

DETLEF MÜLLER-BÖLING/STEFAN HORNPOSTEL

Fehlinterpretationen und Vorurteile

Vom Umgang mit Hochschulrankings und deren Nutzen

Seit zehn Jahren gibt es in Deutschland veröffentlichte Hochschulrankings. Ziel dieser Rankings war mehr Transparenz und Orientierung für die Hochschulwahl der Studierenden. Kritik an diesen Rankings blieb nicht aus und zeigte auch Wirkung. Aber die Kritik hat sich der Kritik zu stellen.

Detlef Müller-Böling, Dr. rer. pol., Univ.-Professor, Methoden der emp. Wirtschafts- und Sozialforschung, Leiter des Centrums für Hochschulentwicklung

Stefan Hornbostel, Referent am Centrum für Hochschulentwicklung

Diejenigen, die Rankings erarbeiten, haben dazugelernt. Das Centrum für Hochschulentwicklung (CHE) in Gütersloh hat zum Beispiel nie die „beste Hochschule“ gekürt. Auch die Leserschaft lässt sich nicht mehr mit einem kleinem Siegereppchen abspeisen. Innerhalb der Wissenschaft allerdings pflegt man die Vorbehalte, die schon vor zehn Jahren formuliert wurden. Es scheint, als sei die Kritik am Hochschulranking inzwischen so selbstreferentiell geworden, dass die Veränderungen und die veröffentlichten Befunde gar nicht mehr wahrgenommen würden. Als exemplarisch für ein solches Vorgehen, kann man den Beitrag von Werner Meinefeld über „Hochschulranking“ in *Forschung & Lehre* 1/2000 betrachten.

Rangplätze und Rangplatzvergleiche

Zur Debatte standen der Studienführer 1999 des CHE für die Fächer Rechtswissenschaft, Mathematik, Informatik und Physik und das Spiegel-Ranking 1999. Beide Untersuchungen präsentieren unter anderem die auf Stichproben basierenden Urteile von Studenten in Form von Ranggruppen, die sicherstellen, dass zwischen der Spitzen- und der Schlussgruppe statistisch signifikante Bewertungsunterschiede bestehen. Das CHE verzichtet völlig auf die Vergabe von Rangplätzen, weil es unsinnig ist, den Fachbereichen Ränge zuzuweisen, die eine Genauigkeit suggerieren, welche durch stichprobenbasierte Befragungsdaten nicht einzulösen ist. Die Zusammenfassung von verschiedenen Fachbereichen in Ranggruppen geht gerade davon aus, dass nicht-signifikante Mittelwertsdifferenzen in der Beurteilung von Leistungsaspekten der Fachbereiche auch keinen Rangunterschied zwischen den Institutionen begründen können. Deshalb werden statistische Vertrauensbereiche um die Fachbereichs-Mittelwerte der jeweiligen Urteile ermittelt und auf dieser Grundlage zunächst eine Spitzen- und eine Schlussgruppe gebildet. Zwischen den Hochschulen dieser beiden Gruppen bestehen erhebliche und statistisch signifikante Urteilsdifferenzen. Dabei werden die Stichprobengröße und die Homogenität der

Urteile am Fachbereich berücksichtigt. Alle Fachbereiche, an denen die Befragten sehr uneinheitlich geantwortet haben und/oder insgesamt ein „durchschnittliches“ Urteil abgeben, fallen in eine mittlere Ranggruppe. Die dort platzierten Fachbereiche unterscheiden sich weder vom Durchschnitt des Faches noch von den benachbarten Fachbereichen in der Spitzen- oder Schlussgruppe statistisch signifikant. Es ist der Sinn des Ranggruppenmodells, die Interpretation auf die Hochschulen in der Spitzen- und Schlussgruppe zu begrenzen. Dabei handelt es sich nicht um eine kosmetische Beigabe zu den klassischen Ranglisten, vielmehr bilden die Ranggruppen im Studienführer des CHE die Basis für eine nutzerspezifische Zusammenstellung von Fachbereichen mit Hilfe einer CD-ROM, die dem Studienführer des CHE beigefügt ist.

Ranggruppen entstehen nur dann, wenn tatsächlich bedeutsame Urteilsunterschiede vorhanden sind. Ihre Größe hängt davon ab, wie stark die Beurteilungen an den Fachbereichen voneinander abweichen. Unter Bedingungen völliger Homogenität würde sich nach dem Ranggruppenmodell nur eine einzige große Mittelgruppe ergeben. Natürlich lässt sich mit einem solchen Modell kein „Sieger“ ermitteln, sondern nur eine Gruppe von Hochschulen, die signifikant besser als der Durchschnitt im Fach beurteilt wird, und eine Gruppe, die signifikant schlechter beurteilt wird. Eine derartige Begrenzung passt aber nicht gut zum verbreiteten Rankingbedürfnis nach *Rangplätzen*, wie im Beitrag von Werner Meinefeld vorgeführt wird: Wenn die publizierten Hitlisten – wie im Falle des CHE-Studienführers – keine Rangplätze enthalten, dann werden sie einfach durch Auszählung im Nachhinein ermittelt, um darin festzustellen, dass die Juristen der Universität Potsdam (Mittelgruppe) aufgrund der studentischen Beurteilungen zwischen dem 9. und 28. Rangplatz positioniert sein könnten. Ein solcher Befund ist nun bestenfalls eine Tautologie, denn nichts anderes besagt die Einstufung in die mittlere Ranggruppe.

Ein ernst zu nehmendes Argument entstände erst, wenn die Hochschulen der Spitzengruppe nach statistischen Krite-

rien auch in der Schlussgruppe positioniert sein könnten. Genau dies schließt die Einordnung in Ranggruppen aber aus. Es wäre also dringend zu wünschen, dass auch empirische Sozialforscher nicht weiterhin gegen potemkinsche Dörfer argumentierten, sondern sich auf die dargebotene Information einließen. Es würde sich dann nämlich sehr schnell zeigen, dass auch ein anderer Kritikpunkt nicht sonderlich stichhaltig ist: So ermittelt Werner Meinfeld, dass zwischen der Spiegel-liste der juristischen Fachbereiche und der Hitliste des CHE-Studienführers Abweichungen von durchschnittlich sieben Rangplätzen auftraten. Abgesehen davon, dass im einen Fall ein Durchschnittswert über 16 Fragen benutzt wird, im anderen Fall direkt ein Gesamturteil erhoben wird, ist dieser Befund nichts sagend, denn entscheidend ist ausschließlich die Ranggruppenzugehörigkeit. Fragt man nach den Ranggruppen, dann zeigt sich, dass 72 Prozent der juristischen Fachbereiche in beiden Studien in die gleiche Ranggruppe eingeordnet wurden und 26 Prozent ihre Position zwischen der Mittelgruppe und einer Extremgruppe wechselten. Lediglich auf einen Fachbereich (Saarbrücken) trifft das zu, was mit Rangplatzvergleichen offenbar suggeriert werden soll: Die widersprüchliche Zuordnung zur Spitzengruppe in der Spiegel-Studie und zur Schlussgruppe in der CHE-Hitliste. Verzichtet man dar-

widersprüchliche Zuordnung in Spitzen- bzw. Schlussgruppe zu. Der Grund ist leicht zu ermitteln: In der Spiegel-Befragung wurde zwischen Lehramts- und Diplomstudierenden nicht unterschieden, in der CHE-Erhebung wurden hingegen nur Mathematikstudenten mit Abschlussziel Diplom befragt. Lehramtsstudierende haben berechtigterweise andere Erwartungen an das Studium als Diplommathematiker. Das schlägt sich in den protokollierten Beurteilungen nieder und erklärt die Differenz von Spiegel- und CHE-Daten.

Was wird gemessen?

Damit ist die Frage aufgeworfen, was mit derartigen Befragungen genau gemessen wird. Auch hier geistern seit den ersten Rankingversuchen immer neue Vermutungen auch durch wissenschaftliche Veröffentlichungen. Gemessen werde „well being“ sagen die einen. Studierende in den Metropolen seien kritischer als in kleineren Orten, sagt Werner Meinfeld und weist darauf hin, dass es etwa für die Bewertung der Bibliothek keinen allgemeinverbindlichen Maßstab gäbe. Man ist bei der Prüfung solcher Fragen aber keineswegs auf Spekulationen angewiesen. Zunächst einmal hat die ganz überwiegende Zahl von empirischen Überprüfungen der Studierendenurteile ergeben, dass Merkmale der Befragten wie Alter, Geschlecht, Abiturnote, Studiendauer, Hochschulwechsel usw. keinen oder (fachspezifisch) nur geringen Einfluss auf das Urteil der Studierenden hat. Das gilt insbesondere für die Daten des CHE-Studienführers. Da sich derartige Einflussfaktoren nie erschöpfend prüfen lassen, kann man sich der Frage danach, was gemessen wird, auch dadurch nähern, dass man die Urteile verschiedener kompetenter Expertengruppen miteinander vergleicht. Neben den Studierenden sind die Professoren sicherlich diejenigen, die kompetent Auskunft geben können und im Verlauf ihrer Berufsbiographie auch überlokale Bewertungsstandards entwickelt haben. Der CHE-Studienführer beschränkt sich daher auch nicht auf die Beurteilungen der Studierenden, sondern informiert ebenso über die Urteile der Professoren zu den untersuchten Sachverhalten und berichtet eine Fülle korrespondierender Fakten.



Warteschlangen an der Humboldt-Universität zu Berlin

Foto: Ausserhofer

auf, den Hitlisten nachträglich Rangplätze anzudichten und betrachtet stattdessen die publizierten Urteilsmittelwerte, dann ergibt sich mit einer Korrelation von .73 eine für sozialwissenschaftliche Daten erstaunlich hohe Übereinstimmung zwischen den Messungen des Spiegels und des CHE in der Rechtswissenschaft.

Derartig hohe Übereinstimmungen findet man nicht in allen Fächern. In der Mathematik korrelieren Spiegel- und CHE-Befunde beispielsweise nur mit .30, und „nur“ 48 Prozent der Fachbereiche sind in die gleiche Ranggruppe sortiert worden, allerdings trifft auch nur auf zwei Fachbereiche die

Wenn also die Vermutung richtig ist, dass in Großstadtuniversitäten überwiegend kritische Studierende anzutreffen sind, in kleineren Städten jedoch eher brave Studierende, dann müssten die Urteile der Professoren – die vermutlich in den Metropolen nicht kritischer sind als ihre Kollegen an anderen Hochschulen - von denen der Studierenden recht deutlich abweichen. Das Gegenteil ist aber der Fall: Studenten- und Professorenurteile über die Bibliotheken korrelierten stark und signifikant (Pearsons Cor. .58). Von 38 juristischen Fachbereichen mit fünf und mehr Professoren lassen sich an 29 Fachbereichen keine signifikanten Urteilsdifferenzen zwischen Pro-

fessoren und Studenten bei der Einschätzung der Bibliotheken finden. Es verwundert daher, dass über die Aussagekraft der Studenturteile spekuliert wird, anstatt die veröffentlichten Daten für eine kritische Prüfung zu nutzen. Offensichtlich urteilen die Studierenden der Rechtswissenschaft durchaus ähnlich kompetent wie ihre Professoren.

Alles eine Frage der Größe der Fakultät?

Hochschulrankings, die nur auf Studenturteile abstellen, kommen immer wieder zu dem Ergebnis, dass kleinere Fachbereiche von den Studierenden besser bewertet werden als große, insbesondere in den stark nachgefragten Fächern wie z.B. Rechtswissenschaften. Zur Erklärung dieses Phänomens wird gern darauf verwiesen – und auch Werner Meinfeld tut dies erneut –, dass den Studierenden an den großen Fakultäten einfach die Nestwärme fehle und sie gar nicht die Qualität der Lehre beurteilten, sondern allenfalls Rahmenbedingungen, die nur indirekt etwas mit Qualität zu tun hätten. Auch hier gilt aber, dass die Jura-Professoren die Lage nicht anders als die Studierenden einschätzen: an großen Fakultäten wird die Lehrsituation signifikant schlechter beurteilt als an kleineren. Auch die Fakten sprechen nicht gerade dafür, dass der Kern des Problems fehlende Nestwärme ist: Die Studiendauer liegt an großen Fakultäten deutlich höher, die Noten im Staatsexamen hingegen unterscheiden sich nicht signifikant von den kleineren Fachbereichen. Noch nachdenklicher stimmt aber die Tatsache, dass die Größe einer Fakultät keineswegs die Qualitätseinschätzungen von Professoren und Studierenden determiniert. Betrachtet man die größeren juristischen Fakultäten mit 400 bis 500 Studierenden im ersten und zweiten Fachsemester, dann stellt man ein erstaunlich großes Urteilsspektrum fest. Die mittleren Studenturteile über die Studiensituation reichen von 1,8 bis 2,8 und die Mittelwerte der Urteile der Professoren zur Lehrsituation von 2,3 bis 3,6. Mit anderen Worten: Auch unter ähnlich widrigen Bedingungen lassen sich sehr unterschiedliche Ergebnisse produzieren. In weniger überfüllten Fächern – wie etwa der Physik – zeigen die Daten des CHE-Studienführers ganz ähnlich, dass es etlichen großen Fachbereichen gelingt, sich sowohl auf den Forschungsindikatoren als auch in der subjektiven Beurteilung durch Studierende und Professoren überdurchschnittlich gut zu positionieren, während andere große Fachbereiche weder das eine noch das andere erreichen.

Was kann ein Hochschulranking leisten?

Die Kritik an Hochschulrankings folgt seit zehn Jahren unverändert einer Argumentation, die folgendermaßen aufgebaut ist: Weil Qualität von Forschung und Lehre schwer oder gar nicht definierbar ist, sind die Rankings fehlerhaft oder ungenau. Und weil das so ist, erzeugen sie fatale Fehlsteuerungen. Implizit verlangen die Kritiker gerade das, was ein Ranking nicht leisten kann, nämlich die auf die Nachkommastelle genaue Quantifizierung der „besten Hochschule“ und eine daran orientierte Ressourcensteuerung. Scheinbar unbemerkt von dieser kritischen Wissenschaftsöffentlichkeit haben sich Rankings inzwischen in eine ganz andere Richtung entwickelt. Sie verzichten gerade auf derartige Scheingenauigkeiten und bieten ein breites fachspezifisches Informationsangebot. Die Beurteilungen der Studierenden sind

im Studienführer des CHE nur eine Information, neben den Beurteilungen der Professoren, den Daten zur infrastrukturellen Ausstattung, zur Studiendauer, zu Abschlussnoten, Betreuungs- und Beratungsangeboten, den Ergebnissen bibliometrischer Messungen der Forschungsleistungen, Drittmittelanalysen und weiteren Informationen. Es ist wenig einleuchtend, die Leistungen eines multidimensionalen Rankings – wie es mit dem CHE-Studienführer vorgelegt wird – auf das Gesamturteil der Studierenden zu reduzieren.

Aufgabe eines Hochschulrankings ist es gerade, die verschiedenen Dimensionen hochschulischer Leistungen auszuleuchten und zwar aus unterschiedlichen Perspektiven, und diese Befunde vergleichend so aufzubereiten, dass ein Informationsbedarf sinnvoll befriedigt werden kann. Das breite Indikatorenprogramm des CHE-Studienführers zeigt im Detail, wie differenziert die Lehr- und Forschungssituation erfasst wurde. Im Ergebnis ergibt sich ein Bild der deutschen Hochschullandschaft, das sich bereits sehr weit von der Fiktion eines einheitlichen Hochschulsystems ohne wesentliche Unterschiede zwischen den einzelnen Institutionen entfernt hat. Es wird sehr deutlich erkennbar, dass nicht nur erhebliche Unterschiede bestehen, sondern auch dass Fachbereiche unterschiedliche Stärken und Schwächen entwickelt haben. Die wenigsten sind einfach gut oder schlecht. Studierende, Studienanfänger und auch die Öffentlichkeit haben ein Recht darauf, über diese unterschiedlichen Studienbedingungen vor Ort und die Leistungsunterschiede zwischen den einzelnen Fachbereichen informiert zu werden. Bisher wählen Studienanfänger einfach die nächstliegende Hochschule. Von dieser Fehlsteuerung wird allerdings wenig gesprochen, stattdessen wird befürchtet, dass ein höheres Maß an rationalem Wahlhandeln bei den Studienanfängern den status quo durcheinanderbringen könnte.

Was tun?

Grotesk wird die Kritik, wenn Werner Meinfeld am Schluss fordert, die Hochschulen sollten selbst das Bedürfnis nach Information und Transparenz in der Gesellschaft befriedigen. Denn genau das tut der CHE-Studienführer. Das CHE, 1994 als Einrichtung der Hochschulen u.a. mit dem expliziten Auftrag gegründet, derartige Transparenz zu schaffen und den Medien ein wissenschaftsadäquates Ranking entgegenzustellen, hat mit den Hochschulen gemeinsam und mit ausgewiesenen Experten das multidimensionale Ranking entwickelt und im STERN/START veröffentlicht. Alle Daten kommen von den Hochschulen, nichts wird publiziert ohne unsere Erlaubnis. STERN/START sind Medienträger, aber nicht Medienmacher. Dass über empirische Indikatoren, die ja bekanntlich stets nur eine partielle Definition eines Phänomens gestatten, gerade unter wissenschaftlichen Gesichtspunkten gestritten werden muss, dass sie weiterzuentwickeln sind auf der Basis empirischer Erkenntnisse ist für uns selbstverständlich. Wir stellen uns gerne der – substantiierten – wissenschaftlichen Kritik.

Anschrift der Autoren

CHE - Centrum für Hochschulentwicklung
Carl-Bertelsmann-Str. 256
33311 Gütersloh