

# Alle auf Platz eins?

## Kritische Anmerkungen zu einem kennzahlengestützten Ranking-System

Von Hartmut Stieger

In der hochschulpolitischen Diskussion der letzten Jahre hat sich „Wettbewerb“ als neues Schlüsselwort etabliert, für manche aber auch als Reizwort, das Ärger auslöst. Mit den im Juli 1985 verabschiedeten „Empfehlungen zum Wettbewerb“ führte der Wissenschaftsrat die Diskussion auf den Punkt: „Wettbewerb ist prinzipiell sinnvoll, wenn und insoweit er die Fähigkeit eines Systems verstärkt, wünschenswerte Leistungen zu erbringen.“ Es geht also um mehr Leistung, um mehr Motivation, letztlich um mehr marktwirtschaftliches Denken und Handeln im Wissenschaftsbetrieb. Markt und Wettbewerb erfordern allerdings eine Bewertung von Kosten und Leistungen, von Aufwand und Ertrag. Sie sind die entscheidende Basis für Gewinn und Verlust, die einfachsten und aussagekräftigsten „Kennzahlen“, wenn man Unternehmen in Marktsystemen vergleichen will.

Ähnliche marktorientierte Vergleiche – allerdings für die Leistungsbewertung von Hochschulen – werden auch mit dem in letzter Zeit so heftig diskutierten „Ranking“ angestrebt. Gemeint sind Rangfolgenlisten, die die Universitäten auf der Grundlage von „Kennzahlen“ in eine ihrer Leistung entsprechende Reihenfolge bringen sollen, weswegen böse Zungen bereits von den „Bundesligatabellen“ der Hochschulen sprechen.

Erhebliche Unruhe löste in diesem Zusammenhang die in Gießen entstandene „Giese-Studie“ aus, in der Ranking mit seinen Möglichkeiten (und Unmöglichkeiten) anhand zahlreicher Beispiele demonstriert wird.<sup>1</sup>

Auf einige nicht ganz unproblematische Aspekte, die in Konstruktion und Anwendung des Ranking begründet liegen, soll im folgenden Beitrag näher eingegangen werden.

In der Diskussion um die Brauchbarkeit eines kennzahlengestützten Ranking-Systems sind vor allem drei Kernfragen zu beantworten:

- (1) Was sind Kennzahlen? Wie lassen sie sich sinnvoll definieren und gegeneinander abgrenzen? Vor allem: Läßt sich

mit ihnen überhaupt Leistung im wissenschaftlichen Sinne messen?

- (2) Sind Rangfolgenlisten geeignet, Leistungsunterschiede (gerecht) wiederzugeben, um Aussagen über „bessere“ oder „schlechtere“ Leistung zu erlauben?
- (3) Welche hochschulpolitischen Schlußfolgerungen werden aus der Stellung bzw. dem Rang der Hochschule in einem solchen Ordnungssystem (oben, unten oder in der Mitte?) gezogen, insbesondere auch im Hinblick auf die Personal- und Mittelzuteilung durch die Landesadministration?

### Kennzahlen und Ranking

Kennzahlen sind für die Hochschulen an sich nichts Neues. Im Gegenteil, sie sind zahlreich und vielfältig in Gebrauch, am häufigsten in Form von Personal- oder Ausstattungsrelationen wie zum Beispiel „Zahl der eingeschriebenen Studenten je Wissenschaftler“, „Hauptnutzfläche je Student“ etc., aber auch in der Weise, daß sie (wenigstens tendenziell) eine Aussage über die „Verfahrensqualität“ zulassen. Dazu gehören beispielsweise „Schwundquoten“, die angeben, wieviel Prozent der Studenten aus dem Studium erfolglos ausscheiden, ihr Studium also vorzeitig abbrechen, Aussagen über die „mittlere Studiendauer“ als Indikator für die Zügigkeit des Studierens (und/oder der Qualität der Lehr- und Studienorganisation). Aber auch auf dem Forschungssektor bürgern sich mehr und mehr Leistungsindikatoren ein, zum Beispiel „Zahl der Veröffentlichungen in Fachzeitschriften je Professur“, „Zahl der Buchpublikationen“ oder sogenannte „Zitationsindizes“, die angeben, wie häufig Wissenschaftler in anderen Publikationen zitiert werden usw.

In gewisser Weise neu, zumindest für die Bundesrepublik – in den Vereinigten Staaten, in denen Hochschulen sehr viel mehr als bei uns auch Wirtschaftsbetriebe sind, ist derartiges seit langem gang und gäbe – ist der Versuch zu sehen, solche Indikatoren in eine Rangfolge zu bringen, in der die Hochschulen auf die „Plätze“ verwiesen werden, um auf einen Blick erkennen zu können, wer der Beste, der Zweitbeste... usw. ist, aber auch, zum Leidwesen der Betroffenen, wer die Schlechtesten sind.

### Konkurrenzprinzip als Pate

Pate steht hier das leistungsstimulierende Konkurrenzprinzip unserer markt- und wettbewerbsorientierten Wirtschaftsordnung. Üblicherweise mißt man den Betriebserfolg an der Höhe des Gewinns. Aber gerade das ist bei Hochschulen nicht möglich.

Kennzahlen-Ranking soll, so stellt man sich vor, die an sich fehlende Marktorientierung ersetzen bzw. im universitären Rahmen überhaupt erst möglich machen. Ranking soll wissenschaftliche Leistung, vor allem im Vergleich zwischen den Hochschulen (bzw. deren Disziplinen), nicht nur transparenter machen und damit den Wettbewerb fördern. Es soll ihn geradezu herausfordern und leistungsstimulierende Effekte auslösen, natürlich mit dem Ziel, die Ranking-Position immer weiter zu verbessern. Im Extrem kämpft jeder mit jedem um die Spitzenposition. Um es überspitzt zu formulieren: Alle also auf Platz eins?

Ohne hier auf die Frage einzugehen, ob Kennzahlen überhaupt in der Lage sind, Leistung im wissenschaftlichen Sinne wirklichkeitsentsprechend widerzuspiegeln (wogegen manches einzuwenden wäre), soll hier auf einige typische Konstruktionsmerkmale des Ranking eingegangen werden, die für sich genommen schon ausreichen, die gewünschte Aussage über Leistung und Leistungsvergleich in Frage zu stellen, zumindest aber sehr zu erschweren, die im allgemeinen aber nicht oder nur wenig beachtet werden.

Worum geht es?

### Qualität durch Quantität bestimmbar?

Bei „Rangfolge-Listen“ gibt es immer einen ersten und letzten Platz. Alle übrigen Positionen liegen dazwischen und erhalten eine fortlaufende Rangziffer. Das mag banal klingen, ist es aber nicht, wie sich gleich zeigen wird.

Das Charakteristische, aber auch das Problematische solcher Einordnungen zwischen dem ersten und dem letzten Platz ist nämlich vor allem zweierlei:

Erstens: Eine quantitative Aussage, die Größe einer Kennzahl, soll zu einer qualitativen Aussage herangezogen werden, wie gut eine Hochschule im Vergleich zu den anderen Hochschulen ist; mit anderen Wor-

<sup>1</sup> Giese, E.: Leistungsmessung wissenschaftlicher Hochschulen in der Bundesrepublik Deutschland. In: Beiträge zur Hochschulforschung, Heft 4 – 1986, S. 399–446



ten: Qualität soll durch Quantität bestimmbar gemacht werden.

Zweitens: Die (wie auch immer definierten) absoluten Kennzahlen werden in eine ordinale Reihenfolge gebracht, wodurch ihre Leistungsunterschiede im Grunde aber wieder aufgelöst, man könnte fast sagen, zum Verschwinden gebracht werden. Am Ende gibt es immer einen ersten, zweiten, dritten etc. Platz, und zwar ganz unabhängig von der wirklichen Höhe der Leistung, die dazu geführt hat.

Damit ist ein statistisches Phänomen des Ranking angesprochen, das man die „Relativität der Rangordnung“ nennen könnte; ein Phänomen, das durch die unterschiedliche Streuung der Kennzahlenwerte entsteht. Es fehlt der feste Punkt, die objektive Bezugsebene, an der sich die vergleichende Bewertung orientieren könnte.

**Ein Beispiel**

	Fall 1	Fall 2	Fall 3	Fall 4
1.	104	140	144	145
2.	103	130	143	129
3.	102	120	142	103
4.	101	110	141	101
	100	100	100	100
6.	99			
7.	98			
8.	97			
9.	96			
6.		90	59	90
7.		80	58	89
8.		70	57	87
9.		60	56	56

Was damit gemeint ist, macht das folgende Beispiel klar. Angenommen, das allgemein anerkannte mittlere Niveau eines wissenschaftlichen Leistungsstandards betrage 100. (Es spielt keine Rolle, wie dieser „Leistungsstandard“ definiert ist.) Dann gibt es mehrere extreme Positionen der Hochschu-

len in der Rangliste, die allerdings durch die ordinale Ordnung der Positionsziffern zwischen dem ersten und dem letzten Platz alle gleichbehandelt werden. Der Einfachheit halber soll das Problem an vier Fällen extremer Positionierung von neun Hochschulen dargestellt werden (s. Abbildung).

Im ersten Fall streuen die neun Hochschulen dicht um das mittlere Niveau von 100, mit Werten zwischen 96 und 104. Im zweiten Fall liegen sie erheblich weiter auseinander, zwischen 140 und 60, verteilen sich aber noch recht regelmäßig um den Mittelwert von 100. Im dritten Fall besteht die Rankingliste eigentlich aus zwei Gruppen. Die einen schneiden über alle Maßen gut ab (von 144 bis 141), die anderen über alle Maßen schlecht (von 59 bis 56). Fall vier zeigt eine ähnliche Verteilung nur völlig ungleichmäßig. Mehreren schlecht abscheidenden Hochschulen stehen dort ganz wenige, übergute Hochschulen gegenüber. Aber auch hier beträgt der Mittelwert wie in allen anderen Fällen 100.

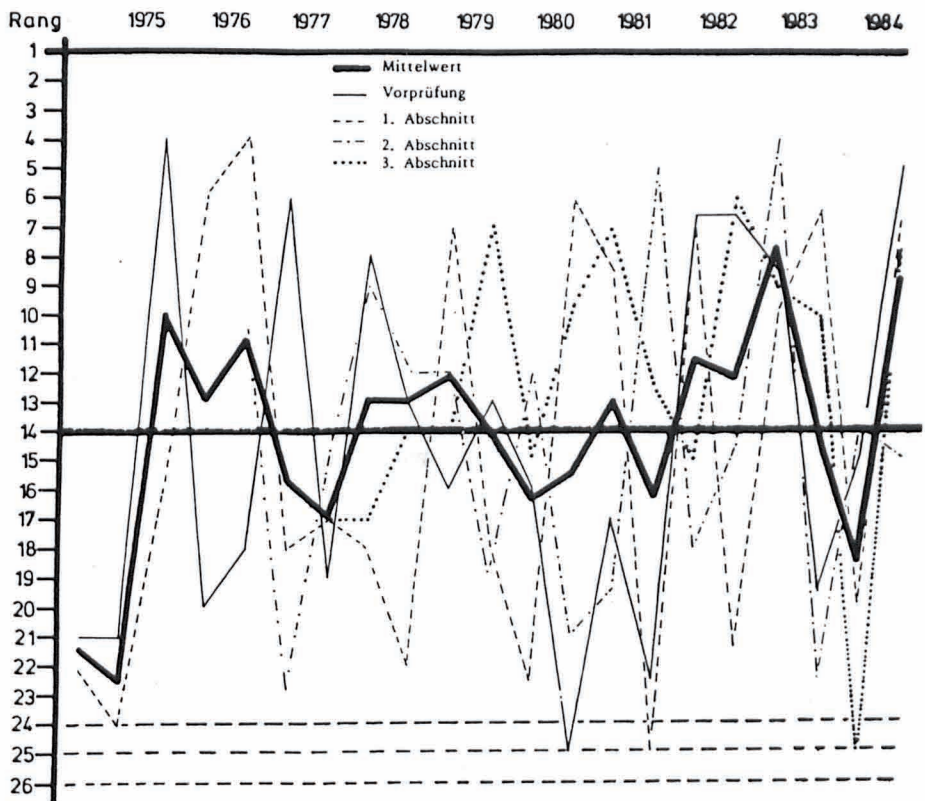
Natürlich gibt es zahlreiche andere Ranking-Konstruktionen. Aber stets – und das sollte hier gezeigt werden – wird eine Positionsfolge von eins bis neun ausgewiesen und es kommt immer zum gleichen Durchschnittswert von 100 als mittlerem Leistungsstandard.

Es gibt also immer einen ersten und einen letzten Platz – unterschiedslos und ganz sicher nicht im gleichen Sinne „leistungsrecht“. Das macht der Vergleich der einzelnen Ranking-Positionen klar: In Fall 1 ist der Erste immer noch schlechter als der Vierte in Fall 2; dafür ist in Fall 1 der Letzte sogar besser als der Sechste in Fall 2. In Fall 2 ist hingegen der Letzte besser als der Sechste von Fall 3 und dessen Vierter ist immer noch so gut wie der Erste in Fall 2 usw.

Wenn man also die Positionen der Hochschulen untereinander vergleichen will (was ja wohl beim Ranking verschiedener Kennzahlen der Regelfall sein dürfte), wären die Schlußfolgerungen im Hinblick auf die zugrunde liegende Leistung weitgehend falsch oder doch zumindest äußerst ungenau.

Das gleiche Problem kann auch im Zeitablauf eintreten, wenn dieselben Kennzahlen im Zeitablauf miteinander verglichen werden. Es ist zwar nicht immer wahrscheinlich, aber auch nicht ausgeschlossen, daß sich ähnliche Positionsunterschiede von einer Periode zur anderen ergeben. Der Vergleich der vier medizinischen Prüfungsabschnitte im Fachbereich Humanmedizin der Universität Gießen zwischen 1975 und 1984 zeigt dies überaus anschaulich (s. Abbildung unten).

Übersicht : Ergebnisse der vier Prüfungsabschnitte nach der Approbationsordnung für Ärzte zwischen 1975 und 1984



Die Rangfolge gibt an, den wievielten Platz Gießen im Vergleich zu den übrigen 26 Hochschulen der Bundesrepublik erreicht hat. Fettgedruckte Linie = Durchschnitt aus allen vier Prüfungsabschnitten.

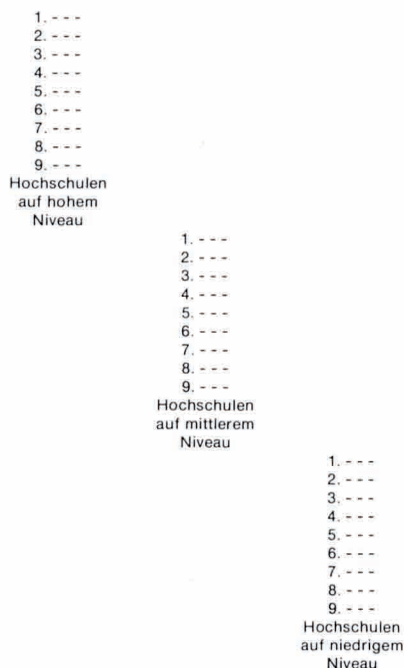


Das wilde Zick-Zack der Verbindungslinien macht deutlich, daß so gut wie alle Rangpositionen, ob gut oder schlecht, von den Gießener Medizinstudenten im Laufe der Jahre erreicht wurden, auch wenn im mehrjährigen Durchschnitt ein sicherlich nicht schlechter 14. Rang erreicht wurde.

### Niveauebene angeben

Hinzu kommt ein weiteres Problem, das sich aus der unterschiedlichen Niveauebene ergibt.

Es ist ein Unterschied, ob sich die skalierten Hochschulen auf hohem, mittlerem oder niedrigerem Niveau um einen Mittelpunkt bewegen, was die folgende Abbildung skizziert:



Sobald man also übergeordneten Standard als Maßstab einbezieht, kommen solche Positionsprobleme hinzu, die die „Leistung“ der Hochschule zusätzlich relativieren. So könnte zum Beispiel die beste Universität Hessens die erste, aber auch die fünf schlechteste in der Bundesrepublik sein. Die schlechteste Hochschule der Bundesrepublik hingegen könnte aber immer noch zum guten oder sogar zum höchsten Standard im europäischen Rahmen zählen.

Nicht so sehr die Rangposition einer zufälligen Rankinggruppe ist also maßgebend für den tatsächlich erreichten Leistungsgrad, sondern das Niveau der gesamten Rankinggruppe im Vergleich zum übergeordneten Leistungssystem (im Bundes-, im Europa- oder im Weltmaßstab). Kennzahlensysteme müßten demnach, um wirklich aussagekräftig zu sein, sowohl die „Streuung“ der Rangpositionen vom ersten bis zum letzten Platz angeben als auch die Ni-

veauebene erkennbar machen, auf der sich das ganze Ranking-System bewegt. Denn: Der letzte Platz auf hohem Niveau ist unter Umständen mehr wert als der höchste Platz auf niedrigerem Niveau.

### Eindeutigkeit und Stabilität

Weitere kritische Anmerkungen ergeben sich aus einer mehr praktischen Sicht der Dinge.

Wenn man schon über Kennzahlen zu einer Positionierung der einzelnen Hochschulen im Ranking kommen will, dann müssen die Kennziffern zumindest eindeutig im Ergebnis sein und möglichst auch stabil bleiben, wenn sich die Eingabedaten nur geringfügig ändern.

Die Eindeutigkeit der Position ist zum Beispiel dann nicht gewahrt, wenn die Kennzahlenberechnung für viele Hochschulen zum exakt gleichen Ergebnis führt. Im Extrem hätten alle die gleiche Kennzahl, wiederum also: alle auf Platz eins?

Gleiche Kennzahlenergebnisse entstehen zumeist dann, wenn sehr kleine Bezugsgrößen die Grundlage der Kennzahlen bilden. Die bereits angesprochene „Giese-Studie“ weist beispielsweise in der Rangliste „DFG-Gutachter 1980–1983“ zwölfmal das Ergebnis 1 aus, was zu den „Rangpositionen“ 34 bis 45 führte, sechsmal das Ergebnis 8 für die Plätze 17 bis 22 und viermal das Ergebnis 3 für Platz 27 bis 30. Aber auch wenn man diese Ergebnisse auf die Zahl der Professuren bezieht, wie das in der Studie im zweiten Schritt geschieht, kommt es zu zahlreichen nahezu gleichen Ergebnissen, was eine Rangunterscheidung unsinnig macht.

Eng verwandt mit diesem Problem ist der Fall mangelnder Differenzierungsmöglichkeiten, so daß sich – wegen der Gleichartigkeit der Kennzahlenergebnisse – Positionsunterschiede nur noch über die zweite oder dritte Stelle hinter dem Komma erreichen lassen. So reicht – um noch einmal die selbe Studie heranzuziehen – die Rangliste „Habilitationenquoten“ von 12,98 bis 0,69 Habilitationen je 100 Professuren. Das heißt aber nichts anderes, als daß je Professur 13 Hundertstel für die beste bzw. 7 Tausendstel Habilitationen (gerundet) für die schlechteste Hochschule als Maßstab für Leistungsunterschiede erhalten müssen. Wenn man sich also bereits in Größenordnungen von Hundertstel und Tausendstel bewegen muß, um überhaupt noch differenzieren zu können, dann verlieren solche Unterscheidungen sicherlich ihren Aussagewert.

In diesem Zusammenhang steht auch die Forderung nach Stabilität der Kennzahlenergebnisse, im Hinblick auf ihre Reagibilität (Reaktionsempfindlichkeit bei Änderung der Grundmenge).

Instabil sind in der Regel kleine Grundmengen. Wenn sich in unserem Beispiel die Zahl der Habilitationen von 1 auf 2 im Jahr erhöht (was ja leicht geschieht), dann verbessert sich die Kennzahl um 100 Prozent – eine nicht sehr sinnvolle Aussage.

### Vielzahl nicht ausufern lassen

Ausgesprochen wichtig ist auch, die Menge der Kennzahlen, mit denen Leistung gemessen werden soll, nicht ausufern zu lassen. Der Wunsch nach größtmöglicher Genauigkeit und Differenziertheit der Leistungsdarstellung kollidiert zwangsläufig mit der Forderung nach Einfachheit und Überschaubarkeit des Kennzahlensystems. Je stärker die Differenzierung, desto größer die Zahl der daraus resultierenden Kennzahlen, desto unübersichtlicher die Gesamtaussage.

Allein auf dem Forschungssektor arbeitet die erwähnte „Giese-Studie“ mit 13 Kennzahlen, die sich leicht verdreifachen läßt, wenn zum Beispiel bei der Kennzahlengruppe „je Wissenschaftler“ zusätzlich nach „je C4-Professuren“ und nach „je C4-C3-C2-Professuren“ unterschieden wird; eine Differenzierung, die in der Planungspraxis weitgehend üblich ist.

Bolsenkötter nennt in seiner Publikation „Ökonomie der Hochschule“ sogar 26, zum Teil ganz andere Kriterien, nach denen sich forschungsbezogene Leistung messen und bewerten lassen soll.

Nimmt man noch die allgemein üblichen lehr- und studienbezogenen Kennzahlen hinzu, so sind mit Leichtigkeit mehrere Dutzend verschiedene Kennzahlen denkbar, die sich noch einmal nicht unbeträchtlich erhöhen, wenn auch Kostenkriterien einbezogen werden (was an sich auch sinnvoll wäre). Die Menge der möglichen Kennzahlen steigert sich auf einige Hundert, wenn außerdem nach Disziplinen und Hochschularten (Universitäten, Gesamthochschulen, Fachhochschulen, Pädagogischen Hochschulen etc.) differenziert werden soll. Auch das wäre ganz zweckmäßig. Spätestens dann führen sich Kennzahlensysteme von selbst ad absurdum, weil man vor lauter Bäumen den Wald nicht mehr sähe.

### Welche Schlußfolgerungen ziehen?

Zum Schluß noch eine weitere, vielleicht die wichtigste Frage: Welche Schlußfolgerungen sind aus der Position in einer Ranking-Liste zu ziehen? Was kann/sollte die Hochschule (der Fachbereich, der einzelne Wissenschaftler) tun, um die erreichte Rangposition zu verbessern, wenn sie schlecht ist, oder um die Position zu halten, wenn sie gut ist? Diese Frage stellt sich nicht nur hochschulintern, sondern auch und vor allem im



Verhältnis zur übergeordneten Landesadministration, die die finanziellen Mittel zur Verfügung stellt.

In einem störungsfreien Marktsystem werden gute Leistungen belohnt (z. B. durch Gewinne), Minderleistungen aber „bestraft“ (z. B. durch Verluste bis hin zum Konkurs). So funktionieren üblicherweise die Gesetze der Marktwirtschaft.

Für unser föderalistisches, staatlich alimentiertes Hochschulsystem gilt dies wohl nicht: Kennzahlen sollten stets „belohnende“ und nicht „strafende“ Sanktionsmechanismen auslösen. Das ist zwar nicht ganz wettbewerbsgerecht, erscheint aber hochschulpolitisch unumgänglich, wenn Kennzahlen- und Rankingsysteme auch nur den Hauch einer Chance haben sollen, als Instrumente zur Leistungsbewertung von den Betroffenen anerkannt zu werden.

Bei positiver Sanktionierung wären mehrere Möglichkeiten denkbar, die allerdings zu völlig verschiedenen Strategien führen, was die Vergabe der Personal- und Sachmittel als Leistungsanreize betrifft. Das Land könnte,

- (a) die landesbeste Hochschule noch stärker fördern, um auch auf Bundesebene möglichst gut abzuschneiden. Ziel: Das Land wünscht eine Hochschule an der Bundesspitze oder mindestens unter den „top five“ der Bundesrepublik.
- (b) alle Hochschulen des Landes zu fördern, daß möglichst viele zu den besten in der Bundesrepublik gehören. Ziel: Alle Universitäten des Landes befinden sich unter den „top ten“ der Bundesrepublik.
- (c) die in der Rangliste am weitesten unten stehenden Hochschulen besonders intensiv fördern, um ihre Rangposition auf Bundesebene verbessern zu helfen. Ziel: keine Hochschule des Landes in der unteren Hälfte der Bundestabelle.

Aus der Sicht der im Ranking unten stehenden Hochschulen dürften natürlich Alternative (b) und (c) wünschenswert erscheinen. Die guten Hochschulen würden wohl auf (a) bestehen.

Grundsätzlich steht das Land also vor der Frage, ob es seine guten Hochschulen dafür belohnen will, weil sie gut sind, oder ob es seine weniger guten Hochschulen fördern soll, weil sie eben nicht gut sind, damit sie sich leichter nach oben arbeiten können.

Wenig Sinn allerdings macht die Forderung, alle Hochschulen um die Spitzenposition konkurrieren zu lassen. „Alle auf Platz eins“ erscheint nicht nur nicht sinnvoll, sondern logisch ausgeschlossen. Selbst wenn sich alle noch so sehr anstrengen und ein noch so gleichmäßig hohes Niveau erreichen, das Ranking zieht sie unerbittlich wieder auseinander: auf die Positionen zwischen dem ersten und dem letzten Platz.

## Übertriebene Gesetzestreue

(df) Eine Regierung, die der Bevölkerung willkürliche Vorschriften auferlegen wollte, könnte sich in der Bundesrepublik Deutschland wahrscheinlich ungeahnter Zustimmung der Betroffenen erfreuen. Rund 80 Prozent der Bevölkerung stehen nämlich ganz oder zumindest mit Einschränkung hinter der Aussage, daß man einem Gesetz auch dann gehorchen müsse, wenn man dessen Zweck nicht versteht. Das kam bei einer Repräsentativerhebung zum Vorschein, die Dr. Gerlinda Smaus vom Institut für Rechts- und Sozialphilosophie der Universität des Saarlandes mit Unterstützung der Deutschen Forschungsgemeinschaft vorgenommen hat. Beamte, besonders im höheren Dienst, sind offenbar die bravsten Untertanen und bekunden gleich zu 90 Prozent ihre potentielle Folgebereitschaft, berichtet Smaus in ihrem Buch „Das Strafrecht und die Kriminalität in der Alltagssprache der deutschen Bevölkerung“.

Ein eher schwach entwickeltes Rechtsbewußtsein spiegelt sich auch in dem Ergebnis wider, daß nur eine Minderheit Paragraphen darin gerechtfertigt sieht, daß sie „vernünftig“, „gut“ und „moralisch“ sind, während die Mehrheit befindet, daß man Gesetze alleine deshalb befolgen muß, weil sich „alle Menschen an etwas halten müssen“.

## Hypnotiseur will Ärzten helfen

(dpa) – In der Bundesrepublik könnten nach Darstellung des „Verbands der Bühnen-Hypnotiseure“ in Krankenhäusern und in den Arztpraxen viele Operationen schmerzfrei ohne das Risiko durch Narkosemittel vorgenommen werden, wenn die Ärzte geschulte Hypnotiseure einschalten würden. Das Vorstandsmitglied des Verbandes, der auch unter dem Künstlernamen „Don Alfredo“ bekannte Berufs-Hypnotiseur Manfred Knoke, erklärte kürzlich in Stuttgart, in Amerika hätten sich Operationen und lokale Eingriffe unter Hypnose bereits durchgesetzt.

Am Rande der internationalen Stuttgarter Künstlerbörse sagte der 39jährige Hypnotiseur, lediglich in zwei Privatkliniken in der Bundesrepublik und einigen Zahnarztpraxen würden heute regelmäßig Patienten unter Hypnose operiert. Etwa hundert Ärzte beherrschen nach Schätzungen von Knoke derzeit bundesweit die Kunst der Hypnose. Von den rund 150 in der Bundesrepublik auftretenden Hypnotisuren seien allerdings nur die wenigsten in der Lage, auf medizinischem Sektor ihre Aufgabe zu erfüllen. Knoke sagte, die Bundesbürger müßten sich von der Vorstellung trennen, daß Hypnose nur ein Spektakel für Jahrmärkte und Gala-Abende sei.