AutorInnen, aus den Rangfolgen auch die Ursachen für eine gute oder schlechte Platzierung ableiten zu können:

Landesspezifische Ursachen für Erfolg und Misserfolg lassen sich somit benennen, messen und vergleichen. Einzelne Bundesländer können also von den Erfolgen, Misserfolgen und Erfahrungen anderer Länder lernen. Die vorliegende Studie liefert dafür die Grundlage. (Berthold et al. 2003: 22)

Mitunter geben sie sogar Ratschläge für die Zukunft. So berechnet beispielsweise Bertelsmanns Standortranking der Bundesländer anhand von Bruttoinlandsprodukt, Wirtschaftswachstum, Arbeitslosigkeit, Erwerbstätigkeit, SozialhilfeempfängerInnenanteil und Zahl der nicht aufgeklärten Straftaten ›Erfolg‹ und ›Aktivität‹ der Länder, kürt Hamburg zum Spitzenreiter und weiß im Zwischenbericht auch weit mehr als 200 Seiten lang, was in nächster Zeit zu tun ist. Die Empfehlungen sind bemerkenswert konkret:

- 1 Selbstverständlich kann eine kurze Abhandlung wie diese nicht all die Bücher und Seminare abbilden, die zu diesem Thema geschrieben und gehalten werden. Aber vielleicht entdecken Sie ihre Lust an methodischen Fragen – wir können Sie nur ermuntern, auf eigene Faust weiter zu recherchieren! Einen guten Einstieg bietet beispielsweise Krämer (1997).
- 2 Die »Initiative Neue Soziale Marktwirtschaft (INSM)« wird von den Arbeitgeberverbänden der Metall- und Elektro-Industrie finanziert (<http://www.insm.de>). Ihr Botschafter Prof. Dr. Ulrich van Suntum, seit 2002 Geschäftsführender Direktor des »Centrums für angewandte Wirtschaftsforschung Münster (CAWM)«, ist für die Bertelsmann Stiftung seit 1996 u. a. als Autor des Internationalen Standort-Rankings tätig (van Suntum 2004; LobbyControl 2006). In der »Arbeitskommission zum Carl-Bertelsmann-Preis« saß 2005 INSM-Kurator Prof. Dr. Michael Hüther (Bertelsmann Stiftung 2005). Oswald Metzger (B90/Die Grünen), ebenfalls Kurator der Initiative, stellte 2003 eine Bertelsmann-Studie vor, die vor der »Bedrohung durch den demographischen Wandel« warnt (Bertelsmann Stiftung 2003).

Die beschlossene Ortsumgebung um Finkenwerder ist ein weiterer wichtiger Schritt, den Hamburger Airbus-Standort zu stärken. Hamburg steht im Wettbewerb mit Toulouse, wo Airbus ebenfalls seine Flugzeuge montiert. Reglementierungen, etwa bezüglich des Lärmschutzes, könnten Finkenwerder dabei zurückwerfen. Nach dem Gezerre um die Landebahnverlängerung hätte es eine fatale Signalwirkung, wenn Airbus in Hamburg nun neue Hürden in den Weg gestellt würden. (Berthold et al. 2005: 117)

Können die veröffentlichten Rankings ihrem Anspruch tatsächlich gerecht werden? Sind die Instrumente so nützlich, wie sie angepriesen werden? Wie müssen wir diesen Rangfolgen begegnen, die vermeintlich objektiv die Spreu vom Weizen trennen? Anders gefragt: Kann uns eine aus Wirtschaftswachstum und Sozialhilfeanteil berechnete Zahl wirklich dazu raten, in Finkenwerder den Lärmschutz zu vernachlässigen?

Rankings, Ratings, Indizes

Ein Ranking (zu deutsch »Rangordnung«) lässt sich definieren als das Ergebnis einer Sortierung von mehreren vergleichbaren Objekten, die mit einer vergleichenden Bewertung einhergeht. Rangordnungen ermöglichen es beispielsweise, komplexe Informationsangebote nach bestimmten Kriterien zu bewerten und eine Auswahl zu treffen. So werden beispielsweise in einer Suchmaschine die Webseiten automatisch nach ihrer Relevanz sortiert. Andere Ranglisten sind das Ergebnis von aufwendigen Vergleichsstudien oder Bestandteil der Hierarchie in sozialen Systemen (Wikipedia 2006d).

Bei der obigen Definition fällt auf, dass von »vergleichbaren Objekten« gesprochen wird. Hierin ist eine erste Sollbruchstelle vieler diskutierter Rankings zu sehen, auf die wir an späterer Stelle noch genauer eingehen werden.

Im wissenschaftlichen Kontext ist eine Rangfolge oft als sogenannte Rangskala (oder auch Ordinalskala) definiert (Bortz 1993: 21). Diese Skalenform ordnet den Eigenschaften von Objekten Zahlen zu, aus denen sich eine Rangreihe erstellen lässt bzw. ein Rangplatz ergibt. Zu beachten ist bei der Interpretation von Rangskalen, dass der Unterschied zwischen zwei Rangplätzen in der statistischen Wissenschaft nicht genau quantifiziert werden kann. Um das Verhältnis der Objekte zueinander zu beschreiben, sind definitionsgemäß lediglich Aussagen im Rahmen einer sogenannten ›monotonen Transformation‹ gültig. Nehmen wir an, sie hätten ihre drei Lieblingschokoladen in eine Rangreihe gebracht: Am liebsten mögen sie Vollmilch, gefolgt von Trauben-Nuss und schließlich Zartbitter. Aus dieser Reihe ließe sich korrekt ableiten, dass Sie Vollmilch lieber als Trauben-Nuss, und diese lieber als Zartbitter mögen. Daraus ergibt sich zwingend auch die Aussage, dass Sie

Vollmilch lieber als Zartbitter mögen. Es ist jedoch nicht erlaubt, ein Verhältnis abzuleiten, z. B. im Sinne von ›Sie mögen Vollmilch doppelt so gerne wie Zartbitter.‹

Zum ähnlichen Begriff des Ratings lesen wir im Internet-Lexikon Wikipedia: Rating (englisch für »Bewertung«, »Einschätzung«) bezeichnet allgemein ein Verfahren für die Einschätzung von Personen, Gegenständen oder Unternehmen (Wikipedia 2006e). Ratings sind in der Wissenschaft sehr beliebt. Hier sind meist Aussagen auf einer Skala zu bewerten, beispielsweise, wie sehr Sie der Aussage zustimmen würden, ein umgänglicher Mensch zu sein. Im Alltag wird dieses Verfahren z. B. bei den Fernsehsendern zur Ermittlung von Einschaltquoten und bei Produktbefragungen, -einführungen oder der Popularität von PolitikerInnen (als sogenannte Kopfnoten) eingesetzt.

Meist werden mehrere dieser Aussagen zur Beurteilung vorgelegt und daraus ein einziger Wert errechnet, ein sogenannter Index. Ein Index ist in der Statistik eine Maßzahl, bei der eine errechnete Zahl (meist aus verschiedenen zeitabhängigen Komponenten zusammengesetzt) als Abweichung zu einer zeitlich konstanten Bezugsgröße dargestellt wird (vgl. Wikipedia 2006b). Die Berechnungsmethoden von Indizes sind vielfältig und ihre Verwendung ist sehr weit verbreitet. Beispiele sind der Aktienindex (Dow Jones, DAX, usw.), der Preisindex zur Berechnung der Inflation oder der Pearl-Index, zur Errechnung der Erfolgsquote von Verhütungsmitteln, etc. Wir sehen, dass ein Index uns Informationen über den Stand einer gewissen Sachlage geben soll. Allerdings ist in der Regel ein spezifisches Wissen vonnöten, um die gegebene Information sinnvoll interpretieren zu können – der reine Rohwert eines Aktienindex ist ohne Hintergrundwissen wenig informativ. Ohne diese erforderlichen Kenntnisse ist eine Fehlinterpretation nicht mehr auszuschließen.

Zur Berechnung von Indizes, Ratings und Rankings können Daten unterschiedlicher Herkunft herangezogen werden. Meist finden entweder ›harte Zahlen‹ aus amtlichen Statistiken (z. B. die Anzahl verkaufter Computer) Verwendung, oder eher ›weiche Zahlen‹ aus Befragungen (z. B. ›Besitzen sie einen Computer?‹): Die Befragung ist eine wissenschaftliche Methode, um systematisch Informationen über Eigenschaften von Menschen zu gewinnen. Befragungen als Basis für die Analyse- und Dokumentationsarbeit sind ein wichtiges Instrument der empirischen Sozialwissenschaft, in der Psychologie, der Sprachwissenschaft (Sprachatlas, Mundartforschung), in der Volkskunde (Gewährsleute, Oral History), und in der Geschichtswissenschaft (Zeitzeugen, Technikgeschichte, Sozialgeschichte) (vgl. Wikipedia 2006a).

Die Übergänge zwischen diesen Verfahren sind fließend. Das liegt oftmals daran, dass Untersuchungen aus mehreren Teilen zusammengesetzt sind, so dass eine

Befragung, ein Rating und abschließend ein Präferenzranking in einem Instrument vereint werden. Dieses multimodale Vorgehen ist sowohl in der Meinungsumfrage wie auch in der empirischen Sozialforschung weit verbreitet. Auch im diagnostischen Kontext der Psychologie erscheint eine Integration der Verfahren oftmals als sinnvoll. Wie an den Definitionen gerade der Indizes zu erkennen ist, kann eine Einschätzung der Ergebnisse aufgrund eines der Verfahren lediglich dann erfolgen, wenn alle am Zustandekommen des Ergebnisses beteiligten Informationen zugänglich sind. Ein veranschaulichendes Beispiel aus dem Bereich Sport bietet die Fußball-Weltmeisterschaft 2006 in Deutschland. Für einen Unwissenden wäre der dritte Platz der deutschen Nationalmannschaft eine Überraschung gewesen, war das bundesdeutsche Team vor der WM doch nur auf Rang 19 der FIFA-Weltrangliste. Die fehlende Information, die die sportliche Leistung weniger ungewöhnlich erscheinen lässt, ist, dass Deutschland als Gastgeberland keine Qualifikation für die WM durchlaufen musste und daher kaum Weltranglistenpunkte sammelte, denn ein Freundschaftsspiel vergütet die FIFA nur mit einem Viertel der Punkte, die durch ein Qualifikationsspiel zu bekommen wären. Durch diese unterschiedliche Punktevergabe landet eine Mannschaft auch bei guter Leistung nur auf einem hinteren Weltranglistenplatz.

Gewichtungsfaktoren und Fehlinterpretationen

Wir sehen, dass ohne entscheidende Hintergrundinformation eine Fehlinterpretation aufgrund einer Rangreihe vorprogrammiert sein kann. Es ist daher grundsätzlich Vorsicht geboten, wenn die Informationen nicht zugänglich sind, die ein Ergebnis bedingen. Insbesondere bei der Berechnung von Indizes besteht die Gefahr, Ungleiches miteinander zu vermengen. Wir können aus drei Äpfeln, einem Liter Milch und zwei Würstchen problemlos einen Durchschnitt von »zwei Lebensmitteln« errechnen. Ob von diesem Durchschnittswert auch eine VegetarierIn satt wird, müssten wir raten, denn der Information »zwei Lebensmittel« ist nicht (mehr) zu entnehmen, ob es sich nicht vielleicht ausschließlich um Fleisch handelt. Eine VegetarierIn interessiert sich mehr für das Vorhandensein von fleischlosen Lebensmitteln. Um einen solchen Index für sie interessant zu machen, müssten die Eingangswerte dementsprechend unterschiedlich stark berücksichtigt, also gewichtet werden. Zum Beispiel könnte man das Vorhandensein von Obst mit 2, das von Milch mit 1 und das von Fleisch mit 0 multiplizieren – der so berechnete Lebensmittelindex (hier $2 \times 3 + 1 \times 1 + 0 \times 2 = 7$) würde nur die Dinge berücksichtigen, die VegetarierInnen wirklich essen. Diese Gewichtung der eingehenden Faktoren (hier: Obst, Milch und Fleisch im Verhältnis 2:1:0) hat also großen Einfluss auf das Endergebnis, d. h. es genügt zur Interpretation von Indexwerten nicht zu

wissen, welche Variablen berücksichtigt wurden – man muss auch wissen, mit welchem Gewicht sie in die Berechnung eingehen.

Im Beispiel spielt die Menge an Fleisch durch die Gewichtung mit Null gar keine Rolle mehr für das Ergebnis, obwohl sie streng genommen in die Berechnung eingeht, eine Veränderung der Obstwerte würde das Endergebnis dagegen durch die doppelte Gewichtung sehr stark beeinflussen. Dennoch bliebe bei diesem – der Verständlichkeit wegen arg konstruierten – Beispiel das Problem bestehen, ungleiche Dinge miteinander verrechnet zu haben: Drei Früchte, einen Liter Flüssigkeit und zwei abgepackte Einheiten Fleischmasse. Damit bliebe auch zu diskutieren, wie sinnvoll es ist, das Ergebnis überhaupt zu interpretieren. Dies veranschaulicht, wie leicht eine Formel tatsächliche Unterschiede unter den Tisch fallen lässt und welchen Einfluss eine Gewichtung auf das Ergebnis haben kann.

Eine spannende Frage ist hierbei natürlich, wie die Gewichte zustande kommen – im Beispiel haben wir sie nach unserem Gutdünken willkürlich festgelegt. Wenn ein Index wissenschaftlichen Kriterien entsprechen soll, können wir dies selbstverständlich nicht tun. Üblich ist vielmehr, über Korrelationen zu berechnen, wie stark einzelne Faktoren mit dem Gesamtergebnis zusammenhängen und die gefundenen Korrelationskoeffizienten als Schätzung für das Gewicht zu verwenden. Gerade an diesem zentralen Punkt herrscht bei vielen veröffentlichten Rankings und Indizes vollkommene Intransparenz. Streng genommen sind die Ergebnisse damit aus wissenschaftlicher Sicht wertlos, weil eine Geheimhaltung unmöglich macht, das Ergebnis auf Fehler zu prüfen. Auch das Standortranking der Bundesländer der Bertelsmann Stiftung (Berthold et al. 2003; 2005) baut auf Indizes, die nach ihrer Berechnung in eine Rangreihe gebracht werden. Sie bilden sozusagen die zentralen Eckpfeiler des Projekts, dessen statische Fähigkeiten damit gut beschrieben sein dürften. Zur Entstehung der sogenannten ›Erfolgs-‹ und ›Aktivitätsindizes‹ wird erklärt:

Wie sich die einzelnen Länder entwickeln, wird anhand der drei Bereiche ›Einkommen‹ (mit den beiden Zielgrößen Bruttoinlandsprodukt und Wirtschaftswachstum), ›Beschäftigung‹ (Arbeitslosigkeit und Erwerbstätigkeit) sowie ›Sicherheit‹ (Sozialhilfeempfängeranteil für die soziale Sicherheit und Zahl der nicht aufgeklärten Straftaten für die innere Sicherheit) untersucht.

Zur Gewichtung der drei Faktoren heißt es weiter lapidar:

Wanderungsbewegungen und Umfrageerhebungen legen nahe, die drei Bereiche mit 1:1:0,25 zu gewichten. (Berthold et al. 2005: 12)

Eigentlich geht es also fast ausschließlich um Einkommen und Beschäftigung. Das Übergewicht dieser Faktoren veranschaulicht der Ranking-Abräumer Hamburg, der trotz seines vorletzten Platzes in puncto Sicherheit in der Gesamtwertung vorne

bleibt. Was das im Einzelnen für ›Wanderungsbewegungen‹ sind und um welche ›Umfragen‹ es sich handelt, erfahren wir an dieser Stelle nicht.

Bei derart schönen Gewichtungsfaktoren ist jedenfalls davon auszugehen, dass sie über den Daumen gepeilt wurden – wären sie das Ergebnis statistischer Berechnungen, wäre es äußerst unwahrscheinlich, nicht drei ›krumme Zahlen‹ erhalten zu haben. Es ist zudem wirklich spannend, einmal darüber zu sinnieren, welche inhaltliche Aussagekraft z. B. eine Zahl hat, die aus der Anzahl der SozialhilfeempfängerInnen und der Quote³ unaufgeklärter Verbrechen berechnet wird – laut Bertelsmann bildet sie die Sicherheit eines Bundeslandes ab. Überhaupt ist schwer nachvollziehbar, warum genau die drei Faktoren Einkommen, Beschäftigung und Sicherheit verwendet werden, und warum sie sich wie angegeben zusammensetzen. In ihren Erklärungen zum methodischen Ansatz behaupten die Autoren immerhin, die Indizes stünden »in einem theoretisch plausiblen und empirisch abgesicherten Ursache-Wirkungs-Zusammenhang« (Berthold et al. 2003: 22). Die Faktoren machen dabei jedoch nicht den Eindruck, das Ergebnis einer vorangegangenen empirischen Analyse zu sein. Dagegen spricht z. B., dass sowohl Arbeitslosigkeit und Erwerbstätigkeit als auch Bruttoinlandsprodukt und Wirtschaftswachstum nicht unabhängig voneinander sein dürften. Wenn zwei Größen stark miteinander zusammenhängen, erklären sie gemeinsam nicht viel mehr, als eine der beiden es alleine schon täte – man würde daher aus Gründen der Testökonomie auf eine verzichten. Falls diese Überschneidungen nicht berücksichtigt werden, vergrößert sich womöglich der Einfluss eines Faktors künstlich, weil vereinfacht ausgedrückt ein und dieselbe Eigenschaft doppelt in die Rechnung eingeht.⁴ Allerdings ist schon viel früher Obacht angebracht: Was bedeutet eigentlich »theoretisch plausibel«? Aus welcher Theorie wurden die Plausibilitäten abgeleitet? Es ist ja beispielsweise leicht nachzuvollziehen, dass ein Globalisierungskritiker andere Zusammenhänge auf dem Weltmarkt als »theoretisch plausibel« annimmt als eine überzeugte Globalisierungsbefürworterin.

3 Im veröffentlichten Standortranking ist abwechselnd von ›Anzahl‹ oder ›Quote‹ unaufgeklärter Straftaten die Rede. Diese begriffliche Ungenauigkeit erschwert eine Interpretation zusätzlich, denn zwei unaufgeklärte Verbrechen sind sicher ein geringer Wert; falls es jedoch überhaupt die einzigen Verbrechen gewesen sind, läge die Quote bei 100% und damit sehr hoch. Wir wissen auch nicht, ob es sich bei den Verbrechen in einem Land mehrheitlich um Steuerhinterziehung, Fahrerflucht oder Mord handelt.

4 Hierbei handelt es sich zwangsläufig um beispielhafte Spekulationen unsererseits, da im Rahmen des Rankings die wirkliche (Un-)Abhängigkeit der Faktoren nicht untersucht wurde.

Vorformulierbare Ergebnisse

Während sich für den Erfolgsindex noch diverse Statistiken für eine wie auch immer durchgeführte Rechnung heranziehen lassen, möchte der Aktivitätsindex das politische Handeln der Länder messen. Nach einem uns nicht bekannten Schema, das sich auf die undurchsichtige ›theoretische Plausibilität‹ stützt, werden hierbei politische Entscheidungen irgendwie in Punkte übersetzt – Willkür und Absurdität dieses Unterfangens liegen auf der Hand. Die Konstrukteure des Aktivitätsindex stellen diesen dar als einen »zum Erfolgsindex spiegelbildlich konstruierten Indikator der landespolitischen Aktivitäten, die die Position der Länder im Erfolgsindex beeinflussen« (Berthold et al. 2003: 35ff). Diese Aktivitäten seien demnach die Ursache, der gemessene Erfolg die Wirkung. Wieder bleibt vollkommen schleierhaft, auf welche Weise sie zu den berücksichtigten Zielgrößen ihrer »ökonometrischen Schätzung« kamen. So entnehmen wir den Tabellen des Berichts (Berthold et al. 2003: 42–43) lediglich im Ergebnis, dass öffentliche Beschäftigung positiv mit einem geschätzten Gewicht von 18,0 % auf den SozialhilfeempfängerInnenanteil wirke, jedoch negativ mit 16,2 % auf das Bruttoinlandsprodukt, und dass das Bruttoinlandsprodukt selbst wieder ein positiver Wirkfaktor (14,8 %) für die nicht aufgeklärten Straftaten sei.

Offensichtlich muss den Ranking-Autoren also sogar bewusst gewesen sein, dass die Faktoren in verschiedener Weise zusammenhängen; trotzdem war es ihnen nicht der Mühe wert, die gegenseitige Abhängigkeit ihrer Zielgrößen zu untersuchen. Im Gegenteil: Da die im Aktivitätsindex aufgeführten Faktoren als Ursache für die Werte des Erfolgsindex angesehen werden, bedeutet dies im Klartext, dass das Bruttoinlandsprodukt herangezogen wird, um sich selbst vorherzusagen. Die *Spiegelbildlichkeit* der Indizes ist damit in gewisser Weise wörtlich zu verstehen.

Manche Aktivitäten fallen im Sinne der Ranking-Autoren durchweg positiv oder negativ auf: Patentanmeldungen wirken ihren Schätzungen nach grundsätzlich positiv, und zwar sowohl auf die Arbeitslosigkeit, den Erwerbstätigenanteil, das Bruttoinlandsprodukt, das Wirtschaftswachstum und sogar den Anteil der SozialhilfeempfängerInnen, wohingegen die Höhe der Sozialhilfe bzw. Ausgaben für Sozialhilfe in den Zielgrößen Arbeitslosigkeit, Bruttoinlandsprodukt, Wirtschaftswachstum und ungeklärte Verbrechen ausschließlich negativen Einfluss haben. Der Effekt: Wer Patentanmeldungen forciert und an der Sozialhilfe spart, kann sich auf allen Skalen der Bundesländer-Studie nur verbessern.

Eine weitere Überlegung lässt das gesamte Konstrukt ins Wanken geraten: Die Ranking-Autoren machen zu Anfang der Studie deutlich, dass alle Länder direkt vergleichbar sind, schließlich stünden sie alle im selben Wettbewerb miteinander (Berthold et al. 2003: 25). Dennoch heben sie selbst hervor, dass eine Reihe

deutlicher Unterschiede zwischen den Bundesländern existiert. So beschreiben sie z. B. ein starkes Ost-West-Gefälle, insbesondere bei den Arbeitslosenquoten (die Zahl der Beispiele ist beliebig erweiterbar). Hieraus ließe sich schlüssig ableiten, ostdeutsche Länder müssten sich in ihren Aktivitäten vergleichsweise stärker in der Arbeitsmarktpolitik engagieren als westdeutsche, bei denen dieses Problem deutlich weniger gravierend auftritt. Trotzdem werden alle Aktivitäten und Erfolge für alle Bundesländer im Ranking grundsätzlich gleich stark gewichtet: Völlig ungeachtet landesspezifischer Gegebenheiten geht so auch der Bevölkerungsanteil in Großstädten mit 18,1 % negativ in den Aktivitätsfaktor für unaufgeklärte Verbrechen ein.⁵ Bei dieser Berechnung ist es natürlich kein Wunder, dass sich Berlin, Hamburg und Bremen auf den hinteren Rängen wiederfinden.

Aus dieser Perspektive wird offensichtlich, dass das ›Eine Formel für alle‹-Vorgehen schlicht nicht praktikabel ist. Über die Unbrauchbarkeit dieses Ansatzes kann auch das Wettbewerbsargument nicht hinwegtäuschen. Wenn aber die Eigenarten der Bundesländer ins Blickfeld geraten, steht gleichzeitig ihre direkte Vergleichbarkeit zur Diskussion: Das gesamte Ranking verliert seine methodische Basis. Die Studie bemängelt dagegen fehlende Bewegung, wenn sie die Aktivitätsindizes im Verlauf der Zeit miteinander vergleicht:

Waren schon beim Erfolgsindex die Veränderungen nicht extrem, sind sie im Aktivitätsindex noch geringer ausgefallen. Hieraus ergibt sich das Bild eines wenig intensiven Standortwettbewerbs zwischen den Bundesländern.

Selbstkritik ist nicht die Stärke der Verfasser, denn eine alternative Erklärung dieses Befundes, die an dieser Stelle in jedem wissenschaftlichen Artikel diskutiert werden müsste, wäre mangelnde Güte des Messinstruments.

Ein Index, bei dem sich einzelne Unterfaktoren gegenseitig verstärken und wieder aufheben, ohne dass diese Zusammenhänge ordentlich untersucht werden, ist vielleicht eher konstruktionsbedingt unfähig, die gewünschten Veränderungen zu erfassen. Berthold et al. (2003) kommen jedoch zu einem anderen Schluss, zu dem sie überdies keine Alternative zulassen:

Für die anstehende Reform des Föderalismus in Deutschland kann das nur heißen, die Gestaltungsspielräume der Länder zu erhöhen, um so auch mehr Wettbewerb zu ermöglichen.

5 Die Ranking-Autoren begründen dies mit dem statistischen Befund, dass die Kriminalitätsrate in Großstädten durchschnittlich höher liegt als auf dem Land. Bundesländer wie z. B. Hamburg bestehen aber eigentlich zu 100 % aus Großstadt – für die Ranking-Autoren ist dies dennoch kein Grund, von einer Pauschalberechnung abzusehen.

Simplifizierung und Propaganda

Es besteht hier in der Sache kein großer Unterschied, ob die Intelligenz einer Einzelperson, die Qualität einer Hochschule oder der Erfolg einer Industrienation gemessen werden soll. In allen Fällen müssen wir zunächst begründet und so exakt wie möglich definieren, was wir mit Intelligenz respektive Qualität oder Erfolg überhaupt meinen. Haben Sie schon einmal versucht, Intelligenz zu definieren? Falls Sie in dieser kniffligen Frage aushelfen können, wären ihnen die in der Psychologie Forschenden sicher für einen zweckdienlichen Hinweis dankbar, denn diese erst einmal so trivial scheinende Frage führte hier zu einer wachsenden Anzahl von Modellen und regalfüllenden Diskussionen. Jedes Intelligenzmodell bedarf zur Messung selbstverständlich anderer Verfahren, und so gilt noch immer als ehrlichste Definition: »Intelligenz ist das, was ein (bestimmter) Intelligenztest misst.«

Kehren wir also noch einmal zurück zur grundlegenden Frage: Was ist überhaupt ›Erfolg‹? Die Bertelsmann Stiftung weiß es, denn wie bereits erläutert, berechnet sie einen ›Erfolgsindex‹ für Industrienationen und Bundesländer, der fast ausschließlich Einkommen und Beschäftigung berücksichtigt. Aber ließen sich nicht auch andere Faktoren finden, die mindestens gleichsam plausibel als Erfolg definiert werden könnten? Immerhin befinden auch die Autoren des Rankings:

Um zur Aggregation in einem Erfolgsindex geeignet zu sein, müssen Zielgrößen die Lebens- und Standortqualität der Bundesländer möglichst umfassend und repräsentativ abbilden. (Berthold et al. 2003: 27)

Mithin soll also auch die Lebensqualität umfassend abgebildet werden, und man könnte begründet der Auffassung sein, dass dies allein über Einkommen, Beschäftigung und einen verschwindenden Teil Sicherheit nicht gelingt. Wie wäre es beispielsweise mit

- dem prozentualen Anteil von Kindern sogenannter ›bildungsferner Schichten‹ an Hochschulen
- der Anzahl ehrenamtlich engagierter BürgerInnen
- dem durchschnittlichen Geräuschpegel in den Städten
- der Anzahl von Kommunen, deren Verwaltung auf freier Software basiert
- dem Angebot an Sportvereinen, Jugend- und Kulturzentren
- der Zufriedenheit mit den Fahrplänen des ÖPNV
- der durchschnittlichen Entfernung der Wohnungen von Kindern zu ihren Schulen
- der Dichte an instandgehaltenen Sehenswürdigkeiten
- der jährlichen Pro-Kopf-Zahl von Kino-, Theater- und Konzertbesuchen
- der Zahl frei verfügbarer Bücher in öffentlichen Bibliotheken

- dem Stellenwert eines künstlerisch-kreativen Undergrounds
- der öffentlichen Akzeptanz nicht-ehelicher oder gleichgeschlechtlicher Lebensgemeinschaften
- Sterblichkeit und Geburtenrate
- Wasserqualität
- Prozentsatz alternativ gewonnener Energie
- Wahlbeteiligung der Bevölkerung
- ...

Da letztendlich nicht überprüft werden kann, wie sich die einzelnen Werte tatsächlich zusammensetzen, womit auch die Interpretation der Daten im weiteren Verlauf der Studie eigentlich obsolet wird, können ihre Autoren dementsprechend nach eigenem Gutdünken erklären, welche politische Maßnahme ihnen besonders gefiel und was sie sich für die Zukunft an Entscheidungen erhoffen. Mit Empirie und Wissenschaft hat dies nichts zu tun. Diese Simplifizierung komplexer Sachverhalte und Zusammenhänge sowie ihre ›alternativlose‹ Interpretation im Sinne einer nebulösen Theorie kann nur mehr als Propaganda⁶ bezeichnet werden.

Wie eingangs schon erwähnt, hängt der Boom von Rankings eng mit der westlichen Wirtschaftspolitik zusammen. Diese nehmen wir derzeit wahr als dominiert von einem gemeinhin als Wettbewerb bezeichneten Selbstverständnis des Gegeneinanders. Es gilt, sich von MitbewerberInnen abzusetzen, z. B. durch Andersartigkeit, also das Darstellen einer Alternative. Wenn es geglückt ist, eine Lösung oder ein Produkt ›anders‹ herzustellen, muss im nächsten Schritt die potentielle Kundschaft davon überzeugt werden, dass die eigene Alternative auch die bessere ist, sonst bestünde für diese außer Zufall, Neugier oder Langeweile wenig Anlass zu einer (Um-)Entscheidung zugunsten des entsprechenden Angebots.

Nehmen wir ein vereinfachendes Beispiel: Wie würden Sie vorgehen, wenn Sie zum ersten Mal Milch kaufen sollten? Ein Dutzend unterschiedlicher Sorten und Marken verwirren Sie. Vermutlich würden Sie sich in ihrem Bekanntenkreis nach deren Erfahrungen mit Milch informieren. Wenn ihnen dabei eine Mehrheit zum Kauf der Frischmilch ›Glückliche Kuh‹ rät, haben Sie quasi ein eigenes Alltags-Ranking vorliegen, und vermutlich wird es auch die Gestalt ihres Einkaufzettels bestimmen – schließlich war das Ihr zentrales Anliegen. Die Beschäftigung mit einem Ranking bekommt also besonderes Gewicht, wenn Sie sich in einem Entscheidungsprozess, einer Meinungsbildung befinden. Aber wonach haben Sie ei-

6 »Unter Propaganda versteht man die gezielte und organisierte Verbreitung einer Nachricht oder Ideologie, meist im politischen Umfeld. Der Propagandatreibende hat das Ziel, das soziale Handeln und Denken anderer im eigenen Sinn zu beeinflussen. Eine jeweils konkret verbreitete Information kann richtig oder falsch sein.« (Wikipedia 2006c)

gentlich genau gefragt? Welche Milch am besten schmeckt? Welche sich am längsten hält? Welche am günstigsten zu bekommen ist? Welche die hübscheste Verpackung hat oder vom renommiertesten Betrieb stammt? Letztere sind vielleicht Fragen, die Ihnen jetzt nicht in den Sinn gekommen wären, weil Sie sich wenig brauchbare Information davon versprechen. Trotzdem ist es natürlich problemlos möglich, eine Umfrage nur darauf aufzubauen und das Ergebnis als ›Großes Milch-Ranking‹ zu veröffentlichen. Wenn Sie selbst im Milchgeschäft tätig sind, kann das durchaus lukrativ sein. Sie hätten in diesem Fall natürlich wenig Interesse daran, ein Ranking zu veröffentlichen, das eine KonkurrentIn anführt. Wie ließe sich dies verhindern? Sie könnten z. B. die Antwortmöglichkeiten vorgeben und dabei ernsthafte Konkurrenz unerwähnt lassen.

Sie finden das abstrus? Bertelsmann nicht: Das Centrum für Hochschulentwicklung (CHE) veröffentlichte sowohl 1998 also auch 2003 die Ergebnisse beim Meinungsforschungsinstitut forsa in Auftrag gegebener Umfragen, nach denen die Mehrheit der Studierenden in Deutschland die Einführung von Studiengebühren befürworten würde. Es ist sicher kein Zufall, dass diese Umfrageergebnisse zu Zeiten präsentiert wurden, in denen Abertausende von Studierenden durch Proteste auf sich aufmerksam machten. Wie sich herausstellte, sollten in beiden Umfragen lediglich verschiedene Gebührenmodelle bewertet werden – eine Möglichkeit, sich explizit gegen Gebühren auszusprechen, war überhaupt nicht vorhanden (Struben 2004).

Plausibilität und Interpretation

Rankings beginnen also – wie eigentlich alle wissenschaftlichen Studien – mit einer anfänglichen Fragestellung. Wie unser Milch-Beispiel veranschaulichte, kann es bereits von großer Bedeutung sein, wer die Fragestellung aufwirft. Ein sehr interessanter Punkt an Rankings ist daraufhin, wonach und in welcher Form genau gefragt wurde. Die exakte Formulierung einer Frage, oder ihre Einbettung in einen inhaltlichen Kontext, hat dabei größeren Einfluss auf das Ergebnis als man allgemein vermutet. So präsentierten Loftus und Palmer (1974) in einer psychologischen Untersuchung ihren ProbandInnen einen kurzen Film, in dem ein Autounfall zu sehen ist. Anschließend baten sie um eine Schätzung der Geschwindigkeit der Fahrzeuge. Dabei variierten sie das Verb, mit dem sie den Unfallhergang bezeichneten: Nach zufälligem Muster wurden die Versuchspersonen also gefragt, mit welcher Geschwindigkeit die Wagen ineinander »fahren«, »rasten« oder »krachten«.⁷ Obwohl alle ProbandInnen exakt identisches Filmmaterial gesehen hatten, schätzten diejenigen, deren Frage ein schneller oder drastischer klingendes Verb enthielt, im Durchschnitt auch deutlich höhere Geschwindigkeiten. Stellen Sie sich

vor, Sie wären Rechtsanwältin und sollten in einem ähnlichen Fall den Unfallfahrer vertreten – wie würden Sie nach Kenntnis dieser Studie wohl ihre Fragen an eine Zeugin formulieren, um ihren Mandanten möglichst gut dastehen zu lassen?

Der Erkenntnisgewinn dieses relativ einfachen, aber eindrucksvollen Experiments ist enorm. Da nur eine Variable (das Verb) verändert wurde, und dies auch noch zufällig, kann der gefundene Effekt auf die Form der Frage zurückgeführt werden. Dies ist für uns in diesem Kapitel nicht nur konkret von Interesse, weil wir nun wissen, dass auch Formulierungen nachweislich Antworten beeinflussen können. Es ist darüber hinaus auch ein anschauliches Beispiel dafür, wie unter methodisch günstigen Umständen auch recht sicher Ursache-Wirkungs-Zusammenhänge überprüft werden können. Anders ausgedrückt ist der Nachweis von kausalen Zusammenhängen nur in einem Experiment möglich, in dem gezielt einzelne als wirkungsvoll vermutete Komponenten manipuliert werden und die daraus resultierenden Veränderungen beobachtbar sind. Ohne die Möglichkeit zur experimentellen Veränderung der vermeintlichen Wirkkomponenten können wir nicht sicher sein, ob ein beobachteter Effekt tatsächlich von einem bestimmten zuvor aufgetretenen Ereignis verursacht wurde, oder nicht z. B. beide Größen von einem dritten, unbekanntem Ereignis gleichermaßen beeinflusst werden.

Bei Rankings fällt der Nachweis dieser Zusammenhänge schwerer. Wir müssen sogar fragen, ob er auf diese Weise überhaupt möglich ist. Wenn jedoch Universitäten, Bundesländer oder Industrienationen einem Ranking unterliegen, begnügen sich die AutorInnen in der Regel mit ›plausiblen‹ Schlussfolgerungen. Die augenscheinlichen Zusammenhänge werden mit gesundem Menschenverstand interpretiert. So wird meist am interessantesten Punkt – der Interpretation der Ergebnisse – mit verklärender Plausibilität zunichte gemacht, was zuvor zum Teil mit beachtlicher mathematischer Präzision errechnet wurde. Der Verweis auf Zahlen mit ein paar Nachkommastellen erzeugt gewissermaßen das Image von Wissenschaftlichkeit, obgleich die inhaltliche Auslegung bei genauerem Hinsehen nicht viel mehr bleibt als reine Spekulation. Drews (2005) spricht in seiner Doktorarbeit »Gestaltungsmöglichkeiten und Aussagekraft des Rankings von Standorten – unter besonderer Berücksichtigung des Bundesländer-Standortrankings der Bertelsmann Stiftung« ebenfalls von »Ursache-Wirkungs-Zusammenhängen«, die sich aus den »Regressionsergebnissen« des Rankings ablesen ließen (229). Im veröffentlichten Standortranking von 2003, dessen Co-Autor Drews war (Berthold et al. 2003), ist der Begriff »Regression« allerdings nicht auffindbar.⁸

7 »How fast were the cars going when they (contacted/hit/bumped/collided/smashed) each other?«

8 Wir durchsuchten den Volltext der Studie (http://www.bertelsmann-stiftung.de/bst/de/media/xcms_bst_dms_17364_17365_2.pdf) nach »Regression« und ähnlichen Begriffen.

Wie Grundlagenliteratur in Statistik (vgl. z. B. Bortz 1993: 173ff) erklärt, lassen sich aus korrelativen Zusammenhängen (und bei Regressionen handelt es sich etwas vereinfacht ausgedrückt um solche) streng genommen auch überhaupt keine Kausalzusammenhänge ableiten. Dies ist an einem Beispiel schnell verdeutlicht: Im Sommer werden mehr T-Shirts getragen als im Winter. Wenn Sie dazu entsprechende Daten erheben würden, ergäbe sich beispielsweise sicherlich eine statistisch bedeutsame Korrelation zwischen der Ärmellänge der Oberbekleidung und dem Kalendermonat. Dies bedeutet allerdings nicht, dass sie ein T-Shirt tragen, weil gerade August ist – sie tragen es, weil es warm ist; die Temperatur wurde bei der Rechnung jedoch gar nicht beachtet, es liegen dazu vielleicht nicht einmal Daten vor. Korrelationen stellen also lediglich (aber immerhin!) die Stärke eines Zusammenhangs dar, lassen aber keine Ursachenbestimmung zu. Sie lassen eigentlich nicht einmal eine Richtung des Zusammenhangs erkennen – wir könnten aus der obigen Korrelation genauso gut (und falsch) ablesen, dass aus einem Monat deswegen ein August wurde, weil viele Menschen T-Shirts getragen haben. Dass diese Ursacheninterpretation nicht zutreffen kann, leuchtet sofort ein; nur sind derartige Trugschlüsse eben nicht immer so offensichtlich.

Messbarkeit und Validierung

Machen wir uns noch einmal bewusst, dass ein Ranking im Prinzip ein Messinstrument darstellt. Bevor es sich lohnt, die Ergebnisse einer speziellen Messung zu diskutieren, muss geklärt werden, ob denn dem Instrument überhaupt getraut werden kann. Erlauben Sie uns an dieser Stelle einen etwas abstrakten Exkurs. Was einer Rangreihe zugrunde liegt, sind gemessene Daten: Auch wenn wir unsere FreundInnen der Größe nach in einer Reihe aufstellen möchten, müssen wir erst auf geeignete Weise messen, wie groß sie sind. Tut es in diesem Fall vielleicht noch grobes Augenmaß, ist beim Vergleich von Bundesländern oder Universitäten eine abstraktere Form der Messung vonnöten. In der psychologischen Teildisziplin der Diagnostik hat sich über die Jahrzehnte ein sehr robustes Prozedere zur Entwicklung von Messverfahren etabliert. Es werden, grundsätzlich ausgehend von einer begründeten wissenschaftlich-theoretischen Annahme, meist Vorabversionen eines Verfahrens erarbeitet und an einer Unzahl von Versuchspersonen getestet. Durch Vergleiche über mehrere Versuchsreihen und den Einsatz weiterer, bereits anerkannter Verfahren muss das neue Instrument z. B. zeigen, dass es auch tatsächlich das misst, was es zu messen vorgibt (Validität), und dass es dies auch mit der nötigen Zuverlässigkeit tut, also nicht bei jeder Messung etwas anderes⁹ herauskommt (Reliabilität) – bezogen auf die zuvor beispielhaft erwähnte Aufstellung nach

Körpergröße sollte die Reihenfolge unserer FreundInnen immer wieder gleich ausfallen, egal wie oft wir diese Aufstellung vornehmen.

Nur Verfahren, die auf dieser wissenschaftlichen Bühne wiederholt in fraglichen Szenarien ihre Verlässlichkeit unter Beweis stellen konnten, bekommen den Segen für einen Einsatz in der Praxis. Allein die Validität eines Instrumentes ist von grundlegender Bedeutung für seine Nützlichkeit. Wenn wir feststellen wollen, ob wir gesund oder krank sind, das vermeintliche Fieberthermometer jedoch in Wirklichkeit die Außentemperatur anzeigt, treffen wir unsere Entscheidung aufgrund falscher Annahmen. Zwei Stunden später wundern wir uns vielleicht, warum wir auf dem Weg zu einem Meeting einer Ohnmacht erliegen.

Die Bertelsmann Stiftung hält es dagegen nicht für nötig, Belege für die Güte ihrer Verfahren zu veröffentlichen. Doch sind das Festhalten an Überzeugungen, die uns direkt einleuchten (Plausibilität) sowie das Vermehren von Wissen durch Beobachtung und Erfahrung (Empirie) nicht das gleiche. So war es vor längerer Zeit durchaus plausibel davon auszugehen, dass die Erde eine Scheibe ist und sich das Universum um sie dreht. Der frühe Empiriker Galilei konnte durch seine Beobachtungen zeigen, dass diese Annahme nicht den Tatsachen entsprach. Uns allen ist bekannt, welchen Preis er dafür zu bezahlen hatte, das Weltbild seiner Zeit zu korrigieren, mit dem sich die Kirche an der Macht hielt.

Mangelhafte Validierung transformiert gewissermaßen auch die ›empirisch abgesicherten‹ Ergebnisse von Rankings zu einer Glaubensfrage. Dies ist nicht nur zu bedauern, sondern in den vorliegenden Fällen schlichtweg nicht hinnehmbar. Das beschriebene strenge Vorgehen in der experimentellen Psychologie begründet sich nämlich durch die Verantwortung gegenüber allen Mitmenschen, die mit Veröffentlichung eines solchen Verfahrens einhergeht. Auch in der Medizin wäre es unverantwortlich, Diagnoseverfahren zu empfehlen, ohne ihre Zuverlässigkeit nachgewiesen zu haben. Demgegenüber scheint die Bertelsmann Stiftung mit ihren Ranking-Veröffentlichungen zwar darum bemüht, auf das Verhalten von Studierenden, Hochschulleitungen und PolitikerInnen bis auf internationaler Ebene einzuwirken, erbringt jedoch keine nachvollziehbaren und tatsächlich überzeugenden Nachweise für die Vertrauenswürdigkeit der Instrumente, mit denen sie ihren Forderungskatalog erzeugt. Vielmehr sind ihre Studien und die dazugehörigen Pressemitteilungen mit Begriffen der Kategorie ›alternativlos‹ oder ›unbestreitbar‹

9 Genau genommen soll sich das Testergebnis natürlich schon verändern, wenn sich auch die zu messende Größe verändert hat. Ein Fieberthermometer soll also selbstverständlich etwas anderes anzeigen, wenn die Temperatur seit der letzten Messung nachgelassen hat. Aber das Gerät wäre nachvollziehbar nutzlos, wenn wiederholte Messversuche bei unveränderter Temperatur immer wieder deutlich andere Ergebnisse produzieren würden.

durchzogen, was die fehlende objektive Überprüfbarkeit der Befunde mit einer fast schon religiös anmutenden ›einzig wahren Lehre‹ zu kompensieren sucht. Dabei gibt es durchaus auch statistische Indizien, die den hier von der Bertelsmann Stiftung geforderten Liberalismus als leeres Versprechen einer besseren Zukunft erscheinen lassen (Berger und Stein 2004; 2005).

Wir sehen uns auch auf internationaler Ebene mit Rangreihen und Platzierungen konfrontiert, denen oft ein hoher Stellenwert zugeordnet wird, gerade wenn die eigene Gruppe, Ethnie oder Nation positiv abgeschnitten hat – beispielsweise der ›eigene‹ Medaillenspiegel bei Olympischen Spielen. Ein Beispiel für internationale Rangreihen im Alltag sind die Angaben im jährlich erscheinenden Fischer-Weltalmanach (Berié et al. 2003). Hier sind diverse Daten zu allen Staaten der Erde zusammengefasst und oftmals in Rangreihen dargestellt. So belegt z. B. Bosnien-Herzegowina den 125. Weltrang in der Fläche des Landes mit 51.129 Quadratkilometern, und bei der Einwohnerzahl mit 4.060.000 den 120. Rang. Während die Flächenangabe mit heutigen Mitteln realistisch einzuschätzen ist, stellt die Angabe zur Bevölkerung lediglich einen Näherungswert dar. Selbst bei so einer alltäglichen Statistik wie der Bevölkerungszahl werden in gängiger Praxis ungefähre Zahlen genannt, da eine genaue Berechnung bei den jährlichen Schwankungen nicht durchführbar ist. Eine alle zehn Jahre stattfindende Volkszählung überzeugt nicht als probates Mittel für genaue Angaben. So ist die Rangreihe der Bevölkerungszahlen schon deshalb oftmals irreführend, weil sich enorme Dehnungen und Stauchungen in der Skala befinden. Während China und Indien mit jeweils ca. 1 Milliarde Menschen die beiden vordersten Plätze belegen, sind auf Rang 3 mit ca. 285 Millionen Einwohnern die USA – das ist weniger als ein Drittel. Bei den Plätzen 120 (Bosnien-Herzegowina: 4,06 Millionen) und 121 (Costa Rica: 3,8 Millionen) beträgt der Unterschied gerade einmal 260.000. Die Rangwerte bilden also die tatsächlichen Verhältnisse der Bevölkerungszahlen untereinander nur sehr schlecht ab.

Analog zu dem Standortranking der Bundesländer basiert das Internationale Standort-Ranking der Bertelsmann Stiftung (Hafemann/van Suntum 2004) auf zwei uns bekannt vorkommenden Indikatoren, einem Erfolgs- und einem Aktivitätsindex. Sie bestehen in bekannter Manier aus zwei Zielgrößen: ›Arbeitsmarkt‹ (Arbeitslosenquote und Erwerbstätigenzuwachs) und ›Wachstum‹ (Bruttoinlandsprodukt pro Kopf und Potentialwachstum). Auch wenn die Zusammensetzung der Indizes von dem Bundesländerranking abweicht – Sicherheit spielt auf der internationalen Bühne interessanterweise keine Rolle –, stellen sich hier prinzipiell die gleichen Fragen nach theoretischer Basis, Abhängigkeiten der Daten untereinander und Auswahl der als relevant erachteten Faktoren. Die berechneten Indizes sind geeignet, Deutschland als internationales Schlusslicht darzustellen, und wieder

weiß die Stiftung Rat: »Deutschland muss seine Reformbemühungen deutlich verstärken, um nicht langfristig den Anschluss zu verlieren.« Um wiederum beispielhaft einzusteigen: Selbstverständlich findet auch die Arbeitslosigkeit ihren Platz im Ranking der Nationen. Gerade in der Europäischen Union ist dieses Thema sehr brisant, da innerhalb der Staaten erhebliche ökonomische Unterschiede bestehen. Gerne wird übersehen, dass es keinen gemeinsamen Standard zur Berechnung der Arbeitslosenzahlen gibt. Da die innerstaatlichen Systeme sehr unterschiedlich strukturiert sind, ist ein direkter Vergleich der bloßen Zahlen wohl auch nicht sehr ergiebig. Im Falle des Bundesländer-Rankings waren den Autoren auch nicht einmal die offiziellen Arbeitslosenstatistiken genehm, sie wurden um einen ›verdeckten‹ Anteil nach oben korrigiert. Die Autoren des internationalen Rankings stören sich an all dem freilich nicht – Zahl ist Zahl. Mit dem Internationalen Standortranking setzt die Bertelsmann Stiftung konsequent den Weg fort, über fragwürdige Zahlenspiele konkrete politische Handlungsempfehlungen zu verbreiten.

Erzwungene Politikmodelle

Im Oktober 2001 startete mit dem Slogan »Ein Klick geht durch Deutschland« das Projekt »Perspektive Deutschland« – mit dem erklärten Ziel, »eine bislang einmalige Debatte über die Zukunft Deutschlands und seiner Regionen auszulösen.« Das Projekt ist eine Koproduktion der Unternehmensberatung McKinsey und Co., der Zeitschrift *Stern*, dem Internet-Diensteanbieter WEB.de (früher T-Online) und neuerdings dem Zweiten Deutschen Fernsehen.¹⁰ Der Bertelsmann-Konzern ist an »Perspektive Deutschland« über sein Verlagshaus Gruner+Jahr beteiligt, in dem der *Stern* erscheint. Die Marschrichtung des Projekts gab bereits die erste Pressemitteilung (Perspektive Deutschland 2004) vor, in welcher der Deutschland-Chef von McKinsey, Dr. Jürgen Kluge, zitiert wird:

Wir werden die Ergebnisse der Studie aktiv an die Politik herantragen. Deutschland braucht eine intensive öffentliche Diskussion über seine Zukunft – wir dürfen sie nicht allein den Experten¹¹ und Politikern überlassen, die Bürger müssen sie führen!

Im Endeffekt produziert das Projekt aber gar keine Debatten, sondern diskussionslos deren Ergebnisse.

10 Dazu wird als ›Aushängeschild‹ die Schirmherrschaft durch eine politische Größe eingeworben, 2005 war dies Bundespräsident a. D. Richard v. Weizsäcker.

11 Es ist bezeichnend, dass Entscheidungen hier nicht von Experten getroffen werden sollen. In der gleichen Pressemitteilung werden ein paar Absätze später dennoch anonyme »Experten« angeführt, um die eigene Argumentation zu stützen.

»Den Deutschen gehen die Reformen nicht weit genug. Von der Politik fordern sie eine energische Modernisierung des Landes mit mehr Leistung und mehr privater Initiative. Aber sie sind nicht bereit, auf sozialen Ausgleich zu verzichten« – so beginnt die Pressemitteilung zu den Ergebnissen der fünften Auflage der »Bürgerstudie«. Die Beteiligung der BürgerInnen reduziert sich bei »Perspektive Deutschland« jedoch auf das Ausfüllen eines Onlinefragebogens, genauer gesagt das Anklicken vorgegebener Alternativen. So wurde man beispielsweise gebeten, zwischen den beiden folgenden Zukunftsmodellen zu wählen:

Modell A

- Viel mehr Lebensrisiken als heute werden vom Staat getragen
- Die Steuern und Abgaben sind viel höher als heute
- Die sozialen Unterschiede sind viel geringer als heute
- Wer viel leistet, wird dafür viel weniger belohnt als heute
- Die Gesellschaft verändert sich viel langsamer als heute

Modell B

- Viel mehr Lebensrisiken als heute werden von jedem Einzelnen getragen
- Die Steuern und Abgaben sind viel niedriger als heute
- Die sozialen Unterschiede sind viel größer als heute
- Wer viel leistet, wird dafür viel mehr belohnt als heute
- Die Gesellschaft verändert sich viel schneller als heute

Woher kommen diese Modelle? Schließlich basieren sie auf einer ganzen Reihe von Annahmen, z. B. der Annahme, dass sich die Gesellschaft langsamer entwickelt, wenn der Staat mehr Lebensrisiken trägt. Auch wird das erste Modell negativ vorgefärbt, weil in ihm viel Leistende weniger belohnt würden (was auch immer das genau bedeuten soll). Dagegen suggeriert das zweite Modell, dass guter Lohn für gute Leistung zwangsläufig mit einer Verschärfung des sozialen Gefälles einhergeht. Insgesamt ist diese Form der Fragestellung suggestiv – egal, welche Alternative man bevorzugt, beide basieren auf der gleichen ›Weniger Staat ist gut‹-Philosophie. Sie sind völlig ungeeignet, um differenzierende Aussagen über Einstellungen machen zu können, wohl aber, um das unterschwellige, ihnen zugrunde liegende gedankliche Konstrukt zu verbreiten. Wer mit seiner ›freiwilligen‹ Wahl akzeptiert, dass diese Modelle in sich schlüssig sind und keine weiteren Alternativen bestehen, wird eine Debatte über ›alternativlose‹ soziale Einschnitte wohl auch weniger kritisch verfolgen.¹²

Die Methode der »erzwungenen Wahl« erinnert an die bereits erwähnten Studiengebühren-Umfragen des CHE. Die Agentur McKinsey ist keine Unbekannte auf dem Gebiet fragwürdiger Studien. 2004 führte sie gemeinsam mit dem *Spiegel* und AOL ein eigenes Hochschulranking (McKinsey & Company, 2004) durch – interessanterweise ebenfalls eine Kooperation aus Unternehmensberatung, Printmedium und Online-Dienst. Auch hier treten die eigentlichen AkteurInnen beiseite in den Hintergrund und verleihen ihrem Produkt durch die gezielte Wahl externer Fachleute den Anschein der Seriosität – in diesem Fall durch einen wissenschaftlichen Beirat, bestehend aus zwei Universitätsprofessoren.¹³ Ausgewertet hat dennoch die Unternehmensberatung; wie sich später herausstellte, sind die Wissenschaftler mit den Daten überhaupt nicht in Berührung gekommen. Der vom *Spiegel* als Innovation gefeierte Versuch, nicht die Hochschulen selbst, sondern die dort Studierenden zu ranken (Friedmann et al. 2004), erwies sich bei kritischer Betrachtung als methodisches Desaster (Claus 2005; Höding et al. 2004), das der Hochschulreformdebatte vermutlich mehr geschadet als genützt hat. Denn jedes Ranking, egal wie schlecht es auch durchgeführt wurde, wird bei entsprechender Platzierung gerne als politisches Argument eingesetzt. Und falls sich Studierwillige tatsächlich auf Basis dieser Veröffentlichung für eine »gut« platzierte Hochschule entschieden haben sollten, ist ebenso offensichtlich, dass ihnen ein Auswürfeln prinzipiell nicht weniger verlässliche Informationen geliefert hätte.

Kriterien der Hochschulrankings

Das CHE-Hochschulranking ist das umfassendste und detaillierteste Ranking deutschsprachiger Universitäten und Fachhochschulen. Es richtet sich vor allem an Studienanfänger und Studierende, hat sich aber auch an den Hochschulen als fair, informativ und qualifiziert durchgesetzt.

So bewirbt das Centrum für Hochschulentwicklung das eigene Ranking auf seiner Homepage (CHE, 2006a). Die Sinnhaftigkeit von Hochschulrankings wird gerne damit begründet, sie würden Studierwilligen eine große Hilfe bei der Wahl einer Universität sein. In diesem Fall wird der Einfluss der Rankings jedoch offenbar von ihren Autoren überschätzt: Stemmler (2005) korrelierte die Ergebnisse des CHE-Hochschulrankings mit der tatsächlichen Beliebtheit der gerankten Institute¹⁴ –

12 Zu den psychologischen Prozessen, die eine Einstellungsänderung vermuten lassen, vgl. Festinger/Carlsmith (1959).

13 Prof. Dr. Gerhard Arminger, Lehrstuhl für Wirtschaftsstatistik, Universität Wuppertal; Prof. Dr. Manfred Deistler, Institut für Wirtschaftsmathematik, Forschungsgruppe Ökonometrie und Systemtheorie, Technische Universität Wien.

und es zeigte sich kein bedeutsamer Zusammenhang. Allerdings fand er eine Korrelation zwischen der Einwohnerzahl der jeweiligen Stadt und Beliebtheit ihrer Hochschule. Studierende scheinen in der Tendenz also mehr Wert auf das Flair einer Metropole zu legen, als auf den wie auch immer gemessenen Ruf der Bildungsstätte. Politik und Medien konsumieren die vollmundigen Erklärungen des CHE dagegen weitgehend unreflektiert: »Als wenig umstritten gilt das Ranking des Centrums für Hochschulentwicklung (CHE), das selbst die deutsche Hochschulrektorenkonferenz empfiehlt.« (Heinisch 2006, erschienen im *Themendienst* des Klett-Verlags)

Im zitierten Beispiel ist der Autorin offensichtlich nicht einmal gewahr, dass die Hochschulrektorenkonferenz das CHE gemeinsam mit der Bertelsmann Stiftung gründete – die Maskerade erweist sich als funktional. Für sein Hochschulranking gibt das CHE nach eigenen Angaben Fragebögen an über 250.000 Studierende, die jeweiligen Fachbereiche, ProfessorInnen und AbsolventInnen in Deutschland, Österreich und der Schweiz. Zudem würden Daten aus Studierenden-, AbsolventInnen- und Prüfungsstatistiken hinzugezogen. Die Hochschulen werden weiter z. B. zur Erhebung von Studiengebühren befragt. ProfessorInnen werden u. a. aufgefordert, fünf Universitäten zu nennen, die sie in ihrem eigenen Fach dem eigenen Kind empfehlen würden. Ausgeschlossen ist hierbei die Nennung der eigenen Universität – und damit meist der einzigen, für die die befragte Person überhaupt einen aktuellen Einblick in die Qualität der Lehre hat.

Das komplette Rankingprodukt setzt sich schließlich aus verschiedenen Schwerpunktrankings zusammen: Dem Hochschulranking, dem Forschungsranking, dem Länderranking und dem AlumniRanking. Während das Hochschulranking verschiedene deutschsprachige Universitäten vergleicht, legt das Forschungsranking den Schwerpunkt auf den bundesweiten Vergleich universitärer Forschungsleistungen. Das Länderranking will die Ergebnisse des Hochschulrankings auf föderaler Ebene aggregieren, und das Alumni-Ranking basiert auf der rückblickenden Bewertung ehemaliger Studierender in Bezug auf die Qualität des Studiums an der jeweiligen Universität, womit also in Erfahrung gebracht werden soll, wie gut oder schlecht die Lehre retrospektiv bewertet wird.

Inwieweit diese Aussagen auch den aktuellen Stand abbilden können, ist selbstverständlich diskussionswürdig. Die Forschungsleistungen der Fachbereiche werden in Form einer bibliometrischen Analyse über die Publikationen erfasst (Berghoff et al. 2006). Für die Fächer wird zunächst in ausgewählten Datenbanken festgestellt, wie viele Publikationen auf die ProfessorInnen eines Fachbereichs

kommen. Diese Publikationen werden dann anhand der Länge des Beitrages und der Anzahl der AutorInnen gewichtet. Ein Beitrag mit 5 Seiten erhält einen Punkt, einer mit 6–10 Seiten zwei, einer mit 11–20 drei und ab 21 Seiten gibt es vier Punkte für einen Beitrag. Für Monographien gibt es immer acht Punkte. Nur, wer sagt uns, dass ein Beitrag mit einem größeren Seitenumfang auch im Maß der Mehrgewichtung einen größeren Beitrag zur Forschung leistet? Und warum sind 21 Seiten genau doppelt so viel wert wie 10 Seiten? Noch verwirrender wird das Ganze, wenn man feststellt, dass die Gewichtung der Länge des Beitrages von der Fachzugehörigkeit abhängt: In der Amerikanistik bekomme ich für meinen 20-Seiten-Beitrag einen ganzen Punkt mehr.

Ein ganz ähnliches Bild zeichnet sich bei der Gewichtung der AutorInnenzahl ab. Die Punkte werden durch die Anzahl der AutorInnen geteilt. Aber leistet ein Artikel einen geringeren Beitrag zur Forschung, weil mehrere AutorInnen an ihm gearbeitet haben? Ist dieser Beitrag bei zwei Verantwortlichen dann pro AutorIn nur noch halb so groß? Betrachtet man die Veröffentlichungsprofile einzelner Fachbereiche genauer, findet sich in vielen Fällen ein ähnliches Muster: Einige wenige Personen veröffentlichen immens viel, während die Mehrheit nur auf vereinzelte Artikel verweisen kann (Schui & Krampen, 2006). Eine einzige ›Gali-onsfigur‹ kann somit nicht selten das Rating des ganzen Instituts bestimmen. Nach Schui & Krampen (2006) macht es daher keinen Sinn, irgendwelche Durchschnittswerte ganzer Institute miteinander zu vergleichen. Die Bibliometrie ist unter der Voraussetzung sorgfältigen Vorgehens als Indikator geeignet, um den fachbezogenen Output einzelner WissenschaftlerInnen zu bewerten – nicht jedoch jenen ganzer Fachbereiche oder gar Universitäten.

Wissenschaft ist ein Prozess, ein Wechselspiel, bei dem ständig eine Vielzahl unterschiedlicher Publikationen als Grundlage neuer Arbeiten dient. Inwieweit die Relevanz oder Qualität einer Arbeit über einen Zitat-Index abgebildet werden kann, ist noch immer Gegenstand von Diskussionen. Vor allem ist das Verfahren aufwendiger als es zunächst scheint – ein reines Abzählen veröffentlichter Artikel ist nicht sinnvoll. Nehmen Sie als Beispiel die Arbeit, in der Wakefield et al. (1998) einen Zusammenhang von Masern-Impfung und Autismus aufzeigten. Der Erstautor wurde Jahre nach der Veröffentlichung des Betrugs überführt (Deer 2004), fast alle Co-AutorInnen distanzieren sich von ihren ›Erkenntnissen‹. Dessen ungeachtet hat der Artikel natürlich schon allein aufgrund nachfolgender Untersuchungen (die den Zusammenhang nicht replizieren konnten) einen hohen Zitat-Index – und wird nun hier noch einmal mehr zitiert.

Im CHE-Länderranking treten erneut die Bundesländer gegeneinander an: »Welches Bundesland hat die Fakultäten mit den meisten Spitzenplatzierungen?« Es basiert auf den Indikatoren ›Studierendenurteil‹ ›durchschnittliche Studiendau-

er, ›Reputation‹ und ›Forschungsleistung‹. Für diese Kriterien wird berechnet, wie häufig die untersuchten Fakultäten des jeweiligen Bundeslandes im Schnitt in der Spitzengruppe platziert sind. Um uns nicht zu wiederholen: Als nunmehr etwas Rankingkritik-erfahrene LeserIn dieses Kapitels bereitet es Ihnen vielleicht Freude, einmal selbst Sinn und Unsinn dieses Ansatzes zu ergründen.

Das CHE bietet auf der eigenen Homepage einen Überblick über einige Konkurrenz-Rankings zum gleichen Thema, und diese Rankings erfahren eine kritische Würdigung. Im Kontrast steht das eigene Ranking noch ein wenig besser da. In der Kritik werden zudem Punkte umschiff, die dem eigenen Projekt ebenfalls anhängen könnten. Aufgrund der schwachen Unterschiede, die das *Spiegel*-Ranking (Friedmann et al. 2004) zwischen den Studierenden unterschiedlicher Hochschulen fand, wird z. B. keinesfalls diskutiert, ob die Methodik vielleicht unzureichend ist, sondern ausschließlich Handlungsbedarf für die Bildungspolitik gefolgert:

Diese Erhebung macht eines deutlich: Solange die Universitäten ihre Studierenden noch nicht selbst auswählen dürfen, werden die Unterschiede in der Qualität der Studierenden an den einzelnen Fachbereichen gering bleiben. Gegenwärtig lassen sich kaum größere Unterschiede ausmachen – am ehesten noch bei den privaten Hochschulen in der BWL, wo die Auswahl bereits heute durch die Hochschulen selbst erfolgt. (CHE, 2006b)

Der Diagnostiker Stemmler (2005) argumentiert jedoch, der Nutzen dieser Selbstauswahl sei zweifelhaft:

Der Gesetzgeber verlangt eine Erhöhung des Anteils der Hochschulauswahl an der Studienplatzvergabe von derzeit 24 % auf 60 %. Die dafür erforderliche Selektion (Quote 17 %) kann ökonomisch allerdings nicht dezentral ausschließlich an den einzelnen Hochschulen erfolgen – vor allen Dingen nicht bei den dann zu erwartenden Mehrfachbewerbungen.

In Beispielrechnungen veranschaulicht er, welche diagnostische Qualität ein Auswahlverfahren haben müsste, um der zentralen Vergabe anhand der Abiturnote überlegen zu sein. Die Anforderungen liegen hierbei so hoch, dass sich bei einem geeigneten Verfahren die Frage stellt, ob der Zugewinn an Information die Kosten rechtfertigt, die es verursacht.

Grenzen der ›Rankability‹

Wenn inzwischen jedes Hochschulranking seine Vor- und Nachteile hat, brauchen wir vielleicht bereits ein Ranking der Rankings? Was bringt uns ein Ranking, wenn wir erst lange recherchieren müssen, wie vertrauenswürdig es wohl ist? Können wir diese Zeit nicht zwischenzeitlich besser verwenden, indem wir uns wieder direkt an

den Hochschulen umsehen und mit Studierenden unterhalten, vielleicht die Fachschaften fragen?

Hochschulpolitische Aktivitäten bzw. ihre AkteurInnen sind derzeit übrigens kein Thema im CHE-Ranking. Warum eigentlich nicht? Schließlich wissen die gewählten StudierendenvertreterInnen von allen Befragten noch am ehesten Bescheid, was aktuell in der Lehre passiert, und zwar aus einer für Studieninteressierte optimalen Perspektive. Doch wie auch in der Diskussion um Studiengebühren, die u. a. damit begründet werden, den Studierenden als ›Kunden‹ mehr Einfluss auf die Lehre geben zu wollen, wird die Alternative demokratischer Mitbestimmung ignoriert. Studentische Mitbestimmung wird seit Jahren sukzessive zurückgefahren – wir erleben gewissermaßen einen schleichenden Wechsel von Wahlzetteln zu Rankingfragebögen.

Ein grundlegendes Problem bleibt die Unklarheit der Faktoren, die tatsächlich die Wahl des Studienortes beeinflussen. Dem kann in letzter Konsequenz auch das CHE-Ranking nicht begegnen: Wenn schon die Einwohnerzahl einer Universitätsstadt bessere Vorhersagen erlaubt, wie viele Bewerbungen an der Hochschule eingehen, dürfte dies deutlich genug sein. Von Seiten des CHE wird dieser merkwürdige Umstand jedoch nicht thematisiert – es spielt offenbar keine Rolle, was zurzeit wirklich entscheidungsrelevant für angehende AkademikerInnen ist. Was bezweckt ein Ranking wirklich, das nachweislich keinen Zusammenhang mit der Größe besitzt, für die es vorgeblich konstruiert wurde? Wir müssen spekulieren. Es könnte durch die massive und meist unreflektierte Medienpräsenz im Sinne einer selbsterfüllenden Prophezeiung zukünftig die Orientierungswirkung erzeugen, die es angeblich bereits hat: Aus dem passiven Ratgeber würde ein aktives Steuerungswerkzeug werden. Wenn sich in ein paar Jahren zeigt, dass ein Zusammenhang zwischen CHE-Ranking und Uniwahl existiert, wird für die Kritiker zu ergründen sein, ob die Methode besser oder lediglich mächtiger wurde. Nicht zufällig wird einigen Rankings auch gerne die Eigenschaft ›einflussreich‹ zugestanden. Das muss nicht zwingend ein Qualitätsurteil sein.

Nun wollten wir, wie zu Beginn angeführt, Rankings nicht per se diskreditieren. Sorgfältig durchgeführt und sinnvoll interpretiert, können sie durchaus von Nutzen sein. Zu einem gewissenhaften Vorgehen gehört allerdings auch immer ein Hinweis auf die Grenzen der Methodik – der Interpretationsspielraum von Rankings ist endlich, und er wird in den vorgestellten Fällen stets zweckdienlich überspannt.

Rankings sind Abbildungen eines Zustands, keine Ursache-Wirkungs-Messung. Die Interpretation, warum ein Ranking so oder so ausgefallen ist, kann nicht durch das Ranking selbst erklärt werden. Außerdem muss das Verfahren vollständig offen liegen, um verständlich und überprüfbar zu sein. Allein die Fülle der veröffentlichten Hochschulrankings, die durch unterschiedliche Methoden zu unterschiedlichen

Ergebnissen kommen, macht auch klar, dass es das ›ultimate Ranking‹ einfach nicht gibt. Da zu Beginn einer Untersuchung genau festgelegt werden sollte, was man überhaupt in Erfahrung bringen möchte, und hieraus das sinnvollste methodische Vorgehen abzuleiten ist, müsste sich bereits hier die grundsätzliche Frage ergeben, ob ein Ranking überhaupt geeignet ist.

Allen in diesem Kapitel besprochenen Rankingprodukten¹⁵ von Bertelsmann liegt die Annahme zugrunde, dass die klassifizierten Objekte – seien es Universitäten, Bundesländer oder Nationen – direkt und quantitativ miteinander vergleichbar sind. Mit dem Annehmen oder Ablehnen dieser Idee steht und fällt das gesamte Unterfangen. Je komplexer die betrachteten Gebilde sind, desto fraglicher wird unseres Erachtens ihre *rankability*. Jede noch so abstrakte Größe lässt sich trotzdem irgendwie in Zahlen übersetzen, Zahlen lassen sich beliebig verrechnen und die Ergebnisse interpretieren. Nur was erfahren wir wirklich von einem Hotelranking, in das der Durchschnitt aller Zimmernummern eingeht?

Brauchbare Informationen lassen sich auch ohne Rangbildung darstellen. Die Computerzeitschrift *c't* genießt in ihrem Segment einen sehr guten Ruf und wird gerne von ›EntscheiderInnen‹ zu Rate gezogen. Die Redaktion testet auch regelmäßig die am Markt befindliche Hardware wie Festplatten, Monitore oder Notebooks. Dabei beschreibt sie ausführlich, wie der Test vonstatten geht und welche Werte wie zu interpretieren sind. Jedoch kürt sie grundsätzlich niemals einen ›Testsieger‹, sondern wiegt für unterschiedliche Anwendungsszenarien die relevanten Pro- und Contra-Argumente der Geräte ab. Denn wahrscheinlich braucht man gar nicht das ›beste Notebook‹, sondern vielmehr eines, das den eigenen Anforderungen am besten entspricht.

Während Rankings gerne als Chance verkauft werden, im Sinne des Wettbewerbs alle Beteiligten zu besseren Leistungen zu motivieren – Bertelsmann bezeichnet dieses Prinzip als »Lernen von den Besten« –, darf die Gefahr nicht übersehen werden, in eine Gruppe bisher relativ gleichwertiger Alternativen durch die willkürliche Auswahl von ›relevanten‹ Faktoren eine Teilgruppe von Verlierern hinein zu definieren. Hier müssen die in Aussicht gestellten Vorteile und ihre negativen Konsequenzen stets gut abgewogen werden. Ist jemand auf dem ›letzten Platz‹ nach objektiven Kriterien wirklich schon ›schlecht‹? Es ließe sich sogar argumentieren, dass die rangfreie Gegenüberstellung von Alternativen mit ihren Vorzügen und Nachteilen im Endeffekt eine konsequentere, zumindest aber sportlichere Umsetzung des Wettbewerbsgedankens wäre.

15 Diese Bezeichnung stammt übrigens nicht von uns, sondern von der Bertelsmann Stiftung.

Ein Ranking greift dem Markt gewissermaßen in die Entscheidung ein – es präsentiert bereits Sieger und Verlierer und leistet so vielleicht einen fragwürdigen Beitrag zum Wettstreit der KontrahentInnen. Ist nicht das Ideal, allen Teilnehmenden bestmögliche Entwicklung zu gewähren? Wie konstruktiv ist der Ansatz, per Rangliste Gewinner und Verlierer zu definieren, und dadurch wie in einer selbsterfüllenden Prophezeiung die Glückspilze zusätzlich zu stützen? In stillen Momenten fragen wir uns, ob die AutorInnen der Bertelsmann Stiftung im privaten Wettbewerb um die eigene Freizeit wohl auch ihren engsten Freundeskreis nach willkürlich festgelegten Kriterien ranken, und den Letztplatzierten damit Anreize geben wollen, in Zukunft vielleicht mehr Humor, weniger Lust auf Diskussionen und keine Katze zu haben.

Literatur

- Aktionsbündnis gegen Studiengebühren/ABS, 2006: *Argumente gegen Studiengebühren – Eine Widerlegung von Behauptungen*. Schriftenreihe des ABS, Heft 2, 7. Auflage.
- Berger, U./Stein, C., 2004: »Das zweite Scheitern des Liberalismus: Hysterie, Regression und Wiederholungszwang«, in: <http://www.heise.de/tp/r4/artikel/19/19062/1.html>. Zugriff am 02.10.2006.
- Berger, U./Stein, C., 2005: »Das zweite Scheitern des Liberalismus: Exportorientierung statt Binnenmarkt-Entwicklung«, in: <http://www.heise.de/tp/r4/artikel/19/19828/1.html>. Zugriff am 02.10.2006.
- Berghoff, S./Giebisch, P./Hachmeister, C.-D./Federkeil, G./Müller-Böling, D./Rölle, D., 2006: »CHE-HochschulRanking 2006 – Vorgehensweise und Indikatoren«, in: http://www.che.de/downloads/Methoden_2006.pdf; Zugriff am 25.09.2006.
- Berié, E./Jansen, S./Kobert, H./Rudloff, F., 2003: *Fischer Weltalmanach 2004*, Frankfurt am Main.
- Bertelsmann Stiftung, 2003: »Zeitbombe Demographie: Die Deutschen werfen der Politik Untätigkeit vor«. Pressemeldung vom 23.05.2003, in: http://www.bertelsmannstiftung.de/cps/rde/xchg/bst/hs.xsl/nachrichten_4632.htm. Zugriff am 04.10.2006.
- Bertelsmann Stiftung, 2005: Carl Bertelsmann-Preis 2005 »Junge Generation und Arbeit: Chancen erkennen – Potenziale nutzen«: Arbeitskommission und Jury, in: http://www.bertelsmann-stiftung.de/cps/rde/xchg/bst/hs.xsl/artikel_3258.htm. Zugriff am 04.10.2006.

- Berthold, N./Fricke, H./Drews, S./Vehrkamp, R., 2003: *Die Bundesländer im Standortwettbewerb 2003*, Gütersloh
- Berthold, N./Fricke, H./Kullas, M., 2005: *Die Bundesländer im Standortwettbewerb 2005*, Gütersloh
- Bortz, J., 1993: *Statistik für Sozialwissenschaftler*, 4. Aufl., Berlin
- CHE, 2006a: »CHE-HochschulRanking 2006«, in: <http://www.che-ranking.de>; Zugriff am 25.09.2006.
- CHE, 2006b: »Hochschulrankings im Überblick: Der Spiegel«, in: <http://www.che.de/cms/?getObject=48&GetName=Der+Spiegel&getLang=de>; Zugriff am 25.09.2006.
- Claus, V., 2005: »Rankings und ›Wo studieren die Besten‹ – Stellungnahme des Vorsitzenden des Fakultätentags Informatik zu Rankings und zur Online-Umfrage Studentenspiegel vom November 2004«, in: *Informatik Spektrum*, 28, 2, 124–128.
- Deer, B., 2004: »The Lancet scandal«, in: <http://briandeer.com/mmr-lancet.htm>; Zugriff am 26.09.2006.
- Drews, S., 2005: *Gestaltungsmöglichkeiten und Aussagekraft des Rankings von Standorten – unter besonderer Berücksichtigung des Bundesländer-Standortrankings der Bertelsmann Stiftung*. Dissertation, Technische Universität Berlin.
- Festinger, L./Carlsmith, J. M., 1959: »Cognitive consequences of forced compliance.«, in: *Journal of Abnormal and Social Psychology*, 58, 203–210.
- Friedmann, J./Hackenbroch, V./Hipp, D./Klawitter, N./Koch, J./Lakotta, B./Mohr, J./Schmitz, C./Thimm, K./Wüst, C., 2004: »Die Elite von Morgen – Wo studieren die Besten?«, in: *Spiegel*, 48, 178–200.
- Hafemann, K./van Suntum, U., 2004: *Internationales Standort-Ranking 2004 – Wachstum und Beschäftigung*, Gütersloh
- Heinisch, C., 2006: »Wenn der Vergleich hinkt – Rankings von Hochschulen und Schulen«, in: *Klett-Themendienst* 35, August 2006, http://www.klettpress-box.de/sixcms/media.php/273/KTD_35_s5-6_298464.pdf; Zugriff am 25.09.2006.
- Höding, M./Michalke, M./Naß, O., 2004: »Die Magie der Zahlen – Die Hochschulverrenkungen von AOL, McKinsey und SPIEGEL«, in: <http://www.students.unimarburg.de/Fachs04/spiegelranking.html>. Zugriff am 16.09.2006.
- Krämer, W., 1997: *So lügt man mit Statistik*, München
- Landtag des Saarlandes, 2005: 13. Wahlperiode, »Protokoll der 19. Sitzung, 13.12.2005«, in: <http://www.landtag-saar.de/dms13/Land019.pdf>. Zugriff am 18.09.2006.

- Landtag Mecklenburg-Vorpommern, 2006: »Drucksache 4/2325 vom 21.06.2006«, in: http://www.landtag-mv.de/dokumentenarchiv/drucksachen/4_Wahlperiode/D04-2_000/Drs042325.pdf. Zugriff am 11.10.2006.
- LobbyControl, 2006: »Standort-Check der Bertelsmann Stiftung«, in: <http://www.lobbycontrol.de/blog/index.php/2006/04/standort-check-der-bertelsmannstiftung>. Zugriff am 04.10.2006.
- Loftus, E. F. / Palmer, J. C., 1974: »Reconstruction of auto-mobile destruction: An example of the interaction between language and memory«, in: *Journal of Verbal Learning and Verbal Behaviour*, 13, 585–589.
- McKinsey & Company, 2004: »Studentenspiegel«, in: <http://www.studentenspiegel.de>. Zugriff am 9.12.2004.
- Müller, P., 2004: »Den Wandel gestalten – Die Zukunft gewinnen. Regierungserklärung des saarländischen Ministerpräsidenten Peter Müller zu Beginn der 13. Legislaturperiode am 03.11.2004«, in: http://www.saarland.de/service_renden_12084.htm. Zugriff am 22.09.2006.
- Müller-Böling, D., 2006: »Sinn und Wirkung von Rankings«, in: *Aviso*, 1, 16–21.
- MVregio, 2006: »Expertenkommission zur Zukunft der Erziehung und Bildung ist konstituiert«, in: <http://www.mvregio.de/mvr/18955.html>. Zugriff am 11.10.2006.
- Perspektive Deutschland, 2004: »Presseinformation: Ein Klick geht durch Deutschland«, in: http://www.perspektive-deutschland.de/files/presse/pm01_basismeldung_zeitung.pdf. Zugriff am 26.09.2006.
- Schui, G./Krampen, G., 2006: »Qualitätssicherung und -entwicklung an Universitäten im Bereich der Forschungsaktivitäten – Systematische Erhebung von Publikationsaktivitäten und internationaler Rezeption« in: 45. Kongress der Deutschen Gesellschaft für Psychologie, Nürnberg.
- Stemmler, G., 2005: »Studierendenauswahl durch Hochschulen: Ungewisser Nutzen«, in: *Psychologische Rundschau*, 56, 2, 125–127.
- Struben, M., 2004: »Bild‹ dir deine Meinungsumfrage! Demoskopie als Demagogie«, in: BdWi / fzs (Hg.): *Studiengebühren, Elitekonzeptionen & Agenda 2010*, Marburg, 25f.
- van Suntum, U., 2004: »Schriftenverzeichnis«, in: http://www.wiwi.uni-muenster.de/insiwo/forschen/pub_ulvs.html. Zugriff am 04.10.2006.
- Wakefield, A. J./Murch, S. H./Anthony, A./Linnell, J./Casson, D. M./Malik, M./Berelowitz, M./Dhillon, A. P./Thomson, M. A./Harvey, P./Valentine, A./Davies, S. E./Walker-Smith, J. A., 1998: »Ileal-lymphoid-nodular hyperplasia

sia, non-specific colitis, and pervasive developmental disorder in children«, in: *The Lancet*, 351, 9103, 637–642.

Wikipedia, 2006a: »Befragung«, in: <http://de.wikipedia.org/wiki/Befragung>. Zugriff 25.09.2006.

Wikipedia, 2006b: »Index«, in: [http://de.wikipedia.org/wiki/Index_\(Statistik\)](http://de.wikipedia.org/wiki/Index_(Statistik)). Zugriff 25.09.2006.

Wikipedia, 2006c: »Propaganda«, in: <http://de.wikipedia.org/wiki/Propaganda>. Zugriff am 05.10.2006.

Wikipedia, 2006d: »Ranking«, in: <http://de.wikipedia.org/wiki/Ranking>. Zugriff am 25.09.2006.

Wikipedia, 2006e: »Rating«, in: <http://de.wikipedia.org/wiki/Rating>. Zugriff am 25.09.2006.